



ENTROPIA, INFORMAÇÃO E QUALIDADE: DE UM PERCEPTRON PARA
AVALIAR SIMILARIDADE ENTRE STRINGS ATÉ UM MODELO DE
CAMPOS CONSERVATIVOS

Paulo Coelho Ventura Pinto

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Luís Alfredo Vidal de Carvalho

Rio de Janeiro
Novembro de 2017

ENTROPIA, INFORMAÇÃO E QUALIDADE: DE UM PERCEPTRON PARA
AVALIAR SIMILARIDADE ENTRE STRINGS ATÉ UM MODELO DE
CAMPOS CONSERVATIVOS

Paulo Coelho Ventura Pinto

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Valmir Carneiro Barbosa, Ph.D.

Prof. Luís Alfredo Vidal de Carvalho, D.Sc.

Prof. Antonio MacDowell de Figueiredo, Dr.Ing.

Prof. Claudio Esperança, Ph.D.

Prof. Luiz Pinguelli Rosa, D.Sc.

Prof. Sérgio Manuel Serra da Cruz, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
NOVEMBRO DE 2017

Pinto, Paulo Coelho Ventura

Entropia, Informação e Qualidade: de um Perceptron para Avaliar Similaridade entre Strings até um Modelo de Campos Conservativos/Paulo Coelho Ventura Pinto. – Rio de Janeiro: UFRJ/COPPE, 2017.

XIV, 202 p. 29, 7cm.

Orientador: Luís Alfredo Vidal de Carvalho

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2017.

Referências Bibliográficas: p. 120 – 159.

1. Teoria da Informação. 2. Incerteza. 3. Entropia.
4. Qualidade de Dados. 5. Modelagem Cognitiva. 6.
Perceptron. 7. Perceptrão. 8. Redes Neurais. 9.
Mitologia. 10. Qualidade. 11. Campos Conservativos.
12. Inteligência Artificial. I. Carvalho, Luís Alfredo Vidal
de . II. Universidade Federal do Rio de Janeiro, COPPE,
Programa de Engenharia de Sistemas e Computação. III.
Título.

*Para minha alma gêmea, Jessica.
Aos meus pais, presentes de
D'us, Paulo & Celina.
À minha sogra, presente de
minha esposa, Carlinda.
Aos meus avós, com quem,
apesar da curta convivência,
muito aprendi, Elza & Joaquim
e Luiz & Generosa.*

Agradecimentos

Aqui estão as pessoas e as instituições – levando-se em conta sua natureza, dimensão, responsabilidade e proximidade – que contribuíram para o meu desenvolvimento pessoal, profissional e acadêmico. Exerceram, do seu modo, algum grau de influência sobre quem sou – seja por gestos, palavras ou ações. Então, uma parte da qualidade de meus julgamentos devo também a elas. Consequentemente, influenciaram as decisões por mim tomadas até o presente momento, o que inclui aquelas que concorreram para a conclusão da presente tese. Infelizmente, esse rol não pode ser exaustivo e, inevitavelmente, incorrerei em omissões. Muito resumidamente, estão nele aquelas que de alguma forma D’us manda com algum propósito: seja num extremo para dar umas boas risadas ou, no outro, para obliterar os obstáculos.

À minha esposa, Jessica, cujo apoio, esforço e obstinação para que eu obtivesse o grau de doutor só é explicado por um amor verdadeiro. Sem ela estes agradecimentos não seriam possíveis – simples assim.

Aos meus pais, o Eng. Paulo Ventura Pinto e a Prof. Celina Coelho de Jesus, pelo amor incondicional que permitiu a realização deste trabalho. Pois nunca me faltou nada, nem material e imaterial – só posso me queixar da fartura de tudo o que é bom. Meu pai, um leão na defesa de seus filhos, nunca deixou um projeto incompleto e sempre utilizando engenho e arte. Minha mãe, diligentemente, impôs uma educação de qualidade e um mundo livre de vicissitudes.

Aos meus irmãos Rodrigo e Jeanine, pelos momentos de alegria e apoio.

Aos professores da COPPE/PESC, os senhores Luis Alfredo Vidal de Carvalho, Sergio Exel Gonçalves e Luis Felipe Magalhães de Moraes, cujos ensinamentos foram fundamentais para abordar o problema desta tese. Em especial, agradeço ao professor Luiz Alfredo pelo estímulo, acolhimento e por ter me apresentado a lei de Weber-Fechner que fora o farol ao longo desta pesquisa e pela presença principalmente nos momentos difíceis – sempre com soluções e palavras amigas.

Aos Professores da Banca de Doutorado e de Qualificação, pelas críticas e sugestões, consubstanciadas no trabalho final aqui apresentado, os senhores Valmir Carneiro Barbosa, Luís Alfredo Vidal de Carvalho, Antonio MacDowell de Figueiredo, Claudio Esperança, Luiz Pinguelli Rosa, Adilson Elias Xavier e Sérgio Manuel Serra da Cruz.

Aos professores da graduação, especialização e mestrado, em especial, Júlio Salek Aude, Ageu Cavalcanti Pacheco Júnior, Sérgio Guedes de Souza, Nelson Quilula Vasconcelos, Josué Laguardia, Raymundo de Oliveira, Herch Moysés Nussenzweig, Jorge de Albuquerque Vieira e Teresinha de Jesus Stuchi.

Aos professores, pelas aulas on-line, L. Susskind e Joseph J. Hughes (o primeiro de estatística mecânica e o segundo de mitologia grega).

Aos colegas do doutorado na COPPE/PESC, Ricardo Mesquita, Renato Cerceau, Elaine Sigette, Evandro Macedo e Vander Luiz Proença da Silva.

À Sociedade Brasileira de Pesquisa Operacional (SOBRAPO), à Sociedade Brasileira de Computação (SBC), e ao Scottish Informatics Programme (SHIP), pelo acolhimento e críticas aos artigos submetidos concernentes à pesquisa aqui desenvolvida.

Aos colegas do SERPRO, enquanto trabalhei como analista de qualidade de processo de desenvolvimento de software. Em especial, ao meu amigo José dos Reis Vitoretto, um grande filósofo, músico e engenheiro. Sem esquecer também de meus colegas Carlos Eduardo, Maurício Moreira e Salvador. E ao gerente Paulo Sérgio Lepore.

À Universidade Federal do Rio de Janeiro, minha alma mater.

À Fundação Oswaldo Cruz (FIOCRUZ), à Fundação Getúlio Vargas, e ao Instituto Padre Leonardo Carrescia.

Às gerentes da Agência Nacional de Saúde Suplementar (ANS), Dra. Ceres Albuquerque e Márcia Frank Piovesan por terem me proporcionado a responsabilidade de solucionar o problema de identificação unívoca de beneficiários de planos de saúde na ANS, no Projeto de Reestruturação do Cadastro de Beneficiários. Problema esse que não tinha recebido solução até o ano de 2010. A oportunidade de trabalhar nesse projeto foi importante no processo seletivo deste doutorado, como também para a formalização do doutorado como exigência de capacitação funcional no âmbito da ANS: objetivando o aprimoramento do exercício de minhas atividades como servidor público federal. Somando-se a isso, a recomendação desses colegas para minha capacitação no International Scottish Health Informatics Programme (SHIP) Conference, no ano de 2011, foi uma grande contribuição para esta pesquisa, pois foi a primeira vez em que tive contato com profissionais na área de record linkage. Congresso esse, ao qual retornei na edição de 2013 também na qualidade de apresentador.

Aos colegas da ANS, Carlos Alexandre Fernandes, Bruno Reis, Jorge Melo, Leonardo Macedo, Flávio Ferreira, Sérgio Costa Junior, Dante Casanova, Claudete Rocha, Renato Cerceau, Simone Schenkman, Andrea Brites, Juliana Dib e Nádia Pinto.

Aos colegas da graduação e mestrado, em especial aos amigos Rodrigo Oliveira de Castilho Rodrigues, Gustavo Adolfo Medeiros, Denilson da Rocha Cândido,

Washington de Jesus Melo, Mauro Leitão, Thalita de Moraes, Alex Magalhães Pinho, Adriano Pupp, Bianca Reis e Ricardo Ribeiro.

À equipe e à iniciativa COPPETEX.

Ao CNPq e CAPES por terem financiado meus estudos no passado, o que permitiu no futuro, o presente trabalho.

“— Dei a volta ao redor de nossos pais, sim, o que é o mesmo que ter dado a volta ao mundo — disse Ganesh, revelando, afinal, as suas razões. Ganesh explicou que seus pais Shiva e Parvati eram considerados como a encarnação do próprio universo, daí podendo ser dito, com a mais absoluta verdade, que ele dera, de fato, a volta em todo o universo.”

Ganesh ganha uma aposta em “As Melhores Histórias da Mitologia Hindu”.

“Não confies tuas mágoas
Por maior que sejam as penas.
Quem suas mágoas descobre
Por suas mãos se condena.”

Dizeres num prato português.

“a Biblioteca Nacional é um mundo; papel deslocado ou mal classificado fica perdido indefinidamente até que um acaso favorável o reconduza à luz”

Carta de **Capistrano de Abreu** para **Afonso de Taunay**, 9 de janeiro de 1915.

ALAN ALDA: (Narration) Knowing what consciousness looks like means you could also search for it in non-human animals. Edelman has little doubt that an animal such as a dog would turn out to possess at least the glimmerings of a consciousness flame.

DR. GERRY EDELMAN: The difference between a dog and you however, or Giulio let’s say, is that if you kick a dog the next time he may bite you or run away, but he doesn’t sit around plotting to remove your professorship!

Gerald Maurice Edelman, em entrevista com **Alan Alda**, em *CHANGING YOUR MIND*, American Scientific Frontiers.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

ENTROPIA, INFORMAÇÃO E QUALIDADE: DE UM PERCEPTRON PARA
AVALIAR SIMILARIDADE ENTRE STRINGS ATÉ UM MODELO DE
CAMPOS CONSERVATIVOS

Paulo Coelho Ventura Pinto

Novembro/2017

Orientador: Luís Alfredo Vidal de Carvalho

Programa: Engenharia de Sistemas e Computação

Informação, enquanto um objeto concreto das ciências exatas e da natureza, tem o trabalho de Claude Shannon, *A Mathematical Theory of Communication*, como referencial teórico basilar. Essa noção de informação nasceu com uma irmã antípoda: a incerteza – chamada de entropia em razão de sua semelhança formal com a entropia na estatística mecânica. Neste trabalho, utiliza-se medidas de entropia para compor um processo de comparação entre palavras usando um único neurônio perceptron como unidade de decisão (um processo de natureza qualitativa modelado quantitativamente). Para produzir a entropia fornecida ao perceptron, o conceito de mitos é desenvolvido – permitindo a contagem de variações de histórias ficcionais e etiológicas sobre diferenças ou similaridades entre as palavras comparadas. Como resultado final, propõe-se a utilização de conceitos clássicos vetoriais – em especial o cálculo de distâncias e de cossenos entre vetores – para comparação funções de probabilidade de uma mesma família.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

ENTROPY, INFORMATION AND QUALITY: FROM A PERCEPTRON TO
ASSESS SIMILARITY BETWEEN STRINGS UP TO A CONSERVATIVE
FIELD MODEL

Paulo Coelho Ventura Pinto

November/2017

Advisor: Luís Alfredo Vidal de Carvalho

Department: Systems Engineering and Computer Science

Information, as a tangible object belonging to the exact sciences, has the work of Claude Shannon, *A Mathematical Theory of Communication*, as its theoretical basilar reference. That notion of information was born with an antipodean sister: uncertainty – called entropy due to its formal similarity to entropy in mechanical statistics. In this work, we use entropy measures to handcraft a process of word comparison using a single perceptron neuron as a decision unit (a qualitative process modeled quantitatively). To produce the entropy supplied to the perceptron, the concept of myths is developed – allowing for the counting of variations of fictional and etiological histories due to differences or similarities between the words compared. In the end, it is proposed the use of classical vector concepts – in particular the calculation of distances and cosines between vectors – to compare probability functions from the same family.

Sumário

1	Apresentação	1
1.1	Introdução	1
1.2	Problema	2
1.3	Motivação	3
1.4	Qualidade e Inteligência Artificial	5
1.5	Por que Teoria da Informação?	7
1.6	Dificuldades	7
1.7	Referenciais Teóricos	8
1.8	Trabalhos Desenvolvidos	9
1.9	Qualidade de Dados	9
1.10	Definições Básicas	9
1.11	Orientações para Leitura	10
2	Informação, Incerteza e Entropia	12
2.1	Introdução	12
2.2	Informação	13
2.3	Incerteza	16
2.4	Entropia	17
2.5	Conclusões	22
3	Da Informação à Incerteza	23
3.1	Introdução	23
3.2	Breve Histórico	24
3.3	Informação de Hartley	25
3.4	Informação de Shannon	29
3.5	Relações entre os Paradigmas de Informação de Hartley e Shannon	38
3.6	Unidades de Informação	41
3.7	Incerteza	41
3.8	Discussão e Conclusão	43

4	Qualidade de Montgomery	45
4.1	Introdução	45
4.2	As Propriedades da Qualidade de Montgomery	46
4.3	Desvio-padrão: uma Medida de Variabilidade	49
4.4	Considerações Finais	49
5	Cálculo de Incertezas I	51
5.1	Introdução	51
5.2	Equilíbrio Químico como Paradigma de Decisão	52
5.3	Médias Ponderadas	55
5.4	Médias, Entropia e Ponderação	57
5.5	Princípios de Procedimentos de Decisão	62
5.6	Discussão e Conclusões	70
6	Uma Mitologia para um Casamento Aproximado de Palavras	72
6.1	Introdução	72
6.2	Objetivos	73
6.3	Trabalhos Relacionados	73
6.4	Definições	74
6.5	Escolha do Neurônio Perceptron	76
6.6	Escolha do Soundex	77
6.7	Uma Baseline para Qualidade entre Strings	81
6.8	Contando Mitos e suas Variações	82
6.9	Casamento Aproximado entre Palavras por um Perceptron	84
6.10	Discussões e Conclusões	89
7	Cálculo de Incertezas II	91
7.1	Introdução	91
7.2	Motivação	91
7.3	Definições Básicas	92
7.4	Partições e Processo de Quantização	93
7.5	Multipolos e Dipolos Qualitativos	95
7.6	Discussão e Conclusões	99
8	Qualidade e Campos Conservativos	100
8.1	Introdução	100
8.2	Motivação	101
8.3	Objetivo	102
8.4	Definições Básicas	102
8.5	Axiomas para Qualidade	103

8.6	Distribuição Normal Unidimensional	104
8.6.1	Diferença de Entropia entre duas Distribuições Normais Unidimensionais	105
8.7	Distribuição Normal Multivariada	107
8.7.1	Diferença de Entropia entre duas Distribuições Normais Multivariadas de Mesma Dimensão	110
8.7.2	Granularidade de Parametrização	111
8.8	Aplicação: Qualions	113
8.8.1	Incerteza derivada de Densidade de Probabilidade: Gaussiana Multidimensional	114
8.8.2	Incerteza derivada de Probabilidade: Variáveis Categóricas	115
8.9	Discussões e Conclusões	116
9	Conclusão	118
9.1	Discussão	118
9.2	Conclusão	119
9.3	Trabalhos Futuros	119
	Referências Bibliográficas	120
A	Comunicação e Mitologia	160
A.1	Introdução	160
A.2	Motivação	161
A.3	Ficcionalismo e Perspectivismo	162
A.4	Comunicação como Fenômeno Natural	163
A.5	Abordagem Mitológica	166
A.5.1	Mitologia	166
A.5.2	Ponderação Mítica	169
A.6	Física Qualitativa de Kleer e Brown: Processos Míticos	170
A.7	Variáveis Qualitativas	173
A.8	Discussão	176
B	Termodinâmica Clássica: Princípio da Máxima Entropia	177
B.1	Introdução	177
B.2	O Modelo	177
B.3	Informações de Otimização	179
B.4	Máxima Entropia	180
C	Campos Conservativos: Conceitos e Notações	185
C.1	Introdução	185
C.2	Definições e Teoremas	185

D Espaço de Termômetros	188
D.1 Introdução	188
D.2 Definições	188
D.3 Propriedades da Soma Vetorial	190
D.4 Propriedades da Multiplicação por Escalar	191
D.5 Espaço Vetorial	193
D.6 Produto Interno e Norma Euclidiana	193
E Outros Estudos	196
E.1 Dados e Programas Programas de Computador	196
E.2 Super-entropia	202

Capítulo 1

Apresentação

MOYERS: Máquinas nos ajudam a perceber a idéia de que queremos um mundo feito à nossa imagem, queremos que ele seja o que pensamos que devia ser.

CAMPBELL: Sim. Mas então chega um momento em que a máquina começa a ditar ordens a você. Por exemplo, eu comprei uma dessas máquinas maravilhosas – um computador. Ora, como lido predominantemente com deuses, foi por aí que identifiquei a máquina: ela me parece um deus do Velho Testamento, com uma porção de regras e nenhuma clemência.

Joseph John Campbell, O Poder do Mito – com Bill Moyers

1.1 Introdução

A ciência da computação é caracterizada pelo desenvolvimento algoritmos, programas, sistemas de softwares e computadores, entretanto possui um conceito que permeia todo esse campo de estudo: a informação.

Ao mesmo tempo, o conceito de informação vem acompanhado com o de incerteza graças ao trabalho de Claude Elwood Shannon, ao final da década de 1940, *A Mathematical Theory of Communication*. Entretanto, o conceito de incerteza não é tão acessível quanto o conceito de informação. Curiosamente, a unidade associada à informação binária, o *bit*, tão bem conhecida nos dias de hoje, é insolitamente uma medida de incerteza também. O esclarecimento desse aparente paradoxo é que ao se pensar em um *bit* logo se considera duas possibilidades de informação (*0* ou *1*, *sim* ou *não*, *ligado* ou *desligado*, por exemplo).

O conceito de qualidade modernamente tem vários significados em diferentes contextos, mas em geral está associado à ideia de valor de um atributo ou melhoria. Nesse sentido amplo, a qualidade se assemelha tanto ao conceito de incerteza e informação: suscita diferentes possibilidades, difícil definir, mas necessária para o bem-estar e êxito.

Sobre o tipo de pesquisa aqui realizada, trata-se de uma pesquisa básica: criação de novos conhecimentos contemplativos e de conhecimentos para aplicação em inteligência artificial para comparação de atributos qualitativos de dados. Também é uma pesquisa exploratória, isto é, busca proporcionar familiaridade com o tema e a construção de hipóteses (DA SILVA e MENEZES, 2005).

Neste capítulo será apresentado o problema, a motivação, os objetivos desta tese e outros tópicos pertinentes.

1.2 Problema

A pergunta que esta tese procura responder:

Como gerar medidas baseadas em entropia para criar racionalmente dados para serem fornecidos a um modelo de inteligência artificial para fins de avaliação de qualidade de dados?

A entropia de um sistema, seja ele um canal de comunicação ou de um sistema termodinâmico, é uma medida, portanto de natureza quantitativa – cuja interpretação mais fundamental é ser uma grandeza de incerteza. Dados, por sua vez, podem ser representações numéricas, ou não.

Ao apresentar uma forma de realizar medidas de entropia e interpretá-la como dados de entrada de um sistema inteligente, estar-se-á pavimentando um caminho para utilização, tanto em processos de aprendizado, de inferência e de classificação.

Por racionalmente, significa que as medidas de entropia não são meramente heurísticas, nem parte do viés de aprendizado do modelo, mas sim obtidas dos dados como uma propriedade e também interpretáveis por seres humanos. Portanto, se é possível fornecer essas medidas como atributos, então um modelo de aprendizagem de natureza numérica pode ser utilizado; seja, por exemplo, para classificação, ou para regressão, efetuada por uma rede neural artificial.

Por exemplo, mesmo que o método de aprendizado envolva algum viés baseado na maximização de entropia, modelos podem receber entropia como dado de entrada (input) – para o modelo isso é indiferente, mas abre um leque de opções. Do ponto de vista da teoria de aprendizado de máquina, uma dimensão extra na modelagem de um problema pode ser redundante ou mesmo viabilizar uma solução (como no problema do *XOR*¹).

¹O problema do aprendizado da função *XOR* pelo perceptron: as classes não são diretamente linearmente separáveis, mas utilização de uma assinatura polinomial quadrática permite que as novas instâncias sejam linearmente separáveis, ver (ABU-MOSTAFA *et al.*, 2012, p. 99-103) e (BRAGA *et al.*, 1998, p. 32-34).

1.3 Motivação

Motivação Teórica

Lavoisier e Laplace projetaram e construíram um calorímetro que poderia medir a quantidade de calor liberado durante mudanças de natureza química ou física (PORTER e NYE, 2003, p. 231) e (TAVARES e PRADO, 2010). Em 1783, com esse aparato – o calorímetro de gelo – puderam mostrar que a quantidade de gelo derretido pelo calor liberado por um *porquinho-da-índia* também era a mesma derretida pelo calor da combustão de carvão para a mesma quantidade de “ar fixo”² produzida. Desse resultado, Lavoisier conclui que a respiração seria um processo de combustão lenta.

Assim, a capacidade de medir quantidades de forma precisa permite: a comparação de sistemas de natureza aparentemente distintas; a especulação de teorias (ou hipóteses) quantitativas sobre os processos físicos e químicos que ocorrem nesses sistemas ou a que são submetidos.

Em verdade, a forma de proceder e a precisão da medição nesse caso foi até mais importante do que o que estava sendo medido, pois na época de Lavoisier o calor não era considerado uma forma de energia, mas sim uma espécie de elemento químico de massa desprezível: o *calórico*.

Motivação Prática

Qualidade é um conceito que evoca diferentes aspectos, em geral, não está relacionada diretamente ao conceito de número. Ainda como conceito, a qualidade era percebida desde os babilônios e egípcios (BORROR, 2009): relacionada à eficiência na produção de bens e à imposição de regras à sociedade, por exemplo. Em (PYZDEK, 2003), qualidade, no contexto dos processos produtivos e sociais, pode ser dividida em quatro Eras de (GARVIN, 1988), na qual uma nova abordagem incorpora a anterior: era da inspeção, era do controle estatístico da qualidade, era da garantia de qualidade e era do gerenciamento estratégico da qualidade.

Em (JURAN e GODFREY, 1999), sobre o ambiente empresarial do século XX, aborda-se a evolução em direção à qualidade total, que engloba as seguintes áreas-chaves: qualidade de produtos (foco nas especificações técnicas), qualidade de processos de produtos (foco nos diagramas de fluxo), qualidade de serviços (foco serviços de suporte ao produto), qualidade de processos de serviços (foco no custo dos serviços), planejamento de negócios (foco nos processos internos de trabalho).

Na engenharia, (TAGUCHI *et al.*, 2004, p. 7 e 173) veem que a qualidade diminui em função de perdas ou prejuízos impingidos à sociedade. Uma de suas preocupações

²Ar fixo: nome arcaico para o gás carbônico

é quantificar atributos qualitativos, em especial, aqueles com relação direta com a engenharia. Consideram que alguns prejuízos devido à falta de qualidade estão relacionados à variabilidade funcional de um produto. Um exemplo de modelo de quantificação desses prejuízos é a *função L de perda de Taguchi*:

$$L = k(y - m)^2 \quad (1.3.1)$$

onde y é a característica real do produto, m é o valor alvo e k é uma constante a ser definida posteriormente, para transformar essa variação em dinheiro (perda financeira).

Em (MITRA, 2016, p. 10 e p. 29), se observa a quantificação dos efeitos de ações de qualidade total no tocante aos processos no ambiente produtivo:

- o conceito de custo e valor de um *design* são uma função da qualidade – não que diz respeito ao custo, este é exponencialmente crescente com a qualidade; e ao valor do bem produzido, o aumento da qualidade passa a não agregar valor significativo após um determinado ponto.
- o custo pela busca da qualidade total é um mínimo global caracterizado pela razoabilidade dos benefícios frente aos custos de investimento em ações de garantia de qualidade e aos prejuízos inerentes a um grau de qualidade admitido pelo negócio (a curva custo da qualidade *versus* nível de qualidade).

Na escola russa, o termo qualimetria se refere à ciência que estuda as técnicas e os problemas de avaliação quantitativa para aferição da qualidade de objetos de qualquer natureza (o que não exclui objetos socioeconômicos, como a qualidade de vida) (AGARD, 1980, p. 343), (LOBANOV, 2013) e (AZGALDOV *et al.*, 2015, p. 67). No ocidente, uma metodologia para quantificação da qualidade de forma genérica é o *Six Sigma*, pois trabalha com múltiplos do desvio-padrão para alguma quantidade observada no escopo da metodologia “Definir, Mensurar, Analisar, Melhorar e Controlar” para processos de negócio e ambiciona manter a taxa de defeitos ou falhas na ordem de 3, 4 partes por milhão (JUGULUM, 2014, p. 25 e 43-54).

Onde dados são um ativo corporativo, a qualidade de dados tem sido considerada como a adequação dos dados para seu uso bem sucedido da implementação de processos corporativos de uma companhia (DAMA, 2009). Em (JUGULUM, 2014, p. 43-54), aborda a quantificação do impacto da qualidade de dados no ambiente produtivo. Nesses contextos, o conceito de qualidade de dados aparece relacionado a programas corporativos de monitoramento de processos para coibir as variações indesejadas: minimizar erros e defeitos.

Em (ISAAC e LYNNES, 2003), a questão da qualidade de dados e impossibilidade prática de se empregar pessoas para inspeções de vastos volumes de dados de

sensoriamento terrestre é a motivação para utilização da inteligência artificial como parte integrante de processos de avaliação da qualidade de dados. (ISAAC e LYNES, 2003) sugerem que os mesmos métodos utilizados para manipular e analisar os dados de sensoriamento sejam utilizados também para realizar a tarefa de avaliação da qualidade desses dados.

Os casos acima mostram que, apesar de o objeto qualidade não estar necessariamente relacionada a uma quantidade específica, a sua quantificação permite uma análise valendo-se de métodos matemáticos relativamente simples e bem conhecidos. Além disso, a quantificação suscita a possibilidade que se encontrar métodos mais simples para sua monetarização das iniciativas de melhoria de qualidade – e até mesmo dos impactos de se abrir mão dessas iniciativas. Finalmente, a quantificação da variabilidade parece ser o primeiro passo na avaliação da qualidade de processos, produtos e serviços.

1.4 Qualidade e Inteligência Artificial

Em textos básicos de física clássica, como também em discussões de sistemas caóticos, o conceito de solução qualitativa está ligado ao de solução aproximada – devido à impossibilidade prática, ou mesmo matemática, de se obter uma solução analítica (exata). Recorre-se, então, para mitigá-las, a métodos numéricos para a obtenção de uma “quase-solução” que preserva “aspectos ou características da solução”, (NUSSENZVEIG, 1994; SAVI, 2006). Tais características são muitas vezes mais relevantes, pois representam comportamentos e tendências, que estariam presentes em uma solução ideal, para resolver um problema de engenharia. Essas soluções aproximadas são tipicamente representadas por gráficos ou por listagem numérica, isto é, que correspondem a comportamentos, curvaturas e outras características da geometria da solução original. Entretanto avaliações qualitativas são uma forma de avaliar o comportamento soluções quantitativas, por exemplo: (JAYNES, 1962) faz uma primeira análise do comportamento da distribuição de Boltzmann meramente baseada na lógica e em propriedades *a priori* desse modelo — no contexto da mecânica estatística.

Por cerca de vinte séculos, as ideias qualitativas de Aristóteles sobre os movimentos dos corpos – como a de um corpo mais pesado cair mais depressa do que uma gota de chuva – foram dogmas (NUSSENZVEIG, 2002, p. 35). Hoje em dia física é sinônimo de modelos matemáticos e muitos cálculos como resultado do seu avanço. A física qualitativa é também um subcampo da inteligência artificial que investiga “como pensar racionalmente sobre sistemas físicos sem precisarmos mergulhar em equações detalhadas e simulações numéricas” (RUSSELL e NORVIG, 2010a). O que é uma postura mais pragmática na busca de modelos mais simples do ponto

de vista numérico: um *trade-off* entre uma quase precisão absoluta e resultados úteis em tempo menor. Como exemplo, FORBUS (1988) utiliza abstrações comuns a nós seres humanos como uma forma de se evitar enumerações de todos os cenários possíveis:

If we are to build a complete qualitative physics, one that covers the breadth and depth of our common sense knowledge of the physical world, we must discover and utilize common abstractions. Generating an ad-hoc model for each scenario is impractical and unreliable. Two such ontological abstractions, *devices* and *processes*, have been widely used in qualitative physics. (...)

Vinte anos mais tarde essa noção se manteve no campo da qualitative reasoning (BREDEWEG e STRUSS, 2004):

Qualitativeness means including only those distinctions in a behavior model that are essential for solving a particular task for a certain system. The goal is to obtain a finite representation that leads to coarse intuitive representation of models and behavior (for example, as finite relation and a state graph, respectively), and efficient algorithms.

(KLEER e BROWN, 1984) incorporam conceitos da termodinâmica para criar modelos na física qualitativa em inteligência artificial: utilização de aproximações de processos quase estáticos, por exemplo.

O próprio conceito de causalidade, segundo (KLEER e BROWN, 1984), não seria algo que a física moderna não oferecesse formalismo. Para (KLEER e BROWN, 1984), a causalidade decorre da percepção uma negociação (ou troca de) de informação entre entidade capazes, também, de processá-la. A casualidade mítica, descrita por eles, é entendida como a possibilidade de uma sequência de estados de um sistema físico entre estados de equilíbrio – tal sequência é chamada de trajetória. O tempo mítico é um intervalo de tempo não-convencional, pois as leis físicas que regem o sistema encontram-se, de alguma forma, violadas. Porém, após o término desse tempo mítico, devido ao intermédio de negociações e interações entre os componentes sistêmicos, tais leis não mais se encontrarão violadas: o sistema volta ao tempo convencional, portanto.

Na inteligência artificial há a questão filosófica conhecida como *problema de qualificação*, que versa sobre a impossibilidade prática de um agente racional considerar todos os detalhes de um problema para tomar uma decisão inteiramente governada pela razão – pois seu raciocínio estaria limitado ao tempo de vida; e à natureza combinatória de se examinar uma lista de precondições para tomada de uma ação. Um agente racional estaria, então, numa situação impraticável: considerar todos os

casos, mas observando as limitações temporais práticas impostas pela efetividade e o *timing* da ação (WEBER, 1989). Em (COSTA e SIMÕES, 2008, p. 412-441), encontram-se exemplos de paradigmas de abordagem qualitativa para se lidar com conhecimento incerto. Como procurar mitigar o problema de qualificação usando exceções e lógicas não-monotônicas. (KRAUSE e CLARK, 1993) abordam o conhecimento incerto usando tanto modelos qualitativos como quantitativos.

1.5 Por que Teoria da Informação?

Um dos primeiros trabalhos em que se busca relacionar os conceitos de qualidade e de entropia é em (SHARMA *et al.*, 1994). Nesse trabalho é feito um paralelo entre um sistema termodinâmico do qual se procura extrair o máximo de benefício com o mínimo de desperdício: transformando insumos de baixa qualidade em produtos com maior qualidade. Tais produtos poderiam ser o resultado (bens, serviços ou mesmo softwares) de qualquer processo. Em verdade, sobre o argumento apresentado em (SHARMA *et al.*, 1994), não se constatou a relação claramente entre entropia e qualidade, excetuando-se a sugestão em si. Entretanto, é mais fácil argumentar que o conceito de qualidade e de entropia são abstratos num primeiro momento.

A palavra qualidade, semântica e etimologicamente, pode evocar o conceito-pergunta “Qual-o-tipo?” (COOPER, 1997; CRESSWELL, 2010; DA CUNHA, 2010; IDA, 1914; OXFORD, 1973). O conceito de resposta eficiente, na teoria da informação, está relacionado ao número médio de perguntas binárias (cuja resposta é *sim* ou *não*) para referenciar o resultado de um experimento aleatório discreto de forma ótima (LEON-GARCIA, 2008, p. 202-212). A estrutura matemática que representa essa forma ótima é a árvore do código de Huffman (COVER e THOMAS, 1991, p. 97-101).

A observação acima sugere existir uma conexão entre relação pergunta-resposta e capacidade de eliminar, selecionar ou decidir eficientemente, mas sem figurar operações tipicamente quantitativas.

1.6 Dificuldades

Em LOGAN (2012), encontra-se a etimologia da palavra *entropia*: uma quantidade de energia (*en*) associada à transformação (*tropos*) irreversível de um sistema termodinâmico. Quando o conceito fora formulado, a entropia não tinha nenhuma conotação de incerteza ou de aleatoriedade. Além disso, essa mistura conceitual é um entrave para qualquer conhecimento científico sobre o papel da entropia – seja nos processos físicos ou como de natureza abstrata. (FERREIRA, 2005) esclarece:

Por deturpação da interpretação de entropia em Termodinâmica Estatística (que exploraremos adiante), é muito comum apresentar-se como seu significado físico a noção de «desordem» ou «aleatoriedade», com exemplos do mundo macroscópico: «a entropia de uma sala desarrumada é maior do que a entropia dessa sala arrumada», ou «a entropia de cartas baralhadas é maior do que a entropia de cartas por estrear», etc. Em análises ao nível microscópico, é também frequente a associação entre aumento de entropia e aumento na «desordem» das moléculas. Lê-se isto em textos reputados de ciências físico-químicas. (...)

(BEN-NAIM, 2007, p.p. 189 e 190), sobre “a fonte de mistério” que é a palavra entropia:

When you encounter, for the first time, a new word such as “entropy”, it conjures up an air of mystery, it has a strange and uneasy effect on you. If you are not a student of physics or chemistry, and by chance hear scientist talking about “entropy”, you certainly feel that this concept is beyond you and *a fortiori* so, when you hear the scientists themselves referring to entropy as a mystery.

Assim, a principal dificuldade está na compreensão do termo entropia nos campos do conhecimento em que figura e garantir coesão e coerência no seu uso ao longo da tese.

1.7 Referenciais Teóricos

Este trabalho tem como principais referenciais teóricos (HARTLEY, 1928), (SHANNON, 1948), (COVER e THOMAS, 1991), (KULLBACK, 1978), (GOLDMAN, 1976), (VITA, 1964), (SCHROEDER, 2000), (MCCARTHY, 1968), (SKINNER, 1960, 1976, 2014), (EKLUND, 2011), (DELLINGER, 2012), (GARDNER, 1995) e (HEINRICH *et al.*, 2007).

(SHANNON, 1948) é o trabalho original que deu origem à interpretação de entropia na mecânica estatística no sentido de incerteza e ao mesmo tempo associado ao conceito de informação (JAYNES, 1957).

(COVER e THOMAS, 1991) abordam estudos da teoria da informação: ciência da computação (complexidade Kolmogorov, correção de erros e compressão de dados), física (termodinâmica), matemática (teoria de probabilidades e estatística), filosofia da ciência (navalha de Occam) e economia (investimentos).

Em (GARDNER, 1995) fica evidenciada a importância do estudo dos incerteza, do ponto filosófico e cognitivo, pois foi a partir dela que Descartes fundou seu método filosófico de investigação (o idealismo):

“[A filosofia] foi cultivada durante muitos séculos pelas melhores mentes que já viveram, e, no entanto, nem uma única coisa pode ser nela encontrada que não seja objeto de disputa” (citado em Wilson 1969, p. 111). Para superar esta incerteza, Descartes desenvolveu um método de dúvida sistemática. Decidiu por em questão toda evidência sobre a qual não tivesse certeza, e verificou que tudo que lhe restou então foram os próprios estados de consciência, suas próprias dúvidas. Em sua capacidade de duvidar, e portanto de pensar, Descartes discerniu um fundamento seguro para edificar uma nova filosofia.

1.8 Trabalhos Desenvolvidos

Anterior a este trabalho foram apresentados trabalhos em simpósios nas áreas de *sistemas de informação, pesquisa operacional e dados para pesquisa na área da saúde* (PINTO *et al.*, 2013a; PINTO e CARVALHO, 2014; PINTO *et al.*, 2013b; PINTO e DE CARVALHO, 2014; PINTO e CARVALHO, 2015).

1.9 Qualidade de Dados

No campo da qualidade de dados, (HEINRICH *et al.*, 2007, 2009) apresentam uma função simples de medida de similaridade entre duas palavras escritas. Essa função fornece um grau de similaridade entre elas parcialmente consistente com o julgamento qualitativo de uma inspeção visual feita por seres humanos. Essa função de similaridade utiliza a distância Levenshtein (desenvolvida no contexto de correção de informações em canais de transmissão binários que admitem, além da troca de símbolos, a deleção e a inserção de símbolos) (LEVENSHTTEIN, 1966). Um aprimoramento do trabalho de (HEINRICH *et al.*, 2007, 2009) será utilizado como um *toy problem*³ para ilustrar uma resposta para o problema desta tese.

1.10 Definições Básicas

Agora, algumas notações serão estabelecidas – elas serão utilizadas ao longo deste trabalho. \sum_i pode denotar $\sum_{i=1}^n$ ou $\sum_{i=1}^{\infty}$ – será assumida a existência do somatório apenas. Analogamente, \prod_i pode denotar $\prod_{i=1}^n$ ou $\prod_{i=1}^{\infty}$ – será assumida a existência do produtório apenas.

³Toy Problem: ver (RUSSELL e NORVIG, 2010a, p. 69)

Definição 1.10.1 (Conjuntos Numéricos). \mathbb{N} , \mathbb{Z} e \mathbb{R} denotam respectivamente os conjuntos dos números naturais (incluindo o zero), inteiros e reais. \mathbb{N}_* , \mathbb{Z}_* e \mathbb{R}_* denotam respectivamente os conjuntos dos números naturais, inteiros e reais – excluindo o zero. \mathbb{N}_* e \mathbb{R}_*^+ denotam respectivamente os conjuntos dos números inteiros e reais que são não-negativos.

1.11 Orientações para Leitura

O capítulo 2 apresenta uma demarcação e articulação entre os conceitos de informação, incerteza e entropia. Ele tem como objetivo ambientar o leitor quanto à natureza interdisciplinar desta pesquisa.

O capítulo 3 apresenta a construção do conceito de quantidade de informação a partir de Nyquist e Hartley até Shannon. O que se faz necessário para contextualizar e dirimir dúvidas sobre o conceito de entropia para uniformizar sua interpretação ao longo deste trabalho.

Observa-se que os capítulos 2 e 3 possuem carácter introdutório, didático e informativo.

O capítulo 4 apresenta um modelo no qual se procura quantificar a qualidade percebida associada a processos estatísticos. Esse modelo foi formulado por (MONTGOMERY, 2009, 2012) (expert com conhecimento na área de controle estatístico da qualidade). Além disso, são caracterizadas algumas anomalias nesse modelo. Ao longo desta tese, será apresentado um modelo que também quantifica o conceito de qualidade, mas que procura contornar tais anomalias.

O capítulo 5 apresenta um modelo de decisão inspirado no equilíbrio químico. No lugar de concentrações de substâncias químicas, são utilizadas medidas de raridade e de surpresa para estabelecimento do um grau de equilíbrio para posterior emissão de julgamento. São mostrados dois modelos de decisão: um que pondera raridades médias e outro, surpresas médias. Também é argumentado que os dois modelos estão relacionados. Entre outras grandezas, são definidos os conceitos de momento qualitativo e momento quantitativo necessários em um julgamento.

O capítulo 6 apresenta o conceito de momento qualitativo que é explicado e aplicado ao matching aproximado de strings usando o paradigma do perceptron. Ilustra como o conceito de momento qualitativo pode ser utilizado num processo de comparação de strings em um sistema decisório composto por um único neurônio perceptron. Apresenta um modelo de mensuração de dúvida baseado em ponderação de incertezas.

O capítulo 7 estende o capítulo 5 para comportar o conceito de entropia diferencial. Apresenta uma interpretação de que se pode realizar medidas de diferenças exatas sem que seja possível estabelecer referenciais absolutos para processos envol-

vendo incerteza.

O capítulo 8 apresenta o modelo de qualidade baseado na interpretação de que a incerteza pode ser uma quantidade que pode ser parametrizada – e interpretada como um campo conservativo. Ilustra o fato de que se pode realizar medidas de qualidade para processos modelados como uma função de distribuição Gaussiana de probabilidades — sendo um dos resultados mais importantes desse trabalho.

O capítulo 9 apresenta as conclusões e trabalhos futuros a respeito dos principais resultados aqui obtidos ao longo desta tese.

Capítulo 2

Informação, Incerteza e Entropia

19 D’us tinha formado do solo todos os animais do campo e todo pássaro do céu. Ele (agora os) trouxe para o homem, para que ele visse como os nomearia. 20 O que quer que o homem chamava a cada coisa viva, permaneceu seu nome. O homem nomeou todo animal de gado e pássaro do céu, assim como todos os animais do campo.

Cap. 8 Gênesis, Torá

Neither is it a small matter these Cypher-Characters have, and may performe: For by this Art a way is opened, whereby a man may expresse and signifie the intentions of his minde, at any distance of place, by objects which may be presented to the eye, and accommodated to the eare: provided those objects be capable of a twofold difference onely; as by Bells, by Trumpets, by Lights and Torches, by the report of Muskets, and any instruments of like nature. But to pursue our enterprise, when you addresse your selfe to write, resolve your inward-infolded Letter into this Bi-literarie Alphabet.

Francis Bacon, Advancement of Learning

2.1 Introdução

Em alguns contextos, as palavras dados e informação são consideradas sinônimos ou intercambiáveis. Correntes de pesquisadores na ciência da computação procuram hierarquizar o conceito de dados e informação, sendo informação o estágio superior ao dos dados, como em (ACKOFF, 1989) – a *pirâmide Dados-Informação-Conhecimento-Sabedoria*. (REZENDE, 2003b, p. 6), baseado nessa pirâmide, considera que “A informação é o dado analisado e contextualizado”.

(KNUTH, 1996a) abordou a diferença conceitual entre *dado* e *informação*. Em breve síntese, para (KNUTH, 1996a), a informação é algo mais intangível ou abstrato

do que o dado. Ou melhor, o dado seria um valor de representação, em uma linguagem formal, de uma informação em especial. Tal representação pode ser aproximada ou exata dessa informação. Essa distinção é reafirmada por (KNUTH, 1996b):

A particular representation of an algorithm is called program, just as we use the word “data” to stand for a particular representation of “information”.

Todavia, a palavra informação ganhou um significado especial ao ser considerada como parte integrante da física moderna, como em (SUSSKIND e LINDESAY, 2005): *An Introduction to Black Holes, Information and the String Theory Revolution: The Holographic Universe*. Curiosamente isso se deve ao trabalho de Claude Shannon na área de telecomunicações: *A Mathematical Theory of Communication*, no qual foi também cunhado o acrônimo *bit* (binary digit).

Neste capítulo procura-se ilustrar sucintamente de que maneira os conceitos de informação e de incerteza se tornaram primeiramente relacionados ao de entropia – este um conceito originalmente da física: a mais fundamental das ciências.

2.2 Informação

(FLORIDI, 2010; LOGAN, 2012) procuram difundir o conceito de informação — principalmente para o público de humanidades. Abordam tangencialmente a relação entre informação e entropia para o público leigo. Entretanto, insistem na questão da multiplicidade de interpretações multifacetadas sobre o que se pode pensar e entender sobre informação – como se destaca “There is no consensus yet on the definition of semantic information” (FLORIDI, 2005).

Na física, o princípio de conservação da informação é considerado um fato científico, tanto na mecânica clássica quanto na mecânica quântica, segundo (SUSSKIND e LINDESAY, 2005, p. 69). Isso indica que, de certa forma, a informação tem um papel preciso e importante nessa ciência.

(LANDAUER, 1996) defende que a informação faz parte, não só da realidade cotidiana, mas da física:

Information is not a disembodied abstract entity; it is always tied to a physical representation. It is represented by engraving on a stone tablet, a spin, a charge, a hole in a punched card, a mark on paper, or some other equivalent. This ties the handling of information to all the possibilities and restrictions of our real physical world, its laws of physics and its storehouse of available parts.

Ainda na física, (JONES e JAKSCH, 2012) não só fornecem uma definição de informação, esta como partícipe da computação, mas também explicitam as suas propriedades:

Classical information processing is performed using bits, which are just two state systems, with the two states called 0 and 1. By grouping bits together we can represent pieces of information, and by manipulating these bits we can perform arbitrary computations. The corresponding basic element in quantum information is the quantum bit, or qubit. This is simply a quantum system with two orthogonal bases states, which we shall call $|0\rangle$ and $|1\rangle$.

Um *bit* não é um padrão, nem um símbolo, mas sim um sistema. A utilização de *1* ou *0* para representar um estado de um *bit* é meramente histórica – pois o termo surgiu na área de telecomunicações e computação como a representação de um sistema capaz de armazenar um estado – ligado ou desligado, como pode ser verificado em (HARTLEY, 1928; MACKENZIE, 1979; NYQUIST, 1924; PIERCE, 1973; SHANNON, 1948). Os estados que um *bit* pode assumir são duas entidades distintas, por exemplo: *A* ou Ω . Portanto, um conjunto de estados de um *bit* podem ser qualquer conjunto com exatamente dois elementos. Assim, segue a definição de *bit*.

Definição 2.2.1 (Bit). *Um bit clássico, ou bit, é um único sistema que assume exatamente um de dois estados possíveis e mutuamente exclusivos. Um sistema de n bits, ou simplesmente n bits, é um único sistema composto por uma sequência de n subsistemas distintos, onde cada subsistema é um bit. Assim, n bits podem assumir exatamente 2^n estados distintos.*

Por exemplo, um sistema composto por 8 *bits* pode ter 2^8 estados distintos. Trocando-se “estados distintos” por informações, um sistema de 8 *bits* pode representar 2^8 informações distintas.

Um padrão de n *bits* é listagem de n estados possíveis. Três exemplos de padrões de 8 *bits* são 01001010, 01000011 ou 10100011. Dois exemplos de padrão de 1 *bit* são **c** e **ç**. Nesse último, a menor diferença (o sinal diacrítico) os torna absolutamente diferentes. Em (MACKENZIE, 1979, p. 12-14), essa diferença conceitual entre *bit* e padrão de *bit* é muito bem ressaltada e exemplificada. Finalmente, não importa se dois padrões de 1 milhão de *bits* diferem em um único símbolo – isso é suficiente para torná-los incomparável e absolutamente distintos. No lugar de padrão de n *bits*, pode-se dizer configuração de n *bits* ou um estado de n *bits*.

Pode-se dizer também que um sistema que seja constituído por n *bits* pode ser uma medida da incerteza ou da ignorância sobre esse sistema.

Porém qual é a relação dessa definição de informação com a utilizada na teoria da informação? Essa relação é absoluta e deve-se aos trabalhos de (HARTLEY, 1928; NYQUIST, 1924; SHANNON, 1948), nos quais eles consideram que cada mensagem distinta seria uma informação distinta – mesmo que a diferença fosse mínima (por exemplo, uma diferença em uma única letra ou pontuação) . Para eles a perspectiva da engenharia era o que importava, como sintetiza (SHANNON, 1948):

Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages.

Os seja, desprezando-se o significado de uma mensagem – resta nela somente a propriedade de se diferenciar de outra mensagem.

Uma interpretação física sobre o estado de um sistema físico dado por (SUSSKIND, 2013):

A state always means the same thing. It means as much as you can possibly know about the system if you were infinitely powerful. (...) There may be limits that have to do with the fundamental rules of physics, such as quantum mechanics, which tells you [that] you can't know as much as you thought you might have liked to know, such as the position and velocity of every particle. But still, a state means as much as can be known about a system by, let's call it, an infinitely powerful observer.

Segundo (SUSSKIND, 2013), a informação está ligada à capacidade de se perceber distinções.

Um exemplo na computação, dessa interpretação de estado, pode ser encontrado em (HOPCROFT e ULLMAN, 1979, p. 14):

The computer itself can be viewed as a finite state system, although doing so turns out not to be as useful as one would like. Theoretically, the state of the central processor, main memory, and auxiliary storage at any time is one of a very large but finite number of states. We are assuming of course that there is some fixed number of disks, drums, tapes, and so on available for use, and that one cannot extend the memory indefinitely. View a computer as a finite state system, however, is not satisfying mathematically or realistically. It places an artificial limited on the memory capacity, thereby falling to capture the real essence of computation. To properly capture the notion of computation we need a potentially infinite memory, though each computer installation is finite. (...)

Observação 2.2.1. Em (SHANNON, 1948), a palavra dados (*data*) aparece somente três vezes: no sentido dos símbolos discretos que eram transmitidos em canal de comunicação – mesmo que os dados que chegassem ao receptor não fossem os mesmos enviados pelo emissor (devido a erros ou ruídos no canal de comunicação). Dados também podem participar do processo de preservar a informação (dados de correção de erros), mas sem necessariamente fazer parte da representação da informação em si.

2.3 Incerteza

O que difere o trabalho de (SHANNON, 1948) dos de (HARTLEY, 1928) e (NYQUIST, 1924) é que, para efeitos de engenharia, deve-se considerar as mensagens prováveis (ou típicas) no lugar de considerar as mensagens possíveis. Para isso a incerteza, no lugar de ser uma função das possibilidades de mensagens, passa a ser uma função da distribuição de probabilidades dos símbolos das mensagens. Por exemplo, há mais incerteza em jogar uma moeda não-viciada do que em jogar uma moeda com probabilidade 10% de resultar em cara (90% em coroa). A incerteza em jogar uma moeda com a probabilidade 0% de resultar cara e 100% de coroa é, obviamente, zero. Isso sugere que a incerteza não depende das informações possíveis (dois resultados possíveis), mas depende da distribuição de probabilidades dos resultados possíveis. Pode-se dizer, entretanto, que a incerteza do lançamento da moeda não-viciada é análoga à incerteza de duas possibilidades e que a incerteza da moeda completamente viciada (com zero incerteza) é análoga à incerteza de uma única possibilidade.

Se a incerteza de b possibilidades for um padrão de unidade de incerteza e se Ω denotar o número de possibilidades, então uma forma razoável de se medir a incerteza associada a Ω possibilidades é calcular $\log_b \Omega$. Observa-se que para qualquer b possibilidades utilizada como padrão de incerteza, excetuando-se a unidade, tem-se que: a incerteza associada a uma única possibilidade é zero e a incerteza é estritamente crescente em Ω .

Exemplo 2.3.1. Caso a incerteza de duas possibilidades seja o padrão de unidade de incerteza então a incerteza associada a Ω possibilidades é $\log_2 \Omega$. Caso $\Omega = 2^{100}$, então a incerteza é $\log_2 2^{100}$. O mesmo que $100 \log_2 2$ (ou 100).

Abordar-se-á, no capítulo 3, a generalização de (SHANNON, 1948) que permite medir a incerteza associada a uma distribuição discreta de probabilidade. Intuitivamente, a incerteza associada ao se jogar uma moeda com probabilidade 10% de resultar em cara (90% em coroa) está entre 0 e 1, para o caso $b = 2$.

(SHANNON, 1948) tinha ciência das conexões da sua teoria com a mecânica estatística: ele associa sua fórmula de incerteza explicitamente à entropia da mecânica

estatística – exemplificando como fonte “*Principles of Statistical Mechanics*” (TOLMAN, 1938), mas sem estabelecer uma relação física. (JAYNES, 1957) atribui ao “A Mathematical Theory of Communication” de (SHANNON, 1948) a interpretação adequada – e mais fundamental – de entropia como incerteza para mecânica estatística.

2.4 Entropia

A termodinâmica é a área da física “que lida com fenômenos associados aos conceitos de temperatura e calor”. Historicamente, as leis da termodinâmica são leis empíricas, isto é, de natureza fenomenológica (NUSSENZVEIG, 1994, p. 250). Um sistema termodinâmico, ou simplesmente sistema, consiste numa quantidade de matéria dentro de um recipiente, cujas paredes podem ser fixas ou móveis. Um sistema fechado é aquele que não troca matéria nem energia com seu exterior. Um sistema termodinâmico pode ser composto por outros sistemas termodinâmicos, ou subsistemas.

A primeira lei da termodinâmica essencialmente versa sobre a conservação de energia do sistema termodinâmico, formulada por Rudolf Clausius em 1875. Sua importância é reconhecer o calor como uma forma de energia (NUSSENZVEIG, 1994, p. 329).

A segunda lei da termodinâmica é a lei de evolução do sistema – formulada também por Clausius, mas dez anos antes da primeira. Essa lei estabelece que existe uma grandeza, denotada por S , que aumenta em um sistema termodinâmico fechado quando esse sofre uma transformação irreversível; ou permanece a mesma quando a transformação é reversível. Se as transformações termodinâmicas fossem sempre reversíveis, em um sistema fechado, essa quantidade teria comportamento análogo ao da energia – no sentido de ser uma quantidade que se conserva. O nome entropia é um neologismo de Clausius para essa quantidade, pois o prefixo grego *en-* sugere uma quantidade que se parece com a energia; e o sufixo *-tropa* que significa transformação: uma quantidade relacionada à transformação do sistema, portanto. (EVANGELISTA, 2014, p. 246) resume bem o espírito do significado: “(...) o que importa, realmente, são duas quantidades: a energia e a entropia!”

A mecânica estatística foi uma tentativa de entender as leis da termodinâmica clássica tendo como base as leis da mecânica clássica. O maior entrave para essa abordagem seria a caracterização de um sistema – composto por um altíssimo número de constituintes (moléculas), mas sem precisar considerar a enormidade de informações que caracterizassem esse sistema mecânico de cada constituinte (posição, momento e energia mecânica de cada molécula de um gás, por exemplo).

A descoberta da existência do movimento browniano na biologia – a observação

experimental da existência de variações perceptíveis e aleatórias de um grão de pólen em meio aquoso, em 1827 – sugeria a irregular distribuição das forças de contato entre o pólen e esse meio (apesar de evidentemente homogêneo) (NUSSENZVEIG, 1994, p. 459-461) – sendo uma evidência experimental da existência de átomos e moléculas. Entretanto, só recebeu um tratamento teórico satisfatório em 1905 (EVANGELISTA, 2014, p. 332).

Ludwig Boltzmann e James Maxwell foram pioneiros em tentar explicar porque um sistema modelado de forma reversível parecia ter um comportamento irreversível, mas usando o paradigma probabilístico. Posteriormente, Josiah Gibbs generalizou o trabalho de Boltzmann, mas em ambos os pressupostos sempre convergiam, de alguma forma, a uma componente aleatória intrínseca ao sistema termodinâmico. Segundo (JAYNES, 1965), o modelo proposto por Boltzmann só era aplicável para gases rarefeitos, no qual as interações entre as moléculas eram desprezíveis; enquanto o modelo proposto posteriormente por Gibbs levava isso em consideração – apesar de mais abstrato (EVANGELISTA, 2014) (o modelo de Gibbs era mais geral e o de Boltzmann um caso particular do último).

Definição 2.4.1 (Mecânica Estatística). *Um formalismo que se destina a explicar as propriedades físicas da matéria em grandes quantidades com base no comportamento dinâmico dos seus constituintes¹ microscópicos (PATHRIA e BEALE, 1996, p. 1). A explicação não recai sobre o constituinte dos sistemas, ou de variação de energia ou de sua posição, mas em função da distribuição de probabilidades de propriedades físicas (EVANGELISTA, 2014, p. 307).*

(EVANGELISTA, 2014) observa que na mecânica estatística o foco não recai sobre os constituintes de um corpo ou em variações de energia e posição, mas sim num número infinito (ensemble) de corpos (sistemas). Para lidar com isso, algumas propriedades macroscópicas são as mesmas para todos os corpos (como a temperatura ou a energia total), ao passo que algumas propriedades microscópicas podem ser totalmente diferentes como a distribuição da posição de corpos microscópicos que o compõe (átomos ou moléculas, por exemplo) e a distribuição de energia entre eles. O que é radicalmente diferente da teoria cinética, na qual choques, trocas de energia entre moléculas são modeladas ou consideradas.

Na mecânica estatística a modelagem de Gibbs, no lugar de considerar uma evolução ao longo do tempo, considerava um único instante de tempo, mas hipoteticamente suficiente, para que as propriedades macroscópicas fossem médias das propriedades macroscópicas de ensembles (ponderada pela probabilidade de cada ensemble). Assim ensembles seriam uma espécie de “cópias mentais” de um dado

¹Da ordem de $6,022 \times 10^{23}$ constituintes, como um 1 mol de um gás monoatômico, que é o número de Avogrado.

sistema – que caracterizam o mesmo macroestado que o sistema original, mas são, naturalmente, os microestados possíveis. Gibbs, ao contrário de Boltzmann, não ignorava os ensembles que não corresponderem à propriedade macroscópica desejada. A abordagem de Boltzmann considerava o sistema de moléculas como uma superposição de sistemas monomoleculares no qual cada um deles teria a mesma propriedade macroscópica. Depois as contribuições de cada um desses modelos eram somadas com o intuito de descrever o sistema na sua integralidade. (JAYNES, 1965) observa, entretanto, que tal modelagem de Boltzmann funciona muito bem para descrever gases com fraca interação intermolecular.

A dinâmica entre os microestados na mecânica estatística tem como pressupostos:

- O sistema muda continuamente de um microestado para outro – com o resultado de que, durante um período razoável de tempo, as estatísticas sobre esses microestados corresponderão a medições diretas ao sistema sobre alguma propriedade do sistema observável (propriedade do macroestado). Tais estatísticas, mais especificamente, corresponderão às médias de propriedades associadas aos microestados pelos quais o sistema passou durante a medição. (PATHRIA e BEALE, 1996, p. 36);
- A hipótese de ergodicidade² que diz que nenhum microestado ficará inacessível e terá alguma probabilidade constante associada a ele (diferente de zero); e que a média estatística de uma propriedade coincide com a média ao longo do tempo por amostragem (JANSEN, 2008, p. 13)(HAYKIN, 2009, p. 586).

Em (EVANGELISTA, 2014, p. 307), Boltzmann chegou à seguinte conclusão no desenvolvimento de suas pesquisas teóricas em 1871:

A grande irregularidade do movimento térmico e a pluralidade das forças que agem sobre um corpo tornam possível que seus átomos, em vista do movimento que chamamos de calor, acesse todas as posições e velocidades que são compatíveis com o princípio da conservação de energia.

Essa observação pode ser considerada o “motor” da aleatoriedade: o que se entende por hipótese ergódica (que até hoje não foi provada, mas utilizada como princípio (JANSEN, 2008, p. 13)).

Os conceitos apresentados a seguir fazem parte do desenvolvimento dos trabalhos de Boltzmann e Gibbs na mecânica estatística e são base para a compreensão de modelos nessa área (EVANGELISTA, 2014; JANSEN, 2008).

²A hipótese de ergodicidade implica a hipótese de estacionariedade, que assume que as estatísticas não dependem do início do tempo da medição

Definição 2.4.2 (Microestado, Macroestado e Multiplicidade). *Um microestado de um sistema é uma configuração específica e mutuamente exclusiva que esse sistema pode assumir. Um macroestado de um sistema é determinado em função do microestado, mas a recíproca não é verdadeira. O número de microestados que correspondem a um macroestado específico é a multiplicidade do macroestado. (SCHROEDER, 2000, p. 50)*

Importante dizer que um microestado é por si só capaz de diferenciar-se de outros microestados. Idem para os macroestados. Entretanto, conhecendo um macroestado pode-se saber o conjunto de microestados, mas estes são indistinguíveis caso se conheça somente o seu macroestado.

Exemplo 2.4.1. O lançamento consecutivo de uma moeda 5 vezes, se obtém uma sequência (portanto a ordem é importante) de um dos possíveis resultados. Cada sequência possível corresponde a um microestado, por exemplo $(Cara, Coroa, Coroa, Cara, Coroa)$, então há 32 (ou 2^5). A soma de todas as ocorrências de coroas em um microestado corresponde, por exemplo, a um macroestado. O número de coroas em 5 lançamentos determina um espaço de seis macroestados: 0,1,2,3,4 e 5. O microestado acima está associado ao macroestado 3 coroas. O microestado $(Coroa, Cara, Coroa, Coroa, Cara)$ é distinto do microestado $(Cara, Coroa, Coroa, Cara, Coroa)$, porém ambos estão relacionados ao macroestado identificado por 3 coroas. A multiplicidade do macroestado 3 coroas é 10.

Exemplo 2.4.2. Seja $N \in \mathbb{N}^+$, cada solução inteira não-negativa para a equação $\sum_{i=1}^n X_i = N$ é um microestado para o macroestado N . A multiplicidade do macroestado N é $\frac{(N+n-1)!}{N! \cdot n!}$.

Como observa (SCHROEDER, 2000, p. 50): conhecendo-se o microestado se conhece o macroestado. Entretanto, o contrário não é verdadeiro.

Na seção anterior, a informação sobre um sistema é uma descrição completa sobre o estado desse sistema e que seja relevante para caracterizá-lo. Dessa descrição pode-se derivar propriedades desse sistema, por exemplo: o número de ocorrências de um determinado símbolo pode dizer, em um contexto, a energia desse sistema. A incerteza surge quando se conhecendo somente algumas propriedades de um sistema pergunta-se quantos estados, ou configurações, são compatíveis com essas propriedades.

Exemplo 2.4.3. Sobre uma quantidade de gás ideal monoatômico, um sistema termodinâmico, cujo o volume, temperatura e a massa constituam propriedades desse sistema, pode-se ter uma estimativa do número de configurações possíveis – por análise combinatória e algumas suposições – que sejam coerentes com tais propri-

idades. Supondo que esse número de configurações coerentes seja W , a incerteza desse gás em bits será $\log_2 W$ (SCHROEDER, 2000, p. 68-72 e 77-79).

Uma grandeza derivada em modelos da mecânica estatística, denotada por H , era uma função exclusiva da distribuição de probabilidades e se comportava com a entropia em termodinâmica, denotada por S . O exemplo 2.4.3 ilustra porque a interpretação da entropia como incerteza se deu. Em especial, tem-se a icônica fórmula (que é uma lei da física) (WRIGHT, 1970):

$$S = k \log W$$

Onde k é uma constante física, \log é logaritmo natural, S é entropia termodinâmica absoluta de um sistema físico.

(MÜLLER e WEISS, 2005, p. 32) informam que Ludwig Boltzmann³ não escreveu a fórmula acima, mas sim o físico Max Planck, apesar de ela estar gravada acima do túmulo de Boltzmann. Como a mudança de base de um logaritmo pode ser feita pela multiplicação de uma constante, é imediato interpretá-la como uma medida de incerteza.

É importante frisar que o conceito de entropia, na termodinâmica clássica, não é um conceito probabilístico nem derivado de probabilidades nem de estatísticas. Originalmente, a variação de entropia é uma espécie de indicador da ocorrência de um processo irreversível para o universo (indica que calor fluiu necessariamente do corpo de maior temperatura para o de menor temperatura, por exemplo).

(FEYNMAN *et al.*, 2005, cap. 44 e 46) explicam que o aumento de entropia de um sistema termodinâmico clássico pode ocorrer pelo aumento de volume de um gás – em um processo adiabático de expansão livre. Suscintamente, no caso de se tentar reverter tal expansão, o calor seria necessariamente liberado para o universo para retornar a um sistema idêntico. (FEYNMAN *et al.*, 2005, cap. 44) argumentam que para o processo de expansão livre, a variação de entropia por molécula pode ser calculada exclusivamente como uma função das razões entre o volume final e inicial do gás. Assim, para uma expansão livre para esse gás, somente o volume extra explicaria o aumento de entropia – sem haver trabalho e sem haver calor envolvidos. Eles fornecem, então, um argumento de graus de liberdade extra (em volume) para cada molécula.

Sobre as unidades de medida na física moderna, (SUSSKIND e LINDESAY, 2005, p. 74):

Entropy and information are naturally measured in “bits”. A bit is the

³Planck que deu a interpretação acabada $S = k \log W$, que é compreensível supondo-se modelos discretos – como ocorrem em (SCHROEDER, 2000), onde se utiliza modelos discretos (mecânica quântica) para descrever a entropia de modelos clássicos – como exemplo a derivação da equação Sackur-Tetrode de entropia para gases ideais.

entropy of a two state system if nothing is known[2]. The numerical value of a bit is $\log 2$. (...)

2.5 Conclusões

O presente capítulo realizou uma discussão articulando a relação existente entre entropia, incerteza e informação nos campos da física e da teoria da informação.

As principais conclusões são: que a informação é aquilo que é utilizado para descrever estados possíveis de sistemas físicos – de forma a se distinguirem um do outro – e permitir a contagem de tais estados. O conceito de incerteza foi abordado: o conceito de *bit* é uma medida de incerteza (pois representa, intrinsecamente, um sistema que pode assumir dois estados).

Muito sucintamente, incerteza é uma grandeza que mensura a possibilidade em potencial, por exemplo: o sistema utilizado para representar um dígito decimal tem mais incerteza associada do que o conceito de dígito binário (no primeiro há dez possibilidades e no segundo há duas).

Por fim, não será objeto de estudo o conceito de informação quântica nesta tese, mas somente a informação clássica. Apesar de os dois conceitos de informação estarem relacionados a sistemas que assumem estados.

Capítulo 3

Da Informação à Incerteza

What do we mean by “understanding” something? We can imagine that this complicated array of moving things which constitutes “the world” is something like a great chess game being played by the gods, and we are observers of the game. We do not know what the rules of the game are; all we are allowed to do is to watch the playing. Of course, if we watch long enough, we may eventually catch on to a few of the rules. The rules of the game are what we mean by fundamental physics. Even if we knew every rule, however, we might not be able to understand why a particular move is made in the game, merely because it is too complicated and our minds are limited. (...) Aside from not knowing all of the rules, what we really can explain in terms of those rules is very limited, because almost all situations are so enormously complicated that we cannot follow the plays of the game using the rules, much less tell what is going to happen next. We must, therefore, limit ourselves to the more basic question of the rules of the game. If we know the rules, we consider that we “understand” the world.

Richard Philips Feynman, The Feynman Lectures on Physics

3.1 Introdução

De certa forma, os estudos de teoria de informação gravitam em torno do conceito de incerteza de variáveis aleatórias. Nesse campo, abordagens axiomáticas para incerteza podem ser vista em (COVER e THOMAS, 1991; KLIR, 2006; KULLBACK, 1978; NATH e KAUR, 1981). (REZA, 1961; SAYOOD, 2005), por outro lado, começam o estudo a partir do conceito de surpresa (ou auto-informação). Em (REZA, 1961; SAYOOD, 2005), a incerteza é a média da “surpresa” dos eventos de um experimento aleatório.

Para uma melhor compreensão sobre o conceito de informação e incerteza – e uniformização conceitual – neste capítulo faz-se uma ponte entre ao trabalho inicial de (HARTLEY, 1928), que não modelou os sistemas de comunicação como

probabilísticos, e o trabalho de (SHANNON, 1948), que levou a aleatoriedade em consideração. A abordagem aqui consiste em reconstruir a argumentação teórica de (SHANNON, 1948) e de (HARTLEY, 1928), mas ao estilo do último. Ambos procuravam quantificar quanta informação estaria sendo transmitida em um processo de comunicação. Essa abordagem procura evidenciar como a informação e a incerteza se tornaram irmãs antípodas na teoria de informação. As unidades de informação *bit*, *nats* e *Harts* são apresentadas. Por fim o conceito de incerteza é abordado – a razão dessa preocupação é que a mesma função matemática é utilizada para realizar medidas de informação e de incerteza (SHANNON, 1948).

3.2 Breve Histórico

(VERDÚ, 1998), no clássico “Fifty Years of Shannon Theory”, faz um breve histórico das pesquisas sobre o conceito de transmissão de informação no âmbito da engenharia e telecomunicações. Esse trabalho, o pesquisador destaca que (NYQUIST, 1924) e (HARTLEY, 1928) estudaram, inicial e matematicamente, os fatores que determinam a “velocidade máxima de transmissão da inteligência¹” em sistemas de sua época:

- um limite finito para a taxa de transmissão de “inteligência”;
- o ruído que reduzia o alcance desse limite.

Tanto (NYQUIST, 1924) quanto (HARTLEY, 1928) desenvolviam o pensamento de que aquilo que era transmitido em si não teria dependência tecnológica nem psicológica. Entretanto, é em (HARTLEY, 1928) que o conceito de transmissão de inteligência é entendido como transmissão de informação. (VERDÚ, 1998) ressalta que o conceito de o que é informação é bem estabelecido na literatura.

Definição 3.2.1 (Informação). *Informação é o resultado de uma seleção entre um número finito de possibilidades. (VERDÚ, 1998, **apud** (HARTLEY, 1928)).*

Em (KULLBACK, 1978)[p. 1-3], a evolução do termo informação é descrito, mas se reconhece que ele entrou no universo da estatística devido a (HARTLEY, 1928).

Pode-se dizer que um sistema está recebendo mais informação que outro porque aquele consegue efetivamente selecionar mais mensagens – para um mesmo período de tempo em consideração. Foi (SHANNON, 1948) que caracterizou o problema fundamental da comunicação:

¹Inteligência, para (NYQUIST, 1924), era a capacidade que um caractere tinha de se diferenciar de outros em um alfabeto específico: quanto maior o alfabeto, então maior a inteligência de cada caractere. Se um alfabeto fosse composto por 27 caracteres e outro por 32, um caractere transmitido do último teria mais inteligência do que o primeiro.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.

A razão do trabalho realizado por (SHANNON, 1948) é que nos anos de 1940 foi a necessidade de se incorporar na teoria proposta por Hartley *tradeoffs* entre taxa de transmissão, *bandwidth*, *signal to noise ratio* e confiabilidade (VARSHNEY, 2004; VERDÚ, 1998).

3.3 Informação de Hartley

Em *Transmission of Information*, (HARTLEY, 1928) se dedicou ao problema de **quantificar a informação** transmitida entre sistema de comunicação. (HARTLEY, 1928) desenvolve suas ideias em dois passos: no primeiro analisa a questão envolvendo um alfabeto utilizado para a transmissão, que será traduzido no alfabeto do interessado pela mensagem. Depois a analisa entre dois sistemas com alfabetos distintos (um produz a mensagem e o outro recebe a mensagem para o interessado).

Recepção com Alfabetos Primário e Secundário

Inicialmente, (HARTLEY, 1928) modela a recepção da informação por um sistema de comunicação em dois passos: no primeiro alfabeto primário (zeros e uns, por exemplo) para a transmissão (envio e recepção) da mensagem e outro secundário (letras, espaços e pontuações, por exemplo) para a representação da mensagem (inteligível pelo interessado da mensagem).

Por hipótese, no alfabeto primário, há n_p símbolos nativos no qual N_p símbolos agrupados representariam um único símbolo no alfabeto secundário de n_s símbolos, assim $n_s = n_p^{N_p}$. Se uma mensagem escrita o alfabeto secundário fosse composta por N_s símbolos, haveria $n_s^{N_s}$ representações possíveis de mensagens distintas de comprimento N_s .

Pode-se então escrever:

$$n_s^{N_s} = (n_p^{N_p})^{N_s} = n_p^{N_p \cdot N_s} \quad (3.3.1)$$

Seja $N = N_p \cdot N_s$, (HARTLEY, 1928) argumenta que N é o número de seleções de símbolos primários que seria necessário para produzir a mesma sequência – caso não houvesse um mecanismo para agrupar os símbolos primários em símbolos secundários. Concluiu, então, que o total de sequências possíveis é n_p^N : independentemente de os

símbolos primários serem grupados no alfabeto secundário para serem interpretados pelo destinatário.

Exemplo 3.3.1. Sejam um alfabeto primário com 2 símbolos e um secundário com 32 símbolos. Assim $n_p = 2$, $N_p = 5$, $n_s = n_p^5 = 32$. Se a mensagem no alfabeto secundário contivesse 100 símbolos, isto é $N_s = 100$, seriam realizadas 100 seleções no alfabeto secundário, ou – equivalentemente – ou 500 seleções no alfabeto primário para se obter a mensagem alfabeto primário correspondente a essa mensagem no alfabeto secundário.

(HARTLEY, 1928) entendeu que a mudança de alfabeto não modificaria a produção da mesma quantidade de informação transmitida, pois o número de mensagens distintas e passíveis de serem interpretadas pelo destinatário seria o mesmo.

Supondo um processo contínuo de comunicação, N iria aumentando conforme mais símbolos fossem transmitidos e recebidos.

Para (HARTLEY, 1928) uma medida da forma n^N não seria adequada para mensurar a informação usando o seguinte argumento – o aumento exponencial pode fazer sentido quando as mensagens recebidas são utilizadas pelo receptor da mensagem – no qual aspectos psicológicos e do contexto² são relevantes. Todavia do ponto de vista de engenharia de comunicações, o armazenamento e a transmissão, são o objetivo final, (HARTLEY, 1928):

A telegraph system finds one ten-word message no more difficult to transmit than the one which preceded it. A telephone system which transmits speech successfully now will continue to do so as long as the system remains unchanged. **In order then for a measure of information to be of practical engineering value it should be of such a nature that the information is proportional to the number of selections.** [*negrito meu*]

Assim, se H_H denotar a medida de informação com as propriedades acima, N for o tamanho da mensagem, K uma constante que depende do número de seleções disponíveis (devido um alfabeto de tamanho n) então para ser consistente com a perspectiva do engenheiro – um modelo para H_H seria:

$$H_H = NK \tag{3.3.2}$$

Assim, H_H captura a noção de que o “esforço” para o armazenamento, a transmissão ou a recepção de uma mensagem deve ser linear no número de símbolos que compõem uma mensagem.

²Contexto: “Eu vi o homem no monte com um telescópio”, em (MARTINS, 1994, p. 43-45), apresenta 4 interpretações possíveis. Fornecendo um contexto, como “Limpei a lente para ver melhor” ou “Falamos sobre astronomia”, seleciona uma das possíveis interpretações.

Emissor e Receptor de Mensagens de Alfabetos Distintos

Sejam dois sistemas de comunicação, por exemplo, chamados de sistema A e sistema B , com alfabetos secundários de tamanhos diferentes. Seja n_A o tamanho do alfabeto no sistema A e n_B , no sistema B . O número de mensagens que o sistema A e B são capazes representar é o mesmo. Seja M esse número de mensagens. Cada mensagem no sistema A é composta por N_A símbolos e no B , por N_B símbolos então:

$$M = n_A^{N_A} \text{ e } M = n_B^{N_B} \quad (3.3.3)$$

Sem perda generalidade, o sistema A é emissor da mensagem e o sistema B é o receptor. Nesse processo de comunicação, não há perdas durante a transmissão (sem ruídos ou falhas).

Como são capazes lidar com exatamente o mesmo número de mensagens, então:

$$n_A^{N_A} = n_B^{N_B} \quad (3.3.4)$$

Seja \log é uma função logaritmo real qualquer, mas cuja base seja não-negativa e diferente da unidade, então:

$$N_A \log n_A = N_B \log n_B \quad (3.3.5)$$

Supondo que a informação transmitida possa ser **quantificada** então (não importando a sua natureza) a mesma **quantidade de informação** transmitida pelo sistema A , denotada por H_{H_A} , seria a mesma recebida pelo sistema B , denotada por H_{H_B} . Então, usando como modelo equação 3.3.2:

$$H_{H_A} = N_A K_A \text{ e } H_{H_B} = N_B K_B \quad (3.3.6)$$

O que permite escrever, pois $H_{H_A} = H_{H_B}$:

$$N_A K_A = N_B K_B \quad (3.3.7)$$

Então pode-se concluir das equações 3.3.5 e 3.3.7, (HARTLEY, 1928):

$$\frac{K_A}{\log n_A} = \frac{K_B}{\log n_B} \quad (3.3.8)$$

$$\frac{K_A}{K_B} = \frac{\log n_A}{\log n_B} \quad (3.3.9)$$

$$\frac{N_A K_A}{N_B K_B} = \frac{N_A \log n_A}{N_B \log n_B} \quad (3.3.10)$$

Substituindo as equações 3.3.6 em 3.3.10:

$$\frac{H_{H_A}}{H_{H_B}} = \frac{N_A \log n_A}{N_B \log n_B} \quad (3.3.11)$$

Que pode ser reescrita como:

$$H_{H_A} = K_0 N_A \log n_A \text{ e } H_{H_B} = K_0 N_B \log n_B \quad (3.3.12)$$

Onde K_0 é uma constante de proporcionalidade. Como não foi feita nenhuma restrição quanto aos valores de n_A e n_B , excetuando-se $n_A \geq 1$ e $n_B \geq 1$, então segue a definição:

Definição 3.3.1 (Informação de Hartley). *Seja K_0 uma constante de proporcionalidade, N tamanho mensagem escrita em um alfabeto e n o tamanho do alfabeto. H_H é quantidade de informação contida da mensagem, tal que:*

$$H_H = K_0 N \log n \quad (3.3.13)$$

Portanto, H_H é uma função do tamanho da mensagem e do tamanho do alfabeto.

Observação 3.3.1. Se o alfabeto contiver um único caractere somente então $H_H = 0$ para qualquer tamanho de mensagem. O que é consistente com a noção de que nenhuma informação está sendo enviada (visto que somente uma mensagem é possível – para qualquer N fixo). Então alfabetos unários não tem utilidade prática.

Observação 3.3.2. A constante K_0 é arbitrária, podendo ser considerada 1. Somando-se a isso, (HARTLEY, 1928) também considerava que a base do logaritmo definiria a unidade de quantidade informação.

Observação 3.3.3. Outra interpretação seria possível: considerar K_0 uma constante que mude a base de log (pois é conhecida), mas não está explicitada.

Observação 3.3.4. Uma questão não abordada por (HARTLEY, 1928), é que, para $K_0 > 0$, se a base de log for maior que a unidade então $H_H > 0$. Porém se for um valor entre 1 e zero, então $H_H < 0$.

Assim, seja b a base logaritmo, admitindo uma valor positivo para a *quantidade de informação* e a partir da expressão 3.3.13 então:

$$H_H = N \log_b n \quad (3.3.14)$$

Onde \log_b é o logaritmo na base b (esta que define a unidade da quantidade de informação) tal que $b > 1$.

É possível definir quantidade de informação média por símbolo numa mensagem, denotada por \bar{H}_H , como:

$$\bar{H}_H = \frac{H_H}{N} \quad (3.3.15)$$

Onde N de número de símbolos da mensagem. A partir de expressão 3.3.14 então

$$\bar{H}_H = \log_b n \quad (3.3.16)$$

Como N é número de símbolos na mensagem, n é o número de símbolos alfabéticos distintos e se M for o total de mensagens distintas de comprimento N então, com base na expressão 3.3.14, obter os três resultados abaixo:

$$M = n^N \quad (3.3.17)$$

$$H_H = \log_b M \quad (3.3.18)$$

$$b^{H_H} = M \quad (3.3.19)$$

3.4 Informação de Shannon

Em *A Mathematical Theory of Communication*, (SHANNON, 1948) percebeu que a medida de informação de (HARTLEY, 1928), apesar de consistente, não capturava a noção que os símbolos em uma mensagem típica tinham frequências diferentes – o que poderia afetar, na prática, a quantidade de informação transmitida. Isso pode ser ilustrado no exemplo 3.4.1.

Exemplo 3.4.1. Sejam quatro fontes de informação, a usando um alfabeto binário ($\{0, 1\}$ ou $\{\heartsuit, \clubsuit\}$, por exemplo), que transmitem mensagens de 100 símbolos. M denota o total de mensagens possíveis para essas fontes. Entretanto, apresentam restrições na frequência de símbolos alfabético conforme abaixo:

Restrições da fonte 1 símbolos ocorrem sem qualquer restrição:

$$M = 2^{100} \approx 1,26765 \times 10^{30}$$

Restrições da fonte 2 todos os símbolos alfabéticos sempre ocorrem com a mesma frequência; Então

$$M = \frac{100!}{50!50!} \approx 1,0089 \times 10^{29}$$

Restrições da fonte 3 o símbolo alfabético mais frequentes ocorrem sempre com frequência de 99 em 100; Então

$$M = \frac{100!}{99!1!} = 100$$

Restrições da fonte 4 somente um símbolo alfabético ocorre. Então

$$M = \frac{100!}{100!0!} = 1$$

O exemplo ilustra, intuitivamente, que quanto “maior” é a restrição então menor o número de mensagens possíveis. Ao mesmo tempo, isso sugere que a quantidade de informação associada a uma mensagem transmitida diminuiria: o universo de mensagens possíveis é menor. As fontes 1 e 4 são consistentes com a medida de informação de Hartley. Para a fonte 1 isso é obvio, mas para a fonte 4 não: a restrição 4 é tão severa que na prática transforma um alfabeto binário em um alfabeto unário.

Isso sugere que, uma vez que restrições afetam o número de mensagens possíveis, a quantidade de informação de fato sendo transmitida é também afetada (pois se há menos mensagens possíveis, um símbolo discrimina entre um número menor de mensagens, logo a quantidade de informação associada a esse símbolo é menor). Todavia, como essas restrições interferiam na quantidade de informação não era clara para (HARTLEY, 1928) – apesar de ele estar ciente de que quantidade de informação estava diretamente relacionada com o número de seleções no alfabeto primário (usado para codificar as mensagens durante o processo de transmissão). Em (HARTLEY, 1928):

When, as in the case just considered, the secondary symbols all involve the same number of primary selections, the relations are quite simple. When a telegraph system is used which employs a non-uniform code they are rather more complicated. (...)

É importante notar que não se pode atribuir uma quantidade de informação intrínseca ao símbolo alfabético – sem entender o número de mensagens possíveis de fato. A quantidade de informação transmitida depende do número de mensagens a serem discriminadas ou selecionadas de fato pelo receptor. Para contornar tais dificuldades de mensuração, (SHANNON, 1948) propôs uma medida estatística. Adiantando, o que (SHANNON, 1948) faz é uma correção ao modelo de (HARTLEY, 1928). Para isso, ele sai do campo determinístico e mecânico do processo de seleção de mensagens e entra no campo probabilístico: a lei dos grandes números operará

para mensagens suficientemente longas no que tange à probabilidade de ocorrência de um símbolo alfabeto.

A partir desse nova perspectiva, segue a expressão 3.3.14 de H_H (pag. 28) reescrita:

$$H_H = -\log_b \left(\frac{1}{n} \right)^N \quad (3.4.1)$$

Em uma modelo probabilístico, e no universo de mensagens de N de um alfabeto em particular, $\left(\frac{1}{n} \right)^N$ denota a probabilidade de um mensagem de N sabendo-se que não há restrições quando a probabilidade de ocorrência de um símbolo: cada símbolo é alfabético equiprovável. Então a probabilidade de cada símbolo é $\frac{1}{n}$, sendo n o tamanho do alfabeto. Essa observação, sobre esse caso especial, fica mais clara quando se observa expressão 3.3.18 e reescrevendo-a:

$$H_H = \log_b M = -\log_b \frac{1}{M} \quad (3.4.2)$$

Ora, $\frac{1}{M}$ é a probabilidade de se selecionar mensagem específica dentre M mensagens equiprováveis. Observando a expressão 3.3.17, que também diz respeito a esse caso em especial, $\frac{1}{M} = \left(\frac{1}{n} \right)^N$.

Fica claro que a quantidade H_H proporcional ao tamanho do alfabeto e à quantidade de mensagens possíveis – observando-se expressão 3.3.14 e 3.4.2. Essa quantidade também fica preservada, se escrita em função das probabilidades dos símbolos alfabéticos e da probabilidade de ocorrer uma mensagem, como acabou de ser observado, para o caso equiprovável.

Como a teoria de (SHANNON, 1948), possui como pressuposto a lei dos grandes números, e que o processo de comunicação produza mensagens com um número suficientemente grande símbolos – sua medida de informação será uma estatística: a informação média por símbolo em uma mensagem. Será observado que essa quantidade será preservada tanto em função do número de mensagens prováveis (ou típicas) quanto em a função das probabilidades dos símbolos alfabéticos de uma mensagem provável (ou típica).

Definição 3.4.1 (Fontes de Informações). *Uma **fonte de informação** produz uma mensagem ou seqüência de mensagens a serem comunicadas. Uma **fonte discreta** gera uma mensagem, símbolo por símbolo. Ela escolherá símbolos sucessivos de acordo com certas probabilidades, dependendo, em geral, das escolhas precedentes, bem como dos símbolos em questão.*

Definição 3.4.2 (Processos Ergódicos). *Em um processo ergódico, cada seqüência produzida pelo processo tem as mesmas em propriedades estatísticas – desde que o*

número de símbolos se aproxime do infinito. Assim, as frequências de letras, as frequências de N -gramas³, obtidas a partir de sequências particulares, tenderão, à medida que os comprimentos das sequências aumentarem, se aproximarem de limites definidos independentemente da sequência particular.

Os processos ergódicos apresentam homogeneidade estatística, isto é, as estatísticas (frequência de símbolos, médias, desvio-padrão, por exemplo) quando computadas para **cada** sequência gerada por um processo ergódico não se obtém, necessariamente, os mesmo resultados, **mas o conjunto dessas sequências para qual tal discrepância ocorre tem probabilidade zero** conforme o número de símbolos de uma sequência aumenta (SHANNON, 1948). Será clarificado mais à frente neste capítulo.

Definição 3.4.3 (Fonte Discreta de Shannon). *Uma fonte discreta de informações cujos símbolos são produzidos por um processo ergódico é um fonte discreta de Shannon.*

Definição 3.4.4 (Mensagens de Shannon). *É mensagem criada a partir de uma fonte discreta de Shannon. Seja N o tamanho da mensagem e N_i no número de ocorrências do i -ésimo símbolo alfabético na mensagem. \mathcal{P}_N e \mathcal{M}_N são, respectivamente, a probabilidade dessa mensagem ser gerada (num universo de mensagens com N símbolos) e seu número de anagramas⁴. p é função de distribuição discreta de probabilidades dos símbolos alfabéticos. p_i é a probabilidade o i -ésimo símbolo alfabético ser produzido pela da fonte discreta de Shannon.*

$$\mathcal{P}_N = \prod_i^n p_i^{N_i} \quad (3.4.3)$$

$$\mathcal{M}_N = \frac{N!}{\prod_i^n N_i!} \quad (3.4.4)$$

Além disso, \mathcal{MP}_N denota a probabilidade de ocorrer de qualquer anagrama dessa mensagem de Shannon, num universo de mensagens com N símbolos, então:

$$\mathcal{MP}_N = \mathcal{M}_N \cdot \mathcal{P}_N \quad (3.4.5)$$

Definição 3.4.5 (Uma Sequência de Shannon). *Uma sequência de mensagens tal que a primeira mensagem é um símbolo da fonte discreta de Shannon é uma sequência de Shannon; e para os demais termos da sequência, o i -ésimo símbolo*

³N-gramas: uma sequência de N itens. Pode ser uma sequência de sons, sílabas, palavras ou letras. Um N-grama de tamanho um é referido como um unigrama, o de tamanho dois é um bigrama (ou, "digrama"); O tamanho três é uma "trigrama" e assim por diante. Ver (MANNING e SCHÜTZE, 1999, p. 191).

⁴Anagramas: são todas as permutações de símbolos alfabéticos sem repetição.

da i -ésima mensagem origina-se dessa fonte discreta de Shannon e o restante da mensagem é idêntico a $(i-1)$ -ésima mensagem dessa sequência de Shannon.

Exemplo 3.4.2. Uma sequência de Shannon de fonte com alfabeto binário $\{0, 1\}$, cujos símbolos ocorrem com a mesma probabilidade: 1, 10, 101, 1011, 10110, 101101, 1011010, 10110100, 101101001, 1011010010, 10110100100, 101101001001, ...

Agora será definida a expressão bem conhecida na teoria da informação, pois ela surgirá naturalmente como resultado de limites.

Definição 3.4.6 (Informação de Shannon). *Seja p uma função de distribuição de probabilidade discreta e $1 < b \in \mathbb{R}$. p_i é a probabilidade de ocorrência de um tipo de símbolo em particular de um alfabeto com n símbolos distintos. A informação de Shannon, denotada por H_S para a distribuição p será definida como:*

$$H_S = - \sum_{i=1}^n p_i \log_b p_i \quad (3.4.6)$$

Definição 3.4.7 (Estimativa Assintótica). *Sejam f e g funções tais que $f : \mathbb{R} \rightarrow \mathbb{R}$ e $g : \mathbb{R} \rightarrow \mathbb{R}$. $f \sim g$ denota que f e g são estimativas assintóticas então:*

$$f \sim g \Leftrightarrow \lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = 1 \quad (3.4.7)$$

Comentário 3.4.1. Estimativas assintóticas permitem, por exemplo, que as funções $f(x)$ e $g(x)$ sejam intercambiáveis em expressões algébricas quando $x \rightarrow +\infty$ (ERDÉLYI, 2010; SETHNA, 2011), caso $f \sim g$.

Teorema 3.4.2. *Seja uma sequência de mensagens de Shannon. N diz respeito tanto ao número de símbolos da mensagem quanto a posição uma mensagem nessa sequência. Além disso, $N = \sum_{i=1}^n N_i$, onde N_i é o número de símbolos do i -ésimo símbolo do alfabeto na N -ésima mensagem dessa sequência.*

$$\lim_{N \rightarrow \infty} \frac{\log_b \mathcal{P}_N}{N} = -H_S \quad (3.4.8)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} = H_S \quad (3.4.9)$$

$$\lim_{N \rightarrow +\infty} \log_b(\mathcal{M}\mathcal{P}_N) = 0 \quad (3.4.10)$$

Corolário 3.4.3. *Seja uma sequência de mensagens de Shannon. N diz respeito tanto ao número de símbolos da mensagem quanto a posição uma mensagem nessa sequência. Além disso, $N = \sum_{i=1}^n N_i$, onde N_i é o número de símbolos do i -ésimo símbolo do alfabeto na N -ésima mensagem dessa sequência.*

$$\mathcal{P}_N \sim b^{-N \cdot H_S} \quad (3.4.11)$$

$$\mathcal{M}_N \sim b^{N \cdot H_S} \quad (3.4.12)$$

$$\mathcal{M}\mathcal{P}_N \sim 1 \quad (3.4.13)$$

O teorema 3.4.2 e corolário 3.4.3 “equivalentes aos 3 e 4 teoremas em (SHANNON, 1948)” espelham exatamente a observação de (SHANNON, 1948) (mas usando 2 como a base do logaritmo e $H = H_S$):

(...) it is possible for most purposes to treat the long sequences as though there were just 2^{HN} of them, each with a probability 2^{-HN} . (...)

Demonstração. (Teorema 3.4.2 e Corolário 3.4.3)

Para a relação 3.4.8, se começa com a definição de sequência de mensagens de Shannon (definição 3.4.4). Seja n a cardinalidade do alfabeto e só depende a fonte de informação. Suponha que para uma N seja o suficiente grande para que o seu número de ocorrências do i -ésimo símbolo alfabético, tal que $N_i \approx N \cdot p_i$ e exato para $N \rightarrow \infty$, onde p_i é a probabilidade do i -ésimo tipo de símbolo alfabético emitido pela fonte de Shannon que criou a mensagem.

Por hipótese, a probabilidade de ocorrência de um símbolo é independente de sua posição na mensagem ou dos símbolos gerados anteriormente, para N suficientemente grande, então a ordem que eles aparecem é irrelevante para o cálculo de \mathcal{P}_N . Sendo também, portanto, idêntica para todos os anagramas de uma mensagem de Shannon, assim:

$$\mathcal{P}_N = \prod_i^n p_i^{N_i} \simeq \prod_i^n p_i^{N \cdot p_i} = \left(\prod_i^n p_i^{p_i} \right)^N \quad (3.4.14)$$

Assim,

$$\log_b \mathcal{P}_N \simeq \log_b \left(\prod_i^n p_i^{p_i} \right)^N = N \log_b \left(\prod_i^n p_i^{p_i} \right) = N \sum_i^n p_i \log_b p_i \quad (3.4.15)$$

$$\frac{\log_b \mathcal{P}_N}{N} \simeq \sum_i^n p_i \log_b p_i \quad (3.4.16)$$

Usando o argumento de homogeneidade estatística, tem-se a igualdade:

$$\lim_{N \rightarrow \infty} \frac{\log_b \mathcal{P}_N}{N} = \sum_i^n p_i \log_b p_i \quad (3.4.17)$$

Valendo-se da definição 3.4.6 então:

$$\lim_{N \rightarrow \infty} \frac{\log_b \mathcal{P}_N}{N} = -H_S \quad (3.4.18)$$

O que prova a relação 3.4.8 do teorema 3.4.2.

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{P}_N}{-NH_S} = 1 \quad (3.4.19)$$

$$\lim_{N \rightarrow +\infty} \frac{\mathcal{P}_N}{b^{-NH_S}} = 1 \quad (3.4.20)$$

O resultado acima é uma estimativa assintótica, pela definição 3.4.7, então:

$$\mathcal{P}_N \sim b^{-N \cdot H_S} \quad (3.4.21)$$

O que prova a relação 3.4.11 do corolário 3.4.3.

Para a relação 3.4.9, também começa valendo-se da definição de mensagem de Shannon (definição 3.4.4). Nessa etapa será usada a aproximação de Stirling⁵ para fatoriais:

$$n! \sim \sqrt{2\pi n} \cdot n^n \cdot e^{-n} \quad (3.4.22)$$

Logo:

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} \cdot n^n \cdot e^{-n}} = 1$$

Por definição:

$$\mathcal{M}_N = \frac{N!}{\prod_i^n N_i!} \quad (3.4.23)$$

Observado que $1 = \sum_{i=1}^n \frac{N_i}{N}$ para a N-ésima mensagem da sequência de Shannon,

⁵A aproximação de Stirling para $n!$ é válida para grande n e é extremamente útil na mecânica estatística. Para mais informação (SETHNA, 2011, p. 7), (SCHROEDER, 2000, p. 65) e (HASSANI, 2000, p. 45).

empregando-se a aproximação de Stirling:

$$\mathcal{M}_N \simeq \frac{\sqrt{2\pi N} \cdot N^N \cdot e^{-N}}{\prod_i^n \sqrt{2\pi N_i} \cdot N_i^{N_i} \cdot e^{-N_i}} = \frac{\sqrt{2\pi N} \cdot N^N \cdot e^{-N}}{\left(\prod_i^n \sqrt{2\pi N_i} \cdot N_i^{N_i}\right) \cdot \left(\prod_i^n e^{-N_i}\right)} \quad (3.4.24)$$

$$\mathcal{M}_N \simeq \frac{\sqrt{2\pi N} \cdot N^N \cdot e^{-N}}{\left(\prod_i^n \sqrt{2\pi N_i} \cdot N_i^{N_i}\right) \cdot e^{\sum_i^n -N_i}} = \frac{\sqrt{2\pi N} \cdot N^N \cdot e^{-N}}{\left(\prod_i^n \sqrt{2\pi N_i} \cdot N_i^{-N_i}\right) \cdot e^{-N}} \quad (3.4.25)$$

$$\mathcal{M}_N \simeq \frac{\sqrt{2\pi N} \cdot N^N}{\prod_i^n \sqrt{2\pi N_i} \cdot N_i^{N_i}} = \frac{\sqrt{2\pi} \cdot N^{N+1/2}}{\prod_i^n \sqrt{2\pi} \cdot N_i^{N_i+1/2}} = \frac{\sqrt{2\pi} \cdot N^{N+1/2}}{(\sqrt{2\pi})^n \cdot \prod_i^n N_i^{N_i+1/2}} \quad (3.4.26)$$

$$\mathcal{M}_N \simeq (\sqrt{2\pi})^{1-n} \cdot \frac{N^{N+1/2}}{\prod_i^n N_i^{N_i+1/2}} = \frac{N^{N+1/2}}{\prod_i^n N_i^{N_i+1/2}} \cdot (\sqrt{2\pi})^{1-n} \quad (3.4.27)$$

$$\log_b \mathcal{M}_N \simeq \log_b \left(\frac{N^{N+1/2}}{\prod_i^n N_i^{N_i+1/2}} \cdot (\sqrt{2\pi})^{1-n} \right) \quad (3.4.28)$$

$$\log_b \mathcal{M}_N \simeq \log_b \left(\frac{N^{N+1/2}}{\prod_i^n N_i^{N_i+1/2}} \right) + \log_b (\sqrt{2\pi})^{1-n} \quad (3.4.29)$$

$$\log_b \mathcal{M}_N \simeq \log_b N^{N+1/2} - \log_b \prod_i^n N_i^{N_i+1/2} + \log_b (\sqrt{2\pi})^{1-n} \quad (3.4.30)$$

$$\log_b \mathcal{M}_N \simeq \log_b N^{N+1/2} - \sum_i^n \log_b N_i^{N_i+1/2} + \log_b (\sqrt{2\pi})^{1-n} \quad (3.4.31)$$

$$\log_b \mathcal{M}_N \simeq (N + 1/2) \log_b N - \sum_i^n (N_i + 1/2) \log_b N_i + \log_b (\sqrt{2\pi})^{1-n} \quad (3.4.32)$$

$$\frac{\log_b \mathcal{M}_N}{N} \simeq \left(1 + \frac{1}{2N}\right) \log_b N - \sum_i^n \left(\frac{N_i}{N} + \frac{1}{2N}\right) \log_b N_i + \frac{\log_b (\sqrt{2\pi})^{1-n}}{N} \quad (3.4.33)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} \simeq \lim_{N \rightarrow +\infty} \log_b N - \sum_i^n \lim_{N \rightarrow +\infty} \frac{N_i}{N} \log_b N_i + \lim_{N \rightarrow +\infty} \frac{\log_b (\sqrt{2\pi})^{1-n}}{N} \quad (3.4.34)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} \simeq \lim_{N \rightarrow +\infty} \log_b N - \sum_i^n \lim_{N \rightarrow +\infty} p_i \log_b N_i + 0 \quad (3.4.35)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} \simeq \lim_{N \rightarrow +\infty} \sum_i^n p_i \log_b N - \sum_i^n \lim_{N \rightarrow +\infty} p_i \log_b N_i \quad (3.4.36)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} \simeq \lim_{N \rightarrow +\infty} \left(\sum_i^n p_i \log_b N - \sum_i^n p_i \log_b N_i \right) \quad (3.4.37)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} \simeq \lim_{N \rightarrow +\infty} \sum_i^n p_i \log_b \frac{N}{N_i} \quad (3.4.38)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} \simeq \sum_i^n p_i \log_b \lim_{N \rightarrow +\infty} \frac{N}{N_i} \quad (3.4.39)$$

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} = - \sum_i^n p_i \log_b p_i \quad (3.4.40)$$

Valendo-se da definição 3.4.6 então:

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} = H_S \quad (3.4.41)$$

O que prova a relação 3.4.9 do teorema 3.4.2.

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{NH_S} = 1 \quad (3.4.42)$$

$$\lim_{N \rightarrow +\infty} \frac{\mathcal{M}_N}{b^{NH_S}} = 1 \quad (3.4.43)$$

O resultado acima é uma estimativa assintótica, pela definição 3.4.7, então:

$$\mathcal{M}_N \sim b^{N \cdot H_S} \quad (3.4.44)$$

O que prova a relação 3.4.12 do corolário 3.4.3.

Para a relação 3.4.10, a prova é quase imediata. Basta somar as relações 3.4.8 e relações 3.4.9, que estão provadas acima, então:

$$\lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N} + \lim_{N \rightarrow \infty} \frac{\log_b \mathcal{P}_N}{N} = H_S - H_S \quad (3.4.45)$$

$$\lim_{N \rightarrow +\infty} \log_b \mathcal{M}_N + \lim_{N \rightarrow \infty} \log_b \mathcal{P}_N = \lim_{N \rightarrow \infty} NH_S - NH_S \quad (3.4.46)$$

$$\lim_{N \rightarrow +\infty} \log_b \mathcal{M}_N + \log_b \mathcal{P}_N = 0 \quad (3.4.47)$$

$$\lim_{N \rightarrow +\infty} \log_b (\mathcal{M}_N \cdot \mathcal{P}_N) = 0 \quad (3.4.48)$$

$$\lim_{N \rightarrow +\infty} \log_b \mathcal{MP}_N = 0 \quad (3.4.49)$$

O que prova a relação 3.4.10 do teorema 3.4.2.

Finalmente,

$$\lim_{N \rightarrow +\infty} \log_b \mathcal{MP}_N = 0 \quad (3.4.50)$$

$$b^j = b^0, \text{ onde } j = \left(\lim_{N \rightarrow +\infty} \log_b \mathcal{MP}_N \right) \quad (3.4.51)$$

$$\lim_{N \rightarrow +\infty} b^{\log_b \mathcal{MP}_N} = 1 \quad (3.4.52)$$

$$\lim_{N \rightarrow +\infty} \mathcal{MP}_N = 1 \quad (3.4.53)$$

$$\lim_{N \rightarrow +\infty} \frac{\mathcal{MP}_N}{1} = 1 \quad (3.4.54)$$

O resultado acima é uma estimativa assintótica, pela definição 3.4.7, então:

$$\mathcal{P}\mathcal{M}_N \sim 1 \quad (3.4.55)$$

O que prova a relação 3.4.13 do corolário 3.4.3.

□

Assim em (SHANNON, 1948) modificou-se a noção de conjunto de mensagens possíveis para conjunto de mensagens prováveis (ou mensagens típicas).

Um fato interessante é a probabilidade das uma mensagem de uma sequência de Shannon tender a assintoticamente a 1, ou $\mathcal{M}\mathcal{P}_N \sim 1$. Pode-se escrever:

$$\mathcal{M}\mathcal{P}_N = \mathcal{M}_N \cdot \mathcal{P}_N = \frac{N!}{\prod_{i=1}^n N_i!} \cdot \prod_{i=1}^n p_i^{N_i} \quad (3.4.56)$$

Definição 3.4.8 (Distribuição de Probabilidade Multinomial). $f(N_1, \dots, N_i, \dots, N_n)$ é, para todo i , a probabilidade do i -ésimo evento – cuja é probabilidade p_i (entre n eventos distintos) – ocorrer N_i vezes em N experimentos idênticos, independentes e sucessivos. Então f é função de distribuição de probabilidade multinomial⁶, tal que (TRIVEDI, 1982, pp. 45-46):

$$f(N_1, \dots, N_i, \dots, N_n) = \frac{N!}{\prod_i N_i!} \cdot \prod_{i=1}^n p_i^{N_i}$$

É digno de nota que a produção de mensagens da fonte de Shannon obedece ao modelo uma população multinomial típica (KULLBACK, 1978, p. 109). Portanto, conforme $N \rightarrow +\infty$, as mensagens, suficientemente longas, produzidas por uma fonte de Shannon tendem a ter as características estatísticas da fonte, porque o conjunto das mensagens prováveis (ou típicas) se concretiza como regra, pois tende ao evento certo. Por outro lado, o conjunto das mensagens atípicas, se esvanece como exceção, pois tende ao evento impossível.

3.5 Relações entre os Paradigmas de Informação de Hartley e Shannon

A expressão 3.3.18 é a medida informação de Hartley, abaixo transcrita:

$$H_H = \log_b M$$

Ela é uma medida de informação absoluta. Pode ser reescrita como:

⁶Probabilidade multinomial: para $n = 2$, essa distribuição é um caso particular: a distribuição binomial, modela N experimentos de Bernoulli sucessivos.

$$b^{H_H} = M$$

A expressão 3.4.9 é a medida de informação de Shannon, abaixo transcrita:

$$H_S = \lim_{N \rightarrow +\infty} \frac{\log_b \mathcal{M}_N}{N}$$

Ela é uma medida de informação média por símbolo. Pode ser reescrita como:

$$b^{N \cdot H_S} \sim \mathcal{M}_N$$

M é um caso particular de \mathcal{M}_N , no qual os símbolos alfabéticos são equiprováveis. Então fica claro que $N \cdot H_S$ é uma correção para H_H . Nesse caso então $H_H = N \cdot H_S$. Outra interpretação é que H_S efetua uma correção para o valor da base b :

$$(b^{H_S})^N \sim \mathcal{M}_N$$

Fazendo $b' = b^{H_S}$ então:

$$b'^N \sim \mathcal{M}_N$$

b' é valor de base corrigida – considerando as restrições impostas pela probabilidade de cada símbolo alfabético.

Agora será demonstrado que b' depende somente da probabilidade dos símbolos alfabéticos. Pela definição 3.4.6 (pag. 33):

$$b' = b^{H_S} = b^{-\sum_{i=1}^n p_i \log_b p_i} = b^{\sum_{i=1}^n \log_b \left(\frac{1}{p_i}\right)^{p_i}} = b^{\log_b \prod_{i=1}^n \left(\frac{1}{p_i}\right)^{p_i}}$$

Portanto,

$$b' = \prod_{i=1}^n \left(\frac{1}{p_i}\right)^{p_i} \tag{3.5.1}$$

Pela relação clássica entre as médias aritmética e geométrica⁷ então:

$$\prod_{i=1}^n \left(\frac{1}{p_i}\right)^{p_i} \leq \sum_{i=1}^n p_i \left(\frac{1}{p_i}\right) \tag{3.5.2}$$

Observando que $1 \leq 1/p_i$,

$$1 \leq \prod_{i=1}^n \left(\frac{1}{p_i}\right)^{p_i} \leq \sum_{i=1}^n p_i \left(\frac{1}{p_i}\right) = n$$

⁷Relação clássica: a média aritmética é maior ou igual à média geométrica (SPIEGEL, 1972, cap. 3, p. 76 e 77).

Portanto,

$$1 \leq b' \leq n \quad (3.5.3)$$

Como n é o número de símbolos alfabéticos, essa correção é de tal que $b' \leq n$. É como se as restrições tivessem o efeito de reduzir a capacidade de discriminar dos símbolos alfabeto: comportando-se como um alfabeto com um número menor de símbolos para uma mensagem típica com mesmo número de símbolos N . Continuando, e observando que $b > 0$:

$$\log_b 1 \leq \log_b b' \leq \log_b n$$

$$0 \leq \log_b b' \leq \log_b n$$

$$0 \leq \log_b b^{H_S} \leq \log_b n$$

Portanto,

$$0 \leq H_S \leq \log_b n \quad (3.5.4)$$

Multiplicando pelo número de símbolos da mensagem então:

$$N \cdot 0 \leq N \log_b b^{H_S} \leq N \log_b n$$

Considerando a expressão 3.3.14 (pag. 28) então:

$$0 \leq NH_S \leq H_H \quad (3.5.5)$$

Por fim, considerando a expressão 3.3.15 (pag. 29) então:

$$0 \leq H_S \leq \bar{H}_H \quad (3.5.6)$$

Propriedades 3.5.1. Com base nas relações 3.5.4, 3.5.5 e 3.5.6, então:

- H_S e \bar{H}_H não podem assumir valores negativos;
- H_S é limitada superiormente por $\log_b n$;
- H_S é uma medida da quantidade de informação por símbolo da mensagem típica limitada superiormente por \bar{H}_H ;
- NH_S é uma medida da quantidade informação para uma mensagem típica de N símbolos e limitada superiormente por H_H .

3.6 Unidades de Informação

Em geral costuma-se no lugar de usar \log_b escrever $k \log$ sem explicitar a base do logaritmo — ou considerar $b = 2$. Quando o alfabeto primário tem dois símbolos, a unidade de informação é chamada de *bit* (dígito binário)⁸, mas se $b = e$ (o número de Euler) será chamada *nat* (unidade natural) — utilizada em trabalhos analíticos, pois simplifica derivações de expressões que envolvam logaritmos (COVER e THOMAS, 1991; KLIR, 2006; KULLBACK, 1978; SHANNON, 1948). Para converter *bits* em *nats*, basta multiplicar por $\ln 2$. Para converter *nats* em *bits*, basta multiplicar por $\log_2 e$. Obviamente, $\ln 2 \cdot \log_2 e = 1$.

Em (IEC80000-13, 2008; SHANNON, 1948; THOMPSON e TAYLOR, 2008), encontram-se unidades de informação universalmente aceitas nos campos da engenharia, da física e da computação. A base 10, por exemplo, corresponde à unidade de informação *Hart* — em homenagem à (HARTLEY, 1928).

Ressalta-se que a unidade não é especificada somente pelo uso do logaritmo, mas também pela natureza do processo observado. Nesse caso, medir a incerteza de uma função de distribuição de probabilidades discreta. Em (NUSSENZVEIG, 1994, p. 211), o logaritmo na base 10 é relativo a unidade *bel*, quando utilizado para medir o nível de intensidade relativa. Dois sons que diferem em 1 *bel*, o de maior intensidade tem uma magnitude 10 vezes maior que a do menor (NUSSENZVEIG, 1994, p. 211) — por óbvio tais medições não são sobre fenômenos modelados probabilisticamente.

3.7 Incerteza

Em (SHANNON, 1948), o termo **escolha** (“choice” em inglês) significa o mesmo que **distribuição discreta de probabilidade**, uma vez que o próprio resultado da experiência probabilística é irrelevante para calcular a incerteza. No entanto, quando (SHANNON, 1948) usa o termo resultado é simplesmente indicar a cardinalidade do domínio da função de distribuição de probabilidade. Cada resultado correspondente a um evento mutuamente exclusivo. Assim, uma função de distribuição discreta de probabilidade e escolha são, nesta seção, intercambiáveis.

(SHANNON, 1948) formulou axiomáticamente uma medida incerteza de uma escolha p com n resultados possíveis, denotada como H — onde p_i é a probabilidade do i -ésimo resultado mutuamente exclusivo. Segundo (SHANNON, 1948) H é uma função de p . A medida H deve satisfazer as três propriedades abaixo:

Propriedade 3.7.1. “ H deve ser contínua em p_i ”.

Propriedade 3.7.2. “Se p_i é igual para os todos resultados, $p_i = 1/n$, então H deve ser monotonamente crescente em função de n . Com eventos igualmente prováveis

⁸bit (binary digit): segundo (SHANNON, 1948), termo sugerido por J. W. Tukey.

há mais escolha, ou incerteza, quando há mais eventos possíveis”.

Propriedade 3.7.3. “Se uma escolha for dividida em duas escolhas sucessivas, o original $H(p)$ deve ser a soma ponderada dos valores individuais de H .”

Teorema 3.7.1. *Seja $k \in \mathbb{R}$ and $k > 0$. A única família de funções que satisfaz as propriedades 3.7.1, 3.7.2 e 3.7.3 é:*

$$H(p) = -k \sum_{i=1}^n p_i \log_2 p_i \quad (3.7.1)$$

Onde k é uma constante de proporcionalidade relacionada à mudança na base da função e ao estabelecimento de uma medida de unidade para incerteza. Para $k = 1$ e o logaritmo na base 2, a unidade de incerteza é o bit por símbolo. Caso $k = \frac{1}{\log_2 e} = \ln 2$, a unidade é o nat por símbolo.

Demonstração. Em (SHANNON, 1948), (JAYNES, 1957), (COVER e THOMAS, 1991) ou (SAYOOD, 2005). □

Obviamente é a mesma expressão que a obtida na seção 3.5 para a informação por símbolo. A diferença é que (SHANNON, 1948) descreveu por axiomas como a incerteza deveria se comportar e obteve a família de funções acima (e garantiu, inclusive, sua unicidade). Segue um exemplo para auxiliar uma breve explanação sobre essas propriedades.

Exemplo 3.7.1. Sejam p , p' , p'' e p''' escolhas. p e p' escolhas cujos eventos são equiprováveis conforme na figura 3.7.1. p' pode ser dividida nas escolhas p'' e p''' equivalentes no que diz respeito às probabilidades globais dos resultados A_1 , A_2 e A_3 , mas observando que o resultado A' é uma pré-condição para alcançar p''' de p'' conforme na figura 3.7.1. Suponha que as escolhas p'' e p''' são realizadas várias vezes, mas sempre esperando os resultados A_1 , A_2 e A_3 (A' não são considerados como resultado final), $2/3$ será o peso para a entropia de p''' e o peso para entropia de p'' é 1 no que diz respeito à incerteza. Portanto, é dizer que p' foi quebrado em p'' e p''' . As árvores na figura 3.7.1 representam a relação entre p e p''' e também fornecem informações adicionais sobre essas e a escolha p'' .

A propriedade 3.7.1 impõe que flutuações suficientemente pequenas nas probabilidades correspondam a flutuações pequenas de incerteza.

A propriedade 3.7.2 é bastante intuitiva: quanto mais possibilidades equiprováveis, maior será a incerteza sobre a escolha: $\log_2(2) = H(p) < H(p') = \log_2(3)$.

A propriedade 3.7.3 garante que $H(p') = \log(3) = H(p'') + \frac{2}{3}H(p''')$. A compreensão dessa propriedade reside no fato de que a modificação de uma sequência de escolhas locais altera a incerteza local, ao mesmo tempo que pode conservar inteiramente a incerteza global. Ou seja, ter mais, ou menos, escolhas em sequência não

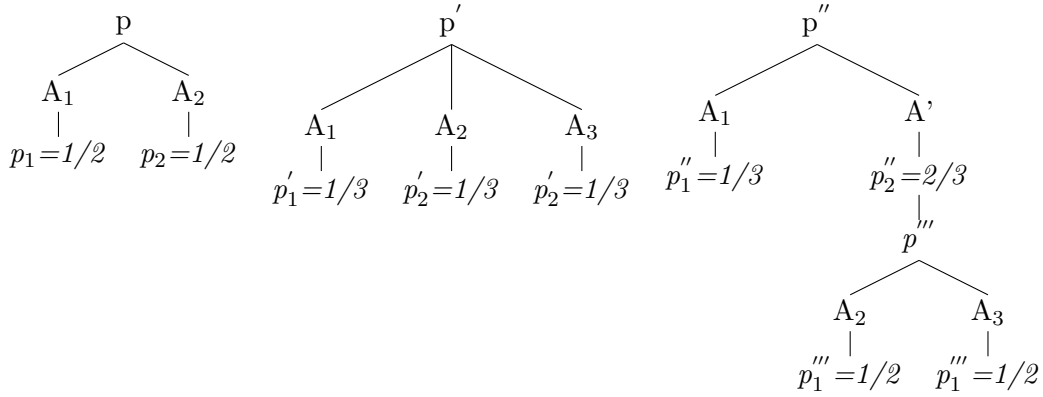


Figura 3.1: Quatro escolhas na representação de árvore: p , p' , p'' e p''' .

implica necessariamente em variação de incerteza. Quando a probabilidade dos resultados finais é inalterada, mais escolhas em sequência necessariamente preservam a média da incerteza global.

A interpretação de incerteza e informação se coadunam quando se aborda o comportamento da fonte de informação em emitir símbolos, pois quanto mais imprevisível for a fonte de informação então maior a diversidade de mensagens estatisticamente similares (mensagens típicas).

Em (SETHNA, 2011) um caso especial da função 3.7.1 onde p é uma distribuição probabilidade de n eventos equiprováveis é chamado de **entropia de contagem**, denotado aqui por H_c :

$$H(p) = k \sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n} = k \log_2 n = H_c(n) \quad (3.7.2)$$

Comentário 3.7.2. Importante notar que H e H_S são obtidos de formas distintas. H é obtido de forma axiomática enquanto H_S surge quando se tenta mensurar a informação por símbolo de uma fonte de informação (na seção anterior). Entretanto, são na prática a mesma fórmula matemática. Quanto mais incerteza estiver associada a uma fonte de informação, então o número de mensagens de Shannon aumenta.

Comentário 3.7.3. Em seu teorema 9, (SHANNON, 1948) frisa que a unidade de H_S é “bits per symbol”. Em textos clássicos como (COVER e THOMAS, 1991; KULLBACK, 1978) omite-se a expressão “per symbol” – o que pode causar uma certa confusão.

3.8 Discussão e Conclusão

Utilizar as propriedades matemáticas da função $H(p) = \sum_{i=1}^n p_i \log p_i$ para compreender o conceito de informação é tão útil quanto estudar as propriedades ma-

temáticas de funções da forma $f(x) = ax^2 + bx + c$ para compreender o movimento dos corpos em queda livre: extremamente difícil e especulativo – senão impossível. Fórmulas e equações dizem respeito a relações entre grandezas sobre os processos estudados – e não o contrário.

Este capítulo procurou abordar os processos nos primórdios das pesquisas de quantificação da informação e da sua relação com a incerteza (entropia). Para isso foi mostrada uma transição entre o conceito de informação de (HARTLEY, 1928) e a correção feita por (SHANNON, 1948).

Finalmente, no que diz respeito à expressão $\sum_{i=1} p_i \log p_i$, ela apareceu em três situações distintas:

- em processos de contagem de mensagens típicas (estatisticamente equivalentes);
- em processos de quantificação da informação; e
- na interpretação axiomática de Shannon para capturar as propriedades de um tipo de incerteza.

Capítulo 4

Qualidade de Montgomery

Qualidade sf. 'propriedade, atributo ou condição das coisas ou das pessoas capaz de distingui-las das outras e de lhe determinar a natureza'. | XVI, calidade XIV | Do lat. *quālitās -ātis* || qualitativo XVIII. Do lat. tard. *quālitātīvus*.

Dicionário Etimológico da Língua Portuguesa.

4.1 Introdução

Em sentido amplo, o conceito de qualidade é difícil definir. Pode-se conjecturar que processos de padronização são aplicáveis a coisas enquanto processos de qualidade são aplicáveis a processos de trabalho (no qual o esforço e destreza humana são importantes). Mesmo sob essa conjectura, a redução de características indesejadas está sempre presente: seja porque se obtém um produto fora das especificações – ou defeituoso – ou seja porque um projeto entregou seu produto com atraso ao cliente. Todavia, aquilo que é invariável é a necessidade de coleta de dados para a consolidação de estatísticas – sejam dados de natureza numérica ou qualitativa. Para estabelecer um denominador comum, o conceito de qualidade diz respeito a processos envolvendo produtos, serviços e pessoas como em (TAGUCHI *et al.*, 2004), (DRUCKER, 2008), (JURAN e GODFREY, 1999), (MITRA, 2016) e (GARVIN, 1988), por exemplo.

O termo *Qualidade de Montgomery*, neste capítulo, é empregado meramente como uma sugestão ao conceito mais formal de uma Teoria sobre Qualidade, ou de uma Hipótese sobre Qualidade, ou de Princípios da Qualidade, ou de Propriedades de Qualidade com base nas definições preconizadas pelo próprio (MONTGOMERY, 2012, p. 6 e 7) – definições estas no primeiro capítulo de seu livro *Introduction to Statistical Quality Control*. Para efeitos deste estudo, mais especificamente, a Qualidade de Montgomery consiste em um *paradigma* composto de propriedades que especificam como grandezas que capturam medidas específicas de qualidade e

de variabilidade devem se relacionar; e versam também sobre como perceber um processo de melhoria de qualidade em função dessas grandezas.

Ao final serão apresentadas algumas *anomalias*¹ em potencial encontradas nesse paradigma. Buscar-se-á conciliar algumas dessas anomalias por um novo paradigma descrito nos capítulos 6 e 8.

4.2 As Propriedades da Qualidade de Montgomery

(MONTGOMERY, 2012, p. 6) propõe três propriedades que caracterizam o conceito de qualidade na produção de bens e serviços. Estas três propriedades estão abaixo:

Propriedade 4.2.1. Qualidade significa adequação para o uso.

Propriedade 4.2.2. Qualidade é inversamente proporcional à variabilidade.

Propriedade 4.2.3. *Melhoria da qualidade* é a redução da variabilidade nos processos e serviços.

A propriedade 4.2.1, segundo o próprio (MONTGOMERY, 2012), trata-se de estar aderente à definição tradicional – da perspectiva que produtos e serviços devam estar dentro das especificações de forma a estarem adequados a seus usuários.

O que chama atenção sobre as propriedades 4.2.2 e 4.2.3 é que – no lugar de definir ou dizer o que é qualidade – elas procuram relacionar o comportamento entre a qualidade e a variabilidade como se fossem quantidades — mesmo que não se saiba o que são realmente, ontologicamente falando.

Se Q_M e V_M são, respectivamente, duas grandezas de qualidade e de variabilidade que satisfazem a propriedade 4.2.2 então:

$$Q_M \propto \frac{1}{V_M} \quad (4.2.1)$$

E, ainda, satisfazem consequentemente a propriedade 4.2.2, pois:

$$dQ_M \propto -\frac{1}{V_M^2} dV_M \quad (4.2.2)$$

Que reescrevendo para evidenciar a variação de variabilidade:

$$dV_M \propto -V_M^2 dQ_M \quad (4.2.3)$$

(MONTGOMERY, 2012, p. 6), entretanto, faz a seguinte ressalva no que tange à propriedade 4.2.2: “We are referring to unwanted or harmful variability. There

¹Anomalias no sentido de (KUHN, 2013, cap. 5).

situations in which variability is actually good.”. Apesar de a propriedade 4.2.2 ser aparentemente coerente do ponto da melhoria de processos, ela cerceia, entretanto, a possibilidade de avaliações relativas a *standards* (padrões) – pois o único critério é a variabilidade. Outro ponto importante é a natureza da variabilidade. Se a variabilidade for resultado de um processo que permita $V_M = 0$, a grandeza Q_M não está definida. Portanto foram identificadas três anomalias no modelo proposto por (MONTGOMERY, 2012). Elas são formalizadas abaixo.

Anomalia 4.2.1 (1ª Anomalia na Qualidade de Montgomery). É possível que processos de melhoria da qualidade possam ser caracterizados pelo aumento da variabilidade.

Anomalia 4.2.2 (2ª Anomalia na Qualidade de Montgomery). É possível que grandezas de qualidade sejam relativas.

Anomalia 4.2.3 (3ª Anomalia na Qualidade de Montgomery). Não contempla modelos em que a variabilidade possa ser nula.

Por que a Qualidade de Montgomery?

(COSTA *et al.*, 2008) e (PYZDEK, 2003) observam que não existe na literatura uma definição única para a qualidade, tabela 4.1.

Tabela 4.1: Qualidade segundo Filósofos da Engenharia de Processos e da Administração

Guru	Definição de Qualidade
Joseph Juran	Adequação para o uso.
Edwards Deming	Atender e, se possível, exceder às expectativas do consumidor.
Philip Crosby	Atender às especificações.
Genichi Taguchi	Diminuição do prejuízo causado à sociedade pela produção, uso e consumo de um produto.

Fonte: (COSTA *et al.*, 2008)

No clássico da administração *Inovação e Espírito Empreendedor*, (DRUCKER, 2008) leva o conceito de qualidade ao extremo da subjetividade:

Ainda mais debilitante é o terceiro mau hábito: a crença em “Qualidade”. Qualidade em um produto ou serviço não é o que o fornecedor insere nele. É aquilo que o cliente obtém dele e está disposto a pagar por isso. Um produto não é “qualidade” só porque seja difícil de fazer e custa muito dinheiro, como os fabricantes tipicamente acreditam. Isso é incompetência. Os clientes pagam somente por aquilo que lhes é de uso e lhes dá valor. Nada mais constitui “qualidade”. (DRUCKER, 2008, p. 312)

No século XIX, (KELVIN, 1889, p. 73 e 74), em *Popular Lectures and Addresses*, mostra com suas próprias palavras como o conceito de qualidade era percebido na ciência do seu tempo:

In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be. (...) (KELVIN, 1889, p. 73 e 74)

(KELVIN, 1889, p. 74 e 75) utilizou a palavra “quality” no sentido de *propriedade intrínseca* do objeto de estudo. Ao mesmo tempo, entendeu que a percepção de uma qualidade, em quanto propriedade, é somente o início do conhecimento – que se tornará conhecimento perfeito quando for representado como número.

O trecho acima é bem conhecido, porém menos conhecida é sua continuação, na qual é exposto um exemplo de conhecimento que seria aquém e investigatório (por ser não-numérico) – no campo da metalurgia e da resistência de materiais de sua época:

I may illustrate by a case in which this first step has not been taken. The hardness of different solids, as precious stones and metals, is reckoned by a merely comparative test. Diamond cuts ruby, ruby cuts quartz, quartz, I believe, cuts glass-hard steel, and glass-hard steel cuts glass; hence diamond is reckoned harder than ruby; ruby, than quartz; quartz, than glass-hard steel; and glass-hard steel, than glass: but we have no numerical measure of the hardness of these, or any other solids. We have, indeed, no knowledge of the moduluses of rigidity, or of the tensile strength, of almost any of the gems or minerals, of which the hardness is reckoned by mineralogists in their comparative scale, beginning with diamond, the hardest of known solids. We have even no reason to believe that the modulus of rigidity of diamond is greater than that of other solids; and we have no exact understanding of what this property of hardness is, nor of how it is related to moduluses of elasticity, or to tensile or shearing strength, or to the quality of the substance in respect of its bearing stresses exceeding the limit of its elasticity. It must, therefore, be admitted, that the science of strength of materials, so all-important in engineering, is but little advanced, and the part of it relating to the so-called hardness of different solids least of all; there being in it no step toward quantitative measurement or reckoning in terms of a definite unit. (KELVIN, 1889, p. 74 e 75)

(MONTGOMERY, 2012) de certa forma tenta quantificar o conceito de qualidade relacionando-o com o de variabilidade. Com isso, em princípio, permitindo que os benefícios de se trabalhar com modelos quantitativos fossem uma possibilidade ao conhecimento – não se restringindo somente a prescrever ou normatizar o que é qualidade.

4.3 Desvio-padrão: uma Medida de Variabilidade

A metodologia *Six Sigma* trabalha com múltiplos do desvio-padrão para alguma quantidade observada no escopo da metodologia “Definir, Mensurar, Analisar, Melhorar e Controlar” para processos de negócio, atividades operacionais e de produtos objetivando baixar a taxa de defeitos, falhas e inconformidades na ordem de quatro parte por milhão (JUGULUM, 2014, p. 25 e 43-54). A ideia fundamental por de trás do *Six Sigma* é o controle estatístico de processos. Inclusive ela tem esse nome, pois se os limites toleráveis demarcarem um intervalo de seis desvios-padrão entorno da média então, virtualmente 100% dos resultados estariam, para fins práticos, dentro da tolerância. Porém, como ressalta (TENNANT, 2001, p. 24-25), uma organização que almeja esse patamar, na prática opera duas vezes abaixo desse limite, isto é 3 Sigma – devido à imprevisibilidade do mundo real.

Na engenharia de qualidade, segundo (TAGUCHI *et al.*, 2004, p. 173), a capacidade de tolerância é preconizada como o comprimento do intervalo de tolerância normalizado por 6 vezes o desvio-padrão da medida observada – a influência do *Six Sigma* é óbvia.

Onde dados são um ativo corporativo, a qualidade de dados tem sido considerada como a adequação dos dados para o uso bem sucedido da implementação de processos corporativos de uma companhia (DAMA, 2009). Em (JUGULUM, 2014), o *Six Sigma* é aplicado aos processos para melhoria da qualidade de dados operacionais e corporativos e, portanto, a utilização de medidas de correlação.

Em estatística e na análise multivariada de dados, o coeficiente de variação (a razão entre a média e o desvio padrão) tem um papel básico para avaliar, respectivamente, a dispersão e a qualidade dos dados objeto de alguma pesquisa.

Os exemplos acima ilustram a importância do desvio-padrão como uma medida básica e clássica para decisões sobre a qualidade de processos produtivos ou que servirão de base para pesquisas científicas.

4.4 Considerações Finais

Esclarece-se que a quantificação de propriedades ou relações qualitativas é mais uma forma de se permitir a criação de novos conhecimentos – e não a única. Por

exemplo, (SHANNON, 1953, 1955) acreditava que a computação numérica seria o oposto do que se deveria esperar de um computador digital: “The shoe is rather on the other foot”. Ele acreditava que os computadores digitais estavam sendo utilizados para cálculos numéricos meramente por questões econômico-financeiras: “bread-and-butter work”. (SHANNON, 1955) volta a utilizar a expressão “tarefas não-numéricas”:

Programming computers to perform such non-numerical tasks is valuable in a number of ways. It widens our understanding of capabilities of this amazingly flexible tool, the general purpose computer; it seems certain that we have only scratched the surface of potentialities of such computing devices and each new application leads to new understanding.

Por outro lado, a capacidade de manter a contabilidade de quantidades é uma parte especial para alguns processos científicos de inferência, como é ilustrado por (FEYNMAN *et al.*, 2005, cap. 4), pelo exemplo dos “blocos de Denis”, ao explicar as propriedades da energia e como ela é utilizada pela física. (FEYNMAN *et al.*, 2005, cap. 4) destacam que não se sabe o que é energia, mas sim como ela se relaciona com os processos físicos: uma quantidade que se conserva.

Capítulo 5

Cálculo de Incertezas I

It may have been only by crude analogy, but some of the Founding Fathers were certainly conscious of the correspondence between their proposed system of constitutional checks and balances and Sir Isaac Newton's system of mechanics, in which the universe was held together and kept in order by a balance of counter action forces set in motion by a Primer Mover, but working endlessly thereafter by a sequence of cause and effect.

Don Krasher Price, *The Scientific Estate*

Dissimular é fingir não ter o que se tem. Simular é fingir ter o que não se tem.

Jean Baudrillard, *Simulacros e Simulações*

5.1 Introdução

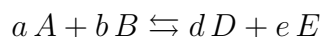
Ao comparar palavras o conhecimento prévio de características do idioma e da gramática em que as palavras foram escritas pode influenciar a decisão em considerá-las similares ou dissimilares. Ou seja, o contexto e o conhecimento linguístico do agente que efetua a comparação pode influenciar a decisão. Como foi visto anteriormente, quando fatores psicológicos são eliminados resta somente, então, a capacidade de uma mensagem discriminar-se das demais: mensagens que seriam similares são consideradas completamente diferentes, mesmo que difiram de um único símbolo alfabético dentre vários. Assim a uma decisão inteligente, no que diz respeito ao processo de comparação, envolverá necessariamente a ponderação de perspectivas diferentes, mas criadas a partir de um único objeto (palavras sofrendo comparação). Tais perspectivas são uma construção com base nos dados e no conhecimento interno do próprio agente inteligente (características linguísticas como a ortografia e a fonética, por exemplo) – caso contrário, seria uma decisão sem articulação e ponderação: ou são iguais, ou são diferentes – sem espaço para um meio-termo.

Neste capítulo será, entretanto, apresentado um paradigma de decisão inspirado no equilíbrio químico, que será utilizado no capítulo 6, para ponderar similaridade e dissimilaridades entre palavras. Portanto este capítulo versará somente sobre propriedades matemáticas do modelo proposto. As grandezas primordiais utilizadas aqui são a surpresa e a raridade, pois têm conexões com a teoria da informação.

5.2 Equilíbrio Químico como Paradigma de Decisão

Em (KOTZ e PURCELL, 1991, cap. 16), o equilíbrio (ou desequilíbrio) de algumas reações químicas reversíveis podem ser expressos como razões ponderadas em função da concentração de produtos e reagentes. Com isso, é possível dizer se uma reação química encontra-se em equilíbrio ou, em caso contrário, evoluirá para uma direção específica de forma a alcançar o equilíbrio. A situação de equilíbrio é aquela na qual, apesar de as reações químicas continuarem a ocorrer, as concentrações de produtos e reagentes permanecem as mesmas. Isso ocorre, pois, tanto produtos como reagentes são gerados na mesma taxa em que são consumidos: o que caracteriza um estado estacionário.

Um exemplo de reação genérica com dois reagentes A e B e dois produtos D e E :



Onde a , b , d e e são, respectivamente, os coeficientes estequiométricos de A , B , D e E . Esses coeficientes são números racionais não negativos e indicam proporcionalidade efetiva de cada substância química presente na reação.

A fórmula que descreve a proporção, no equilíbrio, entre as substâncias químicas envolvidas nessa reação é:

$$K_c = \frac{[D]^d \cdot [E]^e}{[A]^a \cdot [B]^b} \quad (5.2.1)$$

Onde $[D]$, $[E]$, $[A]$ e $[B]$ denotam a concentração das substâncias químicas (em mol/l , por exemplo) e K_c é a constante de equilíbrio¹, tal que $K_c \in \mathbb{R}_*^+$.

Como as concentrações são sempre números não-negativos, a relação 5.2.1 pode ser reescrita² como:

¹A constante de equilíbrio K_c depende da temperatura (KOTZ e PURCELL, 1991, cap. 16), o que é relacionado ao Princípio de Le Chatelier: “a change in any of the factors that determine the equilibrium conditions of a system will cause the system to change in such a manner as to reduce or counteract the effect of the change”.

²Em (CHAPRA e CANALE, 1990, p. 354 e 441), essa estratégia pode ser vista para encontrar coeficientes de equações de potências (modelos de sistemas mecânicos) via regressão para dados

$$\ln K_c = +d \ln[D] + e \ln[E] - a \ln[A] - b \ln[B] \quad (5.2.2)$$

Propriedade 5.2.1. Tendo como referência $\ln K_c$, os coeficientes dos produtos são positivos (+d e +e) e dos reagentes são negativos (-a e -b).

Se um “químico” tivesse que **decidir** a tendência da direção da reação – que não se encontra no equilíbrio – ele precisa calcular o coeficiente de reação Q e compará-lo com K_c :

$$Q = \frac{[D]^d \cdot [E]^e}{[A]^a \cdot [B]^b} \quad (5.2.3)$$

Se $Q < K_c$ então haverá uma diminuição nos reagentes e um aumento nos produtos. Se $Q > K_c$ então haverá uma diminuição nos produtos e um aumento nos reagentes. Se $Q = K_c$ então a reação está em equilíbrio (em um estado estacionário).

Exemplo 5.2.1. Equilíbrio Químico $[SO_3]$, $[SO_2]$ e $[O_2]$ são, respectivamente, as concentrações das substâncias: trióxido de enxofre, dióxido de enxofre e oxigênio. As concentrações são obtidas empiricamente. Seguem a reação química em equilíbrio e o cálculo de K_c (KOTZ e PURCELL, 1991, p. 667):



Os coeficientes estequiométricos de SO_2 e SO_3 é 1, mas é omitido. Então:

$$K_c = \frac{[SO_3]}{[SO_2] \cdot [O_2]^{1/2}} = 2.6 \times 10^{12} \text{ (a } 25^\circ C) \quad (5.2.5)$$

Observação 5.2.1. Os objetos intersubjetivos³ desse modelo são a constante de equilíbrio e os coeficientes estequiométricos. Tais objetos, em sua totalidade, são idealmente invariáveis para a mesma reação química, esteja ela em equilíbrio ou desequilíbrio. Tais objetos intersubjetivos são o conhecimento que se possui sobre a reação, pois permitem o **químico** saber o sentido do progresso de uma reação em particular em função de dados para essa reação: a concentração dos produtos e reagentes.

Espontaneidade das Reações Químicas

No estudo da espontaneidade das reações químicas (KOTZ e PURCELL, 1991, cap. 20), Q/K_c é o quociente de reação química, que mensura a espontaneidade

experimentais.

³Intersubjetividade: segundo (VIEIRA, 2009), no perspectivismo o problema fundamental não é decidir qual visão é a correta ou certa, mas sim permitir a produção de conhecimento. Entretanto, os pontos de vistas justificados devem permitir que sejam compartilhados e testáveis – assim havendo, portanto, a intersubjetividade científica preconizada por (POPPER, 2007, p. 46).

no sentido de formação de produtos. Caso $Q/K_c = 1$, a reação está em equilíbrio. Contudo se uma concentração escassa de produtos se verifica, situação caracterizada por $Q/K_c < 1$, então a reação não está em equilíbrio, e espontaneamente se desenvolverá, ao longo do tempo, para consumir reagentes para formar produtos até que $Q/K_c = 1$. Caso $Q/K_c > 1$, então a reação também não está em equilíbrio, pois existe concentração demasiada de produtos, e espontaneamente a reação se desenvolverá, ao longo do tempo, para consumir produtos para formar reagentes até que $Q/K_c = 1$.

A energia livre de Gibbs, denotada por ΔG é tal que $\Delta G = RT \ln Q/K_c$, onde R é o número de Avogrado e T a temperatura em kelvins. ΔG é uma medida do grau de espontaneidade da reação em direção aos produtos para alcançar o equilíbrio. Segue da interpretação do quociente de reação química (KOTZ e PURCELL, 1991, p. 839):

- $\Delta G \propto \ln Q/K_c$ (pois $R = 6,022 \times 10^{23} \text{ mol}^{-1}$ e $T > 0$);
- $\Delta G = 0$, a reação está em equilíbrio ($Q = K_c$);
- $\Delta G < 0$, a reação é espontânea em direção aos produtos ($Q < K_c$);
- $\Delta G > 0$, a reação não é espontânea em direção aos produtos ($Q > K_c$);

Um ponto importante é que produtos e reagentes são partícipes da reação: ambos passíveis de serem formados e consumidos. Dada as restrições impostas pela reação, o que caracteriza um produto é ser aquilo que se deseja obter e um reagente é ser aquilo que se está disposto a abrir mão.

Analogia como Paradigma de Decisão

Retornando à teoria da informação, observa-se que as grandezas apresentadas no capítulo 3 são necessariamente não-negativas: o número de mensagens (sejam típicas ou possíveis), a probabilidade dos símbolos alfabéticos, a incerteza e a própria quantidade de informação (seja ela média ou absoluta) – ver seção 3.5 (pág. 38).

Entretanto no cálculo da probabilidade do conjunto de mensagens típicas suficientemente longas (produzidas por uma fonte de informação que apresenta homogeneidade estatística) tende a um equilíbrio assintótico. A base para essa afirmação também se encontra destacada em (SHANNON, 1948):

(...) it is possible for most purposes to treat the long sequences as though there were just 2^{HN} of them, each with a probability 2^{-HN} . (...)

Nesta tese, resultado análogo é o do corolário 3.4.3 (pág. 33) – relações 3.4.11 e 3.4.12:

$$\mathcal{P}_N \sim b^{-N \cdot H_S} \text{ e } \mathcal{M}_N \sim b^{N \cdot H_S} \Rightarrow \mathcal{P}_N \cdot \mathcal{M}_N \sim b^{-N \cdot H_S} \cdot b^{N \cdot H_S} = b^{-N \cdot H_S + N \cdot H_S} = b^0 \quad (5.2.6)$$

Continuando, o expoente de $b^{-N \cdot H_S + N \cdot H_S}$ sugere um processo no qual figura a diferença de valores de entropia ponderados (mesmo que por pesos idênticos). Esse expoente é similar em forma ao segundo membro da expressão 5.2.2.

Por fim, resta perguntar se um mesmo objeto, do ponto de vista da teoria de probabilidades, pode originar diferentes medidas de incerteza? A resposta para essa questão é sim, pois (TRIVEDI, 1982, p. 4) explica que o espaço de eventos não é determinado completamente pelo experimento executado, mas também depende do propósito do experimento. Assim, é possível que a partir do mesmo objeto se obtenha medidas de incerteza diferentes, para cada perspectiva objetiva – portanto garantindo a intersubjetividade.

5.3 Médias Ponderadas

Sobre a notação de somatórios e a de produtórios aqui empregada, ver seção 1.10 (pág. 9).

Definição 5.3.1 (Distribuição Discreta de Probabilidades). *Se p é uma distribuição discreta de probabilidades (ddp) então:*

- $\sum_i p_i = 1$
- $\forall i, p_i > 0$

Onde p_i é a probabilidade do i -ésimo evento. Além disso, a grandeza $\frac{1}{p_i}$ é a raridade do i -ésimo evento e a grandeza $\ln \frac{1}{p_i}$ é a surpresa do i -ésimo evento.

O conceito de probabilidade de um evento dispensa apresentação. Entretanto, a raridade de um evento demanda uma discussão. Uma forma de interpretar o inverso de uma probabilidade de um evento é que ela mensura quão raro o evento é. Por exemplo, um evento com probabilidade 0,5 ocorreria, em média, 1 vez a cada 2 eventos. Um evento com probabilidade 0,01 ocorreria, em média, 1 vez a cada 100 eventos. Se a probabilidade de um evento se aproxima de zero, sua raridade se aproxima do infinito. Se sua probabilidade se aproxima de 1 então sua raridade se aproxima de 1 (assumindo o menor valor possível) - sendo um evento certo, portanto é um evento que ocorre 1 vez a cada 1 evento - ocorre sempre!

O conceito de surpresa é bem estabelecido na literatura como em (ROSS, 1998, pp. 438-441) e (SAYOOD, 2005): a ocorrência de evento com baixa probabilidade causa mais surpresa do que um com maior. Se o evento é certo de ocorrer a surpresa é, obviamente, zero; entretanto, a surpresa cresce mais lentamente do que a raridade. A surpresa é medida em *nats* por símbolo, quando a base do logaritmo é e – caso fosse a base 2 seria medida em *bits* por símbolo.

Incerteza de um evento é calculada da mesma forma que a sua surpresa. A diferença é que a incerteza de um evento é atribuída antes de ele ocorrer; a surpresa desse é atribuída depois da ocorrência do evento, como pode ser visto em (ROSS, 1998, p. 438).

Em (SAYOOD, 2005, p. 13-15) a quantidade chamada aqui de surpresa de um evento é denominada auto-informação (self-information). Porque eventos raros, do ponto de vista de eficiência de transmissão de informação, tendem a ser representados com palavras de código maiores que eventos frequentes, deixando as palavras de código menores para eventos com maior probabilidade, assim minimizando o número de escolhas e, portanto, maximizando o número de mensagens finais selecionadas pelo receptor da mensagem – para uma mesma quantidade de símbolos. Esse tipo de atribuição de palavra de códigos ocorre, por exemplo, no código de Huffman para compactação de dados, (SAYOOD, 2005, p. 41) e (SZWARCFITER e MARKENZON, 1994a, p. 294-306).

Definição 5.3.2. *Seja p é uma ddp com N eventos equiprováveis então $p_i = 1/N$ para $1 \leq i \leq N$.*

Definição 5.3.3 (Sequências Reais). *Se x é uma sequência real então $\forall i \in \mathbb{N}_*$, $x_i \in \mathbb{R}$. Se x é uma sequência real não-negativa então $\forall i \in \mathbb{N}_*$, $x_i \in \mathbb{R}_*^+$. Onde x_i é o valor associado do i -ésimo evento.*

Definição 5.3.4 (Sequências Aniquilantes). *Seja x é uma sequência real, x é aniquilante então $\sum_i x_i = 0$. Seja y é uma sequência real não-negativa, y é aniquilante então $\prod_i y_i = 1$.*

Definição 5.3.5 (Média Aritmética Ponderada). *Seja p uma ddp e x uma sequência real. A média aritmética de x ponderada por p é denotada por $\langle x \rangle$ ⁴:*

$$\langle x \rangle = \sum_i p_i \cdot x_i \tag{5.3.1}$$

Se $f: \mathbb{R} \rightarrow \mathbb{R}$ então:

⁴A notação $\langle . \rangle$ é a média em geral na mecânica estatística (SETHNA, 2011, p. 16). Na teoria da informação, um ramo da probabilidade e estatística, a notação para a média é $E[.]$ (COVER e THOMAS, 1991; KULLBACK, 1978)

$$\langle f(x) \rangle = \sum_i p_i \cdot f(x_i) \quad (5.3.2)$$

Definição 5.3.6 (Média Geométrica Ponderada). *Seja p uma ddp e x uma sequência real não-negativa. A média geométrica de x ponderada por p é denotada por $[x]$:*

$$[x] = \prod_i x_i^{p_i} \quad (5.3.3)$$

Se $f: \mathbb{R} \rightarrow \mathbb{R}_*^+$ e x uma sequência real então:

$$[f(x)] = \prod_i (f(x_i))^{p_i} \quad (5.3.4)$$

Para mais informações sobre média geométrica ver (MANFRINO *et al.*, 2010, p. 9), (SMITH, 2005) e (ABRAHAM, 2012, p. 291). Para a relação entre as médias aritmética, geométrica e harmônica, ver (SPIEGEL, 1972, cap. 3)

5.4 Médias, Entropia e Ponderação

Para demonstrar as propriedades 3.5.1, que versam sobre relações entre a informação de Hartley e de Shannon, empregou-se a clássica relação entre a média geométrica e a média aritmética (a relação 3.5.2, pág. 39). Nesta seção, a entropia de Shannon será derivada utilizando-se relação entre a média aritmética e geométrica como ponto de partida – assim garantindo um caráter puramente matemático – o que permite interpretação e uso independentes de domínio de aplicação.

Logo a seguir, as definições de pseudo-entropia e entropia abstrata serão apresentadas. A partir dessas definições, restrições serão acrescentadas a elas até se obter a entropia de (SHANNON, 1948).

Teorema 5.4.1 (Pseudo-Entropia). *Sejam $a, b \in \mathbb{R}_*^+$, $a \neq 1$ e $b \neq 1$ e Se $\log_b [x]$ existe ou $\langle \log_b x \rangle$ existe então:*

$$\log_b [x] = \langle \log_b x \rangle \quad (5.4.1)$$

$$\log_b \left[\frac{1}{x} \right] = -k \cdot \sum_i p_i \cdot \log_a x_i \quad (5.4.2)$$

e ambos existem e $k = \log_b a$

Equivalentemente, a relação 5.4.2 pode ser reescrita como:

$$\log_b \left[\frac{1}{x} \right] = k \cdot \left\langle \log_a \frac{1}{x} \right\rangle \quad (5.4.3)$$

Demonstração.

$$\log_b [x] = \log_b \prod_i x_i^{p_i} = \sum_i \log_b x_i^{p_i} = \sum_i p_i \cdot \log_b x_i = \langle \log_b x \rangle \quad (5.4.4)$$

Por fim,

$$\log_b \left[\frac{1}{x} \right] = \log_b \prod_i \left(\frac{1}{x_i} \right)^{p_i} = \log_b \prod_i x_i^{-p_i} = \sum_i \log_b x_i^{-p_i} = \sum_i -p_i \cdot \log_b x_i = \quad (5.4.5)$$

$$= \frac{\log_b a}{\log_b a} \cdot \sum_i -p_i \cdot \log_b x_i = \frac{\log_b a}{\log_b a} \cdot \sum_i -p_i \cdot \log_b x_i \quad (5.4.6)$$

$$= -\log_b a \cdot \sum_i p_i \cdot \frac{\log_b x_i}{\log_b a} = -\log_b a \cdot \sum_i p_i \cdot \log_a x_i \quad (5.4.7)$$

□

Esse resultado foi chamado de pseudo-entropia, pois tem a forma simbólica da entropia formulada por (SHANNON, 1948) entretanto:

- x_i não são necessariamente probabilidades, pois $x_i > 0$ (não há restrição $x_i < 1$).
- $\log_b [x]$ (ou $\langle \log_b x \rangle$) podem ser negativas dependendo dos valores de b e dos valores de x_i .

Como um refinamento, será imposto que a base tenha um valor fixo positivo. A escolha de b será e , pois facilita operações envolvendo cálculo diferencial e integral.

Corolário 5.4.2 (Pseudo-Entropia Padrão). *Se $\ln [x]$ existe ou $\langle \ln x \rangle$ existe então:*

$$\ln [x] = \langle \ln x \rangle \quad (5.4.8)$$

$$\ln \left[\frac{1}{x} \right] = - \sum_i p_i \cdot \ln x_i \quad (5.4.9)$$

Equivalentemente, a relação 5.4.9 pode ser reescrita como:

$$\ln \left[\frac{1}{x} \right] = \left\langle \ln \frac{1}{x} \right\rangle \quad (5.4.10)$$

Demonstração. Caso particular do teorema 5.4.1, onde $a = b = e$. □

Outro tipo de refinamento, que pode ser imposto é o de que a sequência x seja a função de distribuição discreta de probabilidades, mas permitindo que $b > 0$ e $b \neq 1$.

Corolário 5.4.3 (Entropia Abstrata). *Seja $a, b \in \mathbb{R}_*^+$ e $a \neq 1$ e $b \neq 1$. $[p]$ existe e p é a mesma ddp para calcular $[p]$ então:*

$$\log_b [p] = \langle \log_b p \rangle \quad (5.4.11)$$

$$\log_b \left[\frac{1}{p} \right] = -k \cdot \sum_i p_i \cdot \log_a p_i \quad (5.4.12)$$

e ambos existem e $k = \log_b a$.

Equivalentemente, a relação 5.4.12 pode ser reescrita como:

$$\log_b \left[\frac{1}{p} \right] = k \cdot \left\langle \log_a \frac{1}{p} \right\rangle \quad (5.4.13)$$

Demonstração. Direta do teorema 5.4.1, pois trata-se de um caso particular, no qual $f = p$. \square

O refinamento final impõe as duas restrições, isto é: que a sequência x seja a própria função de distribuição discreta de probabilidades e que $a = b = e$: obtendo-se a entropia de (SHANNON, 1948) medida em *nats*. Um efeito colateral desse refinamento é que se perde a possibilidade de valores negativos para o cálculo da entropia abstrata. Entretanto isso é perdido por conta da escolha da base e como referência e da restrição $0 < x_i < 1$. Nas seções seguintes esse inconveniente será contornado.

Corolário 5.4.4 (Entropia Abstrata Padrão). *Seja p uma ddp. Se $a = b = e$ e $[p]$ é média geométrica de p ponderada por p então:*

$$\ln [p] = \langle \ln p \rangle \quad (5.4.14)$$

$$\ln \left[\frac{1}{p} \right] = - \sum_i p_i \cdot \ln p_i \quad (5.4.15)$$

Equivalentemente, a relação 5.4.15 pode ser reescrita como:

$$\ln \left[\frac{1}{p} \right] = \left\langle \ln \frac{1}{p} \right\rangle \quad (5.4.16)$$

Demonstração. Direta do corolário 5.4.2, pois trata-se de um caso particular, no qual $x = p$. \square

Comentário 5.4.5. Para fins de interpretação:

- $\left[\frac{1}{p} \right]$ é a média geométrica da raridade dos eventos (a raridade do i -ésimo evento é $\frac{1}{p_i}$) ponderada pela ddp p . Essa média será sempre maior que 1, caso não exista evento com probabilidade 1;

- $\left\langle \ln \frac{1}{p} \right\rangle$ é a média aritmética da surpresa dos eventos (a surpresa do i -ésimo evento é $\ln \frac{1}{p_i}$) ponderada pela ddp p . Essa média será sempre maior que 0, caso não exista evento com probabilidade 1.

Para fins de simplificação de cálculos assume-se (COVER e THOMAS, 1991, p. 13):

$$0 \log_b 0 = \lim_{N \rightarrow +\infty} \frac{1}{N} \cdot \log_b \frac{1}{N} = 0$$

Agora serão apresentados os procedimentos de ponderação. Irão demonstrar que uma grandeza é preservada no tocante a ponderação de médias de surpresa e médias de raridade oriundas de processos aleatórios discretos modelados por uma ddp .

Teorema 5.4.6 (Ponderação Pseudo-Entrópica). *Sejam $w_j \in \mathbb{R}$, x^j uma sequência real não-negativa e $[x^j]$ existe e calculada em função da ddp p^j , onde $j \in \mathbb{N}_*$:*

$$\ln \prod_j [x^j]^{w_j} = \sum_j w_j \cdot \langle \ln x^j \rangle \quad (5.4.17)$$

$$\ln \prod_j \left[\frac{1}{x^j} \right]^{w_j} = \sum_j w_j \cdot \left\langle \ln \frac{1}{x^j} \right\rangle \quad (5.4.18)$$

Demonstração.

$$\ln \prod_j [x^j]^{w_j} = \sum_j \ln [x^j]^{w_j} = \sum_j w_j \cdot \ln [x^j] = \sum_j w_j \cdot \langle \ln x^j \rangle \quad (5.4.19)$$

$$\ln \prod_j \left[\frac{1}{x^j} \right]^{w_j} = \sum_j \ln \left[\frac{1}{x^j} \right]^{w_j} = \sum_j w_j \cdot \ln \left[\frac{1}{x^j} \right] = \sum_j w_j \cdot \left\langle \ln \frac{1}{x^j} \right\rangle \quad (5.4.20)$$

□

Corolário 5.4.7 (Ponderação Entrópica Abstrata). *Sejam $w_j \in \mathbb{R}$ e p^j é a mesma ddp para calcular $[p^j]$, onde $j \in \mathbb{N}_*$:*

$$\ln \prod_j [p^j]^{w_j} = \sum_j w_j \cdot \langle \ln p^j \rangle \quad (5.4.21)$$

$$\ln \prod_j \left[\frac{1}{p^j} \right]^{w_j} = \sum_j w_j \cdot \left\langle \ln \frac{1}{p^j} \right\rangle \quad (5.4.22)$$

Demonstração. Direta do teorema 5.4.6, pois trata-se de um caso particular, no qual $x^j = p^j$. □

Corolário 5.4.8 (Entropia de Shannon em *nats*). *Sejam p uma ddp e $H_N(p)$ a entropia de Shannon de p calculada em nats por símbolo:*

$$H_N(p) = \left\langle \ln \frac{1}{p} \right\rangle \quad (5.4.23)$$

Então a constante de conversão de bits em nats é $k = \ln 2$

Demonstração. A entropia de Shannon, via de regra, encontra-se em *bits*.

$\sum_i p_i \cdot \log_2 p_i$ é o valor da entropia de p em *bits por símbolo*. Seja k a constante de conversão então:

$$H_N(p) = -k \cdot \sum_i p_i \cdot \log_2 p_i$$

Por hipótese:

$$\left\langle \ln \frac{1}{p} \right\rangle = -k \cdot \sum_i p_i \cdot \log_2 p_i$$

Pela relação 5.4.16 no corolário 5.4.4:

$$\ln \left[\frac{1}{p} \right] = -k \cdot \sum_i p_i \cdot \log_2 p_i$$

Pela relação 5.4.12 no corolário 5.4.3:

$$-\ln 2 \cdot \sum_i p_i \cdot \log_2 p_i = -k \cdot \sum_i p_i \cdot \log_2 p_i$$

$$\ln 2 = k$$

Então k é a constante de conversão de *bits* em *nats*. □

Corolário 5.4.9 (Entropia de Contagem). *Seja p é uma ddp com Ω eventos equiprováveis então:*

$$H_N(p) = \ln \Omega \quad (5.4.24)$$

Demonstração.

$$H_N(p) = \left\langle \ln \frac{1}{p} \right\rangle = \sum_{i=1}^{\Omega} \frac{1}{\Omega} \ln \frac{1}{\frac{1}{\Omega}} = \frac{\Omega}{\Omega} \ln \frac{1}{\frac{1}{\Omega}} = \ln \frac{1}{\frac{1}{\Omega}} = \ln \Omega \quad (5.4.25)$$

□

Corolário 5.4.10 (Ponderação Entrópica). *Sejam $w_j \in \mathbb{R}$ e p^j é a mesma ddp para calcular $[p^j]$, onde $j \in \mathbb{N}_*$:*

$$\ln \prod_j \left[\frac{1}{p^j} \right]^{w_j} = \sum_j w_j \cdot H_N(p^j) \quad (5.4.26)$$

Demonstração. Direta dos corolários 5.4.7 e 5.4.8. □

A ponderação entrópica pode assumir valores positivos ou negativos, dependendo dos pesos w_i . Caso $w_j > 0$ então $H_N(p^j)$ é análoga à ordem de grandeza da concentração de um produto (aquilo que se deseja maximizar). Caso $w_j < 0$ então $H_N(p^j)$ é análoga à ordem de grandeza da concentração de um reagente (aquilo que se deseja minimizar). Caso $w_j = 0$, $H_N(p^j)$ não tem qualquer influência no resultado. $|w_i|$ indica quão relevante é a concentração de $H_N(p^j)$. É importante enfatizar que a relação 5.4.26 têm características formais similares às das relações 5.2.1 e 5.2.2 no estudo do equilíbrio químico (pag. 52) – excetuando-se pela ausência de uma constante de equilíbrio. Porém isso será tratado na próxima seção.

5.5 Princípios de Procedimentos de Decisão

Finalmente é representado um modelo de ponderação de raridades e outro de ponderação de surpresas. Será argumentado que ambos estão relacionados. Para esses modelos, uma decisão é meramente uma grandeza oriunda de um processo de ponderação. Tal grandeza pode ser fornecida, *a posteriori*, como dado de entrada de um processo inteligente que efetue alguma computação objetivando, por exemplo, efetuar (ou não) uma ação. Serão apresentadas algumas propriedades matemáticas desses modelos, em especial a possibilidade de superposição de decisões, para em outra oportunidade o modelo ser empregado num processo de comparação de palavras (capítulo 6).

O modelo que pondera raridades é chamado de quantitativo e o que pondera surpresas é chamado de qualitativo. Uma motivação para essa escolha de nomenclatura é que os sistemas de decisão quantitativos sempre emitirão como julgamento uma grandeza descrita somente por valores positivos. Porém, os qualitativos emitirão como julgamento uma grandeza que assume valores que podem ser positivos ou negativos (dando a ideia de duas classes opostas, o que sugere uma dicotomia).

Definição 5.5.1 (Sistema de Decisão Quantitativa). *Sejam p e w sequências cujos termos são tais que $w_j \in \mathbb{R}$ e p^j é a mesma ddp para calcular $[p^j]$, onde $1 \leq j \leq n$, $n \in \mathbb{N}_*$. $K_{\text{Qt}} \in \mathbb{R}_*^+$ é a constante do equilíbrio quantitativo. Q_{Qt} é coeficiente de decisão quantitativa, (ou momento quantitativo) tal que:*

$$Q_{\text{Qt}}(w, p) = \prod_j \left[\frac{1}{p^j} \right]^{w_j} \quad (5.5.1)$$

ΔG_{Qt} é a espontaneidade da decisão quantitativa, tal que:

$$\Delta G_{\text{Qt}} = T \ln \frac{Q_{\text{Qt}}}{K_{\text{Qt}}} \quad (5.5.2)$$

Onde T é a pseudo-temperatura e define a base do logaritmo; e $T = 1$ para o

logaritmo natural.

$\frac{Q_{Qt}}{K_{Qt}}$ é a ponderação quantitativa. Dados w e p , se $\frac{Q_{Qt}}{K_{Qt}} > 1$ então Q_{Qt} é uma decisão favorável aos produtos (termos com coeficientes $w_j > 0$); se $\frac{Q_{Qt}}{K_{Qt}} < 1$ então Q_{Qt} é decisão favorável aos reagentes (termos com coeficientes $w_j < 0$). Se $\frac{Q_{Qt}}{K_{Qt}} = 1$ é uma decisão tomada em hesitação.

O **juízo quantitativo**, denotado por Q_{Qt} , é tal que:

$$Q_{Qt} = e^{\Delta G_{Qt}} \quad (5.5.3)$$

S é um sistema forte de decisão quantitativa à pseudo-temperatura T então:

$$S_{Qt} = (Q_{Qt}, Q_{Qt}, K_{Qt})_T \quad (5.5.4)$$

A definição de sistema de decisão quantitativa nada mais é do que uma apropriação do modelo matemático do equilíbrio químico. Nessa apropriação, entretanto, no lugar de utilizar concentrações de substâncias, utiliza a raridade como insumo para determinação de um equilíbrio de natureza abstrata.

Definição 5.5.2 (Sistema de Decisão Qualitativa). *Sejam p e w seqüências cujos termos são tais que $w_j \in \mathbb{R}$ e p^j é a mesma ddp para calcular $[p^j]$, onde $1 \leq j \leq n$, $n \in \mathbb{N}_*$. $K_{Q1} \in \mathbb{R}$ é a constante do equilíbrio quantitativo. Q_{Q1} é coeficiente de decisão qualitativa, (ou momento qualitativo) tal que:*

$$Q_{Q1}(w, p) = \sum_j w_j \cdot H_N(p^j) \quad (5.5.5)$$

ΔG_{Q1} é a espontaneidade da decisão qualitativa, tal que:

$$\Delta G_{Q1} = T \cdot (Q_{Q1} - K_{Q1}) \quad (5.5.6)$$

Onde T é a pseudo-temperatura e define a base do logaritmo; e $T = 1$ para o logaritmo natural (a pseudo-temperatura padrão).

$Q_{Q1} - K_{Q1}$ é a ponderação qualitativa. Dados w e p , se $Q_{Q1} - K_{Q1} > 0$ então Q_{Q1} é uma decisão favorável aos produtos (termos com coeficientes $w_j > 0$); se $Q_{Q1} - K_{Q1} < 0$ então Q_{Q1} é decisão favorável aos reagentes (termos com coeficientes $w_j < 0$). Se $Q_{Q1} - K_{Q1} = 0$ é uma decisão tomada em hesitação.

O **juízo qualitativo**, denotado por Q_{Q1} , é tal que:

$$Q_{Q1} = \Delta G_{Q1} \quad (5.5.7)$$

S é um sistema forte de decisão qualitativa à pseudo-temperatura T então:

$$S_{Q1} = (Q_{Q1}, Q_{Q1}, K_{Q1})_T \quad (5.5.8)$$

A definição de sistema de decisão qualitativa nada mais é do que uma apropriação do modelo matemático do equilíbrio químico. Nessa apropriação, entretanto, no lugar de utilizar concentrações de substâncias, utiliza a surpresa como insumo para determinação de um equilíbrio de natureza abstrata.

Observação 5.5.1. A escolha do logaritmo natural tem alguma parcela de arbitrariedade, mas não impede a mudança da unidade de incerteza – que pode ser modificada arbitrando-se uma outra temperatura coerentemente com a unidade desejada. Porém frisa-se que a utilização da base natural facilita eventuais trabalhos de cunho analítico – então a escolha dessa base também foi pelo bem da parcimônia.

Observação 5.5.2. Como os pesos do momento qualitativo são números reais, mas não são oriundos da medição com base num padrão de medida, então são também números adimensionais. Assim as unidades do momento são as unidades atribuídas ao logaritmo – no caso do logaritmo natural é o *nat* e no do \log_2 é o *bit*. A escolha do *nat* é simplesmente facilitar operações típicas do cálculo diferencial e integral. Para mais informações sobre unidades de medida de informação, ver seção 3.6 na página 41. O corolário 5.4.3 pode ser utilizado para fazer essa conversão de *bits* em *nat* de forma análoga como foi feito no corolário 5.4.8. Assim, para mudar as unidades do momento qualitativo basta mudar a base do logaritmo natural (no corolário 5.4.8) para corresponder a base da unidade desejada.

Nos modelos de decisão apresentados nas definições 5.5.1 e 5.5.2, a pseudo-temperatura é um fator de conversão (de unidade) e, portanto, permite que uma quantidade abstrata seja contabilizada meticulosamente. Essa quantidade abstrata tem como procedimento básico de mensuração a função H_N , que usa como padrão a base natural. Os julgamentos (qualitativos ou quantitativos) são a computação de interesse – são uma forma de saída do processo de computação.

Observação 5.5.3. Ao longo desta tese, na prática, somente julgamentos qualitativos serão utilizados.

Propriedade 5.5.1 (Coeficientes de Decisão). Pelas definições 5.5.1 e 5.5.2:

- Coeficiente Quantitativo de Decisão (Momento Quantitativo): $Q_{\mathbf{Qt}}(w, p)$ é uma ponderação de médias geométricas de frequências – e $Q_{\mathbf{Qt}}(w, p) \in \mathbb{R}_*^+$.
- Coeficiente Qualitativo de Decisão (Momento Qualitativo): $Q_{\mathbf{Ql}}(w, p)$ é uma ponderação de médias aritméticas de surpresas – e $Q_{\mathbf{Ql}}(w, p) \in \mathbb{R}$.

Observação 5.5.4. Sobre o termo “qualitativo de momento”, apesar de ser uma grandeza escalar, cuja unidade é *nat*, é importante dizer que a *raison d’être* desse modelo será sua utilização em processos que envolvam decisão e que as distribuições discretas de probabilidades não precisam estar relacionadas a uma variável aleatória

numérica nem mesmo ordinal. As variáveis qualitativas podem ser nominais (não ordenadas) e ordinais, caso haja uma ordenação (BUSSAB e MORETTIN, 1997). Assim uma variável que apresente uma relação de ordem total, ou parcial, é denominada variável qualitativa ordinal. Caso não apresente uma relação de ordem é dita nominal (BUSSAB e MORETTIN, 1997; PEREIRA, 2004). Tanto para (BUSSAB e MORETTIN, 1997; PEREIRA, 2004; PRIVITERA, 2012), as variáveis quantitativas são classificadas em contínuas ou discretas.

Teorema 5.5.5 (Equivalência entre Sistemas de Decisão). *Sejam $w_j \in \mathbb{R}$ e p^j é a mesma ddp para calcular $[p^j]$, onde $j \in \mathbb{N}_*$. A relação entre $Q_{Qt}(w, p)$ (na definição 5.5.1) e $Q_{Ql}(w, p)$ (na definição 5.5.2) é:*

$$Q_{Qt}(w, p) = e^{Q_{Ql}(w, p)} \text{ e } \ln Q_{Qt}(w, p) = Q_{Ql}(w, p) \quad (5.5.9)$$

Além disso, sejam K_{Qt} e K_{Ql} então:

$$K_{Qt} = e^{K_{Ql}} \text{ e } \ln K_{Qt} = K_{Ql} \quad (5.5.10)$$

$$\forall T \in \mathbb{R}_*^+ : \Delta G_{Qt} = \Delta G_{Ql} \quad (5.5.11)$$

$$\forall T \in \mathbb{R}_*^+ : Q_{Qt} = e^{Q_{Ql}} \text{ e } \ln Q_{Qt} = Q_{Ql} \quad (5.5.12)$$

Demonstração. A prova da relação 5.5.9 é direta do corolário 5.4.10. A prova da relação 5.5.10 é imediata da definição de K_{Qt} e de K_{Ql} – observando-se que as funções logaritmo e exponencial naturais são estritamente crescentes; e uma é a inversa da outra. Seja $T \in \mathbb{R}_*^+$, para relação 5.5.11:

$$\Delta G_{Ql} = T \cdot (Q_{Ql} - K_{Ql}) = T \cdot (\ln Q_{Qt} - \ln K_{Qt}) = T \cdot \ln \frac{Q_{Qt}}{K_{Qt}} = \Delta G_{Qt}$$

Para relação 5.5.12, como $\Delta G_{Ql} = \Delta G_{Qt}$ e pela definições de Q_{Qt} e Q_{Ql} , expressões 5.5.7 e expressões 5.5.7:

$$\Delta G_{Ql} = \Delta G_{Qt}$$

$$Q_{Ql} = \Delta G_{Qt}$$

$$Q_{Ql} = \ln e^{\Delta G_{Qt}}$$

$$Q_{Q1} = \ln Q_{Qt}$$

e

$$e^{Q_{Q1}} = Q_{Qt}$$

□

O teorema 5.5.5 permite que se trabalhe com sistemas lineares no formato usual (usando-se somas e multiplicação) no lugar de utilizar, respectivamente, multiplicações e potenciações (típicas do paradigma do equilíbrio químico). O apêndice D exemplifica um campo vetorial n-dimensional com soma e produto por escalar como base nas multiplicações e potenciações típicas do \mathbb{R}_*^+ .

Teorema 5.5.6 (Superposições Qualitativas). *Sejam $Q = \sum_i \gamma_i Q_{Q1i}$ e $\Delta G = \sum_i \gamma_i \Delta G_{Q1i}$, $K = \sum_i \gamma_i K_{Q1i}$, $Q = \sum_i \gamma_i Q_{Q1i}$ tais que: Q_{Q1i} é o coeficiente de decisão qualitativa do i -ésimo sistema de decisão qualitativa; ΔG_{Q1i} é a espontaneidade do i -ésimo sistema de decisão qualitativa; K_{Q1i} é a constante do equilíbrio qualitativo do i -ésimo sistema de decisão qualitativa; e Q_{Q1i} é o julgamento qualitativo do i -ésimo sistema de decisão qualitativa; Seja T é a pseudo-temperatura comum a todos os sistemas de decisão qualitativa; e $\gamma_i \in \mathbb{R}$. Então, à pseudo-temperatura T :*

- Q é o coeficiente de decisão qualitativa para algum sistema de decisão qualitativa;
- K é a constante do equilíbrio quantitativo para algum sistema de decisão qualitativa;
- ΔG é a espontaneidade de algum sistema de decisão qualitativa; e também é uma espontaneidade em função de Q e K ;
- Q é o julgamento qualitativo de algum sistema de decisão qualitativa;

Demonstração. Cada Q_{Q1i} , por definição, é um somatório no qual cada termo é o produto de um número real e um valor de entropia $Q_{Q1i} = \sum_j w'_{ij} H_N(p^{ij})$, então:

$$Q = \sum_i \gamma_i Q_{Q1i} = \sum_i \gamma_i \sum_j w'_{ij} H_N(p^{ij}) = \sum_i \sum_j \gamma_i w'_{ij} H_N(p^{ij})$$

Fazendo $w''_{ij} = \gamma_i \cdot w'_{ij}$ então:

$$Q = \sum_i \sum_j w''_{ij} H_N(p^{ij})$$

Seja \mathcal{B} uma bijeção⁵ tal que $\exists k \in \mathbb{N}_*$, $k = \mathcal{B}(i j)$, $w_k = w''_{i j}$ e $p^k \equiv p^{i j}$ então:

$$Q = \sum_k w_k H_N(p^k)$$

Logo Q é um coeficiente de decisão qualitativa. Continuando,

$$K = \sum_i \gamma_i K_{\mathbf{Q}i} \in \mathbb{R}$$

Logo K é uma constante de equilíbrio quantitativo. Finalmente,

$$\Delta G = \sum_i \gamma_i \Delta G_{\mathbf{Q}i} = \sum_i \gamma_i T \cdot (Q_{\mathbf{Q}i} - K_{\mathbf{Q}i}) = T \cdot \left(\sum_i \gamma_i Q_{\mathbf{Q}i} - \sum_i \gamma_i K_{\mathbf{Q}i} \right)$$

$$\Delta G = T \cdot (Q - K)$$

Logo, ΔG é a espontaneidade da decisão de um sistema de decisão qualitativa – em função de Q , K e T ; Por fim, $\mathcal{Q} = \sum_i \gamma_i \mathcal{Q}_{\mathbf{Q}i} = \sum_i \gamma_i \Delta G_{\mathbf{Q}i} = \Delta G$ – então, \mathcal{Q} é o julgamento qualitativo de um sistema de decisão qualitativa – em função de ΔG .

□

A propriedade de *superposição* diz que se um sistema é submetido a diferentes estímulos então a resposta sistêmica a cada estímulo em particular pode ser computada individualmente; e os resultados somados para se obter a resposta total (CHAPRA e CANALE, 1990, p. 255). A propriedade de *proporcionalidade* é a de que a resposta, de um estímulo em particular, é uma computação que consiste na multiplicação do estímulo por uma quantidade constante – tal contribuição irá compor a resposta total (CHAPRA e CANALE, 1990, p. 255). A superposição e a proporcionalidade são propriedades necessárias de sistemas lineares (CHAPRA e CANALE, 1990, p. 255). No caso de um *sistema de decisão qualitativa*, cada estímulo é uma medida de incerteza (obtida em função de uma *ddp*).

Teorema 5.5.7 (Superposição de Sistemas Fortes de Decisão Qualitativa). *Sejam T uma pseudo-temperatura, S_i o i -ésimo sistema forte de decisão qualitativa à pseudo-temperatura T , $\gamma_i \in \mathbb{R}$ é o peso que o i -ésimo sistema forte de decisão qualitativa e $S = \sum_i \gamma_i S_{i\mathbf{Q}i}$ tal que:*

⁵A bijeção \mathcal{B} existe, pelo teorema Schroeder-Bernstein em (KRANTZ, 2004, p. 26), proposição 1.2 em (KRANTZ, 2004, p. 30) e (KRANTZ, 2004, p. 36 e 37): se dois conjuntos são contáveis ou têm a mesma cardinalidade então há uma bijeção entre seus elementos.

$$\sum_i \gamma_i S_{i\mathbf{QI}} \equiv \left(\sum_i \gamma_i \mathcal{Q}_{\mathbf{QI}i}, \sum_i \gamma_i Q_{\mathbf{QI}i}, \sum_i \gamma_i K_{\mathbf{QI}i} \right)_T \quad (5.5.13)$$

Então S é um sistema forte de decisão qualitativa à pseudo-temperatura T . Diz-se então que S é um sistema de decisão por superposição.

Demonstração. Pelo teorema 5.5.6, $\sum_i \gamma_i \mathcal{Q}_{\mathbf{QI}i}$, $\sum_i \gamma_i Q_{\mathbf{QI}i}$ e $\sum_i \gamma_i K_{\mathbf{QI}i}$ são, à pseudo-temperatura T , respectivamente, o julgamento qualitativo para algum sistema de decisão qualitativa; um coeficiente de decisão para algum sistema de decisão qualitativa; e uma constante de equilíbrio para algum sistema de decisão qualitativa. Além disso, $\sum_i \gamma_i \mathcal{Q}_{\mathbf{QI}i}$ é definido em função de $\sum_i \gamma_i Q_{\mathbf{QI}i}$ e $\sum_i \gamma_i K_{\mathbf{QI}i}$. Então, pela definição 5.5.2, S é um sistema forte de decisão qualitativa à pseudo-temperatura T . \square

Teorema 5.5.8. *Seja S é um sistema forte de decisão qualitativa à pseudo-temperatura T ; e também é um sistema de decisão por superposição tal que:*

$$S_{\mathbf{QI}} = \left(\sum_i \gamma_i \mathcal{Q}_{\mathbf{QI}i}, \sum_i \gamma_i Q_{\mathbf{QI}i}, \sum_i \gamma_i K_{\mathbf{QI}i} \right)_T \quad (5.5.14)$$

Então:

$$S_{\mathbf{Qt}} = \left(\prod_i \mathcal{Q}_{\mathbf{Qt}i}^{\gamma_i}, \prod_i Q_{\mathbf{Qt}i}^{\gamma_i}, \prod_i K_{\mathbf{Qt}i}^{\gamma_i} \right)_T \quad (5.5.15)$$

Demonstração. Pelo teorema 5.5.5, se pode reescrever essa relação 5.5.14 como:

$$S_{\mathbf{QI}} = \left(\sum_i \gamma_i \ln \mathcal{Q}_{\mathbf{Qt}i}, \sum_i \gamma_i \ln Q_{\mathbf{Qt}i}, \sum_i \gamma_i \ln K_{\mathbf{Qt}i} \right)_T \quad (5.5.16)$$

$$S_{\mathbf{QI}} = \left(\sum_i \ln \mathcal{Q}_{\mathbf{Qt}i}^{\gamma_i}, \sum_i \ln Q_{\mathbf{Qt}i}^{\gamma_i}, \sum_i \ln K_{\mathbf{Qt}i}^{\gamma_i} \right)_T \quad (5.5.17)$$

$$S_{\mathbf{QI}} = \left(\ln \prod_i \mathcal{Q}_{\mathbf{Qt}i}^{\gamma_i}, \ln \prod_i Q_{\mathbf{Qt}i}^{\gamma_i}, \ln \prod_i K_{\mathbf{Qt}i}^{\gamma_i} \right)_T \quad (5.5.18)$$

Pelo teorema 5.5.5, que relaciona sistemas de decisão quantitativa e qualitativa, e pela definição 5.5.1 de $S_{\mathbf{Qt}}$ então:

$$S_{\mathbf{Qt}} = \left(\prod_i \mathcal{Q}_{\mathbf{Qt}i}^{\gamma_i}, \prod_i Q_{\mathbf{Qt}i}^{\gamma_i}, \prod_i K_{\mathbf{Qt}i}^{\gamma_i} \right)_T \quad (5.5.19)$$

\square

Os pesos w_i , tanto para os sistemas de decisão quantitativa quanto os de decisão qualitativa, dizem respeito a um processo que pondera particularidades da mesma perspectiva (ou do mesmo julgador). Já os pesos γ_i dizem respeito à ponderação de sistemas com alguma capacidade de influenciar o resultado final (a síntese de diversos julgadores). Pode-se dizer que o compromisso dos pesos w_i é com a coerência imediata de uma parcela de um único julgamento: dão um sentido básico (a consistência de um único julgamento). Por outro lado, os pesos γ_i estão ligados à combinação de julgamentos: dão uma direção a um conjunto de julgamentos (fornecem a coerência para um conjunto de julgamentos). No capítulo 7, os pesos w_i serão utilizados para a inclusão da entropia diferencial nos modelos aqui propostos.

Corolário 5.5.9 (Coeficiente de Decisão de Contagem). *Seja $Q_{\mathbf{QI}}(w, p)$ um coeficiente de decisão qualitativa e p^j uma ddp com Ω_j eventos equiprováveis. $Q_C(w, \Omega)$ é o coeficiente de decisão de contagem tal que $Q_{\mathbf{QC}}(w, \Omega) = Q_{\mathbf{QI}}(w, p)$, então:*

$$Q_{\mathbf{QC}}(w, \Omega) = \sum_j w_j \cdot \ln \Omega_j \quad (5.5.20)$$

Demonstração. Direto do corolário 5.4.9 com base na relação 5.5.5 da definição 5.5.2. □

Definição 5.5.3 (Sistema Momento Qualitativo). *Um sistema momento qualitativo à pseudo-temperatura $\log_b \mathbf{e}$ é um sistema forte de decisão qualitativa S , tal que $K_{\mathbf{QI}} = 0$ e $T = \log_b \mathbf{e}$.*

Definição 5.5.4 (Sistema Momento Qualitativo de Contagem). *Um sistema momento qualitativo discreto é um sistema momento qualitativo à pseudo-temperatura $\log_b \mathbf{e}$ tal que $Q_{\mathbf{QI}} = Q_{\mathbf{QC}}(w, \Omega)$.*

Definição 5.5.5 (Sistema Momento Qualitativo em bits). *Um sistema momento qualitativo em bits é um sistema momento qualitativo à pseudo-temperatura $\log_b \mathbf{e}$ tal que $b = 2$.*

Corolário 5.5.10 (Sistema Momento Qualitativo de Contagem em bits). *Seja S um sistema momento qualitativo de contagem e, também, um sistema momento qualitativo em bits. Então:*

$$Q_{\mathbf{QI}} = \sum_j w_j \log_2 \Omega_j \quad (5.5.21)$$

Demonstração. Como S é um sistema momento qualitativo de contagem então $Q_{\mathbf{QI}} = Q_{\mathbf{QC}}(w, \Omega)$; e um sistema momento qualitativo em bits então $T = \log_b \mathbf{e}$ e $K_{\mathbf{QI}} = 0$. Pela definição 5.5.2, $Q_{\mathbf{QI}} = \Delta G_{\mathbf{QI}}$.

Assim, $Q_{\mathbf{QI}} = T_{\mathbf{QI}}(Q_{\mathbf{QI}} - K_{\mathbf{QI}}) = \log_2 \mathbf{e} (Q_{\mathbf{QI}} - 0) = \log_2 \mathbf{e} Q_{\mathbf{QI}}$.

Continuando, $\mathcal{Q}_{\mathbf{Q1}} = \log_2 \mathbf{e} \sum_j w_j \ln \Omega_j = \sum_j w_j \log_2 \mathbf{e} \ln \Omega_j$.

$$\mathcal{Q}_{\mathbf{Q1}} = \sum_j w_j \frac{\ln \mathbf{e}}{\ln 2} \ln \Omega_j = \sum_j w_j \frac{1}{\ln 2} \ln \Omega_j = \sum_j w_j \frac{\ln \Omega_j}{\ln 2} = \sum_j w_j \log_2 \Omega_j$$

□

Propriedade 5.5.2. Como os sistemas momento qualitativos de contagem são, por definição, sistemas de decisão qualitativa então eles podem ser combinados por superposição para se obter outros sistemas de decisão.

Aparentemente, o conceito de sistema momento qualitativo de contagem se desprende do paradigma probabilístico, pois se baseia na contagem de eventos. Entretanto, isso é conciliado quando se considera cada evento igualmente provável aos demais eventos contados. Pode-se interpretar o fenômeno das possibilidades, enquanto discretamente mensurável, como dado indutor de incertezas (uma grandeza contínua). Outro ponto de vista é o de que as possibilidades são um caso especial de probabilidades, mas sob a ótica do *Princípio de Razão Suficiente*, como em (JAYNES, 1957):

This problem of specification of probabilities in cases where little or no information is available, is as old as the theory of probability. Laplace's "Principle of Insufficient Reason" was an attempt to supply a criterion of choice, in which one said that two events are to be assigned equal probabilities if there is no reason to think otherwise.

Definição 5.5.6. *Sejam, somente, Ω possibilidades e a incerteza $\log_b \Omega$, onde $b > 1$. Então, $\log_b \Omega$ é a incerteza induzida na base b por Ω possibilidades. Quando a base está subentendida no contexto, omite-se "na base b ". Equivalentemente, Ω possibilidades induzem a incerteza $\log_b \Omega$.*

Assim pode-se dizer que um sistema momento qualitativo pondera exclusivamente incertezas induzidas por possibilidades distintas.

5.6 Discussão e Conclusões

Essencialmente, neste capítulo foram apresentados os conceitos de raridade e de surpresa e como estão relacionados pelo conceito fundamental de probabilidade. A raridade pode ser fruto de um processo de contagem objetivo para estimar as probabilidades de um evento. A raridade de um evento é o inverso da sua probabilidade e, ao mesmo tempo, considerado um atributo desse evento. A surpresa de

um evento nada mais é que o logaritmo da raridade podendo também ser considerado um atributo. Tanto a raridade média como a surpresa média dos resultados de um experimento estão relacionadas, respectivamente, pela média geométrica e média aritmética — usando-se como peso de ponderação a própria probabilidade dos eventos.

O paradigma do equilíbrio químico foi utilizado como inspiração para lidar com o fato de que a surpresa média e a raridade média não podem ser valores negativos assim como as concentrações de uma substância química numa reação não pode ser negativa — mas o que não impede de conceber processos abstratos onde suas contribuições concorram para evidenciar algum tipo de equilíbrio ou desequilíbrio. O tal paradigma é uma forma de sistema de *pesos-e-contrapesos* cujo resultado pode ser um insumo para um processo de decisão — a ser definido *a posteriori*. Esse sistema de *pesos-e-contrapesos* é evidente num sistema momento qualitativo.

No capítulo 6, sistemas momento qualitativo serão utilizados — e estarão associados à efetividade do conhecimento de um agente inteligente. Esse agente será *engendrado* para decidir em função de medidas de incerteza (derivadas de dados do problema) e do peso dessas incertezas (seu conhecimento interno).

Finalmente, ressalta-se que os mecanismos de ponderação e decisão aqui apresentados, em especial o de sistemas de decisão qualitativa, somente foi concebido para ponderar incertezas de distribuições de probabilidade — que podem modelar somente processos aleatórios discretos. A ponderação de incertezas em função de processos contínuos — que podem ser modelados por funções de densidade de probabilidade — será o objeto do capítulo 7.

Capítulo 6

Uma Mitologia para uma Casamento Aproximado de Palavras

É uma das inovações da linguagem humana, o encorajar a narração de estórias, e, por conseguinte, a imaginação criativa. A descoberta científica é semelhante à narração explanatória de estórias, à feitura de mitos, e à imaginação poética. O crescimento da imaginação realça, é claro, a necessidade de algum controle, tais como, na ciência, a crítica interpessoal – a amistosa cooperação hostil entre os cientistas que é baseada, parte na competição e parte no objetivo comum de lugar perto da verdade. (...)

Karl Raimund Popper, *Lógica das Ciências Sociais*

..., the ancient Greeks believed that the goddess **Aphrodite** was born from the splashing down of the severed genitalia of the sky god, **Uranus**. The ancient Greeks also believed that the goddess Aphrodite was the daughter of **Zeus**, the grandson of Uranus, and his wife, **Dione**. Obviously, to the 20th century mindset, these cannot be true. They are mutually exclusive. But not to the ancient Greeks. The ancient Greek tendency, the ancient Roman tendency, is to reconcile these things. I can predict confidently that it will probably, at some point or another, drive you crazy, but it's part of the territory. Please bear with me.

Joseph J. Hughes, Classical Mythology Class – Missouri State University

6.1 Introdução

Neste capítulo será modelado o *toy problem* para ilustrar a utilização de entropia como dado de entrada num processo inteligente de tomada de decisão por um único

neurônio *perceptron*. Isso será feito através da combinação e ponderação de medidas de distância entre *cadeias de caracteres* (ou strings) produzidas a partir de um par de palavras para efetuar um casamento aproximado de strings – um caso particular da igualdade entre strings, mas que leva em conta que tenham sofrido algum grau de comprometimento (BAEZA-YATES e NAVARRO, 1998; NAVARRO, 2001).

As *strings* aqui utilizadas representam um diminuto grupo de pares de palavra da Europa Ocidental no alfabeto latino (palavras inglesas e portuguesas), e parte delas apresentam erros ortográficos e/ou tipográficos (HERZOG *et al.*, 2007, p. 115-121, 131-135). Esse grupo de palavras foi escolhido para ilustrar o conceito proposto e, eventualmente, evidenciar peculiaridades, pois trata-se de uma prova de conceito.

Neste capítulo são desenvolvidos os conceitos do *A Mythology for an Approximate Word Matching: Entropy and Quality* de (PINTO e CARVALHO, 2015), sob a luz do arcabouço conceitual do capítulo 5, em especial: sistema momento qualitativo de contagem em *bits*, definição 5.5.4, (página 69).

Ao final, restará um único neurônio perceptron engendrado para classificar pares semelhantes de palavras como **matched** e pares diferentes como **not matched**.

6.2 Objetivos

Evidenciar que as medidas de entropia (incertezas) podem ser matéria-prima de um processo de decisão – não apenas uma quantidade a ser maximizada, ou minimizada, seja como viés de aprendizado; e não seja parte como princípio do *design* do algoritmo de decisão. Mais especificamente, objetiva-se fornecer evidências sobre a possibilidade do emprego do uso direto de entropia (em *bits*) como dados de entrada (*input data*) de um sistema de classificação para problema de casamento aproximado de strings por meio de um único neurônio perceptron.

6.3 Trabalhos Relacionados

Na literatura, há a existência de diversos algoritmos e processos para mensurar a diferença entre strings, são exemplos: a distância Hamming, a distância Levenshtein, a distância Damerau–Levenshtein e distância Jaro–Winkler. Em (COHEN *et al.*, 2003; DA SILVA *et al.*, 2007; FREIRE *et al.*, 2009; HERZOG *et al.*, 2007; NAVARRO, 2001; PISKORSKI *et al.*, 2009), se ilustra o emprego dessas distâncias entre strings em processos de casamento aproximado de cadeias (approximate string matching). Em (TAN *et al.*, 2009, pp. 78-100), métodos para mensuração de diferenças e similaridades entre dados de diferentes naturezas são apresentados.

Utilizar redes neurais para realizar casamento aproximado de *strings* não é algo novo, por exemplo: (TAMBOURATZIS, 2001) apresenta uma proposta de utilização

de grupos de redes neurais para tratar, com rapidez, processos de casamento de strings *on-line* e *off-line* aplicados a textos (para indicar pontos no texto onde são passíveis de edições potenciais); e (MIRZAEI *et al.*, 2007) utilizam uma rede neural Hopfield usando a técnica de *decaying self-feedback*. (OSMIC *et al.*, 2011) apresentam um modelo para o casamento de strings modelado como um sistema neuro-fuzzy. A entropia é empregada como um critério de otimização no processo de solução de problemas como em (PELAGOTTI e PIURI, 1997); ou como nas *Boltzmann machines* – modelo matemático que é a base do método de aprendizado descrito em (HAYKIN, 2009, p. 598-604); ou como em (NG *et al.*, 2003) para *prunning* de neurônios da rede; ou em (ATICK, 1992) como critério de pré-processamento de dados de entrada. (OSMIC *et al.*, 2011) apresentam paradigma *neuro-fuzzy* para o casamento de strings.

(CAMARENA-IBARROLA e CHAVEZ, 2006) fazem comparações entre peças musicais pela geração de um histograma de entropia (*entropygram*): cada peça é convertida, por método de amostragem, em uma *string* – para posteriormente às strings (referentes a cada peça) serem aplicados algoritmos de busca e alinhamento. Porém não fornecem medidas de entropia diretamente ao modelo, mas sim strings.

6.4 Definições

Definição 6.4.1 (Módulo, Máximo e Mínimo). *Sejam $x, y \in \mathbb{R}$. $|x| = \sqrt{x^2}$. $\max(x, y)$ denota o maior entre x e y . $\min(x, y)$ denota o menor entre x e y .*

Definição 6.4.2. $\Phi(z)$ é um algoritmo que retorna o valor $|z|$, se z é um número e em caso contrário, retorna símbolo $+\infty$.

A razão do algoritmo Φ é simplesmente detectar divisões por zero. O retorno do símbolo $+\infty$ será o inverso de uma grandeza que representa a distância zero.

Definição 6.4.3 (String). *Uma string \aleph é uma sequência finita de símbolos pertencentes a um conjunto chamado alfabeto. Se uma string não possui símbolos, ela é chamada de string vazia e denotada por ϵ . O comprimento (ou tamanho) de uma string \aleph , denotado por $|\aleph|$, é o total do número de símbolos do alfabeto que compõe a string. $\aleph[i]$ denota o i -ésimo símbolo alfabético (da esquerda para a direita) da string \aleph . Se $i < 1$ ou $|\aleph| < i$ então $\aleph[i]$ não está definido.*

Exemplo 6.4.1. Operações de Inserção, Substituição e Deleção

Inserção: Caro \rightarrow Ca?ro \rightarrow Carro (inserção do caracter **r** entre o **a** e o **r**)

Substituição: Carro \rightarrow C?rro \rightarrow Corro (substituição do caracter **a** por **o**)

Deleção: Carro \rightarrow Ca?ro \rightarrow Caro (deleção do caracter **r** entre o **a** e o **r**)

Observa-se que para cada operação de inserção, existe uma operação de deleção que a desfaz (e vice-versa). Para cada operação de substituição, existe exatamente outra operação de substituição que a desfaz.

Exemplo 6.4.2. Operações de Inserção, Substituição e Deleção

Deleção: Caro \leftarrow Ca?ro \leftarrow Carro (deleção do caracter **r** entre o **a** e o **r**)

Substituição: Carro \leftarrow C?rro \leftarrow Corro (substituição do caracter **o** por **a**)

Inserção: Carro \leftarrow Ca?ro \leftarrow Caro (inserção do caracter **r** entre o **a** e o **r**)

A distância Levenshtein é também conhecida como distância de edição. (STEPHEN, 1994, p.p. 39 e 40) observa que essa distância não é uma métrica, pois a desigualdade triangular não é garantida, apesar de ela ser não-negativa, simétrica, e zero somente para strings idênticas. Seu desenvolvimento foi inspirado na correção de informações em canais de transmissão binários que admitem – além da substituição de símbolos – a deleção e a inserção de símbolos como em (LEVENSHTEIN, 1966). A implementação do algoritmo utiliza técnicas de programação linear e comparação dos prefixos de strings (NAVARRO, 2001; STEPHEN, 1994). Assim, ela é mais genérica que a distância Hamming, que só permite substituições para transformar uma *string* em outra (e ambas devem ter o mesmo comprimento).

Definição 6.4.4 (Distância Levenshtein). $L(\aleph_1, \aleph_2)$ denota a distância Levenshtein entre \aleph_1 e \aleph_2 , que é o menor número possível de operações de edição (inserções, deleções ou substituições) necessárias para se transformar \aleph_1 em \aleph_2 , ou vice-versa (NAVARRO, 2001).

O resultado da distância $L(\aleph_1, \aleph_2)$ pode ser interpretado como o menor número de instruções que um programa precisa para a transformação de \aleph_1 em \aleph_2 (ou vice-versa). Uma sequência dessas operações de edição pode ser entendida como um programa *ad hoc* para efetuar tal transformação.

O teto para o valor da distância de Levenshtein entre duas strings é o tamanho da maior *string*, porque sempre se pode, por uma sequência de operações de substituições, e seguida de deleções, transformar a maior *string* na menor. Um argumento análogo também é válido no que tange à transformação da menor *string* na maior, mas no lugar de deleções emprega-se inserções.

Definição 6.4.5 (Número de Igualdades). A função $I(\aleph_1, \aleph_2)$ denota o número de igualdades entre \aleph_1 e \aleph_2 tal que:

$$I(\aleph_1, \aleph_2) = \max(|\aleph_1|, |\aleph_2|) - L(\aleph_1, \aleph_2) \quad (6.4.1)$$

O número de igualdades é o número de operações de edição em potencial que não foram realizadas. Se não foram realizadas substituições, inserções nem deleções de caracteres é porque os caracteres partícipes dessas operações já se encontravam adequadamente na *string* a ser transformada.

Definição 6.4.6. *Sejam \aleph_1 e \aleph_2 strings. $L_n(\aleph_1, \aleph_2)$ é distância Levenshtein normalizada, tal que:*

$$L_n(\aleph_1, \aleph_2) = \Phi \left(\frac{L(\aleph_1, \aleph_2)}{\max(|\aleph_1|, |\aleph_2|)} \right). \quad (6.4.2)$$

Definição 6.4.7. *Sejam \aleph_1 e \aleph_2 strings. $u(\aleph_1, \aleph_2)$ é função indicador de diferença entre \aleph_1 e \aleph_2 tal que:*

$$u(\aleph_1, \aleph_2) = \begin{cases} 0 & , \text{ se } L(\aleph_1, \aleph_2) = 0 \\ 1 & , \text{ senão} \end{cases} \quad (6.4.3)$$

A função de indicação de diferença u será útil para expressar um grau de incerteza pelo simples fato de um agente inteligente saber que $\aleph_a \neq \aleph_b$.

Exemplo 6.4.3. Sejam \aleph_1 , \aleph_2 e \aleph_3 strings tais que $\aleph_1 = abaabcba$, $\aleph_2 = \epsilon$ e $\aleph_3 = c$ então $\aleph_1[6] = c$, $|\aleph_1| = 8$, $|\aleph_2| = 0$, $|\aleph_3| = 1$, $\max(|\aleph_1|, |\aleph_3|) = 8$, $\min(|\aleph_1|, |\aleph_2|) = 0$, $L(\aleph_1, \aleph_3) = 7$, $I(\aleph_1, \aleph_1) = 8$, $I(\aleph_1, \epsilon) = 0$, $I(\aleph_1, \aleph_3) = 1$, $\Phi(10) = 10$, $\Phi(-3) = 3$ e $\Phi\left(\frac{-10}{0}\right) = \infty$.

6.5 Escolha do Neurônio Perceptron

Em (OTIS, 2011, p. 68), Santiago Ramón y Cajal, séculos XIX e XX, em seus estudos anatômicos do sistema nervoso, chegou a comparar esse sistema a uma rede de comunicação (de telégrafos). Sobre os neurônios cerebelares do tipo Purkinje, Cajal os descreve como “*something like the way that telegraph posts support de conducting wire*”. Sobre a célula nervosa, de forma genérica, ele sugeriu, em seus estudos iniciais, que “*The nerve cell provides an apparatus for the reception of currents*” e “*manifested by the dendritic expansions and cell body; an apparatus for transmission, represented by the prolonged axis cylinder; and an apparatus for apportionment or distribution, represented by the arborization of the nerve terminal.*”

Independentemente de o próprio Cajal ter rejeitado sua própria metáfora do telégrafo como aplicável ao sistema nervoso por esse ser um sistema biológico de grande plasticidade – como declara o próprio Cajal (OTIS, 2011, p. 68), não resta dúvida que a engenharia de telecomunicações do seu tempo o tenha influenciado na formulação de suas hipóteses científicas.

Tal metáfora seria um argumento por analogia^{1 2 3}, pois se especula a fisiologia tendo como base a anatomia (pela semelhança de formas infere-se a função). O que é reforçado pelo conhecimento prévio de que tanto o sistema nervoso como uma rede de telégrafos tinham a função de “comunicar” – o primeiro valendo-se de processos biológicos e o segundo de processos puramente elétricos.

Em 1943, o primeiro modelo matemático descritivo do neurônio biológico foi proposto por Warren McCulloch (psiquiatra e neuro-anatomista) e Walter Pitts (matemático) inspirados na anatomia e fisiologia básica do neurônio biológico (BRAGA *et al.*, 1998, p. 2). Frisa-se que McCulloch e Pitts não se concentraram no aprendizado, mas na funcionalidade.

O perceptron de Rosenblatt, criado em 1958, é uma rede neural artificial que possui a propriedade de um classificador linear e um algoritmo de aprendizado. É extremamente simples e pode modelar conceitos binários (sim ou não). Desde então, segundo (RUSSELL e NORVIG, 2004, p. 728), apesar dos avanços e de pesquisas com interesse mais abstratos das propriedades das redes neurais artificiais e do desenvolvimento das redes Bayesianas, o aprendizado por redes neurais artificiais é uma das formas mais populares e efetiva – merecedoras de estudo por si só.

Portanto, historicamente associado a um modelo biológico, o perceptron é um modelo matemático simples, apresenta alguma flexibilidade no aprendizado. É bem conhecido no meio científico e, ao mesmo tempo, tem uma relação com o conceito de transmissão de informação e classificação de dados. O próprio perceptron é também objeto de estudo na teoria da informação (MACKAY, 2005, cap. 40).

No que tange aos objetivos desta pesquisa, quanto mais simples for uma rede perceptron, a eficácia e a eficiência de uma classificação estarão mais relacionadas ao processo de construção das instâncias apresentadas à rede neural artificial do que às características de projeto e de particularidades da rede em si.

6.6 Escolha do Soundex

O código soundex foi desenvolvido para o problema de correspondência devido a diferentes grafias de nomes de pessoas nos registros de censo nos Estados Unidos da América, mas funciona bem para nomes da Europa ocidental (BLACK, 2010), e

¹Argumento por analogia: observa-se que tanto um objeto x e um objeto y possuem n propriedades em comum, além disso, é fato que y possui a uma outra propriedade \mathbf{G} . Assim se infere que x provavelmente tem a propriedade \mathbf{G} (NOLT e ROHATYN, 1991, pp. 429-434).

² O argumento por analogia também pode estabelecer um começo para pesquisa, mesmo de cunho investigativo e subjetivo, como o estudo de reações para inferência de estados emocionais, ou a existência deste como em *The expression of the emotions in man and animals* de (DARWIN, 1897), uma pesquisa de psicologia comparada.

³(NOLT e ROHATYN, 1991, p. 375) informam que é preciso ter cuidado para não se incorrer em conclusões baseadas em analogia defeituosa: uma falácia indutiva associada ao raciocínio analógico. Isso porque, segundo os mesmos, o argumento analógico tem uma carga indutiva altíssima.

era frequentemente utilizado para codificar sobrenomes (KNUTH, 1998, p. 394) no início do século XX (HERZOG *et al.*, 2007, p. 115). Com um código soundex de um nome, se pode utilizá-lo para buscar registros em cadastros – minimizando os efeitos de diferenças gráficas ou pequenos erros ortográficos. O objetivo é obter todos os registros com nomes que “soam” ou que se parecem com o nome procurado; e na prática obter um “canonical index code” (SNAE, 2007). A definição 6.6.1 descreve a variação do algoritmo soundex utilizado neste trabalho.

Definição 6.6.1 (Algoritmo Soundex). *Seja \aleph uma string no alfabeto latino. $Sx(\aleph)$ é o código soundex de \aleph e também é uma string, mas sempre com quatro símbolos. $Sx(\aleph)$ é formado pela primeira letra de \aleph e por mais três dígitos decimais justapostos, conforme os seguintes passos, para uma string não vazia (KNUTH, 1998):*

1. *Mantenha a primeira letra de \aleph e remova todas as outras ocorrências das seguintes letras: **A, E, H, I, O, U, W, Y**;*
2. *Substitua por números as letras restantes (aquelas após a primeira letra), conforme⁴:*
 - **B, F, P, V**: 1
 - **C, G, J, K, Q, S, X, Z**: 2
 - **D, T** : 3
 - **L**: 4
 - **M, N**: 5
 - **R**: 6
3. *Se duas ou mais letras com o mesmo código estavam adjacentes no nome original (antes do passo 1), omita todas, exceto a primeira.*
4. *Converta para a forma “letra, dígito, dígito, dígito” adicionando zeros à direita (se houver menos de 3 dígitos); ou descartando-se dígitos mais à direita (se houver mais de três).*

Caso $Sx(\epsilon)$ então retornar a string ?????;

Para detalhes históricos e computacionais do algoritmo soundex, ver (STEPHEN, 1994, pp. 53-56), (KNUTH, 1998, p. 394) e (NAVARRO, 2001).

Exemplo 6.6.1. Códigos Soundex:

⁴Com base em (RUSSELL, 1923) e (TURLEY e WOOD, 2006, pp. 301 e 302), o código 1 soundex faz referência aos sons categorizados como labiais e labiodentais. O código 2, aos guturais e as sibilantes. O código 3, aos dentais mudas. O código 4, aos palatino-fricativos. O código 5, aos labionasais e aos dento ou língua-nasais. O código 6, aos fricativo dentais.

- $Sx(UFRJ) = U162$
- $Sx(COPPE) = C100$
- $Sx(SERPRO) = S616$
- $Sx(Ventura) = V536$
- $Sx(Pinto) = P530$
- $Sx(Carvalho) = C614$
- $Sx(Coelho) = C400$
- $Sx(John) = J500$
- $Sx(Doe) = D000$
- $Sx(D) = D000$
- $Sx(Different) = D165$
- $Sx(Diferente) = D165$
- $Sx(\epsilon) = \text{????}$

Comentário 6.6.1. A razão desse tratamento para a *string* vazia, neste trabalho, é que ela representa ausência de qualquer informação. Uma *string* com caractere, por exemplo, D receberia o código soundex $D000$, assim como a *string* Doe . A ausência de informação deve ter um código diferente de qualquer outro código que:

- não desse pistas do seu comprimento;
- fosse, ao mesmo tempo, equivalente à outra *string* vazia.

Além disso, a regra da primeira letra do código soundex não poderia ser satisfeita para a *string* vazia. Então se optou em usar a *string* $????$, que é a maior possível para 4 dígitos e não é composta por caracteres alfabéticos.

Em (STOCK e STOCK, 2013), observa-se que o algoritmo soundex tem, pelo menos, duas desvantagens, no ponto de vista da recuperação de informação:

- não aplica a codificação à primeira letra – assim $Kraft$ e $Craft$, por exemplo, têm códigos diferentes (cego para esses homófonos evidentes);
- ignora que combinações de letras podem ter sons diferentes, em função da própria palavra – como em $ought$ e $plough$, por exemplo.

(HERZOG *et al.*, 2007, p. 119) abordam outros “lapsos” existentes Soundex.

Apesar de não ser perfeito, o algoritmo soundex aumenta as chances de se encontrar um nome em um de seus “disfarces” (KNUTH, 1998, p. 395). Uma explicação para isso é dada por (RUSSELL, 1923) sobre a *raison d’être* desse algoritmo:

There are certain sounds which form the nucleus of the English language, and these sounds are inadequately represented merely by the letters of the alphabet, as one sound may sometimes be represented by more than one letter or combination of letters, and one letter or combination of letters may represent two or more sounds.

Because of this, a great many names have two or more different spellings which in an alphabetical index, or an index which separates names according to the sequence of their contained letters in the alphabet, necessitates their filing in widely separated places.

With the use of certain primary sounds and by grouping such sounds as may be represented by the same letters, I have devised a phonetic arrangement which is an important characteristic of my improved index, and with the use, of which I am enabled to group names, not according to their spelling but in accordance with their sounds or phonetic values.

A escolha do algoritmo soundex, neste trabalho, é devido a sua simplicidade conceitual e fácil implementação (que inclusive já faz parte das bibliotecas de grandes sistemas de gerenciamento de bancos de dados, como o Oracle^{TM5}, ou o Postgres⁶). Para uma pesquisa teórica, o algoritmo soundex por ser simples e previsível (não possui um dicionário externo, por exemplo), torna a análise dos resultados bem mais direta – o que possibilita evidenciar as influências dos conceitos desenvolvidos. Em outras palavras, utilizar um método mais sofisticado poderia tornar mais difícil a percepção do impacto da modelagem. É importante ter em mente que o espectro de aplicação dos conceitos não é necessariamente restrito ao casamento aproximado de strings.

O algoritmo soundex procura obter uma espécie de arquétipo da palavra oral que a palavra escrita procura descrever, mas que pode ter sido corrompida pelo agente que a registrou, ou um emissor iletrado, ou erros ortográficos, ou desconhecidos. Entretanto, há o pressuposto tácito de que existe o melhor esforço para registrar a palavra no processo de registro da palavra – o que não quer dizer ótimo.

Análoga e metaforicamente, o papel do algoritmo do soundex é como o do Sol próximo ao horizonte no procedimento descrito em (EDMUNDS e FREETH, 2011), no campo da arqueologia:

On archeological digs, it is standard practice to take aerial photographs of site shortly after sunrise or just before sunset when the sun is low in the sky. The shadows cast by the glancing sunlight enhance faint surface irregularities or texture differences on the land, making underlying physical structures visible.

⁵Soundex no Oracle: https://docs.oracle.com/cd/B19306_01/server.102/b14200/functions148.htm

⁶Soundex no Postgres: <https://www.postgresql.org/docs/9.1/static/fuzzystrmatch.html>

Intuitivamente, objetos semelhantes e no mesmo lugar projetariam sombras semelhantes para uma mesma fonte de luz posicionada sempre no mesmo ponto. Assume-se aqui que os dados considerados semelhantes entre si, quando submetidos ao mesmo processo computacional, podem produzir um efeito semelhante, mas não necessariamente um idêntico (os sistemas caóticos são contraexemplos, mas são descartados). Um exemplo disso são a ocorrência de palavras homógrafas e homófonas.

O código soundex *D165*, por exemplo, pode ser entendido como um efeito para *Different* e *Diferente* – uma espécie de “sombra” fonética dessas strings. O processo que deu origem a esse efeito procurava selecionar atributos fonéticos característicos que permitisse uma busca pela similaridade marcante, mas se esquecendo de nuances. Lembrando que o código soundex não é uma transcrição fonética, mas uma forma de agrupar palavras consideradas semelhantes.

No campo do aprendizado de máquina, ou reconhecimento de padrões, o paralelo com o processo de extração de características e código soundex de uma *string* é inevitável. A extração de características é uma transformação que mapeia um ponto x de um espaço \mathbf{n} dimensional em um ponto y de espaço \mathbf{m} dimensional, onde $\mathbf{m} < \mathbf{n}$ (HAYKIN, 2009, p. 40). Esse processo realiza uma redução dimensional que, em tese, tende a simplificar e acelerar os processos de aprendizado para classificação de instâncias – por meio de redes neurais artificiais, por exemplo – mas que adaptam a essência delas para a efetividade da classificação.

Comentário 6.6.2. A diferença do método que será desenvolvido neste trabalho é que não será necessário descartar o objeto que deu origem à sombra, porque tanto o objeto quanto sua sombra são strings. O código soundex de uma *string* pode ser considerado como uma perspectiva possível da própria *string* – uma caricatura possível.

6.7 Uma Baseline para Qualidade entre Strings

No campo da qualidade dos dados, (HEINRICH *et al.*, 2007, 2009) acreditam que quantificar a qualidade dos dados é essencial para planejar as medidas de qualidade de maneira econômica. Apresentaram uma métrica simples buscado relacionar o conceito de qualidade entre palavras discrepantes.

Definição 6.7.1. $Q_H(\aleph_1, \aleph_2)$ é métrica de correção entre duas strings \aleph_1 e \aleph_2 , tal que (HEINRICH *et al.*, 2007):

$$Q_H(\aleph_1, \aleph_2) = 1 - L_n(\aleph_1, \aleph_2) \quad (6.7.1)$$

A função 6.7.1 expressa quantitativamente um grau de similaridade entre \aleph_1 e \aleph_2 que também pode ser consistente com um julgamento qualitativo feito devido a

uma inspeção visual deste par de strings. Se $Q_H(\aleph_1, \aleph_2) = 1$ é uma correspondência tal que $\aleph_1 = \aleph_2$; Se $Q_H(\aleph_1, \aleph_2) = 0$ então é, sem dúvida, um não-correspondência tal que $I(\aleph_1, \aleph_2) = 0$. Se $0 < Q_H(\aleph_1, \aleph_2) < 1$, então é consistente com um grau crescente de qualidade de 0 até 1 – em que é considerado tanto o número de operações de edição como o tamanho das strings comparadas.

Curiosamente, a primeira tentativa de (HEINRICH *et al.*, 2007) foi a seguinte:

$$Q'_H(\aleph_1, \aleph_2) = \frac{1}{L(\aleph_1, \aleph_2) + 1} \quad (6.7.2)$$

Apesar de (HEINRICH *et al.*, 2007) não fazerem menção a (MONTGOMERY, 2012), não é difícil perceber que a função 6.7.2 procura contornar uma manifestação da anomalia 4.2.3 (pág. 47).

Claramente (HEINRICH *et al.*, 2007) identificam o número de operações de edição entre as duas strings como uma grandeza de variabilidade — e considerando tacitamente, no modelo deles, que a qualidade seria uma grandeza absoluta e inversamente proporcional a essa variabilidade. Assim, quanto mais operações de edição houvesse então as strings difeririam e a qualidade absoluta deveria diminuir. Porém como esse modelo é incompatível com uma variabilidade nula (ocorre quando as strings são idênticas), (HEINRICH *et al.*, 2007) acrescentaram a unidade ao denominador como fator de correção (expressão 6.7.2) para contornar uma eventual divisão por zero.

Todavia, essa função (expressão 6.7.2) foi excluída pelo seguinte fato experimental: $Q'_H(\textit{Eisssonhour}, \textit{Eisenhower}) = Q'_H(\textit{Bird}, \textit{Hunt}) = 0,200$. Este fato mostra que esta função não está em conformidade com a propriedade 4.2.1. Ela é também descartada por (HEINRICH *et al.*, 2007) por essa óbvia falta de eficácia. A tabela 6.3, página 89, lista alguns valores para a função 6.7.1 quando é aplicada a outro par de palavras.

6.8 Contando Mitos e suas Variações

Um agente inteligente pode conjecturar ao comparar dados por que deve aceitá-los ou não, baseado em um sistema de verificações, pesos e contrapesos de incertezas, no lugar de utilizar lógica ⁷, pois o agente pode não ter acesso aos registros históricos do processo. Tais registros históricos rastrearíamos transformações sofridas pelos dados de volta à fonte no mundo real – o que explicaria tais transformações. No lugar disso, tal agente pode criar “novas histórias” ou mitos^{8 9}. Nos estudos clássicos, um

⁷Lógica: no sentido de salvaguarda da verdade.

⁸Mitos: Para mais detalhes sobre mitos ver seção A.5, página 166.

⁹Sobre objetos ficcionais mentais na filosofia medieval, ver (KLIMA, 2010) e (KLIMA *et al.*, 2007, cap. 7)

mito é um relato de criação em um estilo narrativo que tenta explicar “como algo que é assim, mas não era o caso, começou a ser” - como “nova realidade” por causa do esquecimento de “história verdadeira” (BRANDÃO, 2011).

Para a construção de mitos, ou novas histórias, que justificam a similaridade e a diferença entre strings, serão apresentadas algumas definições. Eles serão utilizadas nos processos de ponderação.

Definição 6.8.1. *Uma sequência minimal de operações para \aleph_1 e \aleph_2 , é uma sequência de operações de edição para transformar a string \aleph_1 em \aleph_2 ou \aleph_2 em \aleph_1 , tal que o número de operações é o menor possível. Portanto, $L(\aleph_1, \aleph_2)$ é o número de operações para essa sequência.*

Definição 6.8.2. *Uma sequência maximal de igualdades para \aleph_1 e \aleph_2 , é uma sequência de operações de igualdade entre caracteres de strings, respectivamente, \aleph_1 e \aleph_2 em que o número de igualdades é o maior e há uma bijeção $f(i)$ tal que para cada $\aleph_1[i] = \aleph_2[j]$ desta sequência então $j = f(i)$. Então $I(\aleph_1, \aleph_2)$ é o maior número de igualdades para essa sequência.*

Definição 6.8.3. $\Omega^-(\aleph_1, \aleph_2)$ é o número de permutações de uma sequência minimal de operações para \aleph_1 e \aleph_2 . Então $\Omega^-(\aleph_1, \aleph_2) = L(\aleph_1, \aleph_2)!$.

Definição 6.8.4. $\Omega^+(\aleph_1, \aleph_2)$ é o número de permutações de uma sequência maximal de igualdades para \aleph_1 e \aleph_2 . Então $\Omega^+(\aleph_1, \aleph_2) = I(\aleph_1, \aleph_2)!$.

Exemplo 6.8.1. Para transformar a string *TIP* em *PIT*, é necessário pelo menos duas operações de edição distintas (denotadas por op_1 e op_2), então $2!$ (dois fatorial) possibilidades:

Permutações de op_1 e op_2		
<i>TIP</i>	$\xrightarrow{op_1}$	<i>TIT</i> $\xrightarrow{op_2}$ <i>PIT</i>
<i>TIP</i>	$\xrightarrow{op_2}$	<i>PIP</i> $\xrightarrow{op_1}$ <i>PIT</i>

Tabela 6.1: Permutações de uma sequência minimal de operações para *TIP* e *PIT*.

Exemplo 6.8.2. O número de permutações de uma sequência maximal de igualdades para $\aleph_1 = \textit{back}$ e $\aleph_2 = \textit{bak}$ é $3!$ (Seis) permutações.

Definição 6.8.5. *Com base no resultado de uma igualdade entre duas strings, pode-se acreditar que uma das strings é a fonte original da outra. Pode-se considerar exclusivamente uma possibilidade ($\aleph_1 = \aleph_2$); Ou duas possibilidades (\aleph_1 foi transformada em \aleph_2 ou \aleph_2 foi transformada em \aleph_1). $\Omega^*(\aleph_1, \aleph_2)$ denota essa contagem, tal que:*

Permutações de $(\aleph_1[1] = \aleph_2[1], \aleph_1[2] = \aleph_2[2], \aleph_1[4] = \aleph_2[3])$
$(\aleph_1[1] = \aleph_2[1], \aleph_1[2] = \aleph_2[2], \aleph_1[4] = \aleph_2[3])$
$(\aleph_1[1] = \aleph_2[1], \aleph_1[4] = \aleph_2[3], \aleph_1[2] = \aleph_2[2])$
$(\aleph_1[2] = \aleph_2[2], \aleph_1[1] = \aleph_2[1], \aleph_1[4] = \aleph_2[3])$
$(\aleph_1[2] = \aleph_2[2], \aleph_1[4] = \aleph_2[3], \aleph_1[1] = \aleph_2[1])$
$(\aleph_1[4] = \aleph_2[3], \aleph_1[1] = \aleph_2[1], \aleph_1[2] = \aleph_2[2])$
$(\aleph_1[4] = \aleph_2[3], \aleph_1[2] = \aleph_2[2], \aleph_1[1] = \aleph_2[1])$

Tabela 6.2: Permutações de uma sequência maximal de igualdades para *back* e *bak*.

$$\Omega^{\otimes}(\aleph_1, \aleph_2) = 2^{u(\aleph_1, \aleph_2)} \quad (6.8.1)$$

Onde u é a função indicador de diferença, ver definição 6.4.7.

Análise 6.8.1. Por construção, Ω^+ endossa semelhanças entre strings. Similarmente, Ω^- endossa dissimilaridades. Ω^{\otimes} é um detector de diferença baseado na suposição de um processo de comunicação pré-existente. Como Ω^+ e Ω^- são fatoriais, elas são sensíveis ao tamanho das strings. Estas três funções são a matéria-prima para medir incertezas a serem ponderadas no processo de correspondência de strings. Os valores de Ω^+ , Ω^- e Ω^{\otimes} podem ser usados para corroborar, ou contrapor, um processo de aceitação (ou rejeição) envolvendo pesagens. Tal processo teria uma natureza retórica, mas sem ser irracional. Cada permutação de uma sequência é análoga a uma variação do mesmo mito.

6.9 Casamento Aproximado entre Palavras por um Perceptron

Definição 6.9.1 (Dipolo Qualitativo para Strings). *Sejam*

- S um sistema momento qualitativo de contagem em bits, cujo momento qualitativo de contagem é $Q_{\mathbf{Q1}}(W, \Omega)$.
- σ_1 e σ_2 : duas strings;
- $w \in \mathbb{R}$

Tais que:

- $\Omega = (\Omega_1, \Omega_2, \Omega_3)$, onde $\Omega_1 = \Omega^+(\sigma_a, \sigma_b)$, $\Omega_2 = \Omega^-(\sigma_a, \sigma_b)$ e $\Omega_3 = \Omega^{\otimes}(\sigma_a, \sigma_b)$.
- $W = (1, -1, -1)$;

Seja $Q_{\mathbf{Q1}}$ o julgamento qualitativo de S , $Q(\sigma_a, \sigma_b)$ é um dipolo qualitativo para as strings σ_a e σ_b tal que:

$$Q(\sigma_a, \sigma_b) = \mathcal{Q}_{\mathcal{Q}} \quad (6.9.1)$$

Análise 6.9.1. Observando o corolário 5.5.10 e a definição de $Q(\aleph_a, \aleph_b)$ então:

$$Q(\aleph_a, \aleph_b) = \log_2 \Omega^+(\aleph_a, \aleph_b) - \log_2 \Omega^-(\aleph_a, \aleph_b) - \log_2 \Omega^\otimes(\aleph_a, \aleph_b) \quad (6.9.2)$$

Reorganizando:

$$Q(\aleph_a, \aleph_b) = \log_2 \frac{\Omega^+(\aleph_a, \aleph_b)}{\Omega^-(\aleph_a, \aleph_b)\Omega^\otimes(\aleph_a, \aleph_b)} \quad (6.9.3)$$

Fica evidente que um dipolo qualitativo para strings fornece uma grandeza que resulta da ponderação de incertezas devido às semelhanças ($\Omega^+(\aleph_a, \aleph_b)$) e diferenças ($\Omega^-(\aleph_a, \aleph_b)$) entre as strings \aleph_a e \aleph_b . Além disso, ele considera também a incerteza devido ao fato de não se saber o sentido da transformação ($\Omega^\otimes(\aleph_a, \aleph_b)$), caso $\aleph_a \neq \aleph_b$, ver definição 6.8.5.

Definição 6.9.2 (Perceptron). *Para $n \in \mathbb{N}_*$ e $n > 0$, sejam $x = (x_1, \dots, x_i, \dots, x_n) \in \mathbb{R}^n$ uma instância, $\gamma = (\gamma_1, \dots, \gamma_i, \dots, \gamma_n) \in \mathbb{R}^n$ o vetor de pesos sinápticos e $\Sigma(\gamma, x)$ a função de campo induzido local tais que:*

$$\Sigma(\gamma, x) = \sum_{i=1}^n \gamma_i x_i. \quad (6.9.4)$$

$\nu(\gamma, x)$ é o neurônio perceptron definido como:

$$\nu(\gamma, x) = \begin{cases} +1 & \text{se } \Sigma(\gamma, x) \geq 0 \\ -1 & \text{senão} \end{cases} \quad (6.9.5)$$

Definição 6.9.3 (Casamento de Palavras e Dúvida). *Seja Q um dipolo qualitativo para strings. Seja $\nu_Q(\gamma, x)$ um único neurônio perceptron tal que:*

- A instância x para \aleph_1 e \aleph_2 é definida como:
 - $x = (Q(\aleph_1, \aleph_2), Q(Sx(\aleph_1), Sx(\aleph_2)), 1)$;
- Os pesos sinápticos definidos como: $\gamma = (\frac{1}{2}, 1, 0)$;
- $\nu_Q = +1$ e $\nu_Q = -1$ correspondem, respectivamente, à classificação **matched** e **not-matched** para o par \aleph_1 e \aleph_2 ;
- O campo induzido local de $\nu_Q(\gamma, x)$ será denotado por $\Sigma_Q(\gamma, x)$.

$D_Q(\aleph_1, \aleph_2)$ é o grau de dúvida da classificação para \aleph_1 e \aleph_2 tal que:

$$D_Q(\aleph_1, \aleph_2) = \Phi \left(\frac{1}{\Sigma_Q(\aleph_1, \aleph_2)} \right) \quad (6.9.6)$$

Assim, por construção:

$$\nu_Q(\gamma, x) \equiv \begin{cases} +1 & \text{se } \frac{1}{2} Q(\aleph_1, \aleph_2) + 1 Q(Sx(\aleph_1), Sx(\aleph_2)) \geq 0 \\ -1 & \text{senão} \end{cases} \quad (6.9.7)$$

Análise 6.9.2. Intuitivamente, ν_Q pondera incertezas ortográficas favorecendo incertezas “fonéticas” no processo de classificação. Tais incerteza são calculadas a partir do número das possibilidades de se converter uma *string* na outra; de se dizer que são similares e de se decidir qual é a string original (supondo uma seja reprodução parcial da outra). Pela definição 6.9.1, pelo corolário 5.5.10 e pela propriedade 5.5.2, $\Sigma_Q(\aleph_1, \aleph_2)$ é a superposição de dois julgamentos qualitativos de contagem em *bits*. Como consequência, $\Sigma_Q(\aleph_1, \aleph_2)$ é também a superposição da espontaneidade de duas decisões qualitativas.

Análise 6.9.3. $D_Q(\aleph_1, \aleph_2)$ é um indicador de quão inseguro deve ser a classificação. Suponha $\aleph_Q(\aleph_1, \aleph_2) \approx 0$ então, neste caso, pequenas mudanças em uma das strings podem ser cruciais para considerar o par **matched** ou **not-matched**. Se $\Sigma_Q(\aleph_1, \aleph_2) = 0$, embora uma decisão seja tomada, então um grau de dúvida é “**fora de escala**”.

Com base em (LIMA, 1993), a distância entre um ponto $x \in \mathbb{R}^n$ e um plano Π definido pela equação $\sum_{i=1}^n \gamma_i \cdot x_i = b$ é dada por:

$$d(x, \Pi) = \frac{|\sum_{i=1}^n \gamma_i x_i - b|}{\sqrt{\sum_{i=1}^n \gamma_i^2}} \quad (6.9.8)$$

Assim, a distância entre uma instância x do plano de decisão Π_γ do perceptron da definição 6.9.2 é:

$$d(x, \Pi_\gamma) = \frac{|\sum_{i=1}^n \gamma_i x_i|}{\sqrt{\sum_{i=1}^n \gamma_i^2}} \quad (6.9.9)$$

Como $\sqrt{\sum_{i=1}^n \gamma_i^2}$ é uma constante positiva:

$$d(x, \Pi_\gamma) \propto \left| \sum_{i=1}^n \gamma_i x_i \right| \quad (6.9.10)$$

Obviamente, qualquer instância cuja distância de Π_γ é zero pertence ao plano de decisão do perceptron. Uma faceta da tomada de decisão do neurônio perceptron é que, em tese, para poder classificar, ele aprendeu (de alguma forma) uma região de natureza dicotômica – que pode consubstanciar num conjunto instâncias que não

podem ser classificadas, exceto por padrão.

Pela relação 6.9.10, a magnitude do campo induzido local, para instâncias linearmente separáveis, é diretamente proporcional à distância de uma instância ao plano de decisão. Portanto, é justo dizer que o inverso desta distância é uma quantidade que indica coerentemente um grau de insegurança da decisão – usando como plano de decisão de referência Π_γ . É importante dizer que não se trata de uma medida de eficiência ou eficácia, mas sim quão dificultosa foi a decisão, o que sugere ser um parâmetro da qualidade da decisão em si, com base no plano de decisão e a instância a ser classificada.

Para esse perceptron, essa grandeza é calculada pela função $D_Q(x, \gamma)$, cuja unidade de medida é a mesma que o inverso da unidade do campo induzido local. Assim, a unidade de dúvida, calculada por $D_Q(x, \gamma)$, é o bit^{-1} – visto que o campo induzido local é um julgamento qualitativo de contagem em *bits*. Além disso,

Comentário 6.9.4. Como sugestão, o clássico algoritmo de aprendizagem supervisionada do perceptron pode ser empregado para obter os pesos de um momento qualitativo. Se as instâncias de treinamento não são linearmente separáveis, uma estratégia que combina o “pocket algorithm” com múltiplos treinamentos (o reinício com diferentes pesos sinápticos escolhidos aleatoriamente) podem ser utilizados nessa busca, como em (ABU-MOSTAFA *et al.*, 2012). O ADALINE¹⁰ também pode ser utilizado, uma vez que não há nem diferença morfológica nem funcional do perceptron. A única diferença reside no algoritmo de treinamento, pois o primeiro valor utiliza o campo induzido local e o segundo valor de classificação para calcular o erro. Ambos os modelos de classificação de aprendizagem neurais podem permitir uma busca por pesos sinápticos que atendam a eficácia da classificação.

Uma Pequena Experimentação

A tabela 6.3, na página 89, é o resultado de uma pequena experimentação para ilustrar quantitativamente os resultados de cálculos sobre a correspondência aproximada de strings usando os conceitos engendrados nas seções anteriores. Não se trata de um experimento em *stricto senso*, mas sim uma prova de conceito: cujo objetivo é obter valores tabelados de algumas das grandezas partícipes no casamento aproximado de duas palavras. Frisa-se que o modelo computacional de casamento neste capítulo é um mero protótipo concebido artesanalmente.

A tabela 6.4, na página 90, é mais detalhada. As palavras nessa tabela são especialmente das línguas portuguesa e inglesa – a maioria inglesa, entretanto. Os valores para as colunas L , L_n , Q_H , Ω^+ , Ω^- , Ω^* , Σ_Q , D_Q e ν_Q são obtidos em funções das colunas \aleph_1 e \aleph_2 (o par de strings a ser classificado como **matched**, ou

¹⁰ADALINE: Adaptive Linear Neuron, ver (BRAGA *et al.*, 1998).

not-matched). A coluna Q_H é a métrica apresentada (HEINRICH *et al.*, 2007) – usada como baseline. As colunas Σ_Q , D_Q e ν_Q são respectivamente os valores do campo local induzido, da métrica de dúvida e da decisão perceptron de acordo com a definição 6.9.3.

(HEINRICH *et al.*, 2007) não forneceram explicitamente um limite para aceitação ou rejeição para Q_H , porém o par 19º na tabela 6.3 também foi avaliado como $Q_H = 0.600$ em (HEINRICH *et al.*, 2007). No que diz respeito à qualidade dos dados, foi considerado aceitável por (HEINRICH *et al.*, 2007). Assim, como uma baseline para o leitor, $Q_H \geq 0.600$ seria um **matched** (caso contrário um **not-matched**) para (HEINRICH *et al.*, 2007).

Ambos os métodos, isto é Q_H e ν_Q , foram eficazes em rejeitar pares em que figuravam palavras cuja codificação soundex era totalmente diferente entre si: os 19º e 22º. Quando os pares eram idênticos, ambos os métodos também foram igualmente eficazes, 7º, 13º, 18º, 23º e 36º. Para os pares deste parágrafo, o menor e maior valores para D_Q foram 0.056 e 0.218.

Aborda-se um caso em que ambos os métodos, ν_Q e Q_H , não foram bem sucedidos. O 9º par de strings, *pile* e *mile*, apresentou $Q_H = 0.750$ e $\nu_Q = +1$, o que pode ser considerado um falso positivo para ambos os métodos. Essa ocorrência era esperada, pois não é feita análise semântica.

Do 31º par ao 36º, existem apenas comparações com o nome de origem inglesa *Jessica* com suas versões homofônicas portuguesas, com exceção do par 36º, que é um **matched** pela igualdade léxica. Para estes 6 pares, ν_Q classifica 5 como correspondente enquanto Q_H , 4.

Para o 40º par ($\aleph_1 = Jessica$ e $\aleph_2 = Celina$) ambos são nomes próprios que têm a mesma sequência de letras de vogais (*e-i-a*) e $D_Q = 0,197$ o que é consistente com o propósito original do algoritmo de codificação soundex para privilegiar sons consonantais – o resultado para esse par é **not-matched**.

Observa-se um erro grosseiro de classificação para o 12º par. Isso se deve ao fato de $Sx(I) = I000$ e $Sx(Z) = Z000$. Ou seja, o fato de o código soundex completar com zeros sons vocálicos faltantes aumentou falsamente a contribuição fonética.

Para $D_Q \geq 1$, existem quatro pares: 8º, 14º, 32º e 39º. É digno de nota que, ou são pares de homófonos, cuja ortografia é bastante diferente, ou são de palavras bem mal escritas.

Por fim, duas explicações computacionais para eventuais discrepâncias são:

- a baixa capacidade do algoritmo soundex para lidar com vogais;
- o algoritmo de distância de edição empregado considera uma letra alfabética e sua versão marcada com sinal diacrítico¹¹ tão diferentes quanto duas letras

¹¹Exemplos de letras e exemplo de suas versões com sinal diacrítico: *a/ã, c/ç, a/à, e/é e u/ü*.

alfabéticas distintas. Em outras palavras: a baixa capacidade de criar mitos figurando marcas diacríticas e sons de vocálicos.

Tabela 6.3: Resultados da Experimentação: strings \aleph_1 e \aleph_2 ; baseline $Q_H(\aleph_1, \aleph_2)$, campo induzido local $\Sigma_Q(\aleph_1, \aleph_2)$, o grau de dúvida $D_Q(\aleph_1, \aleph_2)$, perceptron $\nu_Q(\aleph_1, \aleph_2)$.

N°	\aleph_1	\aleph_2	Q_H	Σ_Q	D_Q	ν_Q
1 ^o	cocho	coxo	matched	4,877	0,205	matched
2 ^o	nós	noz	not-matched	3,585	0,279	matched
3 ^o	xeque	cheque	matched	2,877	0,348	matched
4 ^o	cacei	cassei	matched	5,877	0,170	matched
5 ^o	cacei	xeque	not-matched	-2,368	0,422	not-matched
6 ^o	conselho	concelho	matched	10,235	0,098	matched
7 ^o	conselho	conselho	matched	12,235	0,082	matched
8 ^o	ship	cheap	not-matched	0,292	3,419	matched
9 ^o	pile	mile	matched	2,377	0,421	matched
10 ^o	hear	here	not-matched	4,085	0,245	matched
11 ^o	replace	replacements	not-matched	6,781	0,147	matched
12 ^o	I	Z	not-matched	1,085	0,922	matched
13 ^o	I	I	matched	4,585	0,218	matched
14 ^o	I	eye	not-matched	-0,208	4,819	not-matched
15 ^o	kill	queue	matched	-4,953	0,202	not-matched
16 ^o	weak	week	matched	5,377	0,186	matched
17 ^o	one	own	not-matched	3,585	0,279	matched
18 ^o	one	one	matched	5,877	0,170	matched
19 ^o	Eissonhour	Eisenhower	matched	6,538	0,153	matched
20 ^o	complement	compliment	matched	13,320	0,075	matched
21 ^o	complement	Eissonhour	not-matched	-16,980	0,059	not-matched
22 ^o	complement	Eisenhower	not-matched	-16,980	0,059	not-matched
23 ^o	Eisenhower	Eisenhower	matched	15,480	0,065	matched
24 ^o	fir	fur	matched	4,585	0,218	matched
25 ^o	tip	pit	not-matched	-2,000	0,500	not-matched
26 ^o	plain	plane	matched	4,877	0,205	matched
27 ^o	write	right	not-matched	-3,792	0,264	not-matched
28 ^o	whole	hole	matched	3,377	0,296	matched
29 ^o	Eissonhour	ε	not-matched	-16,980	0,059	not-matched
30 ^o	Bird	Hunt	not-matched	-3,792	0,264	not-matched
31 ^o	Jessica	Gécika	not-matched	-1,868	0,535	not-matched
32 ^o	Jessica	Gecika	not-matched	0,085	11,770	matched
33 ^o	Jessica	Jecika	not-matched	5,085	0,197	matched
34 ^o	Jessica	Jecica	matched	7,038	0,142	matched
35 ^o	Jessica	Jesica	matched	8,831	0,113	matched
36 ^o	Jessica	Jessica	matched	10,735	0,093	matched
37 ^o	Jessica	Jessca	matched	5,831	0,172	matched
38 ^o	Jessica	Gessca	matched	1,453	0,688	matched
39 ^o	Jessica	Gesca	not-matched	-0,500	2,000	not-matched
40 ^o	Jessica	Celina	not-matched	-5,085	0,197	not-matched

6.10 Discussões e Conclusões

Os objetivos foram alcançados especialmente com o perceptron artesanal, mas derivado do momentum qualitativo de contagem, com capacidade de tomada de decisão para um problema de correspondência aproximado - com uma grandeza do grau de dúvida para uma classificação (**matched** ou **not-matched**).

A contagem de mitos não só deu outra interpretação a um caso particular de casamento aproximado de strings, mas não restringiu a abordagem a um algoritmo específico. Em vez disso, sugere a potencialização de eficácia de outros algoritmos diferentes do soundex, como por exemplo o *metaphone* e o *NYSIIS*¹²¹³ – caso permitam contagens de semelhanças e dissimilaridades.

Por fim, ao incorporar também elementos fonéticos (e não só ortográficos) o perceptron, em princípio, efetua uma comparação mais abrangente, ao mesmo tempo que sugere um método que leva em consideração efeitos psicológicos, pois aspectos fonéticos são dependentes do conhecimento do agente inteligente. Pesquisas sobre o grau de dúvida devem ser realizadas.

Tabela 6.4: Resultados Detalhados da Experimentação com o Perceptron

N°	N_1	N_2	$S_x(N_1)$	$S_x(N_2)$	$ N_1 $	$ N_2 $	L	L_n	Q_H	Ω^+	Ω^-	Ω^*	Σ_Q	D_Q	ν_Q
1°	cocho	coxo	C200	C200	5	4	2	0,400	0,600	6	2	2	4,877	0,205	+1
2°	nós	noz	N200	N200	3	3	2	0,667	0,333	1	2	2	3,585	0,279	+1
3°	xeque	cheque	X200	C200	5	6	2	0,333	0,667	24	2	2	2,877	0,348	+1
4°	cacei	cassei	C200	C200	5	6	2	0,333	0,667	24	2	2	5,877	0,170	+1
5°	cacei	xeque	C200	X200	5	5	5	1,000	0,000	1	120	2	-2,368	0,422	-1
6°	conselho	concelho	C524	C524	8	8	1	0,125	0,875	5040	1	2	10,235	0,098	+1
7°	conselho	conselho	C524	C524	8	8	0	0,000	1,000	40320	1	1	12,235	0,082	+1
8°	ship	cheap	S100	C100	4	5	3	0,600	0,400	2	6	2	0,292	3,419	+1
9°	pile	mile	P400	M400	4	4	1	0,250	0,750	6	1	2	2,377	0,421	+1
10°	hear	here	H600	H600	4	4	2	0,500	0,500	2	2	2	4,085	0,245	+1
11°	replace	replacements	R142	R142	7	12	5	0,417	0,583	5040	120	2	6,781	0,147	+1
12°	I	Z	I000	Z000	1	1	1	1,000	0,000	1	1	2	1,085	0,922	+1
13°	I	I	I000	I000	1	1	0	0,000	1,000	1	1	1	4,585	0,218	+1
14°	I	eye	I000	E000	1	3	3	1,000	0,000	1	6	2	-0,208	4,819	-1
15°	kill	queue	K400	Q000	4	5	5	1,000	0,000	1	120	2	-4,953	0,202	-1
16°	weak	week	W200	W200	4	4	1	0,250	0,750	6	1	2	5,377	0,186	+1
17°	one	own	O500	O500	3	3	2	0,667	0,333	1	2	2	3,585	0,279	+1
18°	one	one	O500	O500	3	3	0	0,000	1,000	6	1	1	5,877	0,170	+1
19°	Eissonhour	Eisenhower	E256	E256	10	10	4	0,400	0,600	720	24	2	6,538	0,153	+1
20°	complement	compliment	C514	C514	10	10	1	0,100	0,900	362880	1	2	13,320	0,075	+1
21°	complement	Eissonhour	C514	E256	10	10	10	1,000	0,000	1	3628800	2	-16,980	0,059	-1
22°	complement	Eisenhower	C514	E256	10	10	10	1,000	0,000	1	3628800	2	-16,980	0,059	-1
23°	Eisenhower	Eisenhower	E256	E256	10	10	0	0,000	1,000	3628800	1	1	15,480	0,065	+1
24°	fir	fur	F600	F600	3	3	1	0,333	0,667	2	1	2	4,585	0,218	+1
25°	tip	pit	T100	P300	3	3	2	0,667	0,333	1	2	2	-2,000	0,500	-1
26°	plain	plane	P450	P450	5	5	2	0,400	0,600	6	2	2	4,877	0,205	+1
27°	write	right	W630	R230	5	5	4	0,800	0,200	1	24	2	-3,792	0,264	-1
28°	whole	hole	W400	H400	5	4	1	0,200	0,800	24	1	2	3,377	0,296	+1
29°	Eissonhour	ε	E256	???	10	0	10	1,000	0,000	1	3628800	2	-16,980	0,059	-1
30°	Bird	Hunt	B630	H530	4	4	4	1,000	0,000	1	24	2	-3,792	0,264	-1
31°	Jessica	Gécika	J220	G220	7	6	5	0,714	0,286	2	120	2	-1,868	0,535	-1
32°	Jessica	Gecika	J220	G220	7	6	4	0,571	0,429	6	24	2	0,085	11,770	+1
33°	Jessica	Jecika	J220	J220	7	6	3	0,429	0,571	24	6	2	5,085	0,197	+1
34°	Jessica	Jecica	J220	J220	7	6	2	0,286	0,714	120	2	2	7,038	0,142	+1
35°	Jessica	Jesica	J220	J220	7	6	1	0,143	0,857	720	1	2	8,831	0,113	+1
36°	Jessica	Jessica	J220	J220	7	7	0	0,000	1,000	5040	1	1	10,735	0,093	+1
37°	Jessica	Jessca	J220	J200	7	6	1	0,143	0,857	720	1	2	5,831	0,172	+1
38°	Jessica	Gessca	J220	G200	7	6	2	0,286	0,714	120	2	2	1,453	0,688	+1
39°	Jessica	Gesca	J220	G200	7	5	3	0,429	0,571	24	6	2	-0,500	2,000	-1
40°	Jessica	Celina	J220	C450	7	6	4	0,571	0,429	6	24	2	-5,085	0,197	-1

Fonte: (PINTO e CARVALHO, 2015).

¹²New York State Identification and Intelligence System Phonetic Code: código de índice canônico semelhante ao Soundex, difere deste na medida em que retém informações sobre a posição das vogais na palavra codificada convertendo todas as vogais para a letra A (SNAE, 2007).

¹³Para informações sobre o metaphone e o NYSIIS ver (SNAE, 2007)

Capítulo 7

Cálculo de Incertezas II

Apliquei meu espírito a um estudo atencioso e à sábia observação de tudo que se passa debaixo dos céus: Deus impôs aos homens esta ocupação ingrata. Vi tudo o que se faz debaixo do sol, e eis: tudo vaidade, e vento que passa. O que está curvado não se pode endireitar, e o que falta não se pode calcular. Eu disse comigo mesmo: Eis que amontoei e acumulei mais sabedoria que todos os que me precederam em Jerusalém. Porque meu espírito estudou muito a sabedoria e a ciência, e apliquei o meu espírito ao discernimento da sabedoria, da loucura e da tolice. Mas cheguei à conclusão de que isso é também vento que passa. Porque no acúmulo de sabedoria, acumula-se tristeza, e que aumenta a ciência, aumenta a dor.

Eclesiastes, 1:13-18

7.1 Introdução

No capítulo 5 foi introduzido o conceito de momento qualitativo de decisão. Entretanto, somente foram abordadas as distribuições de probabilidades discretas. Neste capítulo, serão apresentadas considerações para a utilização das distribuições que modelam variáveis aleatórias contínuas, mas observando-se critérios que preservam a entropia de Shannon, pois a entropia diferencial pode ser negativa.

7.2 Motivação

(SHANNON, 1948) apresenta uma função para o cálculo da entropia para variáveis aleatórias contínuas: a entropia diferencial. Porém, (SHANNON, 1948):

In the continuous case the entropy can be considered a measure of randomness *relative to an assumed standard*, (...) The entropy of a continuous distribution can be negative. The scale of measurements sets an

arbitrary zero corresponding to a uniform distribution over a unit volume. A distribution which is more confined than this has less entropy and will be negative.

(SAYOOD, 2005) e (HAYKIN, 2009), por outro consideram a entropia diferencial um bem por si – como objeto matemático.

Uma propriedade, ressaltada por (COVER e THOMAS, 1991; SHANNON, 1948), que é preservada, por analogia, para entropia diferencial, é a de que seu valor máximo é limitado pelo logaritmo do volume que contém resultados do experimento aleatório. A entropia diferencial, seja atribuindo valores negativos ou positivos, permite estabelecer um ordenamento entre funções de densidades de probabilidades no que diz respeito ao seu grau de uniformidade.

Porém como a entropia diferencial pode assumir valores negativos isso viola as propriedades 3.5.1. O que dificulta, num primeiro momento, sua utilização como momento qualitativo (os sinais dos pesos não mais identificam se a contribuição de entropia seria análoga a um produto ou a um reagente) – tornando o modelo de decisão proposto no capítulo 5 numa heurística que pondera grandezas de naturezas distintas. Imagine tentar pesar em uma balança de pratos $1Kg$ de tremoços e $-2kg$ de tremoços, como se desenvolve um processo dessa natureza?

Todavia, em (SAYOOD, 2005, p. 205 e 206), (COVER e THOMAS, 1991, p. 227), (HAYKIN, 2009, cap. 10) e (LEON-GARCIA, 2008, p. 206), observa-se um processo de relacionar a *entropia diferencial* com a *entropia* que consiste numa *quantização* para *quantificação* da incerteza. Esse processo será utilizado para estender ao modelo de decisão do capítulo 5 à entropia diferencial – ou adaptá-la ao processo de decisão consistente e coerentemente com o objetivo inicial do modelo.

7.3 Definições Básicas

Para se garantir uma generalidade maior ao trabalho, começar-se-á a partir de vetores de variáveis aleatórias contínuas e a definição de entropia diferencial será apresentada.

Definição 7.3.1 (Função Densidade de Probabilidade Conjunta). *Seja \vec{X} um vetor de variáveis aleatórias contínuas em $\mathcal{D} \subset \mathbb{R}^k$. $\rho_{\vec{X}}: \mathcal{D} \rightarrow \mathbb{R}^+$ é a função densidade de probabilidade conjunta de \vec{X} tal que:*

$$\int_{\mathcal{D}} \rho_{\vec{X}}(\vec{r}) dv = 1 \tag{7.3.1}$$

Ou alternativamente,

$$\int_{\mathcal{D}} \rho_{\mathbb{X}}(\vec{r}) dv = \int \cdots \int_{\mathcal{D}} \rho_{\mathbb{X}}(r_1, \dots, r_k) dr_1 \dots dr_k = 1 \quad (7.3.2)$$

Definição 7.3.2 (Entropia Diferencial). *Seja $\vec{\mathbb{X}}$ um vetor de variáveis aleatórias contínuas com função de densidade de probabilidade conjunta $\rho_{\mathbb{X}}$ cujo domínio é $\mathcal{D} \subset \mathbb{R}^k$. $H_{dif}(\vec{\mathbb{X}})$ é a entropia diferencial de $\vec{\mathbb{X}}$ tal que:*

$$H_{dif}(\vec{\mathbb{X}}) = \int_{\mathcal{D}} \rho_{\mathbb{X}}(\vec{r}) \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r})} dv \quad (7.3.3)$$

Ou alternativamente,

$$H_{dif}(\vec{\mathbb{X}}) = \int \cdots \int_{\mathcal{D}} \rho_{\mathbb{X}}(r_1, \dots, r_k) \ln \frac{1}{\rho_{\mathbb{X}}(r_1, \dots, r_k)} dr_1 \dots dr_k \quad (7.3.4)$$

Em (LAZO e RATHIE, 1978), encontra-se uma tabela com a entropia diferencial de diversas funções de densidades de probabilidade. Em (BISHOP, 2006, p. 685-693), (COVER e THOMAS, 1991, p. 486-487), estão disponíveis o cálculo de entropias e/ou entropias diferenciais, respectivamente, para funções de distribuições e de densidades de probabilidades. Analisar-se-á somente, neste trabalho, a distribuição gaussiana multivariada a título de ilustração no capítulo 8.

7.4 Partições e Processo de Quantização

Na teoria da informação, a principal forma de relacionar a *entropia* e a *entropia diferencial* é calcular entropia de uma função de distribuição contínua por meio da quantização do domínio de integração (COVER e THOMAS, 1991, pp. 228 e 229), (HAYKIN, 2009, cap. 10) e (LEON-GARCIA, 2008, p. 206). Entretanto, essa abordagem diz que é possível uma aproximação, mas a quantidade de entropia calculada – para efeitos práticos – aumenta quanto menor for o intervalo de quantização¹.

Definição 7.4.1 (Partição de \mathbb{R}^k). *Seja a tupla $(f, \mathbb{R}^k, \square)$ então \square é uma partição de \mathbb{R}^k tal que:*

- $\mathbb{R}^k = \bigcup_{i=1}^{\infty} \square_i$, onde $\square_i \neq \emptyset$ é *i*-ésima sub-região (um paralelepípedo retangular *k*-dimensional arbitrário);
- se $i \neq j$ então $\square_i \cap \square_j = \emptyset$;
- $\square_i v = \int \cdots \int_{\square_i} dr_1 \dots dr_k$ e $\square_i v \neq 0$;
- $f: \mathbb{N}_* \rightarrow \mathbb{R}^k$ tal que $f(i) \in \square_i$

¹Em (COVER e THOMAS, 1991, p. 229), “theorem 9.3.1.” sobre “n-bit quantization of a continuous random variable”

- $\|\square_i\|$ é o comprimento da maior distância possível entre dois pontos pertencentes à \square_i .
- $\exists i \in \mathbb{N}_*, \forall j \in \mathbb{N}_* \|\square\| = \|\square_i\|$ então $\|\square\| \geq \|\square_j\|$.

A partição de \mathbb{R}^k indica que regiões mutuamente exclusivas são definidas de forma a abranger todo o \mathbb{R}^k . Cada região \square_i tem um número real associada a ela pela função f e esse número está necessariamente contido nessa região. $\|\square\|$ será utilizado para estabelecer um limite para a distância máxima entre pontos dentro de uma região dessa partição, o que conseqüentemente limita o volume multidimensional de cada região, $\square_i v$ – visto que é um paralelepípedo multidimensional. Outro fato importante é que não foi feita nenhuma suposição sobre a utilização dessa partição num processo de integração.

Definição 7.4.2 (Partição de \mathcal{D}). *Seja \square uma partição de \mathbb{R}^k . Δ é uma partição de \mathcal{D} em função de \square então:*

- $\mathcal{D} \subset \bigcup_i \Delta_i$, onde $\Delta_i \neq \emptyset$; e \bigcup_i é $\bigcup_{i=1}^n$ ou $\bigcup_{i=1}^{\infty}$;
- se $i \neq j$ então $\Delta_i \cap \Delta_j = \emptyset$;
- $f(j) \in \mathcal{D}$ se, e somente se, $\exists i, \Delta_i = \square_j$ e $\vec{r}_i = f(j)$ e $\Delta_i v = \square_j v$;
- $\|\Delta\| = \|\square\|$;

Δ é uma partição com propriedades derivadas da partição de \mathbb{R}^k , mas podendo ser mais restrita dependendo do conjunto \mathcal{D} , pois \mathcal{D} seleciona as regiões \square_j . A associação de uma partição Δ_i a uma única partição \square_j ocorre para construir enumeração (finita ou infinita) ordenada pelos números naturais (a partir do número 1). Disso se obtém o elemento \vec{r}_i como representante de Δ_i , com base em $f(j)$. Finalmente, $\|\Delta\|$ estabelece o valor máximo para a distância de dois em uma região Δ_i , portanto limita $\Delta_i v$. $\|\Delta\|$ será utilizado em limites infinitesimais.

Definição 7.4.3 (Entropia Diferencial com Função de Densidade de Probabilidade Conjunta Riemann Integrável). *Seja \vec{X} um vetor de variáveis aleatórias contínuas com função de densidade de probabilidade conjunta $\rho_{\mathbb{X}}$ cujo domínio é \mathcal{D} , onde $\mathcal{D} \subset \mathbb{R}^k$. Seja um processo de discretização governado por Δ . Se $\rho_{\mathbb{X}}(\vec{r}) \cdot \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r})}$ é Riemann integrável² então:*

$$H_{dif}(\vec{X}) = \int_{\mathcal{D}} \rho_{\mathbb{X}}(\vec{r}) \cdot \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r})} dv = \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r}_i)} \Delta_i v \quad (7.4.1)$$

² Riemann integrável: ver (COVER e THOMAS, 1991, p. 229)

Definição 7.4.4. *Seja Δ uma partição de \mathcal{D} em função de \square . Se Δ é isocórica (isovolumétrica) então:*

- $\forall i, j \Delta_i v = \Delta_j v = \Delta v$.

Onde $\Delta v \in \mathbb{R}^+$.

Definição 7.4.5. *Um processo de quantização governado por Δ então:*

- Δ é uma partição de único \mathcal{D} em função de \square ;
- $\|\Delta\| \rightarrow 0$.

7.5 Multipolos e Dipolos Qualitativos

Agora serão apresentados os conceitos necessários– e a articulação entre eles – para que se efetue as adaptações do modelo de decisão para a entropia diferencial.

Definição 7.5.1 (Entropia de uma Variável Aleatória Contínua Quantizada). *Seja $\vec{\mathbb{X}}$ um vetor de variáveis aleatórias contínuas com função de densidade de probabilidade conjunta $\rho_{\mathbb{X}}$ cujo domínio é \mathcal{D} , onde $\mathcal{D} \subset \mathbb{R}^k$. Seja um processo de quantização governado por Δ . p é uma Δ aproximação de uma distribuição discreta de probabilidade de $\vec{\mathbb{X}}$ tal que:*

$$p_i = \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \quad (7.5.1)$$

Então $H_N(\vec{\mathbb{X}}_\Delta)$ é a entropia discreta de $\vec{\mathbb{X}}$ governada por Δ tal que:

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} H_N(p) \quad (7.5.2)$$

Onde $H_N(p)$ é definida no corolário 5.4.8: entropia em nats. Que pode ser reescrita como:

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} \sum_k p_k \cdot \ln \frac{1}{p_k} \quad (7.5.3)$$

Teorema 7.5.1 (Relação Entropia e Entropia Diferencial). *Seja $H_N(\vec{\mathbb{X}}_\Delta)$ a entropia discreta de $\vec{\mathbb{X}}$ governada por Δ . Δ é isocórica. $\rho_{\mathbb{X}}(\vec{r}) \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r})}$ é Riemann integrável cujo domínio é \mathcal{D} , onde $\mathcal{D} \subset \mathbb{R}^k$. Seja um processo de quantização governado por Δ . p é uma Δ aproximação de uma distribuição discreta de probabilidade para $\vec{\mathbb{X}}$ então:*

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} T(\Delta) \quad (7.5.4)$$

Onde $T(\Delta)$ denota a função de transbordo de entropia de $\vec{\mathbb{X}}_\Delta$. $T(\Delta)$ é tal que, para Δ isocórica:

$$T(\Delta) = \ln \frac{1}{\Delta v} \quad (7.5.5)$$

Assim, o transbordo não depende de $\rho_{\mathbb{X}}$.

Demonstração.

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} \sum_i p_k \cdot \ln \frac{1}{p_k} \quad (7.5.6)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v} \quad (7.5.7)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \left(\ln \frac{1}{\rho_{\mathbb{X}}(\vec{r}_i)} + \ln \frac{1}{\Delta_i v} \right) \quad (7.5.8)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r}_i)} + \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\Delta_i v} \quad (7.5.9)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r}_i)} \Delta_i v + \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\Delta_i v} \quad (7.5.10)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r}_i)} \Delta_i v + \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\Delta_i v} \quad (7.5.11)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = \int_{\mathcal{D}} \rho_{\mathbb{X}}(\vec{r}) \ln \frac{1}{\rho_{\mathbb{X}}(\vec{r})} dv + \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\Delta_i v} \quad (7.5.12)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\Delta_i v} \quad (7.5.13)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \cdot \ln \frac{1}{\Delta v} \quad (7.5.14)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} \ln \frac{1}{\Delta v} \cdot \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \quad (7.5.15)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} \ln \frac{1}{\Delta v} \cdot \lim_{\|\Delta\| \rightarrow 0} \sum_i \rho_{\mathbb{X}}(\vec{r}_i) \cdot \Delta_i v \quad (7.5.16)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} \ln \frac{1}{\Delta v} \cdot \int_{\mathcal{D}} \rho_{\mathbb{X}}(\vec{r}) dv \quad (7.5.17)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} \ln \frac{1}{\Delta v} \cdot 1 \quad (7.5.18)$$

$$H_N(\vec{\mathbb{X}}_\Delta) = H_{dif}(\vec{\mathbb{X}}) + \lim_{\|\Delta\| \rightarrow 0} \ln \frac{1}{\Delta v} \quad (7.5.19)$$

□

A hipótese de Δ isocórica não é nova em teoria da informação, mas ela será explorada não aqui como uma ilustração para mostrar uma relação entre a *entropia diferencial* e a *entropia*, mas sim como o meio de interpretação. Essa divisão do

domínio da função de densidade de probabilidade pode ser vista em (BISHOP, 2006; COVER e THOMAS, 1991; HAYKIN, 2009; SAYOOD, 2005).

Definição 7.5.2 (Momento Qualitativo Diferencial Quantizado). *Sejam $w_j \in \mathbb{R}$ e \vec{X}^j um vetor de variáveis aleatórias contínuas, onde $j \in \mathbb{N}_*$. Um processo de quantização de \mathcal{D} governado por Δ , onde \mathcal{D} é domínio de $\rho_{\mathbb{X}^j}$. K_Q é coeficiente de decisão qualitativo quantizado tal que $K_M \in \mathbb{R}$. $Q(w, \vec{X}_\Delta)$ é momento qualitativo diferencial quantizado de decisão tal que:*

$$Q(w, \vec{X}_\Delta) = \sum_j w_j \cdot H_N(\vec{X}_\Delta^j) \quad (7.5.20)$$

Definição 7.5.3 (Momento Qualitativo Diferencial Multipolar). *Seja $Q(w, \vec{X}_\Delta)$ um momento qualitativo diferencial quantizado de decisão tal que:*

- $Q(w, \vec{X}_\Delta) = \sum_j w_j \cdot H_N(\vec{X}_\Delta^j)$;
- Para todo j , $\rho_{\mathbb{X}^j}(\vec{r}) \cdot \ln \frac{1}{\rho_{\mathbb{X}^j}(\vec{r})}$ é Riemann integrável;
- w é uma sequência aniquilante, isto é $\sum_i w_i = 0$ e $w_i \neq 0$.

Então $Q(w, \vec{X})$ é momento qualitativo diferencial de decisão multipolar tal que $Q(w, \vec{X}) = Q(w, \vec{X}_\Delta)$.

Corolário 7.5.2 (Momento Qualitativo Diferencial Multipolar). *Seja $Q(w, \vec{X}_\Delta)$ momento qualitativo diferencial de decisão multipolar então:*

$$Q(w, \vec{X}) = \sum_j w_j \cdot H_{dif}(\vec{X}^j) \quad (7.5.21)$$

Demonstração.

$$Q(w, \vec{\mathbb{X}}) = Q(w, \vec{\mathbb{X}}_\Delta) \quad (7.5.22)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot H_N(\vec{\mathbb{X}}_\Delta^j) \quad (7.5.23)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot (H_{dif}(\vec{\mathbb{X}}^j) + \lim_{\|\Delta\| \rightarrow 0} T(\Delta)) \quad (7.5.24)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot H_{dif}(\vec{\mathbb{X}}^j) + w_j \cdot \lim_{\|\Delta\| \rightarrow 0} T(\Delta) \quad (7.5.25)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot H_{dif}(\vec{\mathbb{X}}^j) + \sum_j w_j \cdot \lim_{\|\Delta\| \rightarrow 0} T(\Delta) \quad (7.5.26)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot H_{dif}(\vec{\mathbb{X}}^j) + \left(\sum_j w_j \right) \cdot \lim_{\|\Delta\| \rightarrow 0} T(\Delta) \quad (7.5.27)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot H_{dif}(\vec{\mathbb{X}}^j) + 0 \cdot \lim_{\|\Delta\| \rightarrow 0} T(\Delta) \quad (7.5.28)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot H_{dif}(\vec{\mathbb{X}}^j) + 0 \quad (7.5.29)$$

$$Q(w, \vec{\mathbb{X}}) = \sum_j w_j \cdot H_{dif}(\vec{\mathbb{X}}^j) \quad (7.5.30)$$

□

Qual é o papel dos pesos aniquilantes? Na prática, é o de remover o transbordo do processo de quantificação do domínio da função da densidade de probabilidades. Ao mesmo tempo, eles preservam a porção de entropia que é necessária e relevante para a tomada de uma decisão útil. O processo de quantização escolhido, o isocórico, não depende de conhecimento prévio das funções de distribuição de probabilidade, teorema 7.5.1 – o que pode ser considerado um princípio de equidade.

Os pesos de um momento qualitativo diferencial multipolar serão como uma lembrança de que a contabilidade e a coerência da ponderação das incertezas de variáveis aleatórias foram preservadas – apesar do processo de quantização – para interpretar a entropia diferencial como uma quantidade de incerteza.

Definição 7.5.4 (Bipolo Qualitativo Diferencial). *Seja $Q(w, \vec{\mathbb{X}})$ um momento qualitativo diferencial de decisão multipolar tal que:*

$$Q(w, \vec{\mathbb{X}}) = \sum_{j=1}^2 w_j \cdot H_{dif}(\vec{\mathbb{X}}^j) \quad (7.5.31)$$

Então $Q(w, \vec{\mathbb{X}})$ é um dipolo qualitativo diferencial de decisão.

Definição 7.5.5 (Dipolo Qualitativo Diferencial). *Seja $Q(w, \vec{\mathbb{X}})$ um Bipolo qualitativo diferencial tal que:*

$$Q(w, \vec{X}) = \sum_{j=1}^2 w_j \cdot H_{dif}(\vec{X}^j) \quad (7.5.32)$$

E como $w_1 \neq 0$ e $w_1 + w_2 = 0$ então $w_1 = -w_2$. Sejam $w_1 = \mathbf{w}$, $\vec{Y} \equiv \vec{X}^2$, $\vec{Y}^0 \equiv \vec{X}^1$ e:

$$Q(\mathbf{w}, \vec{Y}, \vec{Y}^0) = \mathbf{w} \cdot (H_{dif}(\vec{Y}) - H_{dif}(\vec{Y}^0)) \quad (7.5.33)$$

$Q(\mathbf{w}, \vec{Y}, \vec{Y}^0)$ é um dipolo qualitativo diferencial de decisão tal que $Q(\mathbf{w}, \vec{Y}, \vec{Y}^0) = Q(w, \vec{X})$.

Conceitualmente, não existe nenhuma diferença entre um dipolo e um bipolo. A diferença é meramente formal.

7.6 Discussão e Conclusões

Um dipolo qualitativo diferencial remete a um resultado que é diferença exata entre duas grandezas sem interpretação de medida absoluta, mas que tem interpretação exata de uma diferença entre medidas absolutas hipotéticas.

O momento qualitativo diferencial multipolar é uma forma geral de ponderação para processos aleatórios modelados por funções de variáveis aleatórias contínuas. Um fator que permite essa interpretação é que Δ tem que ser compartilhado para pelo menos um grupo de variáveis aleatórias contínuas, cujos domínios é o mesmo, para se preservar a interpretação de que um resultado do experimento aleatório corresponde a mesma informação (a referência a mesma i -ésima sub-região do domínio).

Para fins práticos, se se deseja manter a interpretação da entropia enquanto uma concentração de incerteza por símbolo, então recomenda-se buscar pelo menos um dipolo qualitativo diferencial.

Portanto conseguiu-se preservar a natureza da interpretação do modelo de decisão – desde que seja incluído nele a hipótese isocórica e os pesos aniquilantes. Portanto a interpretação do resultado de uma decisão permanece a mesma, pois o mecanismo de decisão para distribuições discretas de probabilidades foi mantido.

Capítulo 8

Qualidade e Campos Conservativos

And the lord stood upon Tiamat's hinder parts,
And with his merciless club he smashed her skull.
He cut through the channels of her blood,
And he made the north wind bear it away into secret places...
Then the lord rested, gazing upon her dead body,...and devised a cunning plan.
He split her up like a flat fish into two halves;
One half of her he stablished as a covering for heaven.
He fixed a bolt, he stationed a watchman,
And bade them not to let her waters come forth.
He passed through the heavens, he surveyed the regions thereof,
And over against the Deep he set the dwelling of Nudimmud.
And the Lord measured the structure of the Deep...

Enuma Elish, babylonian creation myth (fourth tablet)

8.1 Introdução

Neste capítulo serão exercitados os conceitos aplicáveis aos capítulos 5 e 7, em especial observando as seguintes propriedades: a entropia ser uma quantidade (indiferentemente de ter sido obtida de uma distribuição discreta ou contínua); e que funções de distribuição (objetos matemáticos) possuem parâmetros para defini-las de forma exata. Como consequência, a entropia é também uma função desses parâmetros. Para as distribuições que diferem somente pela atribuição de valores de parâmetros tem-se uma família de distribuições.

Se tais parâmetros estão, ou se considera estarem, num domínio contínuo pode-se entender a entropia como uma imagem funcional dos valores desses parâmetros – o

que suscita a ideia de campo escalar para tais parâmetros. Assim a entropia pode variar em função desses parâmetros e então se obter objetos matemáticos como o vetor gradiente da entropia.

Este capítulo será uma breve análise sobre o impacto na incerteza e na qualidade percebidas quando ocorrem mudanças nos parâmetros de um modelo probabilístico. As distribuições normal (uni- e multidimensionais) serão utilizadas como modelo, pois são utilizadas para descrever processos de natureza probabilística nos mais diversos ramos do conhecimento científico e da engenharia.

8.2 Motivação

Curiosamente a entropia diferencial da distribuição normal unidimensional (ou Gaussiana clássica) é uma função exclusiva do desvio padrão – inclusive presente no trabalho seminal de (SHANNON, 1948). Fatos que indicam uma possível generalização derivada do conceito de entropia (enquanto uma medida de incerteza) para indicadores e métricas de qualidade de processos.

Em (JAYNES, 1965), há uma discussão sobre a natureza da entropia na física como um conceito antropomórfico – tanto do ponto de vista matemático quanto fenomenológico – como uma medida da ignorância humana sobre um fenômeno:

Thermodynamics knows of no such notion as the “entropy of a physical system”. Thermodynamics does have the concept of the entropy of a *thermodynamic* system; but a given physical system correspond to many different thermodynamic systems; a given system corresponds to many different thermodynamic system.

Consider, for example, a crystal of Rochelle salt. For one set of experiments on it, we work with temperature, pressure, and volume. The entropy can be expressed as some function $S_0(T, P)$. For another set of experiments on the same crystal, we work with temperature, the component e_{xy} of the strain tensor, and the component P_z of electric polarization; the entropy as found in these experiments is a function $S_0(T, e_{xy}, P_z)$. It is clearly meaningless to ask, “What is the entropy of the crystal?” unless we first specify the set of parameters which define its thermodynamic state.

(...) However, we can always introduce as many new degrees of freedom as we please. For example, we might expand each of the strain tensor in a complete orthogonal set of functions $\varphi(x, y, z)$

$$e_{ij}(x, y, z) = \sum a_{Ijk} \varphi_k(x, y, z)$$

and by a sufficiently complicated system of levels, we could vary each of the first 1000 expansion coefficients a_{ijk} independently. Our crystal is now a thermodynamic system of over 1000 degrees of freedom; (...)

From this we see that entropy is an anthropomorphic concept, not only in the well-know statistical sense that it measures the extent of human ignorance as to the microstate. *Even at the purely phenomenological level, entropy is an anthropomorphic.* For it is a property, not of the physical system, but of the particular experiments you or I choose to perform on it.

A razão dessa afirmação é que a entropia é uma medida de desconhecimento sobre um sistema, com base em algumas informações de algumas variáveis sistêmicas, como Σ , por exemplo. Se todas as variáveis que modelam o experimento fornecessem a percepção exata sobre o estado em que o sistema se encontra, ele seria exatamente determinado: o conceito de entropia colapsa (JAYNES, 1965). O que é razoável, pois ela está atrelada ao conceito de incerteza sobre os estados de um sistema termodinâmico e à distribuição de probabilidades de um experimento aleatório na teoria da informação.

8.3 Objetivo

Uma distribuição normal é caracterizada por sua média e matriz de covariância, entretanto, sua entropia é uma função dessa matriz somente. O sistema, ou experimento, é nesta seção uma distribuição normal multivariada. Entretanto, ao se admitir um modelo probabilístico, admite-se que não se conhece tudo a fim de caracterizar os estados do sistema – caso contrário seria um sistema determinístico.

8.4 Definições Básicas

Na literatura, famílias de distribuições de probabilidade são caracterizadas pela presença de um ou mais parâmetros. Assim, um membro de uma família é uma entre as possíveis variedades de formas – mas cada variedade é precisa e unicamente determinada pelo valor de seus parâmetros.

Nesta seção e pelo bem da simplicidade, a entropia de uma distribuição (discreta ou contínua) de probabilidade, será denotada por \mathcal{H} . Haverá como pressuposto que a entropia diferencial já fora previamente calculada em função dos parâmetros. Assim, o objeto de análise será a variação de entropia (ou de incerteza) conforme a variação dos parâmetros.

Para informações sobre definição de gradiente e notações para tratar campos vetoriais e campos escalares pregados neste capítulo, ver apêndice C (pag. 185).

8.5 Axiomas para Qualidade

(MONTGOMERY, 2012) acredita que a qualidade é inversamente proporcional à variabilidade indesejada e conclui que, com essa definição, se a variabilidade de uma característica importante no produto diminui então a qualidade aumenta. As três propriedades da qualidade preconizadas por Montgomery encontram-se na seção 4.2, página 46 desta tese.

Os axiomas aqui apresentados, entretanto, procuram contornar as anomalias do sistema de propriedades de Montgomery. A maior diferença será, que no lugar de relacionar a qualidade à variabilidade, a qualidade e a incerteza serão relacionadas. Entretanto, nesta seção, o conceito de incerteza é usado em *lato sensu*, isto é: não é necessariamente relacionada à incerteza da teoria da informação.

Existem três oportunidades, apesar de o conceito de incerteza ser um tanto difuso: nos capítulos anteriores se ilustrou dois tipos de incerteza na teoria da informação – a entropia e entropia diferencial; também se ilustrou um mecanismo de ponderação inspirado no equilíbrio químico capaz de tratar ambas da mesma forma; e a última, e não menos importante, é que a incerteza de distribuições normais multidimensionais é uma função da matriz de covariância, que é uma generalização do desvio-padrão, que é, por sua vez, uma medida de variabilidade.

Axioma 8.5.1. *Seja $d\mathcal{H}$ um infinitésimo de incerteza, $d\mathcal{Q}$ um infinitésimo de qualidade, $w \in \mathbb{R}$ peso de subjetividade então:*

$$d\mathcal{Q} = w d\mathcal{H} \tag{8.5.1}$$

A intensidade da variação infinitesimal de qualidade é diretamente proporcional à intensidade da variação infinitesimal de incerteza. Se $w > 0$ então a qualidade aumenta com o aumento de incerteza. Se $w < 0$ então a qualidade diminui com o aumento de incerteza. Se $w = 0$ então a variação de qualidade é indiferente à variação de incerteza.

No axioma acima, entretanto, não se considera que a qualidade seja inversamente proporcional à variabilidade, mas sim que a variação de qualidade seja proporcional à variação de incerteza e que a qualidade aumenta ou diminui de acordo com o sinal aritmético de w (infinitesimalmente). Com isso será possível contemplar as situações nas quais, por exemplo, o aumento de variabilidade é desejável para caracterizar o aumento de qualidade.

Axioma 8.5.2. *Seja \mathcal{H} um campo escalar. \mathcal{H} é um campo potencial de incerteza se $\Delta\mathcal{H}$ denota uma diferença de incerteza entre dois pontos nesse campo.*

Corolário 8.5.3. *Caso $\nabla\mathcal{H}$ exista para todo domínio de \mathcal{H} , o axioma 8.5.2 é equivalente a:*

$$\oint_C \nabla \mathcal{H} \cdot d\boldsymbol{\theta} = 0 \quad (8.5.2)$$

O axioma 8.5.2 captura a noção de que: é possível medir a incerteza de um sistema e que essa incerteza é exatamente conhecida e especificada por um conjunto de parâmetros. Ao mesmo tempo, admite-se existir diferenças de incerteza, mas sua intensidade absoluta não pode ser exatamente calculada ou conhecida. Sucintamente, a incerteza nesse modelo também é uma função de estado, que é caracterizada pelos parâmetros em $\boldsymbol{\theta}$ (independentemente do caminho percorrido entre os estados inicial e final).

Definição 8.5.1. *Seja Q uma grandeza que denota uma ponderação de qualitativa, então:*

$$Q = \sum_{i=1}^n w_i \int_C \nabla \mathcal{H}_i \cdot d\boldsymbol{\theta} \quad (8.5.3)$$

Onde \mathcal{H}_i é uma grandeza que satisfaz os axiomas 8.5.1 e 8.5.1.

O axioma acima captura a noção de que se pode ponderar diferenças de incertezas para gerar um valor numérico: o valor da decisão qualitativa. Cada termo do somatório é necessariamente uma diferença de incertezas com um peso – então não concebe falar sobre incertezas absolutas.

O modelo acima mostra potencialidade para lidar com incertezas de processos descritos por variáveis aleatórias contínuas. No capítulo 7, observou-se que a diferença de *entropia diferencial* pode ter a mesma interpretação que a diferença de *entropia*, considerando-se a hipótese isocórica: corolário 7.5.2 e definição 7.5.5. Entretanto, a abordagem de campos conservativos acima não contempla o caso multipolar.

Ao final deste capítulo, será argumentado que esse modelo proposto não impede o uso de variáveis aleatórias discretas.

8.6 Distribuição Normal Unidimensional

As definições nesta seção foram extraídas mormente de (BISHOP, 2006; COVER e THOMAS, 1991; SHANNON, 1948). Aqui será realizada a apresentação dos conceitos utilizando-se a distribuição normal unidimensional para ambientar o leitor para o caso mais genérico posteriormente abordado.

O objetivo é evidenciar a relação matemática entre pequenas variações nos parâmetros de uma distribuição normal unidimensional e as variações na qualidade descrita por um modelo de decisão qualitativa.

Definição 8.6.1 (Distribuição Normal Unidimensional). *Uma distribuição normal unidimensional é denotada por $\mathcal{N}(\mu, \sigma^2)$, onde μ e σ são, respectivamente, a média e o desvio-padrão dessa distribuição. Seja $\mathbb{X}_{\mathcal{N}}$ uma variável aleatória contínua definida por $\mathcal{N}(\mu, \sigma^2)$. A função de densidade de probabilidade de $\mathbb{X}_{\mathcal{N}}$ é denotada por $\rho_{\mathbb{X}_{\mathcal{N}}}(\mathbf{x})$ e tal que:*

$$\rho_{\mathbb{X}_{\mathcal{N}}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8.6.1)$$

Onde $\mathbf{x} \in \mathbb{R}$.

Em (BISHOP, 2006, p. 24 e 688), $1/\sigma$ é também chamado de precisão.

Teorema 8.6.1 (Entropia de uma Distribuição Normal Unidimensional). *Seja $\mathcal{N}(\mu, \sigma^2)$ uma distribuição normal multivariada e $H_{dif}(\mathbb{X}_{\mathcal{N}})$ a sua entropia diferencial (em nats) então*

$$H_{dif}(\mathbb{X}_{\mathcal{N}}) = \ln(\sqrt{2\pi e} \cdot \sigma) \quad (8.6.2)$$

Demonstração. Em (HAYKIN, 2009, p. 483), (COVER e THOMAS, 1991) ou propriedade 6 em (SHANNON, 1948), □

Que $H_{dif}(\mathbb{X}_{\mathcal{N}})$ pode ser reescrita como:

$$H_{dif}(\mathbb{X}_{\mathcal{N}}) = \ln(\sqrt{2\pi e}) + \ln(\sigma) \quad (8.6.3)$$

Percebe-se também que a média não é um fator que influencia essa medida de incerteza de $\mathcal{N}(\mu, \sigma)$ e $\ln(\sqrt{2\pi e})$ é uma constante aditiva.

8.6.1 Diferença de Entropia entre duas Distribuições Normais Unidimensionais

Sejam duas Gaussianas, $\mathcal{N}(\mu, \sigma)$ e $\mathcal{N}(\mu_0, \sigma_0)$ que definem, respectivamente, $\mathbb{X}_{\mathcal{N}}$ e $\mathbb{X}_{\mathcal{N}_0}$ então:

$$H_{dif}(\mathbb{X}_{\mathcal{N}}) - H_{dif}(\mathbb{X}_{\mathcal{N}_0}) = \ln(\sqrt{2\pi e} \cdot \sigma) - \ln(\sqrt{2\pi e} \cdot \sigma_0) = \ln(\sigma) - \ln(\sigma_0) \quad (8.6.4)$$

Fazendo $\mathcal{Q}(\mu, \sigma) = w \cdot (H_{dif}(\mathbb{X}_{\mathcal{N}}) - H_{dif}(\mathbb{X}_{\mathcal{N}_0}))$ então:

$$\nabla \mathcal{Q} = \nabla(w \cdot H_{dif}(\mathbb{X}_{\mathcal{N}}) - w \cdot H_{dif}(\mathbb{X}_{\mathcal{N}_0})) \quad (8.6.5)$$

$$\nabla \mathcal{Q} = w \cdot \nabla H_{dif}(\mathbb{X}_{\mathcal{N}}) - w \cdot \nabla H_{dif}(\mathbb{X}_{\mathcal{N}_0}) \quad (8.6.6)$$

Como $H_{dif}(\mathbb{X}_{\mathcal{N}_0})$ não é parâmetro de \mathcal{Q} , mas trata-se de uma constante, então:

$$\nabla \mathcal{Q} = w \cdot \nabla H(\mathbb{X}_{\mathcal{N}}) \quad (8.6.7)$$

Para simplificar a escrita, o argumento $\mathbb{X}_{\mathcal{N}}$ será omitido. Então

$$\nabla \mathcal{Q} = w \nabla H \quad (8.6.8)$$

Importante notar que $\nabla \mathcal{Q}$ aponta para a direção onde \mathcal{Q} cresce mais rapidamente. Se $w > 0$ então cresce na direção de maior incerteza; se $w < 0$ então decresce mais rapidamente na direção de maior incerteza;

$$\nabla \mathcal{Q} = \frac{\partial H}{\partial \mu} \hat{\mathbf{e}}_{\mu} + \frac{\partial H}{\partial \sigma} \hat{\mathbf{e}}_{\sigma} \quad (8.6.9)$$

$$\nabla \mathcal{Q} = 0 \cdot \hat{\mathbf{e}}_{\mu} + \frac{w}{\sigma} \cdot \hat{\mathbf{e}}_{\sigma} \quad (8.6.10)$$

$$\nabla \mathcal{Q} = \frac{w}{\sigma} \hat{\mathbf{e}}_{\sigma} \quad (8.6.11)$$

Se $d\boldsymbol{\theta}$ for um infinitésimo vetorial tal que

$$d\boldsymbol{\theta} = d\mu \hat{\mathbf{e}}_{\mu} + d\sigma \hat{\mathbf{e}}_{\sigma} \quad (8.6.12)$$

Onde $\hat{\mathbf{e}}_{\mu}$ e $\hat{\mathbf{e}}_{\sigma}$ são vetores ortonormais entre si; e correspondem, cada um, aos parâmetros de $\mathcal{N}(\mu, \sigma)$ que definem $\mathbb{X}_{\mathcal{N}}$.

Seja $d\mathcal{Q} = \nabla \mathcal{Q} \cdot d\boldsymbol{\theta}$ então

$$d\mathcal{Q} = \nabla \mathcal{Q} \cdot d\boldsymbol{\theta} = \left(\frac{w}{\sigma} \hat{\mathbf{e}}_{\sigma} \right) \cdot (d\mu \hat{\mathbf{e}}_{\mu} + d\sigma \hat{\mathbf{e}}_{\sigma}) \quad (8.6.13)$$

$$d\mathcal{Q} = \left(\frac{w}{\sigma} d\mu \hat{\mathbf{e}}_{\sigma} \cdot \hat{\mathbf{e}}_{\mu} + d\sigma \frac{w}{\sigma} \hat{\mathbf{e}}_{\sigma} \cdot \hat{\mathbf{e}}_{\sigma} \right) \quad (8.6.14)$$

$$d\mathcal{Q} = \left(\frac{w}{\sigma} d\mu 0 + d\sigma \frac{w}{\sigma} 1 \right) \quad (8.6.15)$$

$$d\mathcal{Q} = w \cdot \frac{1}{\sigma} d\sigma \quad (8.6.16)$$

Podendo ser reescrita como:

$$\mathcal{Q} = w \ln(\sigma/\sigma_0) \quad (8.6.17)$$

Análise 8.6.2. Seja $\mathcal{N}(\mu, \sigma)$ um modelo cujo desvio-padrão σ é a métrica de variabilidade e a média μ é o padrão de adequação para o uso. Globalmente, \mathcal{Q} sempre

aumenta quando σ diminui, e vice-versa, pois $\sigma > 0$ – para $w < 0$. Localmente, σ possui um efeito inercial em ΔQ para $\Delta\sigma$ pequeno. Então, Q é parcialmente consistente com a proposição 4.2.2 no sentido que Q sempre decresce com o aumento de variabilidade, mas Q não é inversamente proporcional à variabilidade.

O conceito de redução da variabilidade é chave em processos de melhoria da qualidade, ver proposição 4.2.3. O conceito de μ de uma distribuição normal unidimensional seria análogo a satisfazer a propriedade 4.2.1 de Montgomery. Supondo que um processo de qualidade almeja diminuir a variabilidade, então maiores seriam as chances de se obter um valor adequado para uso num intervalo previamente definido – desde que μ pertença sempre a esse intervalo.

É digno de nota, que no campo das ciências cognitivas, as expressões 8.6.16 e 8.6.17 são análogas à lei de Weber-Fechner (apenas de uma perspectiva formal). Esta observação sugere uma possível conexão com a qualidade diferencial de $\mathcal{N}(\mu, \sigma^2)$, entropia e estudos sobre a percepção sensorial. Em (NUSSENZVEIG, 1994, p. 211), a lei de Weber-Fechner é considerada uma “lei” empírica da psicofísica na qual um estímulo é proporcional ao logaritmo de “excitação” sensorial estudada. Se a analogia fosse estendida ao campo semântico dessa lei, σ seria análogo à excitação e σ_0 seria ao limiar da percepção de variabilidade (a menor intensidade de variabilidade que pode ser percebida). Para mais informações sobre a lei de Weber-Fechner, ver (MASIN *et al.*, 2009).

8.7 Distribuição Normal Multivariada

A generalização da densidade normal unidimensional para várias dimensões tem um papel fundamental em vários campos de conhecimento. Dados reais, em geral, dizem respeito a mais de uma dimensão e a densidade normal é frequentemente uma aproximação útil para a verdadeira distribuição de uma população, por exemplo numa análise exploratória de dados – em especial no que diz respeito a medidas de dispersão. Além disso, será possível verificar a generalização de um conceito de variação de qualidade com base em uma medida de variabilidade mais genérica.

O objetivo é evidenciar como uma pequena variação nos parâmetros de uma distribuição normal multivariada afeta uma decisão qualitativa.

Definição 8.7.1 (Média Ponderada (ou Esperança) de uma Variável Aleatória Contínua). *Seja \vec{X} um vetor de variáveis aleatórias contínuas. $E[\vec{X}]$ denota a média de \vec{X} , X_i denota a i -ésima variável aleatória contínua de \vec{X} , $E[X_i]$ denota a média de X_i , $\rho_{\vec{X}}$ denota a função de densidade de probabilidade de \vec{X} e ρ_{X_i} denota a função de densidade de probabilidade de X_i tais que:*

$$E[\vec{\mathbb{X}}] = \int \cdots \int_{-\infty}^{+\infty} (x_1, \dots, x_i, \dots, x_n) \cdot \rho_{\vec{\mathbb{X}}}(x_1, \dots, x_i, \dots, x_n) dx_1, \dots, dx_i, \dots, dx_n \quad (8.7.1)$$

$$E[\mathbb{X}_i] = \int_{-\infty}^{+\infty} x_i \cdot \rho_{\mathbb{X}_i}(x_i) dx_i \quad (8.7.2)$$

Definição 8.7.2 (Covariância). *Sejam \mathbb{X} e \mathbb{Y} variáveis aleatórias contínuas. $\text{cov}(\mathbb{X}, \mathbb{Y})$ denota a covariância entre \mathbb{X} e \mathbb{Y} tal que*

$$\text{cov}(\mathbb{X}, \mathbb{Y}) = E[(\mathbb{X} - E[\mathbb{X}])(\mathbb{Y} - E[\mathbb{Y}])] \quad (8.7.3)$$

A covariância é uma medida não normalizada da intensidade de como duas dimensões de uma variável aleatória vetorial se comportam juntamente, sendo estimada através do produto dos desvios da média para cada dimensão.

Definição 8.7.3 (Matriz de Covariância e outras definições). *$\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são, respectivamente, o vetor da média e a matriz de covariância do vetor de variáveis aleatórias $\vec{\mathbb{X}} = (\mathbb{X}_1, \dots, \mathbb{X}_n)$ tais que:*

- $\boldsymbol{\mu} = (\mu_1, \dots, \mu_i, \dots, \mu_n)$, onde $\mu_i = E[\mathbb{X}_i]$, para $1 \leq i \leq n$

$$\bullet \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_{n,n} \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Onde $\sigma_{i,j} = \text{cov}(\mathbb{X}_i, \mathbb{X}_j)$, para $1 \leq i \leq n$ e $1 \leq j \leq n$.

Além disso, $\boldsymbol{\Sigma}^*$ denota o determinante de $\boldsymbol{\Sigma}$ e $\boldsymbol{\Sigma}^{-1}$ denota a matriz inversa de $\boldsymbol{\Sigma}$. A covariância entre uma dimensão e ela mesma é a sua variância. As variâncias são encontradas na diagonal principal de $\boldsymbol{\Sigma}$. σ_i^2 é a variância da i -ésima dimensão e $\sigma_i^2 = \sigma_{i,i}$. σ_i é o desvio-padrão da i -ésima dimensão e $\sigma_i = \sqrt{\sigma_{i,i}}$ ou, equivalentemente, $\sigma_i = \sqrt{\sigma_i^2}$.

Para fins de simplicidade e parcimônia, se considerará apenas as distribuições de probabilidade para as quais existam o vetor da média e a matriz de covariância.

Propriedade 8.7.1. *Por construção, $\boldsymbol{\Sigma}$ é uma matriz simétrica, pois*

$$\sigma_{i,j} = \text{cov}(\mathbb{X}_i, \mathbb{X}_j) = \text{cov}(\mathbb{X}_j, \mathbb{X}_i) = \sigma_{j,i} \quad (8.7.4)$$

Propriedade 8.7.2. *$\boldsymbol{\Sigma}$ é uma matriz positiva definida. O que implica $\boldsymbol{\Sigma}^* > 0$; ou, é uma matriz positiva semidefinida, o que implica $\boldsymbol{\Sigma}^* \geq 0$, (RENCHEE e SCHALJE, 2008, p. 88-90).*

Essa propriedade dispensa trabalhar com o módulo de Σ^* . Ou seja, $\Sigma^* = |\Sigma^*|$. O que será muito útil no caso desse determinante ser argumento da função logarítmica.

Observação 8.7.3. Para efeitos práticos, neste trabalho, será considerado que Σ seja uma matriz positiva definida, o que do ponto de vista de análise, evita o tratamento de divisões. A razão disso é que quando Σ especifica uma distribuição normal multivariada com pouca intercorrelação, então $\Sigma^* > 0$. Em caso contrário, $\Sigma^* = 0$, então haveria muita multicolinearidade (intercorrelação) entre as variáveis aleatórias que descrevem essa distribuição (RENCHEER e SCHAALJE, 2008, p. 90) – o que indica que o processo descrito poderia ser representado por um número menor de variáveis do ponto de vista da aleatoriedade.

Definição 8.7.4 (Distribuição Normal Multivariada). *Uma distribuição normal multivariada é denotada por $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, onde $\boldsymbol{\mu}$ e Σ são, respectivamente, o vetor da média e a matriz de covariância de $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. Além disso, $\boldsymbol{\mu} \in \mathbb{R}^n$ e $\Sigma \in \mathbb{R}^{n \times n}$. Seja $\vec{X}_{\mathcal{N}_n}$ um vetor das variáveis aleatórias contínuas definidas por $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. A função de densidade de probabilidade $\rho_{\vec{X}_{\mathcal{N}_n}}(\mathbf{x})$ é tal que:*

$$\rho_{\vec{X}_{\mathcal{N}_n}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \Sigma^*}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (8.7.5)$$

Onde $\mathbf{x} = (x_1, \dots, x_n)$ e $(\mathbf{x} - \boldsymbol{\mu})^T$ é a matriz transposta da diferença matricial $(\mathbf{x} - \boldsymbol{\mu})$. Em (BISHOP, 2006, p. 25 e 688), Σ^{-1} é também chamada de matriz de precisão.

Observação 8.7.4. A matriz de covariância generaliza a noção de variância para múltiplas dimensões (RENCHEER e SCHAALJE, 2008, p. 77): a variabilidade de um conjunto de pontos aleatórios no espaço em duas, ou mais dimensões, não pode ser completamente caracterizada por um único número.

Teorema 8.7.5 (Entropia de uma Distribuição Normal Multivariada). *Seja $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$ uma distribuição normal multivariada e $H_{dif}(\vec{X}_{\mathcal{N}_n})$ a sua entropia diferencial então*

$$H_{dif}(\vec{X}_{\mathcal{N}_n}) = \ln \sqrt{(2\pi e)^n \Sigma^*} \quad (8.7.6)$$

Demonstração. Teorema 9.4.1 em (COVER e THOMAS, 1991, p. 230). \square

Propriedade 8.7.6. *A entropia da distribuição normal multivariada não depende do $\boldsymbol{\mu}$, assim como no caso da distribuição normal unidimensional.*

8.7.1 Diferença de Entropia entre duas Distribuições Normais Multivariadas de Mesma Dimensão

Essencialmente a análise feita para a distribuição normal unidimensional será refeita: como a qualidade varia em função dos parâmetros. Entretanto, serão observados dois casos: com maior granularidade em que Σ^* é o parâmetro e outro, com menor granularidade, onde os parâmetros são as covariâncias.

Definição 8.7.5 (Dipolo Diferencial Padronizado). *Sejam $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\mathcal{N}_n^0(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ distribuições normais multivariadas. \mathcal{Q} é uma decisão qualitativa tal que*

$$\mathcal{Q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = w(H_{dif}(\vec{\mathbb{X}}_{\mathcal{N}_n}) - H_{dif}(\vec{\mathbb{X}}_{\mathcal{N}_n^0})) \quad (8.7.7)$$

Onde $w \in \mathbb{R}$ e $w \neq 0$.

Nesse caso, \mathcal{Q} que é um dipolo diferencial padronizado por \mathcal{N}_n^0 .

Então o dipolo diferencial padronizado 8.7.7 pode ser reescrito como:

$$\mathcal{Q} = w \left(\ln \sqrt{(2\pi e)^n \Sigma^*} - \ln \sqrt{(2\pi e)^n \Sigma_0^*} \right) \quad (8.7.8)$$

$$\mathcal{Q} = w \left(\ln \sqrt{(2\pi e)^n} + \ln \sqrt{\Sigma^*} - \ln \sqrt{(2\pi e)^n} - \ln \sqrt{\Sigma_0^*} \right) \quad (8.7.9)$$

$$\mathcal{Q} = w \left(\ln \sqrt{\Sigma^*} - \ln \sqrt{\Sigma_0^*} \right) \quad (8.7.10)$$

O que resulta, sucintamente, em

$$\mathcal{Q} = w \frac{1}{2} \ln \frac{\Sigma^*}{\Sigma_0^*} \quad (8.7.11)$$

Como pode ser observado acima, a diferença de entropia diferencial entre duas distribuições normais multivariadas, ambas com o mesmo número de dimensões (a saber: n), é proporcional ao logaritmo natural da razão entre suas matrizes de covariância. Esse resultado informa que a matriz de covariância tem uma relação direta com a mensuração de incerteza em um processo que pode ser modelado com tal distribuição.

O papel da distribuição \mathcal{N}_n^0 é ser um referencial para calcular essa diferença. A distribuição \mathcal{N}_n^0 funciona como um processo que é escolhido para ser um marco zero, ou melhor, um padrão para medir as diferenças de qualidade entre processos. Intuitivamente, processos gaussianos equivalentes a \mathcal{N}_n^0 seriam aqueles cujos determinantes são iguais a Σ_0^* , o que sugere a possibilidade de superfícies isoqualitativas.

8.7.2 Granularidade de Parametrização

Agora ilustrar-se-á como a qualidade é afetada por pequenas variações dos parâmetros de uma distribuição normal multivariada. Então essencialmente serão evidenciadas as expressões particulares para a distribuição normal multivariada, observando-se:

$$d\mathcal{Q} = \nabla \mathcal{Q} \cdot d\boldsymbol{\theta} \quad (8.7.12)$$

A partir do resultado 8.7.11, fazendo $\boldsymbol{\theta} = \Sigma^{\otimes}$ e lembrando que Σ_0^{\otimes} é uma constante então

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} d\Sigma^{\otimes} \quad (8.7.13)$$

Caso $\Sigma = [\sigma_{1,1}]$, ou seja equivalente à distribuição normal unidimensional, então:

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\sigma_{1,1}} d\sigma_{1,1} \quad (8.7.14)$$

Que pode ser reescrita – usando a notação da distribuição normal unidimensional – fazendo-se $\sigma_{1,1} = \sigma^2$:

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\sigma^2} d\sigma^2 \quad (8.7.15)$$

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\sigma^2} \cdot 2\sigma d\sigma \quad (8.7.16)$$

$$d\mathcal{Q} = w \cdot \frac{1}{\sigma} d\sigma \quad (8.7.17)$$

Voltando para o caso multidimensional, a partir da expressão 8.7.13, fazendo $\boldsymbol{\theta} = \Sigma$ e lembrando que Σ_0^{\otimes} é uma constante, para calcular $d\mathcal{Q}$ é necessário saber como $d\Sigma^{\otimes}$ se comporta localmente em função das covariâncias. Se for levado em conta que Σ^{\otimes} é uma função das covariâncias é preciso escrever a diferencial:

$$d\Sigma^{\otimes} = \nabla \Sigma^{\otimes} \cdot d\boldsymbol{\theta} \quad (8.7.18)$$

Onde,

$$d\boldsymbol{\theta} = \sum_{j=1}^n \sum_{i=1}^j \hat{\mathbf{e}}_{\sigma_{i,j}} d\sigma_{i,j} \quad (8.7.19)$$

$$\nabla \Sigma^{\otimes} = \sum_{j=1}^n \sum_{i=1}^j \frac{\partial \Sigma^{\otimes}}{\partial \sigma_{i,j}} \hat{\mathbf{e}}_{\sigma_{i,j}} \quad (8.7.20)$$

Observação 8.7.7. É importante notar que os dois somatórios acima levam em consideração o fato de Σ ser uma matriz simétrica (o que reduz o número de graus de liberdade no que tange ao número de covariâncias se comparado ao número de elementos da matriz de covariância).

Assim, a expressão 8.7.13 pode ser reescrita como

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \nabla \Sigma^{\otimes} \cdot d\boldsymbol{\theta} \quad (8.7.21)$$

Obviamente,

$$\nabla \mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \nabla \Sigma^{\otimes} \quad (8.7.22)$$

Continuando,

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \left(\sum_{j=1}^n \sum_{i=1}^j \frac{\partial \Sigma_1^{\otimes}}{\partial \sigma_{i,j}} \hat{\mathbf{e}}_{\sigma_{i,j}} \right) \cdot d\boldsymbol{\theta} \quad (8.7.23)$$

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \sum_{j=1}^n \sum_{i=1}^j \frac{\partial \Sigma_1^{\otimes}}{\partial \sigma_{i,j}} d\sigma_{i,j} \quad (8.7.24)$$

Se as dimensões do vetor $\mathbb{X}_{\mathcal{N}_n}$ não são correlacionadas – isto é $\sigma_{i,j} = 0$, se $i \neq j$ – então

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \sum_{i=1}^n \frac{\partial \Sigma^{\otimes}}{\partial \sigma_{i,i}} d\sigma_{i,i} \quad (8.7.25)$$

Como Σ é uma matriz triangular superior e inferior, devido à hipótese de não-correlação, então $\Sigma^{\otimes} = \prod_{k=1}^n \sigma_{k,k}$. O que permite escrever

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\prod_{i=1}^n \sigma_{i,i}} \cdot \sum_{i=1}^n \frac{\partial \prod_{k=1}^n \sigma_{k,k}}{\partial \sigma_{i,i}} d\sigma_{i,i} \quad (8.7.26)$$

Observando-se que $\frac{\partial \prod_{k=1}^n \sigma_{k,k}}{\partial \sigma_{i,i}} = \frac{\prod_{k=1}^n \sigma_{k,k}}{\sigma_{i,i}}$ então

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \frac{1}{\prod_{i=1}^n \sigma_{i,i}} \cdot \sum_{i=1}^n \frac{\prod_{k=1}^n \sigma_{k,k}}{\sigma_{i,i}} d\sigma_{i,i} \quad (8.7.27)$$

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \sum_{i=1}^n \frac{1}{\sigma_{i,i}} d\sigma_{i,i} \quad (8.7.28)$$

Como σ_i^2 é a variância da i -ésima dimensão e $\sigma_i^2 = \sigma_{i,i}$ então

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \sum_{i=1}^n \frac{1}{\sigma_i^2} d\sigma_i^2 \quad (8.7.29)$$

$$d\mathcal{Q} = w \cdot \frac{1}{2} \cdot \sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot 2\sigma_i d\sigma_i \quad (8.7.30)$$

$$d\mathcal{Q} = \sum_{i=1}^n w \cdot \frac{1}{\sigma_i} d\sigma_i \quad (8.7.31)$$

Fazendo $d\mathcal{Q}_i = w \cdot \frac{1}{\sigma_i} d\sigma_i$ (que é, também, a variação para o caso unidimensional),

$$d\mathcal{Q} = \sum_{i=1}^n d\mathcal{Q}_i \quad (8.7.32)$$

Análise 8.7.8. $d\mathcal{Q}$, para a hipótese de não-correlação, é igual a soma da variação de qualidade – se cada dimensão fosse observada isoladamente e para a mesma constante w . Se $w = 1$, é o mesmo que somar a variação de incerteza (ou entropia) em *nats* de cada dimensão. O que equivale dizer que, ao não haver correlação em um processo modelado por uma distribuição normal multivariada, então toda variação de incerteza pode ser inferida pela variação de incerteza de cada dimensão, como no caso unidimensional – porém, o peso w das dimensões deve ser o mesmo. Por fim, observa-se que por variação entende-se variação infinitesimal.

8.8 Aplicação: Qualions

Aqui será apresentado o conceito de qualion, que nada mais é do que o gradiente da entropia (nos parâmetros de uma distribuição de probabilidade) multiplicado por um escalar. A ideia do qualion é que ele é um vetor com a mesma direção da incerteza. Ao mesmo tempo, informa quão intensa é essa variação de qualidade localmente.

Definição 8.8.1. *Seja \mathcal{Q} uma grandeza de qualidade que obedece o axioma 8.5.1. Além disso, \mathcal{Q} é uma função de $\boldsymbol{\theta}$. O vetor $\vec{q}_{\boldsymbol{\theta}}$ é um qualion em \mathcal{Q} , se $\vec{q}_{\boldsymbol{\theta}} = \nabla \mathcal{Q}(\boldsymbol{\theta})$.*

Corolário 8.8.1. *Seja $\vec{q}_{\boldsymbol{\theta}}$ um qualion em \mathcal{Q} . Então $\vec{q}_{\boldsymbol{\theta}} = w \cdot \nabla \mathcal{H}(\boldsymbol{\theta})$, $w \in \mathbb{R}_*^+$.*

Demonstração. Direta do axioma 8.5.1,

$$\vec{q}_{\boldsymbol{\theta}} = \nabla \mathcal{Q}(\boldsymbol{\theta}) = \nabla (w \cdot \mathcal{H})(\boldsymbol{\theta}) = w \cdot \nabla \mathcal{H}(\boldsymbol{\theta})$$

□

Corolário 8.8.2. *Sejam $\vec{q}_{\boldsymbol{\theta}_1}$ e $\vec{q}_{\boldsymbol{\theta}_2}$ qualions em \mathcal{Q} ; e ϕ o ângulo entre $\vec{q}_{\boldsymbol{\theta}_1}$ e $\vec{q}_{\boldsymbol{\theta}_2}$.*

$$\cos \phi = \frac{\nabla \mathcal{H}(\boldsymbol{\theta}_1) \cdot \nabla \mathcal{H}(\boldsymbol{\theta}_2)}{|\nabla \mathcal{H}(\boldsymbol{\theta}_1)| |\nabla \mathcal{H}(\boldsymbol{\theta}_2)|}$$

Demonstração.

$$\cos \phi = \frac{\vec{q}_{\theta_1} \cdot \vec{q}_{\theta_2}}{|\vec{q}_{\theta_1}| |\vec{q}_{\theta_2}|} = \frac{w \cdot \nabla \mathcal{H}(\theta_1) \cdot w \cdot \nabla \mathcal{H}(\theta_2)}{|w \cdot \nabla \mathcal{H}(\theta_1)| |w \cdot \nabla \mathcal{H}(\theta_2)|} = \frac{w^2 \cdot \nabla \mathcal{H}(\theta_1) \cdot \nabla \mathcal{H}(\theta_2)}{|w|^2 \cdot |\nabla \mathcal{H}(\theta_1)| |\nabla \mathcal{H}(\theta_2)|}$$

Então,

$$\cos \phi = \frac{\nabla \mathcal{H}(\theta_1) \cdot \nabla \mathcal{H}(\theta_2)}{|\nabla \mathcal{H}(\theta_1)| |\nabla \mathcal{H}(\theta_2)|}$$

□

Ou seja, $\cos \phi$ não depende do peso w . Por outro lado, há outra interpretação possível: w pode ser a constante de mudança de base do logaritmo, que define a unidade de informação, para \mathcal{H} – para $w > 0$. Assim, mesmo mudando-se a unidade também não haveria mudança do cosseno.

8.8.1 Incerteza derivada de Densidade de Probabilidade: Gaussiana Multidimensional

Usando a definição de qualion, se Σ for a matriz de covariância de uma gaussiana multidimensional e fazendo $\theta = \Sigma$ então, com base na expressão 8.7.22, na página 112:

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \nabla \Sigma^{\otimes}$$

Exemplo 8.8.1 (Gaussiana Bidimensional). Seja $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}$. Onde $-1 < \rho < 1$. Onde σ_x^2 , σ_y^2 e ρ^2 são, respectivamente, a variância em x , a variância em y e o coeficiente de correlação ao quadrado. Então $\Sigma^{\otimes} = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$. Fazendo $\gamma^2 = 1 - \rho^2$. Então $\Sigma^{\otimes} = \sigma_1^2 \sigma_2^2 \gamma^2$.

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \nabla \Sigma^{\otimes} \quad (8.8.1)$$

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \frac{1}{\sigma_x^2 \sigma_y^2 \gamma^2} \cdot \nabla (\sigma_x^2 \sigma_y^2 \gamma^2) \quad (8.8.2)$$

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \frac{1}{\sigma_1^2 \sigma_2^2 \gamma^2} \cdot \left(\sigma_y^2 \gamma^2 \hat{\mathbf{e}}_{\sigma_x^2} + \sigma_x^2 \gamma^2 \hat{\mathbf{e}}_{\sigma_y^2} + \sigma_x^2 \sigma_y^2 \hat{\mathbf{e}}_{\gamma^2} \right) \quad (8.8.3)$$

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \left(\frac{1}{\sigma_x^2} \hat{\mathbf{e}}_{\sigma_x^2} + \frac{1}{\sigma_y^2} \hat{\mathbf{e}}_{\sigma_y^2} + \frac{1}{\gamma^2} \hat{\mathbf{e}}_{\gamma^2} \right) \quad (8.8.4)$$

Onde $0 < \sigma_x^2$, $0 < \sigma_y^2$ e $0 < \gamma^2 \leq 1$.

A escolha de γ^2 como uma dimensão no lugar de ρ traz mais uniformidade a expressão que modela o qualion do sistema, entretanto, uma vez conhecido o qualion a comparação pode ser realizada – seja medindo ângulos, intensidade e/ou distâncias vetoriais, por exemplo. γ^2 , por construção, é um coeficiente de não-correlação.

A qualidade aqui proposta é uma função da incerteza, pelo produto de uma constante real não-nula, faz a título de interpretação as seguintes observações.

Exemplo 8.8.2 (Gaussiana Bidimensional). Seja $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$. Onde $-1 < \rho < 1$. Onde σ_x^2 , σ_y^2 e ρ^2 são, respectivamente, a variância em x , a variância em y e o coeficiente de correlação ao quadrado. Então $\Sigma^{\otimes} = \sigma_1^2\sigma_2^2(1 - \rho^2)$.

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \frac{1}{\Sigma^{\otimes}} \cdot \nabla \Sigma^{\otimes} \quad (8.8.5)$$

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \frac{1}{\sigma_x^2\sigma_y^2(1 - \rho^2)} \cdot \nabla(\sigma_x^2\sigma_y^2(1 - \rho^2)) \quad (8.8.6)$$

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \cdot \left(\sigma_y^2(1 - \rho^2)\hat{e}_{\sigma_x^2} + \sigma_x^2(1 - \rho^2)\hat{e}_{\sigma_y^2} + \sigma_x^2\sigma_y^2(-2\rho)\hat{e}_{\rho} \right) \quad (8.8.7)$$

$$\vec{q}_{\Sigma} = w \cdot \frac{1}{2} \cdot \left(\frac{1}{\sigma_x^2}\hat{e}_{\sigma_x^2} + \frac{1}{\sigma_y^2}\hat{e}_{\sigma_y^2} + \frac{2\rho}{\rho^2 - 1}\hat{e}_{\rho} \right) \quad (8.8.8)$$

Onde $0 < \sigma_x^2$, $0 < \sigma_y^2$ e $-1 \leq \rho \leq 1$.

Os dois exemplos acima ilustram que um mesmo processo, mas observado por parâmetros diferentes resulta em qualions diferentes.

8.8.2 Incerteza derivada de Probabilidade: Variáveis Categóricas

A partir de variáveis aleatórias categóricas não ordinais, como o nome de um país ou o estado civil, não é possível se obter estatísticas como médias ou desvios-padrões. Entretanto, a incerteza associada a essa variável pode ser calculada, pois só depende da função de distribuição de probabilidades.

Um exemplo de problemas de otimização envolvendo o princípio da máxima entropia é aquele em que se obtém a distribuição de Boltzmann. Ela é a solução¹ para (JAYNES, 1962, 1957; SUSSKIND, 2013) e (SCHMALIAN, 2012, p. 18-23):

- maximizar $h = -\sum_{i=1}^n p_i \ln p_i$ tal que $\forall i : p_i > 0$.

¹Ver apêndice B para desenvolvimento da solução.

• sujeito às restrições:

1. $1 = \sum_{i=1}^n p_i$;
2. $\mu = \sum_{i=1}^n e_i p_i$, onde $\mu, e_1, \dots, e_i, \dots, e_n \in \mathbb{R}$ são *dados* do problema.

Assim, a distribuição de Boltzmann modela um experimento em que p maximiza a incerteza e satisfaz a média μ (observando que cada e_i é valor assumido pelo i -ésimo resultado). Como o domínio de h não admite valores negativos nem nulos para qualquer p_i , a restrição 1 obriga a $0 < p_i < 1$.

Uma técnica utilizada para resolver esse tipo de problema são os multiplicadores de Lagrange. Nesse problema em particular, da distribuição de Boltzmann, se existir uma solução, essa será única. Uma parte importante dessa técnica é obter ∇h .

$$\nabla h = \sum_{i=1}^n \left(\ln \frac{1}{p_i} - 1 \right) \hat{\mathbf{e}}_{p_i} \quad (8.8.9)$$

Para qualquer distribuição p , haverá um valor para ∇h , mas nem todos os pontos do domínio de ∇h são de fato uma distribuição de probabilidades. Além disso, o vetor ∇h tem a direção e sentido do maior aumento de h .

Se p for uma distribuição de *probabilidade discreta*, fazendo $\theta = p$ e considerando $\mathcal{H} = h$ então:

$$\vec{q}_p = w \cdot \nabla h \quad (8.8.10)$$

$$\vec{q}_p = w \cdot \sum_{i=1}^n \frac{\partial h}{\partial p_i} \hat{\mathbf{e}}_{p_i} \quad (8.8.11)$$

Como $h = -\sum_{i=1}^n p_i \ln p_i$. Então $\frac{\partial h}{\partial p_i} = -\ln p_i - 1$,

$$\vec{q}_p = w \cdot \sum_{i=1}^n \left(\ln \frac{1}{p_i} - 1 \right) \hat{\mathbf{e}}_{p_i} \quad (8.8.12)$$

Percebe-se que se $\vec{q}_p = \vec{0}$ então $\forall i, p_i = 1/e$. Assim, para fins práticos $\vec{q}_p \neq 0$.

8.9 Discussões e Conclusões

A descrição em campo escalar convida explorar propriedades geométricas da entropia, enquanto incerteza, da perspectiva local do gradiente da incerteza em função dos parâmetros. O que sugere métricas e medidas que incentivam a comparação entre distribuições variáveis aleatórias, que compartilham os mesmos parâmetros. Os qualions nada mais são que o gradiente da entropia enquanto unidade concei-

tual e interpretativa da qualidade. Ângulos e distâncias entre qualions, podem ser trivialmente calculados.

É possível sugerir estudos para comparação com a “métrica” bem conhecida e sucedida em teoria da informação: a *distância Kullback-Leibler* – ou entropia relativa. O conceito de qualion captura o comportamento local de funções dos parâmetros que mensuram a qualidade, além de permitir operações de comparação através das operações clássicas de vetores: em especial o produto interno, o módulo e a distância entre vetores.

Definição 8.9.1 (Distância Kullback-Leibler (ou Entropia Relativa)). *Sejam p e q duas distribuições de probabilidade. $D_{\text{KL}}(p||q)$ denota a distância Kullback-Leibler de q para p tal que:*

$$D_{\text{KL}}(p||q) = \sum_i p_i \log \frac{p_i}{q_i} \quad (8.9.1)$$

Sejam \vec{X} e \vec{Y} duas distribuições de probabilidade. $D_{\text{KL}}(\vec{X}||\vec{Y})$ denota a distância Kullback-Leibler de \vec{Y} para \vec{X} tal que:

$$D_{\text{KL}}(\vec{X}||\vec{Y}) = \int_{\mathcal{D}} \rho_{\vec{X}}(\vec{r}) \log \frac{\rho_{\vec{X}}(\vec{r})}{\rho_{\vec{Y}}(\vec{r})} dr. \quad (8.9.2)$$

A entropia relativa não aparece no texto original de (SHANNON, 1948), mas tem papel de destaque na teoria da informação, pois – apesar de ser uma “pseudo-distância”² – mensura a quantidade de informação que uma variável aleatória contém da outra (COVER e THOMAS, 1991, p. 18). Ela não foi explorada neste trabalho, pois ela não é presente na literatura básica de física que aborda a entropia na termodinâmica. (WELLER-FAHY *et al.*, 2015) fornecem uma lista de distâncias para a distribuição de probabilidades (como também outras distâncias de interesse em aprendizado de máquina).

As medidas de variação de qualidade aqui propostas, são inspiradas nas experiências profissionais de (MONTGOMERY, 2012). Elas preconizam que medidas de qualidade estão associadas a variabilidade dos valores observados. A ideia então foi usar a entropia, pois permite utilizar uma única unidade (que não depende das unidades do desvio-padrão) e, como consequência, possibilita a comparação entre sistemas distintos, mas que possam ser modelados por funções de uma mesma família de distribuição de probabilidade. Além disso, o conceito de qualion permite que distâncias e ângulos possam ser calculados sem o inconveniente da propriedade de pseudo-distância.

²A pseudo-distância: a distância Kullback-Leibler não é uma medida verdadeira de distância entre distribuições, pois não é simétrica e não satisfaz à desigualdade triangular (COVER e THOMAS, 1991, p. 18).

Capítulo 9

Conclusão

Seldom do more than a few of nature's secrets give way at one time. It will be all too easy for our somewhat artificial prosperity to collapse overnight when it is realized that the use of a few exciting words like information, entropy, redundancy, do not solve all our problems.

Claude Elwood Shannon, The Bandwagon

9.1 Discussão

Como pôde ser observado ao longo desta tese, inicialmente formulou-se um sistema de interpretação entre os conceitos abstratos de probabilidade e raridade de um evento, inspirado no paradigma de equilíbrio químico. A medida de entropia formulada por Shannon, pode (e deve ser) interpretada como uma concentração de incerteza média por símbolo (ou resultado). Tal concentração pode ser utilizada para abordar problemas de decisão: escolhendo-se os pesos (ou contribuições) dessas concentrações para a decisão. Para fins práticos, os pesos podem ser escolhidos utilizando-se um perceptron para um problema simples de classificação: o que vincula sua escolha à efetividade de uma decisão. Apesar de o conceito de entropia diferencial ser diferente do de incerteza (ambos formulados por Shannon), tais conceitos podem ser conciliados utilizando a hipótese isocórica. A diferença entre medidas de entropia, conjuntamente com a hipótese isocórica, permite a ponderação de entropias diferenciais mantendo-se a mesma interpretação. Tal interpretação permite abordar a questão como diferença de funções escalares – o que levou à investigação de um campo vetorial conservativo. Este, por sua vez, permite explorar a geometria local da incerteza em função dos parâmetros de interesse. O enfoque da abordagem em campos conservativos não são as funções de distribuição (ou densidade) de probabilidades, mas sim a *geometria* das superfícies da função de *entropia*: tanto para processos contínuos quanto para discretos. Assim, o estudo mais se aproxima

das formas de abordagem utilizadas no campo da física do que no da teoria da informação, como consequência: a presença de vetores e ângulos – que não são vistos corriqueiramente em estatística ou em teoria da informação (talvez pelo predomínio da bem sucedida *distância Kullback-Leibler*) e similares. A razão disso foi a tentativa de buscar uma interpretação para a *entropia diferencial* para operar sempre como uma quantidade.

O conceito de *qualion*, que é simplesmente o gradiente da entropia multiplicado por um escalar, permite comparar membros de uma família de funções de probabilidade usando resultados clássicos da álgebra linear e vetorial.

Um fato importante sobre o conceito de qualion é que ele é definido diretamente do conceito de entropia (incerteza) – ao contrário da *distância Kullback-Leibler*.

9.2 Conclusão

O objetivo desta tese foi mostrar, com o auxílio do conceito de mitos e ponderações qualitativas, que a entropia pode ser racionalmente fornecida a um neurônio perceptron para efetuar um pequeno processo de casamento aproximado de palavras – mesmo que não se tenha efetuado um treinamento propriamente dito. Ao mesmo tempo, se desenvolveu um framework conceitual que pode ser aplicado à problemas em que a incerteza possa ser obtida a partir de modelos probabilísticos. Um exemplo disso é que no lugar de relacionar qualidade e variabilidade procurou-se relacionar qualidade com incerteza – o que deu origem à modelagem por campos conservativos de entropia.

9.3 Trabalhos Futuros

- Estudo comparativo entre a *distância Kullback-Leibler* e distâncias e cossenos de ângulos entre qualions – tanto analítico como numérico.
- Busca por variáveis análogas ao volume e temperatura em modelos para comparação de strings (visto que o conceito de entropia é comum). Na literatura volume e temperatura são propriedades, respectivamente, extensivas e intensivas. As intensivas não dependem do volume, já as extensivas sim). Dois indicativos disso são: os ensembles da termodinâmica estatística e a percepção de que o determinante da matriz de covariância sugere tanto uma medida extensiva (volume da região com alta probabilidade de ocorrência de um evento) quanto intensiva (uma medida resumo da distribuição).
- Um estudo mais consistente sobre as relações entre a física qualitativa e mitologia (iniciado no apêndice A).

Referências Bibliográficas

- ABRAHAM, B., 2012, *Quality improvement through statistical methods*. Statistics for Industry and Technology. Waterloo, Ontario, Canada, Birkhäuser. ISBN: 978-1-4612-1776-3.
- ABU-MOSTAFA, Y. S., MAGDON-ISMAIL, M., LIN, H., 2012, *Learning from data: a short course*. USA, AMLBook.com. ISBN: 978-1-60049-006-4.
- ACKOFF, R. L., 1989, “From data to wisdom”, *Journal of Applied Systems Analysis*, v. 16, pp. 3–9.
- AGARD, 1980, *Aeronautical multilingual dictionary – AGARD (Dictionnaire Aéronautique Multilingue)*. Neuilly-sur-Seine, France, NATO/AGARD.
- AGGARWAL, C. C., YU, P. S., 2009, “A Survey of Uncertain Data Algorithms and Applications”, *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 5 (May), pp. 609–623. ISSN: 1041-4347. doi: 10.1109/TKDE.2008.190.
- AKERLOF, G., 1970, “The Market for ”Lemons”: Quality Uncertainty and the Market Mechanism”, *The Quarterly Journal of Economics*, v. 84, n. 3, pp. 488–500.
- ANS, 2012a, *Glossário temático: saúde suplementar*. 2 ed. Brasília, Editora MS. ISBN: 978-85-334-1971-1. Disponível em: <http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Materiais_por_assunto/saudesup_glossario_site-1.pdf>.
- ANS, 2012b. “Plano Diretor de Tecnologia de Informação 2012 – 2015”. [online], Brasil, agostob. Disponível em: <http://www.ans.gov.br/images/stories/A_ANS/Transparencia_Institucional/Prestacao-de-Contas/Contratos_de_Gestao/PDTI_ANS_2012-2015.pdf>. Aprovado pelo C2i em 25 de julho de 2012, Aprovado pela DICOL em 01 de agosto de 2012 (accessado 4 de Junho 2013).

- ANS, 2012c, *Sistema de Informações de Beneficiários - SIB/XML - Críticas dos campos de dados cadastrais de beneficiários do SIB - versão 2.4*. ANS – Agência Nacional de Saúde Suplementar, 10 de setembro de 2012.
- ANS, 2013a, *Relatório de Atividades 2012*. ANS – Agência Nacional de Saúde Suplementar, a. www.ans.gov.br (acessado em novembro de 2013).
- ANS, 2010a, “Relatório do Cruzamento de Dados SIB x CNIS”. In: *Relatório de Gestão 2009*, Agência Nacional de Saúde Suplementar, pp. 200–209, a. Disponível em: <http://www.ans.gov.br/images/stories/A_ANS/Transparencia_Institucional/Prestacao-de-Contas/Relatorios_de_gestao/Relatorio_Gestao_2009.pdf>.
- ANS, 2011a, “Relatório do Cruzamento de Dados do CADSUS, CPF e CNIS com o Cadastro de Beneficiários do SIB”. In: *Relatório de Gestão 2010*, Agência Nacional de Saúde Suplementar, pp. 238–240, a.
- ANS, 2013b, “Conquistas da ANS em 2010–2012”. In: *Relatório de Gestão 2010-2012*, Agência Nacional de Saúde Suplementar, p. 115, b. Disponível em: <http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Materiais_por_assunto/relatoriodegestao_2010_2012.pdf>.
- ANS, 2012d, “Objetivos e Metas Institucionais e/ou Programáticos”. In: *Relatório de Gestão 2011*, Agência Nacional de Saúde Suplementar, pp. 24–25 e 39–40, d.
- ANS, 2009a, *Instrução Normativa – IN Nº 35, de 03 de abril de 2009 da Diretoria de Desenvolvimento Setorial*. Agência Nacional de Saúde Suplementar, Brasil, abril.
- ANS, 2011b, *Instrução Normativa – IN Nº 46, de 25 de março de 2011 da Diretoria de Desenvolvimento Setorial*. Agência Nacional de Saúde Suplementar, Brasil, março.
- ANS, 2012e, *Instrução Normativa – IN Nº 50, de 25 de setembro de 2012 da Diretoria de Desenvolvimento Setorial*. Agência Nacional de Saúde Suplementar, Brasil, setembro.
- ANS, 2010b, *Relatório de Gestão 2009*. Agência Nacional de Saúde Suplementar, b. Disponível em: <http://www.ans.gov.br/images/stories/A_ANS/Transparencia_Institucional/Prestacao-de-Contas/Relatorios_de_gestao/Relatorio_Gestao_2009.pdf>. www.ans.gov.br (acessado em novembro de 2013).

- ANS, 2011c, *Relatório de Gestão 2010*. Agência Nacional de Saúde Suplementar, março. Disponível em: <http://www.ans.gov.br/images/stories/A_ANS/Transparencia_Institucional/Prestacao-de-Contas/Relatorios_de_gestao/relatorio_gestao_2010.pdf>. Acesso: 29/04/2013.
- ANS, 2011d, “Anexo V - Relatório do Cruzamento de Dados do CADSUS, CPF e CNIS com o Cadastro de Beneficiários do SIB.” In: *Relatório de Gestão 2010*, Agência Nacional de Saúde Suplementar, março. Disponível em: <http://www.ans.gov.br/images/stories/A_ANS/Transparencia_Institucional/Prestacao-de-Contas/Relatorios_de_gestao/relatorio_gestao_2010.pdf>. Acesso: 29/04/2013.
- ANS, 2012f, *Relatório de Gestão 2011*. Agência Nacional de Saúde Suplementar, f.
- ANS, 2013c, *Relatório de Gestão 2012*. Agência Nacional de Saúde Suplementar, c.
- ANS, 2009b, *Resolução Normativa – RN Nº 187, de 9 de março de 2009*. Agência Nacional de Saúde Suplementar, Brasil, março.
- ANS, 2011e, *Resolução Normativa – RN Nº 250, de 25 de março de 2011*. Agência Nacional de Saúde Suplementar, Brasil, março.
- ANS, 2012g, *Resolução Normativa – RN Nº 295, de 9 de maio de 2012*. Agência Nacional de Saúde Suplementar, Brasil, maio.
- ARAÚJO, W., 2005, *A Gênese do pensamento galileano*. LIVRARIA DA FÍSICA. ISBN: 978-85-7861-004-3.
- ARISTOTLE, ACKRILL, J., 1963, *Categories and de interpretatione*. Clarendon Aristotle Series. NY, USA, Clarendon Press. ISBN: 0-19-872086-6.
- ARISTÓTELES, BINI, E., 2010, *Órganon*. Clássicos Edipro. 2 ed. Bauru, SP, Edipro. ISBN: 978-85-7283-643-2.
- ASPRAY, W., CAMPBELL-KELLY, M., ENSMENGER, N., et al., 2013, *Computer: a history of the information machine*. The sloan technology series. 3 ed. Boulder, CO, USA, Westview Press. ISBN: 978-0-8133-4590-1.
- ATICK, J. J., 1992, “Entropy Minimization: a Design Principle for Sensory Perception?” *International Journal of Neural Systems*, v. 03, n. supp01, pp. 81–90. doi: 10.1142/S0129065792000413.

- AYERS, M., 2000, *Locke – idéias e coisas*. Coleção grandes filósofos. São Paulo, SP, Editora Unesp. ISBN: 85-7139-272-2.
- AZGALDOV, G. G., KOSTIN, A. V., OMISTE, A. E. P., 2015, *The ABC of qualimetry: the toolkit for measuring immeasurable*. -, Ridero. ISBN: 978-5-4474-2248-6.
- BAEZA-YATES, R., NAVARRO, G., 1998, “Fast Approximate String Matching in a Dictionary”. In: *In Proc. SPIRE’98*, pp. 14–22. IEEE Computer Press.
- BAIS, F. A., FARMER, J. D., 2007, “The physics of information”, *arXiv*. Disponível em: <<http://arxiv.org/abs/0708.2837>>.
- BAROFSKY, I., 2012, “Can quality or quality-of-life be defined?” *Quality of Life Research*, v. 21, n. 4, pp. 625—631. ISSN: 1573-2649. doi: 10.1007/s11136-011-9961-0. Disponível em: <<http://dx.doi.org/10.1007/s11136-011-9961-0>>.
- BARROW, J. D., 1998, *Impossibility: the limits of science and the science of limits*. Oxford, UK, Oxford University Press. ISBN: 0-19-513082-0.
- BARTHEM, R. B., 1995, *Tratamento e análise de dados em física experimental*. Cadernos didáticos. 3 ed. Ilha do Fundão, RJ, Brasil, Cadernos Didáticos UFRJ.
- BASTOS, R., DA COSTA, E. A., 2012, “Colocando ordem nos complexos”, *Colloquium Exactarum*, v. 4 (Jan-Jun), pp. 33–38. doi: 10.5747/ce.2012.v04.n1.e043.
- BATES, H. W., 1864a, “XXXII. Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera: Heliconidæ.” *Transactions of the Linnean Society of London*, v. 23, n. 1, pp. 495–566. Disponível em: <<http://www.biodiversitylibrary.org/item/86346>>.
- BATES, H. W., 1905, “The naturalist on the Amazons – a record of the adventures, habits of animals, sketches of brazilian and indian life, and aspects of nature under the equator, during eleven years of travel”. cap. Chapter VIII, pp. 198–201, London, UK, George Routledge & Sons. Disponível em: <<https://archive.org/details/naturalistonamaz00bate>>. From Remington Kellogg Library of Marine Mammalogy – Smithsonian Institution.
- BATES, H. W., 1864b. “XXXII. Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera: Heliconidæ.” b. Disponível em: <<http://www.biodiversitylibrary.org/item/86346>>.

[//biodiversitylibrary.org/page/26846077](http://biodiversitylibrary.org/page/26846077)>. Plate LV – NOT IN COPYRIGHT.

BATES, H. W., 1864c. “XXXII. Contributions to an insect fauna of the Amazon Valley : Lepidoptera : Heliconidae”. c. Disponível em: <<http://biodiversitylibrary.org/page/26846078>>. Plate LVI – NOT IN COPYRIGHT.

BATINI, C., CAPPIELLO, C., FRANCALANCI, C., et al., 2009, “Methodologies for Data Quality Assessment and Improvement”, *ACM Comput. Surv.*, v. 41, n. 3 (jul.), pp. 16:1–16:52. ISSN: 0360-0300. doi: 10.1145/1541880.1541883. Disponível em: <<http://doi.acm.org/10.1145/1541880.1541883>>.

BAUDRILLARD, J., 1991, *Simulacros e simulações*. Antropos. Lisboa, Portugal, RELOGIO D’AGUA. ISBN: 972708141X. Tradutora: Maria João da Costa Pereira.

BEKENSTEIN, J. D., 2003, “Informação no Universo Holográfico”, *Scientific American Brasil*, , n. 16 (Setembro), pp. 42–50.

BEN-NAIM, A., 2007, *Entropy demystified: the second law reduced to plain common sense: with seven simulated games: $S=k\log W$; Expanded ed.* New Jersey, NJ, World Scientific. ISBN: 981-283-225-4.

BERGADHL, M., EHLING, M., ELVERS, E., et al., 2007, *Handbook on Data Quality Assessment Methods and Tools*. European Commission EUROSTAT, Weinsbaden, Germany. Disponível em: <<http://www.paris21.org/sites/default/files/dataqualityassessment-eurostat.pdf>>.

BERTSEKAS, D. P., 2014, *Constrained optimization and lagrange multiplier methods*. Belmont, MA, USA, Athena Scientific. ISBN: 1-886529-04-3.

BERTSEKAS, D. P., TSITSIKLIS, J. N., 2000. “Introduction to Probability”. Lecture Notes Course 6.041-6.431, M.I.T. FALL.

BERTSEKAS, D., RHEINBOLDT, W., 2014, *Constrained optimization and lagrange multiplier methods*. Computer science and applied mathematics. Elsevier Science. ISBN: 97-814-83-260-471.

BIDLACK, C. R., WELLMAN, M. P., 2010, “Exceptional Data Quality using Intelligent Matching and Retrieval”, *AI MAGAZINE*, v. 31, n. 1, pp. 65–74. ISSN: 0738-4602. doi: <http://dx.doi.org/10.21918/aimag.v31i1.2280>.

Disponível em: <<http://www.aaai.org/ojs/index.php/aimagazine/article/download/2280/2136>>.

- BILENKO, M., MOONEY, R. J., 2002, *Learning to Combine Trained Distance Metrics for Duplicate Detection in Databases*. Relatório técnico, Artificial Intelligence Lab - University of Texas at Austin.
- BINI, E., 2010, “Aristóteles: Sua Obra”. In: Vieira, J. L., Vieira, M. N. (Eds.), *Órganon*, Clássicos Edipro, 2 ed., Edipro, p. 17:33, Bauru, SP. ISBN: 978-85-7283-643-2.
- BISHOP, C. M., 2006, *Pattern recognition and machine learning*. Information science and statistics. New York, NY, USA, Springer. ISBN: 978-0387-31073-2.
- BLACK, P. E., 2008. “Levenshtein distance”. in *Dictionary of Algorithms and Data Structures* [online], Paul E. Black, ed., U.S. National Institute of Standards and Technology, 14 August. Disponível em: <<http://xlinux.nist.gov/dads//HTML/Levenshtein.html>>. (accessed 4 June 2013).
- BLACK, P. E., 2010. “‘soundex’, in Dictionary of Algorithms and Data Structures [online], Vreda Pieterse and Paul E. Black, eds.” December. (accessed 4 January 2015) Available from: <http://www.nist.gov/dads/HTML/soundex.html>.
- BORCHARDT, G., 1994, *Thinking between the lines: computers and the comprehension of causal descriptions*. Artificial intelligence. Cambridge, MA, MIT Press. ISBN: 9780262023740.
- BORGES, J. L., 1998, “The Library of Babel”. In: *Jorge Luis Borges – Collected Fictions*, Penguin, pp. 112–118, New York, NY. ISBN: 0-670-84970-7. Translated by Andrew Hurley.
- BORROR, C. M., 2009, “Quality philosophies and foundations”. In: *The certified quality engineer handbook*, 3 ed., cap. 1, ASQ Quality Press. ISBN: 978-0-87389-745-7.
- BOUTROUX, É., 2000, *Aristóteles*. Biblioteca de filosofia. 4 ed. Rio de Janeiro, RJ, RECORD. ISBN: 85-01-05749-5.
- BOYD, S., VANDENBERGHE, L., 2004, *Convex optimization*. New York, NY, USA, Cambridge University Press. ISBN: 978-0-521-83378-3.

- BRAGA, A. D. P., DE LEON F. DE CARVALHO, A. P., LUDERMIR, T. B., 1998, *Fundamentos de redes neurais artificiais*. Rio de Janeiro, RJ, Brasil, DCC/IM, COPPE/Sistemas, NCE/UFRJ. Trabalho apresentado na 11^a Escola de Computação.
- BRANDÃO, J. S., 2011, *Mitologia grega, vol. 1*. 23 ed. Petrópolis, RJ, Brasil, Vozes. ISBN: 978-85-326-0407-1.
- BREDEWEG, B., STRUSS, P., 2004, “Current Topics in Qualitative Reasoning.” *AI Magazine*, v. 24, n. 4, pp. 13–16.
- BURNS, E. M., 1964, *História da civilização ocidental – do homem das cavernas até a bomba atômica*, v. 1. 2 ed. Pôrto Alegre, Rio Grande do Sul dos Estados Unidos do Brasil, Editôra Globo S.A. Tradução de Lourival Gomes Machado, Lourdes Santos Machado e Leonel Vallandro. Revista e atualizada com a quarta edição americana.
- BUSSAB, W. O., MORETTIN, P. A., 1997, *Estatística básica*. 4^a ed. São Paulo, SP, ATUAL. ISBN: 85-7056-716-3, 9-788570-567161.
- CALAZANS, N. L. V., 1998, *Projeto lógico automatizado de sistemas digitais seqüenciais*. Rio de Janeiro, RJ, DCC/IM, COPPE/Sistemas, NCE/UFRJ.
- CAMARENA-IBARROLA, A., CHAVEZ, E., 2006, “On Musical Performances Identification, Entropy and String Matching (Best Paper Award, Second Place)”, *LECTURE NOTES IN COMPUTER SCIENCE*, v. 1, n. 4293, pp. 952–962. ISSN: 0302-9743.
- CAMPBELL, J., 2004, *The hero with a thousand faces*. Bollingen series XVII. Princeton, NJ, USA, Princeton University Press. ISBN: 0-691-11924-4.
- CAMPBELL, J., 1990, *O poder do mito – com Bill Moyers*. 25 ed. São Paulo, SP, Palas Athena. ISBN: 85-7242-052-5. Tradução de Carlos Felipe Moisés.
- CAPPIELLO, C., FRANCALANCI, C., PERNICI, B., 2004, “Data Quality Assessment from the User’s Perspective”. In: *Proceedings of the 2004 International Workshop on Information Quality in Information Systems, IQIS ’04*, pp. 68–73, New York, NY, USA. ACM. ISBN: 1-58113-902-0. doi: 10.1145/1012453.1012465. Disponível em: <<http://doi.acm.org/10.1145/1012453.1012465>>.

- CARPINETTI, L. C. R., MIGUEL, P. A. C., GEROLAMO, M. C., 2011, *Gestão da qualidade ISO 9001:2008: princípios e requisitos*. 4 ed. São Paulo, SP, Atlas. ISBN: 978-85-224-6504-0.
- CARROLL, L., 2011, *Alice's adventures in wonderland*. Mineola, NY, USA, Calla Editions. ISBN: 1-60660-014-1.
- CASTIÑEIRAS, J., CRISPINO, L. C. B., MATSAS, G. E. A., et al., 2006, “Singularidade e Informação.” *Scientific American Brasil*, pp. 70–75.
- CHAHIN, A., CUNHA, M. A., KNIGHT, P. T., et al., 2004, *e-gov.br: a próxima revolução brasileira : eficiência, qualidade e democracia : o governo eletrônico no Brasil e no mundo*. Prentice Hall. ISBN: 85-87918-93-1.
- CHAPRA, S. C., CANALE, R. P., 1990, *Numerical methods for engineers*. Applied Mathematics. 2 ed. Singapore, McGraw-Hill International Editions. ISBN: 0-07-100412-2.
- CHRISTEN, P., 2008, “Automatic Record Linkage Using Seeded Nearest Neighbour and Support Vector Machine Classification”. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, pp. 151–159, New York, NY, USA. ACM. ISBN: 978-1-60558-193-4. doi: 10.1145/1401890.1401913. Disponível em: <<http://doi.acm.org/10.1145/1401890.1401913>>.
- Coeli, C. M., de Camargo Jr, K. R., Pinheiro, R. S. (Eds.), 2006, *Relacionamento de bases de dados em saúde*, v. XIV, *Cadernos de saúde coletiva*. NESC UFRJ.
- COHEN, W. W., RAVIKUMAR, P., FIENBERG, S. E., 2003, “A Comparison of String Metrics for Matching Names and Records”. In: *KDD WORKSHOP ON DATA CLEANING AND OBJECT CONSOLIDATION*. American Association for Artificial Intelligence.
- COLETTI, G., DUBOIS, D., SCOZZAFAVA, R., 2013, *Mathematical models for handling partial knowledge in artificial intelligence*. Language of science. Boston, Massachusetts, USA, Springer US. ISBN: 978-1-4899-1424-8.
- CONTE, G. B., 1999, *Latin literature: a history*. A Johns Hopkins Paperback. Baltimore, Maryland, USA, Johns Hopkins University Press. ISBN: 0-8018-6253-1.
- COOPER, J. M., 1997, *PLATO – complete works*. USA, Hackett Pub. ISBN: 978-1-60384-670-7.

- COSTA, A. F. B., EPPRECHT, E. K., CARPINETTI, L. C. R., 2008, *Controle estatístico de qualidade*. 2 ed. São Paulo, Brasil, Atlas. ISBN: 978-85-224-4156-3.
- COSTA, E., SIMÕES, A., 2008, *Inteligência artificial: fundamentos e aplicações*. 2 ed. Lisboa, Portugal, FCA. ISBN: 978-972-722-340-4.
- COVER, T. M., THOMAS, J. A., 1991, *Elements of information theory*. New York, NY, USA, Wiley-Interscience. ISBN: 0-471-06259-6.
- CRESSWELL, J., 2010, *Oxford dictionary of word origins*. Paperback Reference. Oxford, UK, Oxford University Press. ISBN: 978-0-19-954793-7.
- DA CUNHA, A. G., 2010, *Dicionário etimológico da língua portuguesa*. 4 ed. Rio de Janeiro, RJ, Lexikon. ISBN: 978-85-86368-63-0.
- DA SILVA, E. L., MENEZES, E. M., 2005, *Metodologia de pesquisa e elaboração de dissertação*. 4 ed. Florianópolis, UFSC.
- DA SILVA, J. P. L., TRAVASSOS, C., DE VASCONCELLOS, M. M., et al., 2006, “Revisão sistemática sobre encadeamento ou linkage de bases de dados secundários para uso em pesquisa em saúde no Brasil”. pp. 197–224. v. XIV de , COELI *et al.* (2006), junho.
- DA SILVA, R., STASIU, R., ORENGO, V. M., et al., 2007, “Measuring quality of similarity functions in approximate data matching”, *Journal of Informetrics*, v. 1, n. 1, pp. 35 – 46. ISSN: 1751-1577. doi: <http://dx.doi.org/10.1016/j.joi.2006.09.001>.
- DADEROT, 2005. “Grave of Ludwig Boltzmann, physicist, on Zentralfriedhof (Central Cemetery), Vienna, Austria”. Disponível em: https://upload.wikimedia.org/wikipedia/commons/a/ad/Zentralfriedhof_Vienna_-_Boltzmann_B.jpg. By Own work by Daderot [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons.
- DAFFERTSHOFER, A., PLASTINO, A., 2005, “Landauer’s principle and the conservation of information”, *Physics Letters A*, v. 342, n. 3, pp. 213 – 216. ISSN: 0375-9601. doi: <http://dx.doi.org/10.1016/j.physleta.2005.05.058>. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0375960105007784>.

- DALZIELL, A. H., WELBERGEN, J. A., 2016, “Mimicry for all modalities”, *Ecology Letters*, v. 19, n. 6, pp. 609–619. ISSN: 1461-0248. doi: 10.1111/ele.12602. Disponível em: <<http://dx.doi.org/10.1111/ele.12602>>.
- DAMA, 2009, “The DAMA guide to the data management body of knowledge - DAMA-DMBOK”. cap. Data Quality Management, pp. 291–316, Brandy Beach, NJ, USA, Technics Publications, LLC. ISBN: 0977140083, 978-0-9771400-8-4.
- D’AMELIO, V., CHMARRA, M. K., TOMIYAMA, T., 2011, “Early design interference detection based on qualitative physics”, *Research in Engineering Design*, pp. 1–21. doi: 10.1007/s00163-011-0108-7.
- DARWIN, C., 1859, *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. London :John Murray Disponível em: <<http://www.biodiversitylibrary.org/item/122307>>.
- DARWIN, C., 1897, *The expression of the emotions in man and animals*. New York ;D. Appleton and Co.,. doi: <http://dx.doi.org/10.5962/bhl.title.4820>. Disponível em: <<http://www.biodiversitylibrary.org/item/24064>>.
- DE ALMEIDA, N. M., 2001, *Gramática latina – curso único e completo*. 29 ed. São Paulo, SP, Saraiva. ISBN: 978-85-02-00307-1.
- DELLINGER, J., 2012, “Relendo a perspectividade: algumas notas sobre “o perspectivismo de Nietzsche””, *Cadernos Nietzsche*, (00), pp. 127 – 155. ISSN: 2316-8242. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2316-82422012000200008&nrm=iso>.
- DEMARCO, T., 1999, *Management can make quality (im)possible*. Cutter IT Summit.
- DENNETT, D., 1988, “Quining Qualia”. In: Marcel, A., Bisiach, E. (Eds.), *Consciousness in Modern Science*, Oxford University Press.
- DIJKSTRA, E. W., 1968, “Letters to the Editor: Go to Statement Considered Harmful”, *Commun. ACM*, v. 11, n. 3 (mar.), pp. 147–148. ISSN: 0001-0782. doi: 10.1145/362929.362947. Disponível em: <<http://doi.acm.org/10.1145/362929.362947>>.

- DIJKSTRA, E. W., 1972, “The Humble Programmer”, *Commun. ACM*, v. 15, n. 10 (out.), pp. 859–866. ISSN: 0001-0782. doi: 10.1145/355604.361591. Disponível em: <<http://doi.acm.org/10.1145/355604.361591>>.
- DILLON, J. M., 2005, *The heirs of Plato: a study of the old academy (347-274 BC)*. Oxford scholarship. Oxford, UK, Clarendon Press – Oxford. ISBN: 0-19-823766-9.
- DISSANAYAKE, Y. B., PANKOV, A., SHESTAKOV, V., 2004, “Quality control on the basis of entropy determination”, *Transport*, v. 19, n. 2, pp. 51–55. doi: 10.1080/16484142.2004.9637953. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/16484142.2004.9637953>>.
- DO PRADO, F. R. O., DE CARVALHO, J. P. P., 2001, “Números Negativos têm Logaritmos?” *Disciplinarum Scientia. Série: Ciências Exatas*, v. 1, pp. p.99–105. ISSN: 2176-462x. Disponível em: <http://sites.unifra.br/Portals/36/tecnologicas/2001/numeros_negativos.pdf>.
- DOS SANTOS, S. A., 2008. “O Sistema de Informações de Beneficiário da Agência Nacional de Saúde Suplementar”. Dezembro. Disponível em: <<http://www.arca.fiocruz.br/handle/icict/2857>>. Projeto (Requisito parcial para obtenção do título de Especialista em Informação Científica e Tecnológica em Saúde), ICICT/FIOCRUZ (Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz), Rio de Janeiro, RJ, Brasil.
- DOWNING, L., 1998, “The Status of Mechanism in Locke’s Essay”, *Philosophical Review*, v. 107, n. 3, pp. 381–414.
- DRAGULESCU, A., YAKOVENKO, V., 2000, “Statistical mechanics of money”, *The European Physical Journal B - Condensed Matter and Complex Systems*, v. 17, n. 4, pp. 723–729. ISSN: 1434-6028. doi: 10.1007/s100510070114. Disponível em: <<http://dx.doi.org/10.1007/s100510070114>>.
- DRUCKER, P. F., 2008, *Inovação e espírito empreendedor (entrepreneurship): práticas e princípios*. São Paulo, Cengage. ISBN: 85-221-0085-3. Tradução: Carlos J. Malferrari.
- DUNN, H. L., 1946, “Record Linkage”, *American Journal of Public Health*, v. 36, n. 1 (jan.), pp. 1412–1416.

- EDMUNDS, M. G., PROJECT, A. R., FREETH, A., et al., 2006, “The Antikythera Mechanism - Real Progress Through Greek/UK/US Research”. In: Solomos, N. H. (Ed.), *Proceedings of the 7th International of the Hellenic Astronomical Society*, v. 848, *Recent Advances in Astronomy and Astrophysics*, pp. 913–918, October. doi: <http://dx.doi.org/10.1063/1.2348078>.
- EDMUNDS, M., FREETH, T., 2011, “Using Computation to Decode the First Known Computer”, *Computer*, v. 44, n. undefined, pp. 32–39. ISSN: 0018-9162. doi: doi.ieeecomputersociety.org/10.1109/MC.2011.134.
- EGOV.BR., 2005a. “Breve Histórico da Experiência Brasileira em Aplicações de E-GOV”. Novembroa. Disponível em: http://www.governoeletronico.com.br/index.php?option=com_content&task=view&id=20&Itemid=29>. Acesso em: 10/12/2012.
- EGOV.BR., 2005b. “Breve Histórico da Experiência Brasileira em Aplicações de E-GOV”. Novembrob. Disponível em: http://www.governoeletronico.com.br/index.php?option=com_content&task=view&id=20&Itemid=29>. Acesso em: 10/12/2012.
- EKLUND, M., 2011, “Fictionalism”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, fall 2011 ed.
- ELMAGARMID, A. K., IPEIROTIS, P. G., VERYKIOS, V. S., 2007, “Duplicate Record Detection: A Survey”, *IEEE Trans. Knowl. Data Eng.*, v. 19, n. 1, pp. 1–16.
- EMMECHE, C., 1996, *The garden in the machine – the emerging science of artificial life*. Princeton Science Library. Princeton, NJ, USA, Princeton University Press. ISBN: 0-691-02903-2.
- ENGLAND, J. L., 2013, “Statistical physics of self-replication”, *The Journal of Chemical Physics*, v. 139, n. 12, pp. 121923. doi: 10.1063/1.4818538. Disponível em: <http://dx.doi.org/10.1063/1.4818538>>.
- ERDÉLYI, A., 2010, *Asymptotic expansions*. Dover books on mathematics. Mineola, NY, USA, Dover Publications. ISBN: 978-0-486-60318-6.
- ERLICHSON, H., 1999, “Sadi Carnot, ‘Founder of the Second Law of Thermodynamics’”, *European Journal of Physics*, v. 20, n. 3, pp. 183. Disponível em: <http://stacks.iop.org/0143-0807/20/i=3/a=308>>.

- EVANGELISTA, L. R., 2014, *Perspectivas em história da física: da física dos gases à mecânica estatística*, v. 2. Livraria da Física. ISBN: 978-85-7861-291-7.
- FAGIN, R., HALPERN, J. Y., MOSES, Y., et al., 1995, *Reasoning about knowledge*. Cambridge, Massachusetts, USA, MIT Press. ISBN: 0-262-06162-7.
- FAN, W., 2008, “Dependencies Revisited for Improving Data Quality”. In: *Proceedings of the Twenty-seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '08, pp. 159–170, New York, NY, USA. ACM. ISBN: 978-1-60558-152-1. doi: 10.1145/1376916.1376940. Disponível em: <<http://doi.acm.org/10.1145/1376916.1376940>>.
- FELDMAN, S., 2004, “A Conversation with Alan Kay”, *Queue*, v. 2, n. 9 (dez.), pp. 20–30. ISSN: 1542-7730. doi: 10.1145/1039511.1039523. Disponível em: <<http://doi.acm.org/10.1145/1039511.1039523>>.
- FELLEGI, I. P., SUNTER, A. B., 1969, “A Theory for Record Linkage”, *Journal of the American Statistical Association*, v. 64, n. 328 (Dec.), pp. 1183–1210.
- FERREIRA, J. P. M., 2005, “Como interpretar a entropia?” In: *Boletim da Sociedade Portuguesa de Química*, n. 96, II, pp. 38–42. Sociedade Portuguesa de Química, Março. Disponível em: <<http://www.spq.pt/magazines/BSPQ/621/article/30001229/swf>>.
- FEYNMAN, R. P., LEIGHTON, R. B., SANDS, M. L., 2005, *The Feynman lectures on physics*, v. v. 1, *The Feynman Lectures on Physics*. Massachusetts, Addison-Wesley. ISBN: 0-8053-9046-4.
- FEYNMAN, R. P., 2012, “Simetria nas leis da física”. In: *Feynman sobre as leis da física*, 1 ed., Contraponto – PUC-Rio, Rio de Janeiro, RJ. ISBN: 978-85-7866-047-5.
- FILHO, C. F., 2007, *História da computação: O caminho do pensamento e da tecnologia*. Porto Alegre, RS – BRASIL, EDIPUCRS. ISBN: 978-85-7430-691-9.
- FLORIDI, L., 2005, “Is Semantic Information Meaningful Data?” *Philosophy and Phenomenological Research*, v. 70, n. 2, pp. 351–370. ISSN: 1933-1592. doi: 10.1111/j.1933-1592.2005.tb00531.x. Disponível em: <<http://dx.doi.org/10.1111/j.1933-1592.2005.tb00531.x>>.
- FLORIDI, L., 2010, *Information: a very short introduction*. Oxford, UK, Oxford University Press. ISBN: 978-0-19-955137-8.

- FORBES, A. B., 2006, “Measurement uncertainty and optimized conformance assessment”, *Measurement*, v. 39, n. 9, pp. 808 – 814. ISSN: 0263-2241. doi: <http://dx.doi.org/10.1016/j.measurement.2006.04.007>. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0263224106000777>>. Advanced Mathematical Tools for Measurement in Metrology and Testing.
- FORBUS, K. D., 1984, “Qualitative process theory”, *Artificial Intelligence*, v. 24, n. 1, pp. 85 – 168. ISSN: 0004-3702. doi: [http://dx.doi.org/10.1016/0004-3702\(84\)90038-9](http://dx.doi.org/10.1016/0004-3702(84)90038-9).
- FORBUS, K. D., 1988, “Qualitative Physics: Past, Present, and Future”. In: *Exploring Artificial Intelligence*, Morgan-Kaufmann Publishers, Inc., cap. 7, pp. 239–296, San Francisco, California.
- FRANCHINI, A. S., SEGANFREDO, C., 2010, *As melhores histórias da mitologia hindu*. Porto Alegre, RS, Artes e Ofícios. ISBN: 978-85-7421-179-4.
- FREIRE, S. M., GONÇALVES, R. D. C. B., BANDARRA, A. C., et al., 2009, “Análise da efetividade de comparadores de strings para discriminar pares de verdadeiros de pares falsos no relacionamento de registros”. In: *Anais: IX Workshop de Informática Médica*, pp. 2119 – 2128, Bento Gonçalves, RS, Brasil, Julho. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/wim/2009/027.pdf>>.
- GAELZEL, R., 2006, *Física matemática*. Instituto de Física - UFRGS. Disponível em: <https://professor.ufrgs.br/rgaelzer/files/FisMat_complete-aluno.pdf>. Access: 21/02/2017.
- GALINARI, M. M., 2015, “Górgias e o antigo problema da referência: uma interface entre a sofística e a linguística da enunciação”. In: Dias, L. F., Lacerda, P. B. G., Dalmaschio, L. (Eds.), *Enunciação e materialidade linguística*, pp. 92–100, Belo Horizonte, MG, FALE/UFMG. ISBN: 978-85-7758-261-7.
- GARDNER, H., 1995, *A nova ciência da mente - uma história da revolução cognitiva – Vol. 09*. 3 ed. São Paulo, SP, EdUSP. ISBN: 85-314-0145-3.
- GARVIN, D. A., 1988, *Managing quality: the strategic and competitive edge*. Harvard business school. New York, N.Y., US, Free Press. ISBN: 0-02-911380-6.

- GELMAN, A., NOLAN, D., 2002, “You Can Load a Die, But You Can’t Bias a Coin”, *The American Statistician*, v. 56, n. 4, pp. 308–311. doi: 10.1198/000313002605. Disponível em: <<http://dx.doi.org/10.1198/000313002605>>.
- GIBBS, J. W., 1914, *Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics*. Intermediate and advanced mathematics. 180 Varick Street, New York, NY, USA, Dover Publications.
- GOLDMAN, A., 1976, “Discrimination and Perceptual Knowledge”, *Journal of Philosophy*, v. 73, pp. 771–791.
- GOTKIN, K., 2014, “When Computers Were Amateur”, *IEEE Annals of the History of Computing*, v. 36, n. 2 (Apr), pp. 4–14. ISSN: 1058-6180. doi: 10.1109/MAHC.2014.32.
- GRECO, R., 2016, *Curso de direito penal, V.1 - parte geral*. 18 ed. Niterói, RJ, IMPETUS. ISBN: 978-85-7626865-9.
- GRIER, D. A., 2007, *When computers were human*. Princeton, NJ, USA, Princeton University Press. ISBN: 978-0-691-13382-9.
- GRÜNWALD, P., VITÁNYI, P. M. B., 2004, “Shannon Information and Kolmogorov Complexity”, *CoRR*, v. cs.IT/0410002. Disponível em: <<http://arxiv.org/abs/cs.IT/0410002>>.
- GUIL, F., MARÍN, R., 2013, “A Theory of Evidence-based method for assessing frequent patterns”, *Expert Systems with Applications*, v. 40, n. 8, pp. 3121 – 3127. ISSN: 0957-4174.
- GUIMARÃES, I. C. T., 2015, “O conceito de substantivo em uma perspectiva enunciativa”. In: Dias, L. F., Lacerda, P. B. G., Dalmaschio, L. (Eds.), *Enunciação e materialidade linguística*, pp. 19–38, Belo Horizonte, MG, FALE/UFMG. ISBN: 978-85-7758-261-7.
- GURPILHARES, M. S. S., 2004, “As bases filosóficas da gramática normativa: uma abordagem histórica”, *Janus*, v. 1, n. 1, pp. 43–51. ISSN: 1984-7998.
- HAIR, J., ANDERSON, R., TATHAM, R., 2005, *Análise multivariada de dados*. Porto Alegre, RS, Bookman. ISBN: 978-85-7780402-3.
- HAMMING, R. W., 1950, “Error Detecting and Error Correcting Codes”, *Bell System Technical Journal*, v. 29, n. 2, pp. 147–160.

- HAN, J., KAMBER, M., 2006, *Data mining: concepts and techniques*. 2 ed. San Francisco, CA, USA, Morgan Kaufmann Publishers. ISBN: 978-1-55860-901-3.
- HANSEN, H., 2015, “Fallacies”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, summer 2015 ed.
- HARTLEY, R. V. L., 1928, “Transmission of Information”, *Bell System Technical Journal*, v. 7, pp. 535–563.
- HASSANI, S., 2000, “Thermodynamics Using Coins”, *European Journal of Physics*, v. 21, pp. 41–47.
- HAYKIN, S., 2009, “Neural networks and learning machines”. 3 ed., Prentice Hall. ISBN: 978-0-13-147139-9.
- HEINRICH, B., KAISER, M., KLIER, M., 2007, “How to measure data quality? – a metric based approach”. Montreal. 28th International Conference on Information Systems (ICIS).
- HEINRICH, B., KLIER, M., KAISER, M., 2009, “A Procedure to Develop Metrics for Currency and Its Application in CRM”, *J. Data and Information Quality*, v. 1, n. 1 (jun.), pp. 5:1–5:28. ISSN: 1936-1955. doi: 10.1145/1515693.1515697. Disponível em: <<http://doi.acm.org/10.1145/1515693.1515697>>.
- HELPERICH, C., 2006a, *Historia Da filosofia*. 1 ed. São Paulo, SP, Martins Fontes. ISBN: 85-336-2215-5.
- HELPERICH, C., 2006b, “A Filosofia da Idade Média Cristã”. In: *Historia Da Filosofia*, Martins Fontes, pp. 70–115, b. ISBN: 85-336-2215-5.
- HELPERICH, C., 2006c, “Humanismo, Reforma e a Revolução do saber sobre a Natureza”. In: *Historia Da Filosofia*, Martins Fontes, pp. 116–154, c. ISBN: 85-336-2215-5.
- HERZOG, T., SCHEUREN, F., WINKLER, W., 2007, *Data quality and record linkage techniques*. New York, NY, USA, Springer. ISBN: 978-0-387-69505-1.
- HETTIARACHCHI, G. P., HETTIARACHCHI, N. N., HETTIARACHCHI, D. S., et al., 2014, “Next generation data classification and linkage: Role of probabilistic models and artificial intelligence”. In: *Global Humanitarian Technology Conference (GHTC), 2014 IEEE*, pp. 569–576, Oct. doi: 10.1109/GHTC.2014.6970340.

- HOLMES, F. L., 1987, *Lavoisier and the chemistry of life: an exploration of scientific creativity*. History of science and medicine series. Madison, WI, USA, University of Wisconsin Press. ISBN: 978-0-299-09984-8.
- HOLMES, N., 2004, “Data and information as property”, *Computer*, v. 37, n. 5 (May), pp. 92, 90–91. ISSN: 0018-9162. doi: 10.1109/MC.2004.1297256.
- HOPCROFT, J. E., ULLMAN, J. D., 1979, *Introduction to automata theory, languages and computation*. Addison-Wesley series in computer science. USA, Addison-Wesley. ISBN: 0-201-02988-X.
- HOUAISS, 2001, *Dicionário Houaiss da língua portuguesa*. 1 ed. Rio de Janeiro, RJ, Objetiva. ISBN: 85-7302-383-X.
- HUGHES, J., 2013a. “Lecture 1 - Introduction In: Classical Mythology by Joseph Hughes”. Web iTunes, may.a. Disponível em: <<https://itunes.apple.com/br/course/lecture-1-introduction/id655630388?i=1000160205296&mt=2>>. Acesso em: 12/09/2014. A study of Greek and Roman myths and legends as they appear in art, music and literature, especially epic and tragedy. This is the first in a series of lectures on Classical Mythology by Dr. Joseph Hughes.
- HUGHES, J., 2013b. “LLT 121 CLASSICAL MYTHOLOGY – LECTURE 1”. Web page, may.b. Disponível em: <<http://courses.missouristate.edu/josephhughes/myth/TranscriptsWord/Lecture01.doc>>. Acesso em: 25/09/2016. A study of Greek and Roman myths and legends as they appear in art, music and literature, especially epic and tragedy. A transcript from <https://itunes.apple.com/br/course/lecture-1-introduction/id655630388?i=1000160205296&mt=2>.
- HYMAN, A., 1982, *Charles Babbage: pioneer of the computer*. Princeton, NJ, USA, Princeton University Press. ISBN: 0-691-08303-7.
- IDA, M. S. A. M., 1914, “O PURISMO E O PROGRESSO DA LINGUA PORTUGUEZA”. In: *ANNAES DA BIBLIOTHECA NACIONAL NO RIO DE JANEIRO 1916*, v. 30, *ANNAES DA BIBLIOTHECA NACIONAL NO RIO DE JANEIROS*, pp. 65–79, Rio de Janeiro, RJ, outubro. Oficinas Graphics da Bibliotheca Nacional.
- IEC80000-13, 2008, *Quantities and units – Part 13: Information science and technology*. Relatório técnico, The International Organization for Standardization (ISO), 03. Disponível em: <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=31898>. In

IEC 80000-13:2008, names, symbols and definitions for quantities and units used in information science and technology are given. Where appropriate, conversion factors are also given. International Standard IEC 80000-13 has been prepared by IEC technical committee 25: Quantities and units, and their letter symbols in co-operation with ISO/TC 12. This standard cancels and replaces subclauses 3.8 and 3.9 of IEC 60027-2:2005. The only significant change is the addition of explicit definitions for some quantities.

IEEE, 1990, “IEEE Standard Glossary of Software Engineering Terminology”, *IEEE Std 610.12-1990*, (Dec), pp. 1–84. doi: 10.1109/IEEESTD.1990.101064.

INCOR, 2008. “Componentes de Fonetização”. WEB, janeiro. Disponível em: <www.incor.usp.br/spdweb/ccsis/fonetica>. Acesso em: 25/09/2016.

IRN2013, 2013. “IRN - Registo Civil - Composição do nome”. Fevereiro. Disponível em: <http://www.irn.mj.pt/sections/irn/a_registral/registo-civil/docs-do-civil/dar-o-nome/>. Acesso em: 01/03/2013. O Instituto dos Registos e do Notariado, I. P. (IRN, IP), é um instituto público integrado na administração indireta do Estado (Português), dotado de autonomia administrativa, que tem por missão executar e acompanhar as políticas relativas aos serviços de registo, tendo em vista assegurar a prestação de serviços aos cidadãos e às empresas no âmbito da identificação civil e do registo civil, de nacionalidade, predial, comercial, de bens móveis e de pessoas coletivas, bem como assegurar a regulação, controlo e fiscalização da atividade notarial.

ISAAC, D., LYNNE, C., 2003, *Automated Data Quality Assessment in the Intelligent Archive*. White paper prepared for the intelligent data understanding program, Goddard Earth Sciences Data and Information Services Center, January. Disponível em: <http://disc.sci.gsfc.nasa.gov/intelligent_archive/presentations/AutoQualityAssessment.pdf>. Acesso: 21/01/2013.

ISO, 2015. “International Organization for Standardization (ISO)”. Disponível em: <<http://www.iso.org/iso/home/standards.htm>>. Acesso em: 21/09/2015.

- ISO/IEC/IEEE, 2010, “Systems and software engineering – Vocabulary”, *ISO/IEC/IEEE 24765:2010(E)*, (Dec), pp. 1–418. doi: 10.1109/IEEESTD.2010.5733835.
- ISOKAWA, T., NISHIMURA, H., MATSUI, N., 2012, “Quaternionic Multilayer Perceptron with Local Analyticity”, *Information*, v. 3 (November), pp. 756–770. ISSN: 2078-2489.
- JACKSON, A., 2002, “The world of blind mathematicians”, *Notices of the AMS*, v. 49, n. 10 (November), pp. 1246–1251. Disponível em: <<http://www.ams.org/notices/200210/comm-morin.pdf>>.
- JANSEN, H. J., 2008. “Statistical Mechanics”. Web, Oct. Disponível em: <<http://physics.oregonstate.edu/~jansenh/papers/StatMechNotes.pdf>>. Acesso em: 24/06/2013.
- JAPIASSÚ, H., MARCONDES, D., 2006, *Dicionário básico de filosofia*. 4 ed. Rio de Janeiro, RJ, Jorge Zahar Editor. ISBN: 85-7110-095-0.
- JAYNES, 1965, “Gibbs vs Boltzmann Entropies”, *American Journal of Physics*, v. 33, n. 5, pp. 391–398. doi: 10.1119/1.1971557. Disponível em: <<http://dx.doi.org/10.1119/1.1971557>>.
- JAYNES, E., 1991. “The Gibbs paradox”. .
- JAYNES, E. T., 1990, “Probability Theory as Logic”. In: Fougère, P. F. (Ed.), *Maximum-Entropy and Bayesian Methods*, Kluwer, pp. 1+, Dordrecht. Disponível em: <<http://bayes.wustl.edu/etj/articles/prob.as.logic.pdf>>. Presented at the Ninth Annual Workshop on Maximum Entropy and Bayesian Methods, Dartmouth College, New Hampshire, August 14, 1989. In the Proceedings Volume, Maximum Entropy and Bayesian Methods, Paul F. Fougere, Editor, Kluwer Academic Publishers, Dordrecht, Holland (1990). The present version was substantially revised, corrected, and extended 5/1/94.
- JAYNES, E. T., 1962, “Information Theory and Statistiscal Mechanics”. In: Ford, K. W. (Ed.), *Brandeis university summer institute in theoretical physics – lecture notes*, v. 3, pp. 182–217. W. A. BENJAMIN. Disponível em: <<http://bayes.wustl.edu/etj/articles/brandeis.pdf>>.
- JAYNES, E. T., 1957, “Information Theory and Statistical Mechanics”, *Physical Review*, v. 106 (May), pp. 620–630. doi: 10.1103/PhysRev.106.620. Disponível em: <<http://link.aps.org/doi/10.1103/PhysRev.106.620>>.

- JCGM, 2012, *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*, 3rd ed. Joint Committee for Guides in Metrology. Disponível em: <https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf>. 2008 version with minor corrections.
- JONES, J., JAKSCH, D., 2012, *Quantum information, computation and communication*. Cambridge, UK, Cambridge University Press. ISBN: 978-1-107-01446-6.
- JUGULUM, R., 2014, *Competing with high quality data: concepts, tools, and techniques for building a successful approach to data quality*. Wiley. ISBN: 978-1-118-41649-5.
- Juran, J. M., Godfrey, A. B. (Eds.), 1999, *Juran's quality handbook*. 5th ed. New York, NY, USA, McGraw Hill. ISBN: 0-07-034002-X.
- Kahneman, D., Slovic, P., Tversky, A. (Eds.), 1982, *Judgment under uncertainty: heuristics and biases*. Cambridge, UK, Cambridge University Press. ISBN: 0-521-28414-7.
- KANG, H. Y., KWAK, B. M., 2009, “Application of maximum entropy principle for reliability-based design optimization”, *Structural and Multidisciplinary Optimization*, v. 38, n. 4, pp. 331–346. ISSN: 1615-1488. doi: 10.1007/s00158-008-0299-3. Disponível em: <<http://dx.doi.org/10.1007/s00158-008-0299-3>>.
- KAPLAN, A., WASSERMAN, A., 2003, *A Torá Viva*. 2 ed. Brasil, Maayanot. ISBN: 85-85512-55-5.
- KELVIN, W. T. B., 1889, *Popular lectures and addresses*. N. vol. 1, Nature series. NY, USA, Macmillan & Company.
- KLEER, J. D., BROWN, J. S., 1984, “A qualitative physics based on confluences”, *Artificial Intelligence*, v. 24, n. 1, pp. 7 – 83. ISSN: 0004-3702. doi: [http://dx.doi.org/10.1016/0004-3702\(84\)90037-7](http://dx.doi.org/10.1016/0004-3702(84)90037-7). Disponível em: <<http://www.sciencedirect.com/science/article/pii/0004370284900377>>.
- KLEIN, D., 2003. “Lagrange multipliers without permanent scarring”. Disponível em: <<https://people.eecs.berkeley.edu/~klein/papers/lagrange-multipliers.pdf>>. “Permanently in rough draft form, it seems!”

- KLIMA, G., 2010, “Nominalist Semantics”. pp. 159–172. v. 1 de , PASNAU e VAN DYKE (2010). ISBN: 978-0-521-76216-8.
- Klima, G., Allhoff, F., Vaidya, A. J. (Eds.), 2007, *Medieval philosophy – essential readings with commentary*. Oxford, UK, Blackwell. ISBN: 978-1-4051-3564-1.
- KLIR, G. J., 2006, *Uncertainty and information: foundations of generalized information theory*. Hoboken, New Jersey, US, John Wiley & Sons. ISBN: 978-0-471-74867-0.
- KNUTH, D. E., 1996a, “Selected papers on computer science”. cap. 0 – Algorithms, Programs, and Computer Science, USA, Cambridge University Press, a. ISBN: 1-881526-91-7.
- KNUTH, D. E., 1996b, “Selected papers on computer science”. cap. 1 – Computer Science and its Relation to Mathematics, USA, Cambridge University Press, b. ISBN: 1-881526-91-7.
- KNUTH, D. E., 1998, *The art of computer programming – volume 3: sorting and searching*. 2 ed. USA, ADDISON-WESLEY. ISBN: 0-201-89685-0.
- KOTZ, J. C., PURCELL, K. F., 1991, *Chemistry and chemical reactivity*. 2 ed. Orlando, Florida, USA, Sauders College Publishing. ISBN: 0030530873.
- KRANTZ, S., 2004, *Real analysis and foundations*. Studies in advanced mathematics. 2nd ed. Boca Raton, Florida, USA. ISBN: 1-58488-483-5.
- KRAUSE, P., CLARK, D., 1993, *Representing uncertain knowledge: an artificial intelligence approach*. Oxford, UK, Intellect. ISBN: 978-94-011-2084-5.
- KRIPPENDORFF, K., 1986, *Information theory : structural models for qualitative data*. Quantitative applications in the social science. Beverly Hills, Sage Publications. ISBN: 0-8039-2051-2. Disponível em: <<http://opac.inria.fr/record=b1086057>>.
- KUHN, T. S., 2013, *A estrutura das revoluções científicas*. São Paulo, SP, Brasil, Perspectiva. ISBN: 978-85-273-0111-4. Tradução: Beatriz Vianna Boeira e Nelson Boeira.
- KUIPERS, B., 1986, “Qualitative simulation”, *Artificial intelligence*, v. 29, n. 3, pp. 289–338.

- KULLBACK, S., 1978, *Information theory and statistics*. A Wiley publication in mathematical statistics. Gloucester, MA, USA, Wiley. ISBN: 0-8446-5625-9.
- LALANDE, A., 1999, *Vocabulário técnico e crítico de filosofia*. 3 ed. São Paulo, SP, Martins Fontes. ISBN: 85-336-1096-3.
- LANDAUER, R., 1996, “The physical nature of information”, *Physics Letters A*, v. 217, n. 4–5, pp. 188–193. ISSN: 0375-9601. doi: [http://dx.doi.org/10.1016/0375-9601\(96\)00453-7](http://dx.doi.org/10.1016/0375-9601(96)00453-7).
- LAZO, A. V., RATHIE, P., 1978, “On the entropy of continuous probability distributions (Corresp.)”, *IEEE Transactions on Information Theory*, v. 24, n. 1 (Jan), pp. 120–122. ISSN: 0018-9448. doi: 10.1109/TIT.1978.1055832.
- Lee, J. A. N. (Ed.), 1995, *International biographical dictionary of computer pioneers*. 1 ed. London, England, Fitzroy Dearborn. ISBN: 1-884964-47-8.
- LEITHOLD, L., 1994, *O cálculo com geometria analítica*. 3 ed. Rua Joaquim Távora, 776, 04015-0001 São Paulo, SP, Brasil, HARBRA. Tradução de The Calculus with Analytic Geometry, 6th edition, 1990 HarperCollins Publishers.
- LEON-GARCIA, A., 2008, *Probability, statistics, and random processes for electrical engineering*. Upper Sadle River, NJ, USA, Pearson/Prentice Hall. ISBN: 978-0-13-147122-1.
- LEVENSHTEIN, V. I., 1966, “Binary Codes Capable of Correcting Deletions, Insertions and Reversals”, *Soviet Physics Doklady*, v. 10 (fev.), pp. 707.
- LEWIS, C. T., SHORT, C., 1879. “Dătus”. [online]. Disponível em: <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059:entry=datus>>. A Latin Dictionary. Founded on Andrews’ edition of Freund’s Latin dictionary. revised, enlarged, and in great part rewritten by. Charlton T. Lewis, Ph.D. and. Charles Short, LL.D. Oxford. Clarendon Press. 1879. The National Endowment for the Humanities provided support for entering this text. (Accessed 10 February 2015).
- LIGHT, J. S., 1999, “When Computers Were Women”, *Technology and Culture*, v. 40, n. 3, pp. 455–483.

- LIMA, E. L., 1983, “Números Negativos têm Logaritmos?” In: *Revista do Professor de Matemática*, 3, Sociedade Brasileira de Matemática. Disponível em: <<http://rpm.org.br/cdrpm/3/6.htm>>.
- LIMA, E. L., 1993, *Coordenadas no espaço*. Coleção do Professor de Matemática. Rio de Janeiro, RJ, Sociedade Brasileira de Matemática. ISBN: 0082-X.
- LIVIO, M., 2011, *Razão áurea: a história de FI, um número surpreendente*. 6 ed. Rio de Janeiro, RJ, Brazil, Record. ISBN: 978-85-01-06653-4.
- LOBANOV, A. S., 2013, “The basic concepts of qualimetry”, *Scientific and Technical Information Processing*, v. 40, n. 2 (Apr), pp. 72–82. ISSN: 1934-8118. doi: 10.3103/S0147688213020044.
- LOCKE, J., 1999, *Ensaio acerca do entendimento humano*. Os pensadores. São Paulo, SP, Nova Cultural. ISBN: 85-13-00906-7. Anoar Aiex (Tradutor).
- LOGAN, R. K., 2012, *Que é informação?: a propagação da organização na biosfera*. Rio de Janeiro, RJ, Contraponto, PUC-Rio. ISBN: 978-85-7866-046-8.
- MACHADO, J. P., SILVEIRA, D. P. D., SANTOS, I. S., et al., 2008, “Aplicação da metodologia de relacionamento probabilístico de base de dados para a identificação de óbitos em estudos epidemiológicos”, *Revista Brasileira de Epidemiologia*, v. 11 (03), pp. 43 – 54. ISSN: 1415-790X. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2008000100004&nrm=iso>.
- MACKAY, D. J. C., 2005, *Information theory, inference & learning algorithms*. New York, NY, USA, Cambridge University Press. Disponível em: <<http://www.inference.phy.cam.ac.uk/itprnn/book.pdf>>. Version 7.2.
- MACKENZIE, C. E., 1979, *Coded character sets: history and development*. Addison-Wesley Pub (Sd). ISBN: 0-201-14460-3.
- MADNICK, S. E., WANG, R. Y., LEE, Y. W., et al., 2009, “Overview and Framework for Data and Information Quality Research”, *J. Data and Information Quality*, v. 1, n. 1 (jun.), pp. 2:1–2:22. ISSN: 1936-1955. doi: 10.1145/1515693.1516680. Disponível em: <<http://doi.acm.org/10.1145/1515693.1516680>>.
- MAGALHÃES, M. D. A., 2010, *Desafios da gestão de uma base de dados de identificação unívoca de indivíduos: a experiência do Projeto Cartão Nacional de Saúde no SUS*. Tese de Mestrado, Escola Nacional de Saúde

Pública Sergio Arouca, Rio de Janeiro, RJ, Brasil, Janeiro. Disponível em: <<http://www.arca.fiocruz.br/xmlui/handle/icict/2336>>.

- MANFRINO, R., ORTEGA, J., DELGADO, R., 2010, *Inequalities: a mathematical olympiad approach*. Basel, Schweiz, Birkhäuser Basel. ISBN: 978-3-0346-0050-7.
- MANKIW, N. G., 2005, “Introdução à economia – edição compacta”. cap. Dez princípios de economia, São Paulo, SP, Pioneira Thomson Learning. ISBN: 85-221-0394-1, 9788522103942.
- MANNING, C., SCHÜTZE, H., 1999, *Foundations of statistical natural language processing*. Cambridge, MA, MIT Press. ISBN: 9780262133609.
- MARICONDA, P. R., 2006, “Galileu e a ciência moderna”, *Cadernos de Ciências Humanas - Especiaria*, v. 9, n. 16 (jul./dez), pp. 267–292.
- MARTINS, J. P., 1994, *Introdução à programação usando o pascal*. Lisboa, Portugal, Mc Graw-Hill Portugal. ISBN: 972-9241-59-7.
- MARZAL, A., VIDAL, E., 1993, “Computation of Normalized Edit Distance and Applications”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, v. 15, n. 9, pp. 926–932.
- MASIN, S. C., ZUDINI, V., ANTONELLI, M., 2009, “Early alternative derivations of Fechner’s law”, *Journal of the History of the Behavioral Sciences*, v. 45, n. 1, pp. 56–65. ISSN: 1520-6696. doi: 10.1002/jhbs.20349. Disponível em: <<http://dx.doi.org/10.1002/jhbs.20349>>.
- MASON, H. E., MASON, W., 2017. “A la Memoire de J.M. Jacquard”. Disponível em: <<http://collection.sciencemuseum.org.uk/objects/co44695>>. Science Museum Group. A la memoire de J.M. Jacquard. Science Museum Group Collection Online. Accessed July 25, 2017. <http://collection.sciencemuseum.org.uk/objects/co44695>.
- MAUGIN, G. A., 1999, “Introduction: A Post-Duhemian Thermodynamics”. In: *The thermomechanics of nonlinear irreversible behaviors: an introduction*, cap. 1, World Scientific. ISBN: 981-02-3375-4.
- MCCARTHY, J., 1968, “Programs with Common Sense”. In: *Semantic Information Processing*, pp. 403–418. MIT Press.
- MCCARTHY, J., 1979, “Ascribing Mental Qualities to Machines”. In: Ringle, M. (Ed.), *Philosophical Perspectives in Artificial Intelligence*, Humanities

Press, pp. 161–195. Ascribing mental qualities like beliefs, intentions and wants to a machine is sometimes correct if done conservatively and is sometimes necessary to express what is known about its state. We propose some new definitional tools for this: definitions relative to an approximate theory and second order structural definitions.

MCGIBBON, S., 2005, “Standards and Interoperability”, *Baltic IT&I Review*, , n. 1, pp. 16–18. Disponível em: <[http://www.ebaltics.com/doc_up1/McGibbon\(4\).pdf](http://www.ebaltics.com/doc_up1/McGibbon(4).pdf)>.

MEDAWAR, P. B., 2008, *Os limites da ciência*. São Paulo, SP, UNESP. ISBN: 978-85-7139-852-8.

MICHEL, H. E., AWWA, A. A. S., 2010, “Artificial neural networks using complex numbers and phase encoded weights”, *APPLIED OPTICS*, v. 49, n. 10 (April), pp. B71–B82.

MIRZAEI, A., ZABOLI, H., SAFABAKHSH, R., 2007, “A Neural Network String Matcher”. In: Kropatsch, W. G., Kampel, M., Hanbury, A. (Eds.), *Computer Analysis of Images and Patterns: 12th International Conference, CAIP 2007, Vienna, Austria, August 27-29, 2007. Proceedings*, pp. 784–791, Berlin, Heidelberg, Springer Berlin Heidelberg. ISBN: 978-3-540-74272-2.

MITRA, A., 2016, *Fundamentals of quality control and improvement*. Wiley. ISBN: 978-1-118-70514-8.

MÜLLER, I., WEISS, W., 2005, *Entropy and energy: a universal competition*. Interaction of Mechanics and Mathematics. Springer Berlin Heidelberg. ISBN: 978-3-540-24281-9.

MONTGOMERY, D. C., 2009, *Introdução ao controle estatístico da qualidade*. 4 ed. Rio de Janeiro, RJ, LTC. ISBN: 978-85-216-1400-5.

MONTGOMERY, D. C., 2012, *Introduction to statistical quality control*. 7 ed. USA, John Wiley & Sons. ISBN: 978-1-118-14681-1.

MORRIS, R., 2001, *O que sabemos sobre o universo?* Rio de Janeiro, RJ, Jorge Zahar. ISBN: 85-7110-617-7.

MUSEUM, B., 2017a. “Book of the Dead’, Papyrus of Ani (sheet 3): Ani’s Judgment: the scene is the Hall of Judgment. Centrally placed is a balance, holding in its two pans Ani’s heart (on the left) and a feather (on the right) representing Maat, the divine personification of truth and order.

The crossbar of the balance hangs from a feather-shaped peg attached to the upright support, on the top of which squats a small baboon. This creature is a form of the god Thoth, who acts in a different form and with a different duty elsewhere in this "trial". The god Anubis, here shown as a jackal-headed, human-bodied, kneeling deity, described as "he who is in the place of embalming," holds the cord of the right-hand pan, and steadies the plumb bob of the balance. To the right of the balance stands Thoth, here in human form with ibis head; he is the scribe of the gods, and he holds a scribe's palette and a reed brush, ready to note down the results of Ani's interrogation. On a mat behind Thoth sits a monster ready to spring forward to consume Ani's heart if he fails to pass the test. This creature has the head of a crocodile, the forepart of a lion, and the hindquarters of a hippopotamus. At the top of the scene the great gods of Egypt are shown, formally seated on thrones, waiting to deliver judgment: Ra-Horakhty, Atum, Shu, Tefnut, Geb, Nut, Isis and Nephthys, Horus and Hathor, joined by gods personifying the divine word (Hu) and perception (Sia). Other deities observe the proceedings: to the left of the balance, Shay (fate) and, strangely, two birth goddesses, Renenutet and Meskhenet. Ani's soul or 'ba' bird, which will allow him freedom of movement in and out of the tomb after death, perches on a shrine-shaped building, ready to be released if judgment is given in Ani's favor. Into this formidable gathering comes Ani, accompanied by his wife Tutu. They enter from the left, bending forward in proper humility, and Ani mutters the words of Spell 30B of the 'Book of the Dead', which are addressed to his heart in the balance. All, it seems, goes well for Ani; he has qualified for the Afterlife." a. Disponível em: <http://www.britishmuseum.org/collectionimages/AN00684/AN00684654_001_1.jpg>. Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

MUSEUM, C. H., 2017b. "1801: Punched cards control Jacquard loom – Joseph Jacquard's punched cards program patterns on a weaving loom". b. Disponível em: <<http://www.computerhistory.org/storageengine/punched-cards-control-jacquard-loom/>>. [Online; accessed 20-February-2017].

NASA. "National Aeronautics and Space Administration Guidelines for Ensuring the Quality of Information". Disponível em: <http://www.nasa.gov/sites/default/files/517756main_FINAL_NASA_guidelines.pdf>. Acesso em: 12/09/2014.

- NATH, P., KAUR, M., 1981, “Symmetry and the Shannon entropy”, *Information Sciences*, v. 24, n. 2, pp. 111 – 120. ISSN: 0020-0255. doi: [https://doi.org/10.1016/0020-0255\(81\)90014-1](https://doi.org/10.1016/0020-0255(81)90014-1). Disponível em: <<http://www.sciencedirect.com/science/article/pii/0020025581900141>>.
- NAVARRO, G., 2001, “A Guided Tour to Approximate String Matching”, *ACM Computing Surveys*, v. 33, n. 1 (mar.), pp. 31–88. ISSN: 0360-0300. doi: 10.1145/375360.375365. Disponível em: <<http://doi.acm.org/10.1145/375360.375365>>.
- NELSON, R. R., 2004, “The market economy, and the scientific commons”, *Research Policy*, v. 33, n. 3 (April), pp. 455–471. Disponível em: <<http://ideas.repec.org/a/eee/respol/v33y2004i3p455-471.html>>.
- NG, G. S., WAHAB, A., SHI, D., 2003, “Entropy Learning and Relevance Criteria for Neural Network Pruning”, *International Journal of Neural Systems*, v. 13, n. 05, pp. 291–305. doi: 10.1142/S0129065703001637.
- Nolan, L. (Ed.), 2011, *Primary and secondary qualities: the historical and ongoing debate*. Oxford, UK, Oxford University Press. ISBN: 978-0-19-955615-1.
- NOLT, J., ROHATYN, D., 1991, *Lógica*. Coleção Schau. São Paulo, SP, McGraw-Hill. ISBN: 9-07-460872-X.
- NUNES, C. A., PLATÃO, 2001, *Diálogos de Platão - Teeteto Crátilo*. 3 ed. Belém, PA, Editora Universitária UFPA.
- NUSSENZVEIG, H. M., 2003, *Complexidade e caos*. Coleção COPEA. UFRJ. ISBN: 8571082219.
- NUSSENZVEIG, H. M., 1994, *Curso de física básica 2: fluidos, oscilações e ondas e calor*. 2 ed. São Paulo, SP, Brasil, Edgard Blücher. ISBN: 978-85-212-0299-8.
- NUSSENZVEIG, H. M., 2002, *Curso de física básica: mecânica*. 4 ed. São Paulo, SP, Edgard Blücher. ISBN: 978-85-212-0298-1.
- NYQUIST, H., 1924, “Certain factors affecting telegraph speed”, *Bell System Technical Journal, The*, v. 3, n. 2 (abr.), pp. 324–346. ISSN: 0005-8580. doi: 10.1002/j.1538-7305.1924.tb01361.x.
- OF LONDON., L. S., OF LONDON. LIBRARY., L. S., 1791, *Transactions of the Linnean Society of London.*, v. v.1 (1791). London :[The Society], 1791-1875. Disponível em: <<http://www.biodiversitylibrary.org/>>

item/13711>. <http://www.biodiversitylibrary.org/bibliography/683> — Vols. 10-13, 15, issued in 2 pts. each; v. 14, 16, 20, 23-25, 29-30, in 3 pts. each; v. 17-19, 21-22, 26-28, in 4 pts. each. — Vols. 5-21 include section: "Catalogue of the Library of the Linnean Society.— List of members appears in v. 1 and appended to v. 13-14, 17-18.

OLIVA, A., 2003, *Filosofia da ciência*. PAP - Filosofia. Rio de Janeiro, RJ, Brasil, Zahar. ISBN: 85-7110-745-9.

OSMIC, J., SKEJIC, E., SEHIC, Z., 2011, "Model reference fuzzy-neuro text matching system". In: *MIPRO, 2011 Proceedings of the 34th International Convention*, pp. 948–952, May.

OTIS, L., 2011, "Networking: communicating with bodies and machines in the nineteenth century". cap. 2, USA, University of Michigan Press. ISBN: 978-0-472-11213-5.

OXFORD, 1973, *The shorter oxford english dictionary*, v. I & II. Oxford, UK, Clarendon Press - Oxford. ISBN: 0-19-861126-9.

OXFORD2015. "Quality". [online]. Disponível em: <<https://en.oxforddictionaries.com/definition/quality>>. (Accessed 28 February 2015).

PAPATHANASSIOU, M. K., 2010, "Reflections on the antikythera mechanism inscriptions", *Advances in Space Research*, v. 46, n. 4, pp. 545 – 551. ISSN: 0273-1177. doi: <http://dx.doi.org/10.1016/j.asr.2009.10.021>. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S027311770900684X>>. *Advances in Space Environment Research*.

PARRY, R., 2014, "Episteme and Techne". In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, fall 2014 ed.

Pasnau, R., van Dyke, C. (Eds.), 2010, *The cambridge history of medieval philosophy*, v. 1. Cambridge, UK, Cambridge University Press. ISBN: 978-0-521-76216-8.

PATHRIA, R., BEALE, P., 1996, *Statistical mechanics*. 2 ed. Burlington, MA, USA, Elsevier Science. ISBN: 978-0-7506-2469-5.

PÚBLIO, R. N., COUTO, B. R. G. M., VALADÃO, A. F., et al., 2014, "Perfil das solicitações de medicamentos de alto custo ao Sistema Único de Saúde em Minas Gerais", *Revista de Administração Pública*, v. 48 (12), pp. 1567 – 1585. ISSN: 0034-7612. Disponível

em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122014000600010&nrm=iso>.

- PELAGOTTI, A., PIURI, V., 1997, “Entropic Analysis and Incremental Synthesis of Multilayered Feedforward Neural Networks”, *International Journal of Neural Systems*, v. 08, n. 05n06, pp. 647–659. doi: 10.1142/S0129065797000574.
- PEREIRA, J. C. R., 2004, *Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde humanas e sociais*. 3 ed. São Paulo, SP, Brasil, EdUSP. ISBN: 85-314-0523-8.
- PERISSE, G., 2010, *Palavras e origens*. 2 ed. São Paulo, SP, Saraiva. ISBN: 978-85-02-09540-3.
- PETZ, D., 2001, “Entropy, von Neumann and the von Neumann Entropy”. In: Rédei, M., Stöltzner, M. (Eds.), *John von Neumann and the foundations of quantum physics*, pp. 83–96, Dordrecht, Springer Netherlands. ISBN: 978-94-017-2012-0. doi: 10.1007/978-94-017-2012-0_7. Disponível em: <http://dx.doi.org/10.1007/978-94-017-2012-0_7>.
- PHILIPPSON, P., 1944, “Untersuchungen über den Griechischen Mythos”. Zurich, Deutschland, Rhein-Verlag. trad. ital. com o título Origini e forme del mito greco, Boringhieri, Turim 1983, passim.
- PIERCE, J., 1973, “The early days of information theory”, *IEEE Transactions on Information Theory*, v. 19, n. 1 (Jan), pp. 3–8. ISSN: 0018-9448. doi: 10.1109/TIT.1973.1054955.
- PINEDA, J. O. D. C., 2006, *A Entropia Segundo Claude Shannon: o Desenvolvimento do Conceito Fundamental da Teoria da Informação*. Tese de Mestrado, Pontifícia Universidade Católica de São Paulo. Disponível em: <http://www.sapientia.pucsp.br/tde_busca/arquivo.php?codArquivo=2368>.
- PING, W., 2011, “A Brief History of Standards and Standardization Organizations: a chinese Perspective”. In: *East-West Center Working Papers*, n. 117, Economics Series, East-West Center, April. Disponível em: <<http://www.eastwestcenter.org/system/tdf/private/econwp117.pdf?file=1&type=node&id=32840>>.
- PINTO, D., MORGADO, M. C. F., 1997, *Cálculo diferencial e integral de funções de várias variáveis*. Rio de Janeiro, RJ, EDITORA UFRJ. ISBN: 85-7108-204-9.

- PINTO, P., CERCEAU, R., MESQUITA, R., et al., 2013a, “Electronic Checking Process of Governmental Data: A Qualitative Approach”. In: *Proceedings: 4th SHIP Conference: “Exploiting Existing Data for Health Research”*, p. 184, St Andrews, Scotland, August.a.
- PINTO, P. C. V., CARVALHO, L. A. V., 2014, “Record Linkage e Conferência de Dados: uma Avaliação de Metodologias de Cruzamento de Dados Cadastrais Governamentais”. In: *XLVI Simpósio Brasileiro de Pesquisa Operacional*, pp. 671–682, Salvador, BA, Brasil, setembro.
- PINTO, P. C. V., SANTOS, S. A., BARONE, J. A. S., et al., 2011, “Aplicação de Métodos Computacionais para Avaliação da Qualidade das Informações do Cadastro de Beneficiários no SIB/ANS”. In: *Anais: VIII Congresso Brasileiro de Epidemiologia*, São Paulo, SP, Brasil, Novembro.
- PINTO, P. C. V., CERCEAU, R., MESQUITA, R., et al., 2013b, “Conferência Eletrônica de Dados Cadastrais Governamentais por Critérios Qualitativos”. In: *IX Simpósio Brasileiro de Sistemas de Informação: trilhas técnicas*, pp. 803–814, João Pessoa, PB, Brasil, maio. Disponível em: <http://sbsi2013.dce.ufpb.br/pt-br/programacao/artigos#trilha_regular_-_artigos_completos>.
- PINTO, P. C. V., DE CARVALHO, L. A. V., 2014, “Record Linkage e Conferência de Dados: uma Avaliação de Metodologias de Cruzamento de Dados Cadastrais Governamentais”. In: *XLVI Simpósio Brasileiro de Pesquisa Operacional*, pp. 671–682, Salvador, BA, Brasil, setembro.
- PINTO, P. C. V., 2005, *Método de Tableaux Modal para Lógica Modal Epistêmica com Operador de Conhecimento Comum*. Tese de Mestrado, Universidade Federal do Rio de Janeiro. Coordenação dos Programas de Pós-Graduação de Engenharia, Rio de Janeiro, RJ, Brasil, Dezembro.
- PINTO, P. C. V., CARVALHO, L. A. V. D., 2015, “A Mythology for an Approximate Word Matching: Entropy and Quality”. In: *XLVII Simpósio Brasileiro de Pesquisa Operacional*, pp. 2198–2209, Porto de Galinhas, Pernambuco, Brasil, agosto. Disponível em: <<http://www.sbp2015.iltc.br/pdf/142748.pdf>>.
- PIPINO, L. L., LEE, Y. W., WANG, R. Y., 2002, “Data Quality Assessment”, *Commun. ACM*, v. 45, n. 4 (abr.), pp. 211–218. ISSN: 0001-0782. doi: 10.1145/505248.506010. Disponível em: <<http://doi.acm.org/10.1145/505248.506010>>.

- PISKORSKI, J., WIELOCH, K., SYDOW, M., 2009, “On knowledge-poor methods for person name matching and lemmatization for highly inflectional languages”, *Inf. Retr.*, v. 12, n. 3 (jun.), pp. 275–299. ISSN: 1386-4564. doi: 10.1007/s10791-008-9085-5. Disponível em: <<http://dx.doi.org/10.1007/s10791-008-9085-5>>.
- PON, R. K., CÁRDENAS, A. F., 2005, “Data Quality Inference”. In: *Proceedings of the 2Nd International Workshop on Information Quality in Information Systems*, IQIS '05, pp. 105–111, New York, NY, USA. ACM. ISBN: 1-59593-160-0. doi: 10.1145/1077501.1077519. Disponível em: <<http://doi.acm.org/10.1145/1077501.1077519>>.
- POPPER, K. R., 2007, *A Lógica da pesquisa científica*. São Paulo, SP, Cultrix. ISBN: 978-85-316-0236-8. Tradução de Leonidas Hegenberg e Octanny Silveira da Mota.
- POPPER, K. R., 2004, *Lógica das ciências sociais*. Biblioteca tempo universitário. 3 ed. RJ – Rio de Janeiro – Brasil, Tempo Brasileiro. Tradução: Estevão de Rezende Martins, Apio Cláudio Muniz Acquarone Filho e Vilma Oliveira Morais e Silva.
- PORTER, R., NYE, M., 2003, *The Cambridge history of science: Volume 5, the modern physical and mathematical sciences*. Cambridge history ebook collection. Cambridge, UK, Cambridge University Press. ISBN: 0-521-57199-5, 0-521-57243-6.
- PRESSMAN, R. S., 2001, *Software engineering: a practitioner's approach*. Computer science. 5 ed. New York, NY, USA, McGraw-Hill Higher Education. ISBN: 0-07-365578-3.
- PRICE, D. K., 1965, In: *The scientific estate*, pp. 14,24,105 e 282, Cambridge, MA, USA, The Belkanap Press of Harvard University Press. ISBN: 0-674-79485-0.
- PRIESTLEY, H. A., 1990, *Introduction to complex analysis*. Oxford, UK, Clarendon Press - Oxford. ISBN: 0-19-853428-0.
- PRITCHARD, D., 2010, “Relevant Alternatives, Perceptual Knowledge and Discrimination”, *Nouûs*, v. 44, n. 2, pp. 245–268. ISSN: 1468-0068. doi: 10.1111/j.1468-0068.2010.00739.x. Disponível em: <<http://dx.doi.org/10.1111/j.1468-0068.2010.00739.x>>.
- PRIVITERA, G. J., 2012, *Statistics for the behavioral sciences*. Thousand Oaks, CA, USA, SAGE Publications. ISBN: 978-1-4129-6931-4.

- PYZDEK, T., 2003, “I.G Quality Philosophers and Approches (e.g. Juran, Deming, Taguchi, Ishikawa)”. In: Keller, P. A. (Ed.), *Quality engineering handbook (quality and reliability)*, pp. 121–131, New York, NY, US, Marcel Dekker. ISBN: 0-8247-4614-7.
- RAMACHANDRAN, D., CANNY, J., 2008, “The persuasive power of human-machine dialogue”. In: Oinas-Kukkonen, H., Hasle, P., Harjumaa, M., et al. (Eds.), *Persuasive technology: third international conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008. Proceedings*, pp. 189–200, Heidelberg, Berlin, Deutschland, Springer Berlin Heidelberg. ISBN: 978-3-540-68504-3. doi: 10.1007/978-3-540-68504-3_17. Disponível em: <http://dx.doi.org/10.1007/978-3-540-68504-3_17>.
- REALE, G., 2007, *Platão*, v. III, *História da filosofia grega e romana*. São Paulo, SP, Loyola. ISBN: 978-85-15-03304-1. Tradução: Henrique Cláudio de Lima Vaz & Marcelo Perine.
- RENCHER, A. C., SCHAALJE, G. B., 2008, *Linear models in statistics*. 2 ed. Hoboken, New Jersey, USA, Wiley-Interscience. ISBN: 978-0-471-75498-5.
- REZA, F. M., 1961, *An introduction to information theory*. Dover Books on Mathematics Series. New York, NY, USA, Dover. ISBN: 9780486682105.
- Rezende, S. (Ed.), 2003a, *Sistemas inteligentes: fundamentos e aplicações*. Baueri, SP, Brasil, Manole. ISBN: 85-204-1683-7.
- REZENDE, S. O., 2003b, “Sistemas inteligentes: fundamentos e aplicações”. cap. 1. In: , REZENDE (2003a), b. ISBN: 85-204-1683-7.
- RFB, 2014, *CPF - Cadastro de Pessoas Físicas*. Receita Federal Brasileira - Ministério da Fazenda do Brasil, Brasil, janeiro. Disponível em: <<http://idg.receita.fazenda.gov.br/orientacao/tributaria/cadastros/cadastro-de-pessoas-fisicas-cpf>>. Acesso: 08/01/2014).
- RHODES, G., 2006, “The Evolutionary Psychology of Facial Beauty”, *Annual Review of Psychology*, v. 57, n. 1, pp. 199–226. doi: 10.1146/annurev.psych.57.102904.190208. Disponível em: <<http://dx.doi.org/10.1146/annurev.psych.57.102904.190208>>.
- ROSENFELD, L., 2005, *Classical statistical mechanics – notas de aulas de Leon Rosenfeld*. 1 ed. São Paulo, SP, Livraria da Física/CBPF. ISBN: 85-88325-51-9.

- ROSS, S., 1998, *A first course in probability*. Prentice Hall international editions. 5 ed. Upper Saddle River, NJ, USA, Prentice Hall. ISBN: 0-13-896523-4.
- RUSSELL, R., 1923. “Robert c”. abr. 17. Disponível em: <<https://www.google.com/patents/USRE15582>>. US Patent RE15,582.
- RUSSELL, S., NORVIG, P., 2004, *Inteligência artificial: tradução da segunda edição*. 2 ed. Rio de Janeiro, Brasil, Elsevier. ISBN: 85-352-1177-2. Consultoria Editorial e Revisão Técnica: Raul Sidnei Wazlawick, Dr. Tradução: Vandenberg D. de Souza, 11^a Tiragem.
- RUSSELL, S., NORVIG, P., 2010a, *Artificial intelligence : a modern approach*. 3rd ed. Upper Saddle River, NJ, USA, Prentice Hall. ISBN: 978-0-13-604259-4.
- RUSSELL, S., NORVIG, P., 2010b, “Quantifying Uncertainty”. In: *Artificial intelligence : a modern approach*, 3 ed., Prentice Hall Press, Upper Saddle River, NJ, USA, b. ISBN: 0136042597, 9780136042594.
- SANDBACH, F. H., 1959, “Cicero’s Translations from Plato”, *The Classical Review*, v. 9, n. 2, pp. 142–144. Disponível em: <<http://www.jstor.org/stable/702884>>.
- SAVI, M. A., 2006, *Dinâmica não-linear e caos*. Rio de Janeiro, RJ, E-PAPERS. ISBN: 978-85-7650-062-0.
- SAYOOD, K., 2005, “Introduction to data compression”. 3 ed., cap. 1, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. ISBN: 0-12-620862X.
- SCHMALIAN, J., 2012. “Lecture Notes, Statistical Mechanics”. Summer Semester. Disponível em: <<http://www.tkm.kit.edu/downloads/TheoryF2012.pdf>>.
- SCHROEDER, D., 2000, *An introduction to thermal physics*. USA, Addison Wesley. ISBN: 0-201-38027-7.
- SERPRO, 2013, *Relatório de Contratos Firmados com o SERPRO*. SERPRO – Serviço Federal de Processamento de Dados, 04.
- SETHNA, J. P., 2011, “Entropy, order parameters, and complexity”. cap. 5, pp. 82–90, Laboratory of Atomic and Solid State, Corwell University, Ithaca, NY, Clarendon Press - Oxford.
- SHANNON, C. E., 1948, “A mathematical theory of communication”, *Bell System Technical Journal*, v. 27 (July and October), pp. 379–423 and 623–656.

- SHANNON, C. E., 1953, “Computers and Automata”, *Proceedings of the IRE*, v. 41, n. 10 (Oct), pp. 1234–1241. ISSN: 0096-8390. doi: 10.1109/JRPROC.1953.274273.
- SHANNON, C. E., 1955, “Game playing machines”, *Journal of the Franklin Institute*, v. 260, n. 6, pp. 447 – 453. ISSN: 0016-0032. doi: [http://dx.doi.org/10.1016/0016-0032\(55\)90186-1](http://dx.doi.org/10.1016/0016-0032(55)90186-1). Disponível em: <<http://www.sciencedirect.com/science/article/pii/0016003255901861>>.
- SHANNON, C. E., 1958, “Von Neumann’s contributions to automata theory”, *Bulletin of the American Mathematical Society*, v. 64, pp. 123–129.
- SHANNON, C. E., 1956, “The Bandwagon”, *IRE Transactions on Information Theory*, v. 2, n. 1, pp. 3.
- SHAPIRO, S. C., 1992, *Encyclopedia of artificial intelligence*. 2nd ed. New York, NY, USA, John Wiley & Sons, Inc. ISBN: 0471503053.
- SHARMA, H. D., GUPTA, S. D., GUPTA, A. D., 1994, “Entropy, Quality and Wastivity: A Unified View of System Performance”, *Kybernetes*, v. 23, pp. 47–54. doi: 10.1108/03684929410071187.
- SHETTY, J., ADIBI, J., 2005, “Discovering Important Nodes Through Graph Entropy the Case of Enron Email Database”. In: *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD ’05, pp. 74–81, New York, NY, USA. ACM. ISBN: 1-59593-215-1. doi: 10.1145/1134271.1134282. Disponível em: <<http://doi.acm.org/10.1145/1134271.1134282>>.
- SHIELDS, C., 2015, “Aristotle”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, fall 2015 ed.
- SINGLA, P., DOMINGOS, P., 2006, “Entity Resolution with Markov Logic”, *2013 IEEE 13th International Conference on Data Mining*, v. 00, n. undefined, pp. 572–582. ISSN: 1550-4786. doi: doi.ieeecomputersociety.org/10.1109/ICDM.2006.65.
- SKINNER, B. F., 1960, “Pigeons in a pelican”, *American Psychologist*, v. 15, n. 1, pp. 28–37. Disponível em: <<http://www.bfskinner.org/newtestsite/wp-content/uploads/2014/02/Pigeons-in-a-Pelican-pt1.pdf>>.
- SKINNER, B. F., 1972, *Cumulative record: a selection of papers*. Century psychology series. Appleton-Century-Crofts.

- SKINNER, B. F., 1976, *About behaviorism*. Vintage. Knopf Doubleday Publishing Group. ISBN: 0-394-71618-3.
- SKINNER, B. F., 2014, *Verbal behavior*. B. F. Skinner Reprint Series. Cambridge, Massachusetts, B. F. Skinner Foundation. ISBN: 978-0-9899839-0-7.
- SMITH, W., 2005, “Mean”. In: Armitage, P., Colton, T. (Eds.), *Encyclopedia of biostatistics*, USA, John Wiley & Sons. ISBN: 978-0-4700-1181-2. doi: 10.1002/0470011815.b2a15085. Disponível em: <<http://dx.doi.org/10.1002/0470011815.b2a15085>>.
- SNAE, C., 2007, “A Comparison and Analysis of Name Matching Algorithms”, *International Journal of Applied Science. Engineering and Technology*, pp. 252–257.
- SNELL, C., 1968, “Die naturwissenschaftliche Begriffsbildung im Griechischen”. In: Gadamer, H. G. (Ed.), *Um die Begriffswelt der Vorsokratiker*, pp. 21–42, Darmstadt, Deutschland.
- SNIDER, G., 2001, *Measuring the Entropy of Large Software Systems*. Relatório Técnico HPL-2001-221, HP Laboratories, Palo Alto, September. Disponível em: <<http://www.hpl.hp.com/techreports/2001/HPL-2001-221.pdf>>.
- SPIEGEL, M. R., 1972, *Estatística*. Coleção Schaum. 1 ed. Rio de Janeiro, Guanabara, Brasil, Editôra McGraw-Hill do Brasil. Traduzido por Pedro Cosentino.
- STANDARD&POOR`S, 2011. “Guide to Credit Rating Essentials - What are credit ratings and how do they work?” Disponível em: <http://img.en25.com/Web/StandardandPoors/SP_CreditRatingsGuide.pdf>. version 1.4.
- STEINBRUCH, A., WINTERLE, P., 1987, *Álgebra linear*. 2 ed. São Paulo, SP, MacGraw-Hill. ISBN: 0-07-450412-6.
- STEPHEN, G., 1994, *String searching algorithms – v. 6*. Lecture Notes Series on Computing. Sale, MA, USA. ISBN: 978-9-81450186-6.
- STEWART, J., 2007, *Calculus – early transcendentals*. Available 2010 Titles Enhanced Web Assign Series. 6 ed. Belmont, CA, USA, Cengage Learning. ISBN: 978-0-495-01166-8.
- STOCK, W., STOCK, M., 2013, *Handbook of information science*. Knowledge and Information. Berlin, Germany, De Gruyter. ISBN: 978-3-11-023500-5.

- STUDTMANN, P., 2014, “Aristotle’s Categories”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, summer 2014 ed.
- SUSSKIND, L., 2013. “Modern Physics: Statistical Mechanics (Spring 2013)”. Internet. Disponível em: <<http://theoreticalminimum.com/courses/statistical-mechanics/2013/spring>>. Accessed: 01/02/2014.
- SUSSKIND, L., LINDESAY, J., 2005, *An introduction to black holes, information and the string theory revolution: the holographic universe*. Singapore, World Scientific. ISBN: 981-256-083-1.
- SZWARCFITER, J. L., MARKENZON, L., 1994a, *Estruturas de dados e seus algoritmos*. 2 ed. Rio de Janeiro, RJ, Brasil, Livros Tecnicos e Cientificos. ISBN: 85-216-1014-9.
- SZWARCFITER, J. L., MARKENZON, L., 1994b, “Estruturas de dados e seus algoritmos”. 2 ed., cap. Árvores Balanceadas, pp. 127–174, Rio de Janeiro, RJ, Brasil, Livros Tecnicos e Cientificos, b. ISBN: 85-216-1014-9.
- TAGUCHI, G., CHOWDHURY, S., WU, Y., 2004, *Taguchi’s quality engineering handbook*. Wiley-Interscience. ISBN: 0-471-41334-8.
- TAISBAK, C., 2003, *Dedomena: Euclid’s data, or, the importance of being given*. Acta historica scientiarum naturalium et medicinalium. Copenhagen, Denmark, Museum Tusulanum Press. ISBN: 87-7289-815-1.
- TAMBOURATZIS, T., 2001, “String Matching Artificial Neural Networks”, *International Journal of Neural Systems*, v. 11, n. 05, pp. 445–453. doi: 10.1142/S0129065701000874.
- TAN, P.-N., STEINBACH, M., KUMAR, V., 2009, *Introdução ao data mining - mineração de dados*. Rio de Janeiro, RJ - Brasil, Ciência Moderna Ltda. ISBN: 978-85-7393-761=9.
- TAVARES, G. W., PRADO, A. G. S., 2010, “Calorímetro de gelo: uma abordagem histórica e experimental para o ensino de química na graduação.” *Química Nova*, v. 33, pp. 1987 – 1990. ISSN: 0100-4042. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422010000900030&nrm=iso>.
- TAYLOR, J. H., 2010, *Journey through the afterlife: ancient egyptian book of the dead*. Cambridge, MA, Harvard University Press. ISBN: 978-0-674-05750-0.

- TENNANT, G., 2001, *Six sigma: SPC and TQM in manufacturing and services*. Hampshire, England, Gower. ISBN: 0-566-08374-4.
- THIELSCHER, M., 2001, “The Qualification Problem: A solution to the problem of anomalous models”, *Artificial Intelligence*, v. 131, n. 1–2, pp. 1 – 37. ISSN: 0004-3702. doi: 10.1016/S0004-3702(01)00131-X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S000437020100131X>>.
- THOMPSON, A., TAYLOR, B. N., 2008, *Guide for the Use of the International System of Units (SI) – NIST Special Publication 811*. Relatório técnico, National Institute of Standards and Technology, Gaithersburg, MD 20899, US, March. Disponível em: <http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=200349>. Supersedes NIST Special Publication 811, 1995 Edition, April 1995 – Note on 2nd printing: This 2nd printing dated November 2008 of NIST SP811 corrects a number of minor typographical errors present in the 1st printing dated March 2008.
- TOLMAN, R. C., 1938, *The principles of statistical mechanics*. Great Britain, Oxford at the Clarence Press.
- TRIBUS, M., MCIRVINE, E. C., 1971, “Energy and Information”, *Scientific American*, v. 225, n. 3 (set.), pp. 179–188.
- TRIPATHI, A., FARUQUI, M., 2011, “Gender prediction of Indian names”. In: *Students’ Technology Symposium (TechSym), 2011 IEEE*, pp. 137–141, Jan. doi: 10.1109/TECHSYM.2011.5783842.
- TRIVEDI, K. S., 1982, *Probability and statistics with reliability, queuing and computer science applications*. Englewood Cliffs, NJ, USA, Prentice Hall. ISBN: 0-13-711564-4.
- TSALLIS, C., 1995, “Some comments on Boltzmann-Gibbs statistical mechanics”, *Chaos, Solitons & Fractals*, v. 6 (jan.), pp. 539–559. ISSN: 09600779. doi: 10.1016/0960-0779(95)80062-1. Disponível em: <[http://dx.doi.org/10.1016/0960-0779\(95\)80062-1](http://dx.doi.org/10.1016/0960-0779(95)80062-1)>.
- TURING, A. M., 1936, “On Computable Numbers, with an Application to the Entscheidungsproblem”, *Proceedings of the London Mathematical Society*, v. 2, n. 42, pp. 230–265.
- TURLEY, P., WOOD, D., 2006, *Beginning transact-SQL with SQL server 2000 and 2005*. Wrox beginning guides. Indianapolis, Indiana, USA, Wiley. ISBN: 978-0-4717-8016-8.

- TVERSKY, A., KAHNEMAN, D., 1974, “Judgment under Uncertainty: Heuristics and Biases”, *Science*, v. 185, n. 4157, pp. 1124–1131. ISSN: 0036-8075. doi: 10.1126/science.185.4157.1124.
- TZU, S., 2007, *A arte da guerra*. Coleção L&PM pocket. Porto Alegre, RS, L&PM. ISBN: 85-254-1059-4.
- UNDP (Ed.), 2013, *Human Development Report 2013 – The Rise of the South: Human Progress in a Diverse World*. Human Development Report. San Francisco, CA, United Nations Development Programme. ISBN: 978-92-1-126340-4. Disponível em: <http://hdr.undp.org/sites/default/files/reports/14/hdr2013_en_complete.pdf>.
- VALDAR, W. S. J., 2002, “Scoring residue conservation”, *Proteins*, v. 48, n. 2 (ago.), pp. 227–241. ISSN: 1097-0134. doi: 10.1002/prot.10146. Disponível em: <<http://dx.doi.org/10.1002/prot.10146>>.
- VALIANT, L. G., 1984, “A Theory of the Learnable”, *Commun. ACM*, v. 27, n. 11 (nov.), pp. 1134–1142. ISSN: 0001-0782. doi: 10.1145/1968.1972. Disponível em: <<http://doi.acm.org/10.1145/1968.1972>>.
- VARSHNEY, L. R., 2004, “Engineering Theory and Mathematics in the Early Development of Information Theory”. In: *Proceedings of the IEEE Conference on the History of Electronics*, Bletchley Park.
- VERDÚ, S., 1998, “Fifty Years of Shannon Theory.” *IEEE Transactions on Information Theory*, v. 44, n. 6, pp. 2057–2078.
- VIEIRA, J. D. A., 1992, “Astronomia e Meio Ambiente: Uma Visão Semiótica”. In: Guerra, A. T., de Castro, I. E. (Eds.), *Anais do I Workshop de Geociências 1992 UFRJ*, v. 15, Rio de Janeiro - Brasil. Instituto de Geociências, UFRJ / CCMN. Disponível em: <http://www.anuario.igeo.ufrj.br/anuario_1992/vol_15_13%20_20.pdf>.
- VIEIRA, J. D. A., 2009, “Conhecimento e Ficção”, *Kaliopé*, v. 5, n. 9, pp. 59–68. ISSN: 1808-6977.
- VILAS-BÔAS, R. M., 2005, *Introdução ao estudo do direito*. Série Acadêmica. 1 ed. Brasília, Fortium. ISBN: 85-98993-44-1.
- VITA, L. W., 1964, *Introdução à filosofia*. Cultura e ciência. 2 ed. São Paulo, SP, Brasil, Edições Melhoramentos.

- WAND, Y., WANG, R. Y., 1996, “Anchoring Data Quality Dimensions in Ontological Foundations”, *Commun. ACM*, v. 39, n. 11 (nov.), pp. 86–95. ISSN: 0001-0782. doi: 10.1145/240455.240479. Disponível em: <<http://doi.acm.org/10.1145/240455.240479>>.
- WEBER, J. C., 1989, *Principles and Algorithms for Causal Reasoning with Uncertainty*. Tese de Doutorado, Rochester, NY, USA. UMI Order No: GAX89-26062.
- WELLER-FAHY, D. J., BORGHETTI, B. J., SODEMANN, A. A., 2015, “A Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection.” *IEEE Communications Surveys and Tutorials*, v. 17, n. 1, pp. 70–91.
- WESTFALL, R., COHEN, I., 2002, *Newton: textos, antecedentes, comentários*. Rio de Janeiro, RJ, Contraponto – EDUERJ. ISBN: 85-85010-43-7.
- WICKLER, W., 2013, “Understanding Mimicry – with Special Reference to Vocal Mimicry”, *Ethology*, v. 119, n. 4, pp. 259–269. ISSN: 1439-0310. doi: 10.1111/eth.12061. Disponível em: <<http://dx.doi.org/10.1111/eth.12061>>.
- WIDOM, B., 2002, “The Boltzmann Distribution Law and statistical thermodynamics”. In: *Statistical mechanics: a concise introduction for chemists*, cap. 1, Cambridge, UK, Cambridge University Press. ISBN: 0-521-81119-8.
- WIKTIONARY, 2016. “datus — Wiktionary, The Free Dictionary”. Disponível em: <<https://en.wiktionary.org/w/index.php?title=datus&oldid=41076433>>. [Online; accessed 28-February-2016].
- WILLIAMS, B., 2000, *Platão – invensão da filosofia*. Coleção grandes filósofos. São Paulo, SP, Editora Unesp. ISBN: 85-7139-279-X.
- WILLIAMS, M. R., 1992, “Biographies-Joseph Clement: the first computer engineer”, *IEEE Annals of the History of Computing*, v. 14, n. 3, pp. 69–76. doi: 10.1109/85.150083. Disponível em: <<http://dx.doi.org/10.1109/85.150083>>.
- WILLIAMSON, R., TROTTER, H., 2004, *Multivariable mathematics*. 4 ed. Englewood Cliffs, NJ, USA, Pearson Prentice Hall. ISBN: 978-0131235700.
- WRIGHT, P. G., 1970, “Entropy and disorder”, *Contemporary Physics*, v. 11, n. 6, pp. 581–588. doi: 10.1080/00107517008202196.

YAMANO, T., 2001, “A Possible Extension of Shannon’s Information Theory”, *Entropy*, v. 3, n. 4, pp. 280. ISSN: 1099-4300. doi: 10.3390/e3040280. Disponível em: <<http://www.mdpi.com/1099-4300/3/4/280>>.

ZHANG, Z., WANG, Z., 2009, “Empirical study on orderliness evaluation of production system based on aging–quality entropy”, *Production Engineering*, v. 3, n. 1, pp. 95–101. ISSN: 1863-7353. doi: 10.1007/s11740-008-0129-x. Disponível em: <<http://dx.doi.org/10.1007/s11740-008-0129-x>>.

Apêndice A

Comunicação e Mitologia

MEN ACT upon the world, and change it, and are changed in turn by the consequences of their action. Certain processes, which the human organism shares with other species, alter behavior so that it achieves a safer and more useful interchange with a particular environment. When appropriate behavior has been established, its consequences work through similar processes to keep it in force. If by chance the environment changes, old forms of behavior disappear, while new consequences build new forms.

Burrhus Frederic Skinner, Verbal Behavior

A.1 Introdução

Sobre o projeto de sistema de inteligência artificial, (RUSSELL e NORVIG, 2004, p. 5) informam que historicamente há abordagens centradas nos seres humanos e centradas na racionalidade: a primeira seria caracterizada pela pesquisa empírica e confirmação de hipóteses; e a segunda, pela matemática e engenharia.

É patente o desconformo de (RUSSELL e NORVIG, 2004, p. 4) de terem feito uma distinção entre o comportamento humano e racional. Esclarecem que os seres humanos não seriam necessariamente irracionais ou emocionalmente instáveis ou insanos; justificam sua posição atribuindo a irracionalidade humana a algum tipo de imperfeição ou a uma inteligência abaixo da ideal. Por isso pedem ao leitor que leiam *Judgment Under Uncertainty: Heuristics and Biases* de (KAHNEMAN *et al.*, 1982), que aborda sobre alguns erros sistemáticos do raciocínio humano.

Em (KAHNEMAN *et al.*, 1982; TVERSKY e KAHNEMAN, 1974), se observa que os seres humanos não são muito bons em pensar probabilisticamente: recorrem a arquétipos ou modelos pré-determinados para realizar suas inferências quando há incertezas no processo de decisão.

Para utilizar quantidades de incerteza como entrada de um sistema de decisão que pondere as diferenças e similaridades entre duas cadeias de caracteres, ou strings,

será desenvolvido um modelo que sirva como fonte de incerteza construída a partir dessas duas entidades estáticas. Para isso serão apresentados diversos conceitos, dentre os principais: o conceito de comunicação como fenômeno, o de mitologia e o de física qualitativa. A mitologia versa sobre um mecanismo humano de lidar com o desconhecimento sobre o passado para justificar o presente. A física qualitativa, por sua vez, pode se valer dos conceitos míticos para permitir a representação da simulação menos “quantitativa” de sistemas físicos.

A.2 Motivação

Como foi precisamente definido por (SHANNON, 1948), o problema fundamental da comunicação é o de reproduzir em um ponto, exatamente ou aproximadamente, uma mensagem selecionada em outro ponto – já mencionado no capítulo 3. Por outro, é possível também pensar no *problema do cogitar a ocorrência de um processo de comunicação*¹, isto é: uma mensagem pode ter sido fruto de um processo de comunicação de outra mensagem?

Imagine que duas mensagens sejam postas lado a lado. Caso as duas forem idênticas então é mais do que justa a hipótese de se considerar a existência da ocorrência de algum processo de comunicação: visto que uma pode ser a reprodução da outra (mas não necessariamente). Todavia, mas se as duas mensagens forem completamente diferentes, provavelmente, a hipótese de existência de comunicação nem se quer seria cogitada. Ainda assim, e se as mensagens não forem idênticas, mas ainda aparentem ser semelhantes?

Obviamente, cogitar, ou não, a existência de um processo de comunicação pode demandar algum processo de especulação. Deve-se deixar claro que não se leva em conta a possibilidade de que uma das mensagens seja uma forma cifrada da outra (o que demandaria algum tipo de algoritmo de criptografia para garantir essa correspondência), ou que haja um número suficiente grande de pares de mensagens para realizar algum tipo de criptoanálise ou estatística.

A mensagem mais básica na teoria proposta por (SHANNON, 1948) é composta por símbolos oriundos de um alfabeto finito e tal mensagem é também finita (mas com diferentes números de símbolos) – apesar de o sistema receptor ser capaz de recriar um número infinito, mas enumerável, de mensagens distintas. Nada mais

¹O problema de cogitar a existência de comunicação é distinto do reconhecimento de caracteres ou do de aprendizado para esse fim. O problema de reconhecimento de caracteres é um problema de classificação em machine learning. Nele tenta-se obter um processo computacional que infra se um padrão (uma imagem de uma letra representada por uma matriz de zeros e uns) está associado a uma letra alfabética. Existe, portanto, uma relação funcional entre uma instância (o padrão) e o seu conceito (uma letra alfabética). O aprendizado nesse contexto consiste em descobrir uma função, num conjunto de funções possíveis, que faça essa inferência de forma efetiva e com falhas toleráveis para a aplicação em questão – como em (ABU-MOSTAFA *et al.*, 2012)

natural, então, começar a abordar esse problema considerando as mensagens como cadeias de caracteres (strings).

Na literatura, há a existência de diversos algoritmos e fórmulas para mensurar a diferença entre strings, são exemplos: a distância Hamming, a distância Levenshtein, a distância Damerau–Levenshtein e distância Jaro–Winkler. Em (COHEN *et al.*, 2003; DA SILVA *et al.*, 2007; FREIRE *et al.*, 2009; NAVARRO, 2001; PISKORSKI *et al.*, 2009; ?), se ilustra o emprego dessas distâncias entre strings em processos de casamento aproximado de cadeias (approximate string matching).

Outro fator que merece ser considerado, é a percepção de que um processo de comunicação pode envolver tanto um processo de seleção ou de eliminação que é visto, originalmente, em (HARTLEY, 1928):

“In any given communication the sender mentally selects a particular symbol and by some bodily motion, as of his vocal mechanism, causes the attention of the receiver to be directed to that particular symbol. By successive selections a sequence of symbols is brought to the listener’s attention. At each selection there are eliminated all of the other symbols which might have been chosen. **As the selections proceed more and more possible symbol sequences are eliminated, and we say that the information becomes more precise.** For example, in the sentence, “Apples are red”, the first word eliminates other kinds of fruit and all other objects in general. The second directs attention to some property or condition of apples, and the third eliminates other possible colors. It does not, however, eliminate possibilities regarding the size of apples, and this further information may be conveyed by subsequent selections.”

A.3 Ficcionalismo e Perspectivismo

Um pressuposto deste trabalho é que quando duas strings são comparadas espera-se que relações sejam criadas e desfeitas entre elas para se aceitá-las, ou não, como uma sendo o possível fruto de uma reprodução da outra. Essa criação de relações não estaria nas duas strings, mas no processo externo que está propenso a criá-las ou desfazê-las. Porém quão válido, do ponto de vista científico, seria tal processo?

A teoria do conhecimento, ou Gnosiologia, ramo filosófico que aborda os supostos materiais mais gerais do conhecimento – este sendo o seu fenômeno de estudo (VITA, 1964, p. 85 e 89). A Gnosiologia é mais geral que a Epistemologia: enquanto esta se preocupa com os fundamentos dos princípios e leis científicas, aquela associa outras formas de conhecimento. São problemas abordados pela Gnosiologia, por exemplo: a possibilidade do conhecimento, origem do conhecimento e conhecimento da verdade (VITA, 1964, p. 85-111).

O *ficcionalismo* é um conceito-tese sobre a possibilidade do conhecimento, em especial dentro da gnosiologia (VIEIRA, 2009; VITA, 1964) (portanto, engloba também o conhecimento científico, pois este é um caso particular de conhecimento humano). Em (EKLUND, 2011):

Fictionalism about a region of discourse can provisionally be characterized as the view that claims made within that discourse are not best seen as aiming at literal truth but are better regarded as a sort of ‘fiction’.

Uma ficção, que pode ser entendida como útil, mas não mera fantasia dissociada da realidade. Em (LALANDE, 1999, p. 400) uma ficção (representativa) é uma hipótese útil para representar uma lei ou o mecanismo de um fenômeno, sem afirmar sua realidade objetiva. Em (VITA, 1964, p. 94), ela pode ser entendida como a formulação “de pensamentos corretos sobre a realidade à base de representações conscientemente falsas.”

Um exemplo de ficção são os modelos matemáticos mecânicos que envolvem o uso de vetores para descrever a aceleração de um campo gravitacional. Além de ser útil a descrição vetorial do movimento, pois resolve uma gama de problemas, sendo deveras improvável que pessoas percebam pelos seus sentidos um campo gravitacional como uma série de vetores (ou “flexinha” com direção, sentido e propriedades operatórias).

Outro conceito-tese, diferente do *ficcionalismo*, é o perspectivismo que tem em seu cerne (VITA, 1964, p. 93):

(...) a possibilidade de considerar uma coisa e, em geral, o mundo de diversos pontos de vista, todos eles justificados, de tal maneira que cada ponto de vista ofereça uma perspectiva única e, ao mesmo tempo, indispensável acêrca do universo. (..)

Segundo (VIEIRA, 2009), no perspectivismo o problema fundamental não é decidir qual visão é a correta ou certa, mas sim permitir a produção de conhecimento. Entretanto, os pontos de vista justificados devem permitir que sejam compartilhados e testáveis – assim havendo, portanto, a intersubjetividade científica definida em (POPPER, 2007, p. 46).

A.4 Comunicação como Fenômeno Natural

(ENGLAND, 2013), em *Statistical physics of self-replication*, começa:

“Every species of living thing can make a copy of itself by exchanging energy and matter with its surroundings. One feature common to all

such examples of spontaneous “self-replication” is their statistical irreversibility: clearly, it is much more likely that one bacterium should turn into two than that two should somehow spontaneously revert back into one. (...)”

Em (ENGLAND, 2013) , verifica-se que os conceitos de entropia, energia e a capacidade de se autorreplicar são recorrentes nas tentativas de se explicar o que é vida, como em (EMMECHE, 1996). Sobre o estudo pioneiro a respeito de máquinas capazes de se autorreplicar, (SHANNON, 1958) destaca o trabalho de Von Neumann:

“Another branch of automata theory developed by von Neumann is the study of self-reproducing machines – is it possible to formulate a simple and abstract system of ”machines” which are capable of constructing other identical machines, or even more strongly, capable of a kind of evolutionary process in which successive generations construct machines of increasing ”complexity.

A pedra fundamental da biologia moderna, *On the Origin of Species* de (DARWIN, 1859) tinha o título alternativo de *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Nessa obra, a seleção natural foi consagrada como o mecanismo de evolução das espécies e utilizada para explicar a diversidade da vida na Terra. Quando esse trabalho foi escrito, Darwin não tinha conhecimento da existência dos genes. O que chama atenção ao título original é o título alternativo utilizando a palavra preservação — que implicitamente sugere a eliminação das raças (no sentido de espécies) menos favorecidas². A expressão “seleção natural” procura deixar claro que não se trata de uma “seleção pelo homem” – como Darwin destacou nos seus estudos – em especial com a criação de raças de pombos e demais domésticas (DARWIN, 1859).

Assim, é possível pensar que no lugar de haver uma *seleção natural*, haja um processo de *eliminação natural*. É simplesmente uma forma alternativa de ver esse mecanismo: no lugar de os mais “aptos” deixarem descendentes, os menos “inaptos” não chegam a viver o suficiente para tal.

(BATES, 1864a, 1905) percebeu que – em uma vasta região da floresta Amazônica – uma espécie borboleta do gênero *Callithea* possuía um padrão (forma e cor) de asas distinto de outra espécie, mas de mesmo gênero, que habitava na parte ocidental da floresta. Ambas espécies de *Callithea*, em seus habitats, possuíam uma espécie mímica de borboleta (no que tangia ao padrão de asas) – mas ambas do mesmo gênero (*Agrias*). Bates notou que o padrão de asa das borboletas mímicas era coerente com o deslocamento para o Alto Amazonas. As borboletas do gênero

²Menos favorecidas pelas circunstâncias ambientais – importante destacar.

Agrias eram tidas como as mímicas das do gênero *Callithea*, pois Bates acreditava que estas possuíam um gosto ruim para os predadores (devido ao seu forte odor), enquanto aquelas não.

Sobre esse fenômeno, (BATES, 1905) explicou:

When an insect, instead of a dead or inorganic substance, mimicks another species of its own order, and does not prey, or is not parasitic, may it not be inferred that the mimicker is subject to a persecution by insectivorous animals from which its model is free? Many species of insects have a most deceptive resemblance to living or dead leaves; it is generally admitted, that this serves to protect them from the onslaughts of insect-feeding animals who would devour the insect, but refuse the leaf. The same might be said of a species mimicking another of the same order; one may be as repugnant to the tastes of insect persecutors, as a leaf or a piece of bark would be, and its imitator not enjoying this advantage would escape by being deceptively assimilated to it in external appearances.

Ou seja, Bates explicou a espécie mímica como resultado de um processo de seleção natural por predação. Um efeito colateral dessa seleção pode ser considerado um *processo natural de comunicação*: um padrão de asa de uma espécie de borboleta é reproduzido em outra espécie (mas sem ocorrer troca de material genético, por óbvio). Porém no lugar de pensar num processo no qual há um emissor e receptor, Bates conjectura um processo de predação e sobrevivência para justificar sua crença na similaridade. O fato de o padrão da espécie mímica se modificar conforme o padrão da espécie a ser evitada, inclusive confere caráter permanente a esse processo de comunicação (que estampa em uma espécie a aparência da outra).

(DALZIELL e WELBERGEN, 2016) fazem uma apresentação sobre esse processo natural: o mimetismo. É interessante observar que o vocabulário para descrever o processo que resulta no mimetismo inclui termos similares aos da teoria da comunicação. Por exemplo, para classificar algumas espécies envolvidas nesse processo de criação de espécies mímicas há termos como: espécie **receptora**, espécie **emissora** e **sinal** mimético. Todavia o termo espécie receptora, na biologia, está reservado à espécie que realiza a seleção que cria a espécie mímica. A espécie que é a cópia é a mímica (não sendo esta chamada de receptora).

Assim, a noção de que algum tipo de comunicação esteja ocorrendo não é um total absurdo, visto que estudiosos do mimetismo encontram mais facilidade em descrever esse fenômeno valendo-se de vocabulário análogo ao presente na engenharia de telecomunicações. Entretanto, os estudos parecem colocar ênfase nas espécies e processos que realizam a seleção que cria a espécie mímica – com pode ser observado em (DALZIELL e WELBERGEN, 2016; WICKLER, 2013). Portanto, sugere que

onde há processos discretos de seleção (ou eliminação) há uma potencialidade para que comunicações ocorram.

A.5 Abordagem Mitológica

(BATES, 1905) ao tentar justificar a similaridade entre espécies diferentes de borboleta sugere um processo ecológico onde atores figuram em interações na luta pela sobrevivência. Bates, para reforçar essa ideia, considera um processo análogo: insetos que se parecem mais com folhas teriam mais chances de sobreviver do que aqueles que se parecem menos (aos olhos de um predador, por exemplo). Para sustentar a existência do mimetismo tem que se imaginar a existência de um processo ecológico de predação e de eliminação. Obviamente, a espécie mímica deve possuir um potencial para permitir a diversidade de seus indivíduos – caso contrário a eliminação ocorreria de forma indiferente e não existiria uma espécie mímica.

Mudando de situação, se pedir a uma pessoa que explicita a razão de duas *strings* em particular serem similares – um impulso possível pode ser aquele de verbalizar as suas similaridades: a correspondência entre um símbolo idêntico em uma e na outra, levando-se em consideração a ordem das letras nas *strings*. Para justificar diferenças pode-se também recorrer a conjecturas como: erros de digitação ou palavras homófonas ou ambos. No caso de erros, eles podem ser explicitados ou se verbalizar como as diferenças seriam corrigidas.

Tanto as similaridades quanto as diferenças entre duas strings podem ter justificações, mas tais justificações não fazem parte delas e não fazem parte dos processos que a geraram, mas sim do processo de conjectura ou especulação. As justificações podem de alguma forma ser avaliadas: prós e contras em aceitá-las, por exemplo.

Nas duas próximas subseções será apresentado o conceito de mitologia e de ponderação mítica.

A.5.1 Mitologia

Em (BRANDÃO, 2011; HUGHES, 2013a,b), a *mitologia* designa o ramo do conhecimento que se dedica ao estudo dos mitos. A palavra mito (*mythos*), em stricto sensu, é a palavra antiga grega que designava uma estória que era transmitida oralmente. A palavra *logos*, igualmente outra palavra grega antiga, significa uma explicação racional no sentido de ser uma visão geral e coesa de um determinado tópico.

Um mito pode ser classificado com base em (BRANDÃO, 2011; HUGHES, 2013a,b):

- Mito Verdadeiro: ocorre em tempos primordiais, figuram personagens supernaturais, e possui – mas nem sempre – uma função etiológica (explicar por que algo é do jeito que é).
- Lenda³: ocorre em tempos históricos, figuram personagens heroicos e possui uma função histórica (relembrar um fato ou acontecimento, mesmo que de forma inexata).
- Contos Folclóricos⁴: aparecem personagens anti-heroicos e possuem função educativa.

Os mitos gregos poderiam ter versões diferentes e inclusive conflitantes. (HUGHES, 2013a,b) explica que a tentativa de reconciliar essas contradições seria uma forma de compreender o modo de pensar do povo grego da antiguidade clássica:

From these stories, we can deduce, we can arrive at conclusions about who these people were, these ancient Greeks and Romans, who told us these stories. We'll be able to find out what their social values were, what their religious beliefs were. We'll get some pretty good ideas, at any rate. Gender roles. It's all in there. But please keep in mind that these are just stories that are told to entertain and amuse and only incidentally explain things. And that, as such, we can't really expect coherency and consistency from them. If I can give you just one example, the ancient Greeks believed that the goddess **Aphrodite** was born from the splashing down of the severed genitalia of the sky god, **Uranus**. The ancient Greeks also believed that the goddess Aphrodite was the daughter of **Zeus**, the grandson of Uranus, and his wife, **Dione**.

(HUGHES, 2013a,b) segue a analisar a inusitada relação de paternidade acima:

Obviously, to the 20th century mindset, these cannot be true. They are mutually exclusive. But not to the ancient Greeks. The ancient Greek tendency, the ancient Roman tendency, is to reconcile these things. I can predict confidently that it will probably, at some point or another, drive you crazy, but it's part of the territory. Please bear with me.

Outra faceta desse pensamento conciliatório grego da antiguidade clássica pode ser vista em (REALE, 2007, pp. 88 a 90) com a concepção do ser como uma “mistura” de estrutura sempre bipolar (composta por dois elementos antagônicos,

³Lenda (e Saga): sagas também são como lendas, mas dizem respeito a história de um povo.

⁴Contos folclóricos (e Fábulas): Fábulas também são como contos folclóricos, mas seus personagens são animais que atuam como pessoas.

mas não contraditórios) a maneira típica do pensar dos antigos gregos. Em (REALE, 2007, pp. 88 a 90), na teologia do grego antigo observa-se a síntese, ou mistura, de propriedades, ou características, antagônicas nos próprios deuses gregos e entre eles:

(...) como contida na Teogonia de Hesíodo, notar-se-á que, desde a origem, os Deuses e as forças cósmicas se dividiam em duas esferas opostas, encabeçadas pelo Caos e por Gaia e tendo respectivamente, como foi bem observado, as características da “amorfidade” e da “forma” as quais, justamente com essa oposição, resumem a totalidade do real. Também a segunda fase da Teogonia, ou seja com o advento do reino de Zeus e dos Deuses olímpicos, essa concepção fundamental se mostra bem evidente. Os Titãs, derrotados por Zeus, são precipitados no Tártaro que é o “contramundo, oposto polarmente” ao Olímpo. Mas, há mais. Cada um dos Deuses se mostra como um misto de forças com um carácter polarmente oposto. Apolo, por exemplo, tem mesmo como símbolos típicos a lira e o arco com as flechas cruéis; Artemis é virgem e, juntamente, protetora das parturientes, e assim por diante. Além disso, toda divindade tem outra divindade polarmente contraposta, como por exemplo, Apolo tem polarmente contraposto Dionísio; Artêmis tem como polarmente contraposta Afrodite, e assim por diante.

“Por isso Paula Philippson afirmou justamente que “a forma polar” é a estrutura de base da teogonia grega e, em geral, do modo grego de pensar” (REALE, 2007, pp. 88 a 90):

“A forma polar do pensamento vê, concebe, modela e organiza o mundo, como unidade, em pares de contrários. São esses a forma com a qual o mundo se apresenta ao espírito grego e com a qual ele transforma e concebe em ordenamentos e como ordenamentos a multiplicidade do mundo. Esses pares de contrários de forma polar de pensamento são fundamentalmente diferentes dos pares de contrários da forma de pensamento monística ou da dualista, no âmbito das quais os pares se excluem ou então, combatendo-se reciprocamente, se destroem ou, finalmente, reconciliando-se reciprocamente, cessam de existir como contrários [...]. Ao contrário, na forma polar de pensamento, os contrários de um par não somente estão entre si indissolúvelmente unidos, como os pólos do eixo de uma esfera, mas são, na sua mais íntima existência lógica, isto é, exatamente polar, condicionado pela sua oposição: perdendo o pólo oposto eles perderiam o próprio sentido. Esse sentido consiste exatamente no fato de eles, como contrários — do mesmo modo que o eixo que os separa e, no entanto, os une — serem parte de uma unidade maior que não pode ser definida exclusivamente a partir deles: para exprimir-se em termos geométricos, eles são pontos de uma esfera perfeita em si mesmas. Essa forma polar do pensamento informa necessariamente toda

objetivação do pensamento grego. Por isso, foi no seu seio que se formou a visão grega do divino.”

É importante dizer que as propriedades inusitadas podem sugerir uma forma vetorial de interação para a composição do pensamento grego. Porque tais propriedades interagem para compor, no exemplo acima, os Deuses e seu Panteão, mas não são necessariamente opostas nem contraditórias, porém têm características potencializantes e atenuante: formadoras dos seres mitológicos e seu contexto.

A.5.2 Ponderação Mítica

Segundo (BURNS, 1964, p. 53), no final do período que compreende o Médio Império Egípcio (2100 a.C – 1780 a.C), as noções de vida após a morte dos antigos egípcios tinham alcançado sua máxima sofisticação. Um exemplo disso pode ser visto pelo processo de julgamento do egípcio falecido que decidiria o destino de sua alma:

“Com o amadurecimento da teologia, foi, no entanto, adotada uma concepção menos ingênua da vida extraterrena. Acreditava-se então que os mortos deveriam comparecer diante de Osíris para ser julgados de acordo com suas ações na terra. O processo de julgamento compreendia três estágios. No primeiro, exigia-se que o morto declarasse ser inocente de quarenta e dois pecados, inclusive o homicídio, o furto, a mentira, a cobiça, o adultério, a blasfêmia, a ira, o orgulho e a desonestidade em transações comerciais. Após desonerar-se desse rol de vícios o defunto era obrigado a afirmar suas virtudes. Devia confessar que satisfizera as vontades dos deuses, que dera “pão ao faminto, água ao sedento, vestira o despido e dera condução a quem não possuía um barco”. No terceiro e último estágio, o coração do réu era posto na balança em face de uma pena, símbolo da verdade, para se determinar a exatidão do que afirmara. De acordo com a concepção egípcia, o coração representava a consciência, que denunciaria o falso testemunho.”

Ainda segundo (BURNS, 1964), caso o coração não pesasse o mesmo que a pena, a alma do falecido teria um destino de danação. A pesagem do coração⁵ do réu

⁵A Pesagem do Coração ver figura A.1 do Papiro de Ani, região central, (aprox. 1250 a.C): A cena ocorre no Salão de Julgamento. No centro, uma balança que segura em seus dois pratos o coração de Ani (à esquerda) e uma pena (à direita) representando Maat (a personificação divina da verdade e da ordem). A barra transversal da balança pende de uma presilha em forma de pena anexada ao suporte vertical, no topo da qual um pequeno babuíno agachado. Esta criatura é uma forma do deus Thoth, que age de uma forma diferente e com um dever diferente em outro lugar neste “julgamento”. O deus Anúbis, aqui mostrado como uma deidade de cabeça de chacal, de corpo humano, de joelhos, descrito como “Aquele que está no local do embalsamamento”, segura o cordão do prato da direita e estabiliza o prumo do equilíbrio

pode ser vista na figura A.1. Entretanto, em (TAYLOR, 2010, p. 209) não está explicitamente escrito como deveria ser o resultado da pesagem, mas se supõe que haveria um equilíbrio entre a pena e o coração honesto.

O que chama atenção nessa alegoria é a possibilidade de comparar algo abstrato (como a verdade) como se fosse de natureza mecânica ou material, com base numa espécie de máquina de pesagem baseada num princípio de equilíbrio estável. Outro aspecto é que ela não se enquadra como um mito verdadeiro, mas descreve um ritual sobrenatural no qual figuram seres da mitologia egípcia. Ele possuiu uma função etiológica, mas não é sobre o passado, mas sim sobre o futuro em função do passado do falecido.

Nas ciências econômicas, esse tipo de ponderação, mas não tão sobrenatural, é um objeto de estudos, como “confiança”, como pode ser visto em (AKERLOF, 1970) sobre o custo da honestidade:

“Consider a market in which goods are sold honestly or dishonestly; quality may be represented, or it may be misrepresented. The purchaser’s problem, of course, is to identify quality. (...) The cost of dishonesty, therefore, lies not only in the amount by which the purchaser is cheated; the cost also must include the loss incurred from driving legitimate business out of existence.”

A.6 Física Qualitativa de Kleer e Brown: Processos Míticos

(KLEER e BROWN, 1984), juntamente com (FORBUS, 1984; KUIPERS, 1986), são os trabalhos seminais de física qualitativa enquanto subcampo da inteligência artificial (BORCHARDT, 1994, p. 213 e 214). Em (KLEER e BROWN, 1984), observa-se: o conceito de tempo e causalidade míticos; prova como uma explicação; e referências a termodinâmica clássica e a necessidade de um cálculo qualitativo, para modelagem de sistemas físicos. Seguem, *in verbis*, os conceitos-chave para a física qualitativa, segundo (KLEER e BROWN, 1984).

- *Quasistatic approximation.* Most modeling, whether quantitative or qualitative, makes the approximation that behavior at some small time scale is unimportant. In modern thermodynamics, this concept is central to the definition of equilibrium. Until now, qualitative physics has treated this modeling issue in both an ad hoc and tacit manner. In our formulation, quasistatic assumptions play a theoretically motivated and explicit role.



Figura A.1: Pesagem do Coração. Fonte: (MUSEUM, 2017a)

- *Causality*. The behavior of a device is viewed as arising from the interactions of a set of processors, one for each component of the 'device'. The information-passing interactions of the individual components are the cause-effect interactions between the device's components. Within this framework, causal accounts are defined (as interactions that obey certain metaconstraints) and their limitations explored.

- *Mythical causality and mythical time*. Any set of component models makes some assumptions about device behavior (i.e., quasistatic assumptions) and hence cannot, in principle, yield causal accounts for the changes that must occur between equilibrium states of a system. In order to handle this problem we have defined new notions of causality and time (i.e., mythical causality and mythical time) cast in terms of information-passing 'negotiations' between processors of neighboring components.

- *Generalized machines*. Many physical situations can be viewed as some kind of generalized machine, whose behavior can be described in terms of variable values. These variables include force, velocity, pressure, flow, current, and voltage.

- *Proof as explanation*. Physical laws, viewed as constraints, are acausal. We discuss how a logical proof of the solution of a set of constraints is a kind of acausal explanation of behavior.

- *Qualitative calculus*. Qualitative physics is based on a qualitative calculus, the qualitative analog to the calculus of Newton and Leibniz. We define qualitative versions of value, continuity, differential, and integral.

- *Episodes*. Episodes are used to quantize time into periods within which the device's behavior is significantly different.

- *Digital physics*. Each component of a physical system can be viewed as a simple information processor. The overall behavior of the device is produced by causal interactions between physically adjacent components. Physical laws can then be viewed as emergent properties of the universal 'programs' executed by the processors. A new kind of physical law might thus be expressible as constraints on these programs, processors, or the information-flow among them.

Não deixa de ser notável que, de uma forma ou de outra, que (KLEER e BROWN, 1984) vislumbrou a termodinâmica como uma espécie de “musa” para criar modelos de qualitativos da realidade. O próprio conceito de causalidade, segundo o próprio (KLEER e BROWN, 1984), seria algo que a física moderna não oferece formalismo.

Para (KLEER e BROWN, 1984), a causalidade decorre da percepção de uma negociação (ou troca) de informação entre entidades capazes, também, de processá-la. A causalidade mítica descrita por ele pode ser entendida como a possibilidade de uma sequência de estados de um sistema físico entre estados de equilíbrio desse sistema – tal sequência é chamada de trajetória.

O tempo mítico é o conceito mais heterodoxo, pois (KLEER e BROWN, 1984) admite que as leis do sistema possam ser violadas entre certos instantes de tempo, mas, eventualmente, no futuro, pelo intermédio de negociações e interações entre os componentes sistêmicos, elas serão verificadas como válidas.

No que tange à causalidade e ao tempo mítico, a física qualitativa obriga a pensar um sistema não como conjunto de processos nos quais as leis são sempre observadas, mas sim um conjunto de processos que concorre para cumpri-las ou que parecer ser cumpridor delas em alguns instantes. Tal fato sugere que a visão de lei de (KLEER e BROWN, 1984) é mais similar à das leis do direito do que às leis da física. Em (VILAS-BÔAS, 2005, p. 54), “Sendo a norma jurídica considerada como sustentáculo da sociedade, sem a qual a sociedade não se estabilizaria e não se manteria (...)”. Num sistema legal, o código penal estabelece crimes não porque as pessoas sejam avessas a naturalmente não cometê-los. Segundo (GRECO, 2016, p. 2):

A finalidade do Direito Penal é proteger os bens mais importantes e necessários para a própria sobrevivência da sociedade, (...) A pena, portanto, é simplesmente o instrumento de coerção do Direito Penal para a proteção dos bens, valores e interesses mais significativos da sociedade.

Em princípio, as leis no direito, por exemplo, procuram dissuadir os cidadãos de cometer atos que prejudicam a estabilidade do “bom” funcionamento da sociedade. Dependendo do crime, um sistema legal pode até privar a liberdade do criminoso – seja por um tempo (na esperança que não volte a repetir ou se torne inofensivo), ou para sempre (caso não se considere a recuperação possível).

Assim, as visões de leis de (KLEER e BROWN, 1984) na física qualitativa não são as mesmas de um físico, porém são semelhantes às de um legislador. Mas, um legislador pode também se inspirar na física. Segundo (PRICE, 1965, p. 24), a constituição dos Estados Unidos da América – criada durante o Iluminismo – além de ter sido baseada na herança do sistema legal inglês, nas visões filosóficas de Locke e nos interesses econômicos vigentes; ela também teria sido, em parte, influenciada pelo conhecimento científico da época:

It may have been only by crude analogy, but some of the Founding Fathers were certainly conscious of the correspondence between their proposed system of constitutional checks and balances and Sir Isaac Newton’s system of mechanics, in which the universe was held together and kept in order by a balance of counter action forces set in motion by a Primer Mover, but working endlessly thereafter by a sequence of cause and effect.

Por fim, e como já foi dito, a física qualitativa é um subcampo da Inteligência Artificial que investiga “como raciocinar sobre sistemas físicos sem precisarmos mergulhar em equações detalhadas e simulações numéricas” (RUSSELL e NORVIG, 2010a). Troca-se precisão da física pela efetividade, eficiência em resultados práticos (ou pragmáticos) que a engenharia requer, mas sem um volume excessivo de cálculos numéricos. No pior caso, pode-se dizer que a física qualitativa não é sobre o estado da arte da física moderna, mas uma das formas de se pensar em problemas físicos usando-se paradigmas alternativos.

A.7 Variáveis Qualitativas

No século XIX, (KELVIN, 1889, p. 73 e 74), mostra com suas próprias palavras como o conceito de qualidade era percebido na ciência do seu tempo:

In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre

and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be.

(KELVIN, 1889, p. 73 e 74)

Acima percebe-se que Lord Kelvin utilizou a palavra “quality” no sentido de propriedade intrínseca do objeto de estudo. Ao mesmo tempo, conceituou que a percepção de uma qualidade (propriedade) é somente o início do conhecimento – que se tornará conhecimento perfeito quando for representado como números.

O trecho acima é bem conhecido, porém menos conhecida é sua continuação, na qual é exposto um exemplo de conhecimento que seria aquém e investigatório (por ser não-numérico) – no campo da metalurgia e da resistência de materiais:

I may illustrate by a case in which this first step has not been taken. The hardness of different solids, as precious stones and metals, is reckoned by a merely comparative test. Diamond cuts ruby, ruby cuts quartz, quartz, I believe, cuts glass-hard steel, and glass-hard steel cuts glass; hence diamond is reckoned harder than ruby; ruby, than quartz; quartz, than glass-hard steel; and glass-hard steel, than glass: but we have no numerical measure of the hardness of these, or any other solids. We have, indeed, no knowledge of the moduluses of rigidity, or of the tensile strength, of almost any of the gems or minerals, of which the hardness is reckoned by mineralogists in their comparative scale, beginning with diamond, the hardest of known solids. We have even no reason to believe that the modulus of rigidity of diamond is greater than that of other solids; and we have no exact understanding of what this property of hardness is, nor of how it is related to moduluses of elasticity, or to tensile or shearing strength, or to the quality of the substance in respect of its bearing stresses exceeding the limit of its elasticity. It must, therefore, be admitted, that the science of strength of materials, so all-important in engineering, is but little advanced, and the part of it relating to the so-called hardness of different solids least of all; there being in it no step toward quantitative measurement or reckoning in terms of a definite unit.
(KELVIN, 1889, p. 74 e 75)

Digno de nota é a observação de Lord Kelvin – com talvez um quê de frustração – de que não se vislumbrava, em seu tempo, uma mensuração quantitativa para representar a dureza de um material. Ainda sim, se observa que uma relação de ordem é estabelecida pela experimentação que determina a dureza relativa entre materiais. Uma relação de ordem, mas de natureza não-quantitativa, pois não foi possível rotular cada material com sua dureza representada por uma grandeza física (número real) que permitiria a ordenação da dureza e com base no material em si –

ou seja sem necessitar de teste comparativo empírico como outros materiais: não há medição. Lord Kelvin atribuiu essa situação ao estágio pouco avançado da ciência da resistência dos materiais da sua época.

Atualmente, variáveis qualitativas em estatística descritiva são aquelas cujos valores possíveis não representam números, mas sim classes (mesmo que números sejam usados para representar tais classes, mas tais números funcionam como códigos ou rótulos). As variáveis quantitativas são aquelas cujos valores possíveis estão de acordo com uma quantidade observada. As variáveis quantitativas podem, mas não necessariamente, ser medidas em unidades numéricas. Assim ambas as variáveis discretas⁶ e contínuas⁷ podem ser quantitativas (PRIVITERA, 2012).

- qualitativas ordinais: a colocação de um corredor em uma competição, os dias de um mês, grau de instrução formal;
- qualitativas nominais: a cor dos olhos de um indivíduo, um valor booleano (falso ou verdadeiro), as espécies de animais.

Em (BUSSAB e MORETTIN, 1997), as variáveis qualitativas podem ser nominais (não ordenadas) e ordinais, caso haja uma ordenação. Assim uma variável que apresente uma relação de ordem total, ou parcial, é denominada variável qualitativa ordinal. Caso não apresente uma relação de ordem é dita nominal (BUSSAB e MORETTIN, 1997; PEREIRA, 2004). A figura abaixo ilustra um quadro sinótico para a categorização das variáveis em estatística descritiva.

Todavia uma relação de ordem total, ou parcial, pode ser estabelecida entre os possíveis valores que pode assumir.

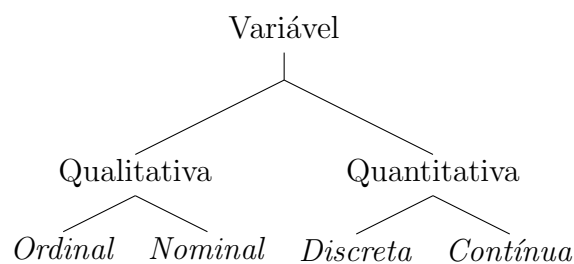


Figura A.2: Classificação das Variáveis Estatísticas (BUSSAB e MORETTIN, 1997)

Para (BUSSAB e MORETTIN, 1997; PEREIRA, 2004; PRIVITERA, 2012), as variáveis quantitativas são classificadas em contínuas ou discretas. Os valores de variáveis discretas devem ser enumeráveis (finitas ou infinitas), ao contrário das variáveis contínuas que não são enumeráveis. Os valores que as variáveis contínuas podem assumir são oriundos de um processo de contagem ou de mensuração.

⁶Variáveis discretas: seus possíveis valores são enumeráveis

⁷Variáveis contínuas: seus possíveis valores são mensuráveis

Um excelente exemplo de variável contínua é a temperatura de um objeto. Pode ser medida pelo intermédio de termômetros e ser maior, menor ou igual a temperatura de outro corpo.

A.8 Discussão

Observa-se que conceitos mitológicos estavam embutidos na física qualitativa, quando se tornou um ramo de estudos na inteligência artificial: a noção de causalidade e tempo míticos.

Os mitos, em *stricto sensu*, têm por característica preencher lacunas de um passado esquecido e satisfazer a curiosidade humana através do discurso. Sistemas mitológicos de crença podem apresentar variações e acomodar versões que podem ser até mutuamente exclusivas – pelo simples fato de serem explicações distintas (o que não exclui serem contraditórias). Por analogia, um mecanismo pode ser sugerido para explicar o porquê strings são similares ou dissimilares: o que sugere o emprego de ficções e de mitologias.

Uma particularidade dos mitos gregos é a bipolaridade: dois polos que presentes em divindades e nas suas relações com o mundo – por “misto de forças com um caráter polarmente oposto”.

A física qualitativa pode ser entendida como um trade-off entre efetividade e precisão: o número é substituído, por exemplo, por processos, regras, provas e dispositivos de forma a se alcançar uma representação não-estática de um sistema físico. Entretanto, deseja manter alguma propriedade operatória, com o cálculo qualitativo.

O mimetismo é um exemplo de que um processo de comunicação pode surgir quando há processos de seleção – sem necessariamente haver consciência, vontade nem engenharia envolvidas. O mimetismo pode ser explicado por um processo de luta pela sobrevivência ao longo do qual um efeito colateral possível é que há espécies de seres vivos que atuam como cópias, outras como copiadas e ainda outras como copiadoras.

O rito da pesagem do coração serve, aqui, como ponto de reflexão sobre a possibilidade da ponderação precisa de grandezas mesmo que originadas de conceitos antropomórficos, como a verdade. Por isso, esse rito parece estar na contramão de ser um exemplo de física qualitativa: se deseja ir de conceitos antropomórficos para o terreno da quantificação e da precisão da ciência vitoriana.

Por fim, tem-se neste capítulo tão somente uma sugestão de arcabouço de mentalidade para o leitor nos próximos capítulos. Nada do que foi apresentado aqui será essencial para o entender do que será desenvolvido nos capítulos seguintes, mas possivelmente será uma ajuda para a sua compreensão.

Apêndice B

Termodinâmica Clássica: Princípio da Máxima Entropia

(...) What's normally called an average quantity in statistics or in probability theory. I've never understood the difference between what statistics is and what probability is. I'm told by some of my friends who are statisticians that there is a difference but I have never understood what it is. I think statistics are actually sets of data, aren't they? (...) Yeah, a real data...

Leonard Susskind, Statistical Mechanics Class – Stanford University

B.1 Introdução

Neste capítulo será apresentado um modelo de maximização da incerteza e preservação da média da energia interna de um sistema termodinâmico usando o *princípio da máxima entropia*. Para isso será utilizado um modelo que calcula **entropia absoluta** de um único sistema termodinâmico. Ao final será ilustrado que há conexões dessa fórmula com a definição de variação de entropia na termodinâmica clássica.

B.2 O Modelo

Definição B.2.1. *O modelo de sistema termodinâmico tem como pressupostos:*

- *O sistema interage com o seu entorno trocando somente energia na forma de calor;*
- *O sistema pode a qualquer momento ter uma quantidade de energia com uma determinada probabilidade;*

- A energia interna do sistema é igual a média da energia dos sistemas prováveis.
- A entropia absoluta do sistema é máxima;
- A distribuição de probabilidade de energia é discreta.

Nesse modelo, não é levado em conta a existência de átomos ou moléculas. Ele tem como objetivo mostrar um modelo de máxima entropia absoluta consistente, pelo menos em forma, com os resultados da termodinâmica clássica, mas do ponto de vista probabilístico. O modelo de distribuição de probabilidade que aqui será derivado é o mesmo em (SCHROEDER, 2000, p.p. 220-241), mas sua derivação é a de (JAYNES, 1957; SUSSKIND, 2013).

Definição B.2.2. *Sejam $\forall i$ $1 > p_i > 0$ a probabilidade de o sistema ter energia a quantidade de energia E_i e $\langle E \rangle$ a média¹ energética do sistema tais que:*

$$\sum_i p_i = 1 \quad (\text{B.2.1})$$

$$\langle E \rangle = \sum_i p_i \cdot E_i \quad (\text{B.2.2})$$

Como a energia interna do sistema, denotada por U , é igual à média da energia do sistema, por definição, então:

$$U = \langle E \rangle \quad (\text{B.2.3})$$

Como o trabalho não é realizado, por definição, então tem-se que variações infinitesimais de energia interna são causadas pelo calor, denotado por Q , dadas por:

$$dU = \delta Q \quad (\text{B.2.4})$$

Isto é: as variações de energia interna são devido a pequenas quantidades de calor que entra no sistema.

Definição B.2.3. *[Entropia Absoluta] $\forall i$, $p_i > 0$, a entropia absoluta S' do sistema é calculada como:*

$$S' = -\mathbf{k}_B \sum_i p_i \cdot \ln p_i \quad (\text{B.2.5})$$

Onde \mathbf{k}_B é uma constante de proporcionalidade não-negativa.

¹A notação $\langle . \rangle$ é a média em geral na mecânica estatística (SETHNA, 2011, p. 16). Na teoria da informação, um ramo da probabilidade e estatística, a notação para média é $E[.]$ (COVER e THOMAS, 1991; KULLBACK, 1978)

B.3 Informações de Otimização

Nessa seção iremos abordar alguns conceitos de otimização para compreender e uniformizar aspectos do método de otimização dos multiplicadores de Lagrange que posteriormente serão abordados, com base em (BERTSEKAS, 2014) e (TAN *et al.*, 2009, an. e).

Propriedade B.3.1 (Entropia Absoluta). Para efeitos de procedimentos de otimização, a função que calcula a entropia absoluta S' tem como propriedades:

- $\forall i, 0 < p_i < 1$;
- Para $\forall i, j$, se $i \neq j$ então $\frac{\partial^2 S'}{\partial p_i \partial p_j} = \frac{\partial^2 S'}{\partial p_j \partial p_i} = 0$ senão $\frac{\partial^2 S'}{\partial p_i^2} = -\frac{\mathbf{k}_B}{p_i}$;
- $\nabla S' = \mathbf{0}$ somente quando para $\forall i, p_i = e^{-1}$.

Definição B.3.1 (Hessiano de uma Função). *Seja $f : R^n \rightarrow R$ contínua e derivável em seu domínio então $H[f]$ denota sua matriz hessiana (ou hessiano de f) tal que:*

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \quad (\text{B.3.1})$$

Teorema B.3.1. *Seja $f : R^n \rightarrow R$ contínua e duas vezes diferenciável em seu domínio então sua expansão de segunda ordem de uma série de Taylor é:*

$$f(\mathbf{x} + \Delta \mathbf{x}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{x}^T \nabla^2 f(\mathbf{x}) \Delta \mathbf{x} \quad (\text{B.3.2})$$

Intuitivamente, o termo $\frac{1}{2} \Delta \mathbf{x}^T \nabla^2 f(\mathbf{x}) \Delta \mathbf{x}$ está relacionado à curvatura de f em \mathbf{x} .

Assim, calculando a matriz Hessiana de S' obtém-se:

$$\nabla^2 S' = \begin{bmatrix} -\frac{\mathbf{k}_B}{p_1} & 0 & \cdots & 0 \\ 0 & -\frac{\mathbf{k}_B}{p_2} & 0 & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{\mathbf{k}_B}{p_n} \end{bmatrix} \quad (\text{B.3.3})$$

É trivial que $\frac{1}{2}\Delta\mathbf{p}^T\nabla^2S'(\mathbf{p})\Delta\mathbf{p} = -\frac{\mathbf{k}_B}{2}\sum_i\frac{(\Delta p_i)^2}{p_i}$. Assim para qualquer incremento $\Delta\mathbf{p} \neq \mathbf{0}$, $\frac{1}{2}\Delta\mathbf{p}^T\nabla^2S'(\mathbf{p})\Delta\mathbf{p} < 0$ para todo domínio de S' . Tal resultado garante que S' é uma função estritamente côncava.

B.4 Máxima Entropia

Análise B.4.1. Maximizar S' sujeito às restrições:

1. $\sum_i p_i = 1$ (Obriga a propriedade de probabilidade)
2. $\langle E \rangle$ (Obriga que as probabilidades sejam coerentes com a média estatística)

A primeira restrição obriga a propriedade de probabilidade e a segunda que as probabilidades sejam coerentes com a média estatística. Como o domínio de S' é $p_i > 0$, para $\forall i$, então p_i será de fato uma probabilidade.

Para resolver esse problema de maximização, bem conhecido na literatura, basta utilizar multiplicadores de Lagrange² com duas restrições e a função objetivo reescritas, respectivamente como (JAYNES, 1965, 1957):

$$G_1 = \sum_i p_i - 1 \tag{B.4.1}$$

$$G_2 = \sum_i p_i \cdot E_i - \langle E \rangle \tag{B.4.2}$$

$$F = -\sum_i p_i \cdot \ln p_i \tag{B.4.3}$$

Essencialmente F é a função que tem que ser maximizada. G_1 e G_2 são restrições – geometricamente definem como é possível se movimentar no domínio de F – ou seja elas definem um subconjunto do domínio de F . É possível perceber que a constante \mathbf{k}_B foi omitida de F se comparada com S' . A razão disso é que um valor máximo de F é também um valor máximo para S' , e vice-versa, visto que $\mathbf{k}_B > 0$. Sobre o método usado, apesar de ele tanto identificar máximos e mínimos 'S' que satisfaçam às restrições – somente máximos serão identificados, pois S' é estritamente côncava (seção B.3) e um máximo local que é global (isto é para todo domínio da função): para cada valor de S' há somente uma curva de nível.

O problema agora consiste em encontrando-se as valorações de α e β , tais que³:

²Multiplicadores de Lagrange com duas restrições ver (STEWART, 2007, p. 938-939). Ver também (BOYD e VANDENBERGHE, 2004)

³Em (BISHOP, 2006, 707-710) e em (STEWART, 2007, p. 938-939), é possível encontrar uma interpretação geométrica para essa equação vetorial.

$$\nabla F = \alpha \cdot \nabla G_1 + \beta \cdot \nabla G_2 \quad (\text{B.4.4})$$

Entretanto, ao longo do desenvolvimento será considerado que seus valores já foram obtidos. Assim, α e β serão considerados novos parâmetros do modelo. Como são soluções, a seguinte relação deve valer:

$$\frac{\partial F}{\partial p_i} = \alpha \cdot \frac{\partial G_1}{\partial p_i} + \beta \cdot \frac{\partial G_2}{\partial p_i} \quad (\text{B.4.5})$$

$$-\ln p_i - 1 = \alpha + \beta \cdot E_i \quad (\text{B.4.6})$$

$$\ln p_i = -(1 + \alpha) - \beta \cdot E_i \quad (\text{B.4.7})$$

$$p_i = \frac{e^{-\beta \cdot E_i}}{e^{1+\alpha}} \quad (\text{B.4.8})$$

Voltando para a restrição 1:

$$\sum_i p_i - 1 = 0 \Rightarrow \sum_i \frac{e^{-\beta \cdot E_i}}{e^{1+\alpha}} - 1 = 0 \quad (\text{B.4.9})$$

$$\sum_i \frac{e^{-\beta \cdot E_i}}{e^{1+\alpha}} = 1 \quad (\text{B.4.10})$$

$$\frac{1}{e^{1+\alpha}} \cdot \sum_i e^{-\beta \cdot E_i} = 1 \quad (\text{B.4.11})$$

$$\sum_i e^{-\beta \cdot E_i} = e^{1+\alpha} \quad (\text{B.4.12})$$

O resultado acima indica que α pode ser obtido como uma função de β , e vice-versa.

Fazendo $Z = \sum_i e^{-\beta \cdot E_i}$ então a equação B.4.8 pode ser reescrita como:

$$p_i = \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \quad (\text{B.4.13})$$

Voltando para a restrição 2:

$$\sum_i p_i \cdot E_i - \langle E \rangle = 0 \Rightarrow \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot E_i - \langle E \rangle = 0 \quad (\text{B.4.14})$$

$$\sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot E_i = \langle E \rangle \quad (\text{B.4.15})$$

Observando-se que:

$$\frac{\partial Z}{\partial \beta} = - \sum_i e^{-\beta \cdot E_i} \cdot E_i \quad (\text{B.4.16})$$

$$- \frac{1}{Z} \cdot \frac{\partial Z}{\partial \beta} = \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot E_i = \langle E \rangle \quad (\text{B.4.17})$$

Pode-se, então, escrever $\langle E \rangle$ em função de Z :

$$\langle E \rangle = - \frac{1}{Z} \cdot \frac{\partial Z}{\partial \beta} = - \frac{\partial \ln Z}{\partial \beta} \quad (\text{B.4.18})$$

Agora resta calcular a entropia absoluta (definição B.2.3 em função da distribuição da probabilidade de estados (equação B.4.13) então:

$$S' = -\mathbf{k}_B \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot \ln\left(\frac{1}{Z} \cdot e^{-\beta \cdot E_i}\right) = -\mathbf{k}_B \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot \left(\ln \frac{1}{Z} + \ln e^{-\beta \cdot E_i}\right) \quad (\text{B.4.19})$$

$$S' = -\mathbf{k}_B \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot (-\ln Z - \beta \cdot E_i) = \mathbf{k}_B \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot (\ln Z + \beta \cdot E_i) \quad (\text{B.4.20})$$

$$S' = \mathbf{k}_B \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot \ln Z + \mathbf{k}_B \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot \beta \cdot E_i \quad (\text{B.4.21})$$

$$S' = \mathbf{k}_B \ln Z \cdot \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} + \mathbf{k}_B \beta \cdot \sum_i \frac{1}{Z} \cdot e^{-\beta \cdot E_i} \cdot E_i = \mathbf{k}_B \ln Z \cdot \sum_i p_i + \mathbf{k}_B \beta \cdot \sum_i p_i \cdot E_i \quad (\text{B.4.22})$$

Finalmente obtém-se:

$$S' = \mathbf{k}_B (\ln Z + \beta \cdot \langle E \rangle) \quad (\text{B.4.23})$$

Assim, o valor da entropia absoluta foi obtido em função das energias internas possíveis E_i e do valor do multiplicador de Lagrange β . O valor médio das energias é um dado do problema. Agora S' será escrita na sua forma diferencial:

$$dS' = \mathbf{k}_B (d(\ln Z) + d(\beta \cdot \langle E \rangle)) \quad (\text{B.4.24})$$

$$dS' = \mathbf{k}_B (d(\ln Z) + d\beta \cdot \langle E \rangle + \beta \cdot d(\langle E \rangle)) \quad (\text{B.4.25})$$

$$dS' = \mathbf{k_B} \left(\frac{\partial \ln Z}{\partial \beta} \cdot d\beta + d\beta \cdot \langle E \rangle + \beta \cdot d(\langle E \rangle) \right) \quad (\text{B.4.26})$$

$$dS' = \mathbf{k_B} \left(\left(\frac{\partial \ln Z}{\partial \beta} + \langle E \rangle \right) \cdot d\beta + \beta \cdot d(\langle E \rangle) \right) \quad (\text{B.4.27})$$

Pela equação B.4.18, $\frac{\partial \ln Z}{\partial \beta}$ pode ser substituído de tal forma:

$$dS' = \mathbf{k_B} ((-\langle E \rangle + \langle E \rangle) \cdot d\beta + \beta \cdot d(\langle E \rangle)) \quad (\text{B.4.28})$$

$$dS' = \mathbf{k_B} 0 \cdot d\beta + \mathbf{k_B} \beta \cdot d(\langle E \rangle) \quad (\text{B.4.29})$$

$$dS' = \mathbf{k_B} \beta \cdot d(\langle E \rangle) \quad (\text{B.4.30})$$

O que pelas equações B.2.3 e B.2.4

$$dS' = \mathbf{k_B} \cdot \beta \cdot dU = \mathbf{k_B} \cdot \beta \cdot \delta Q \quad (\text{B.4.31})$$

Se $S' + S'_0 = S$, onde S_0 é uma constante e $S'_0 > 0$, então, como $dS = \frac{\delta Q}{T}$ e $dS' = dS$ e

$$\beta \cdot \delta Q = \frac{\delta Q}{T} \quad (\text{B.4.32})$$

Assim,

$$\beta = \frac{1}{\mathbf{k_B} \cdot T} \quad (\text{B.4.33})$$

A título de esclarecimento, a interpretação física de $\mathbf{k_B}$ é o fato de ela ser uma constante que converte uma medida de temperatura em Kelvin em uma medida de energia em Joules. A razão disso é que quando se definiram as escalas de temperaturas, usando-se termômetros, ainda não se tinha conhecimento de que a temperatura mensura a energia mecânica cinética (translacional e rotacional) média por molécula do sistema termodinâmico. Isso se reflete no baixíssimo valor dessa constante $\mathbf{k_B} \approx 1,3806503 \cdot 10^{-23} \frac{J}{K}$.



Figura B.1: A Sepultura do físico Ludwig Boltzmann no Cemitério Central, Viena, Áustria. Fonte: (DADEROT, 2005)

Apêndice C

Campos Conservativos: Conceitos e Notações

Imagine a child, perhaps “Dennis the Menace,” who has blocks which are absolutely indestructible, and cannot be divided into pieces. Each is the same as the other. Let us suppose that he has 28 blocks. His mother puts him with his 28 blocks into a room at the beginning of the day. At the end of the day, being curious, she counts the blocks very carefully, and discovers a phenomenal law—no matter what he does with the blocks, there are always 28 remaining! This continues for a number of days, until one day there are only 27 blocks, but a little investigating shows that there is one under the rug—she must look everywhere to be sure that the number of blocks has not changed. (...)

Richard Philips Feynman, The Feynman Lectures on Physics

C.1 Introdução

Neste capítulo serão apresentadas algumas definições e alguns teoremas clássicos sobre campos conservativos e campos vetoriais.

C.2 Definições e Teoremas

No estudo ao cálculo aplicado aos campos vetoriais, um *campo vetorial* associa um vetor a um ponto em \mathbb{R}^n ; um *campo escalar* associa uma escalar¹ a um ponto em \mathbb{R}^n . Dessa forma, tanto os campos vetoriais quanto os campos escalares podem ser modelados como funções. Para efeitos deste trabalho um escalar é um número real.

¹Escalar: um número em \mathbb{R} (um número real) que pode representar uma grandeza, ou magnitude. Entretanto, um escalar pode ser um número em \mathbb{C} (um número complexo) em lato-senso.

Os teoremas e definições nesta seção são bem conhecidos em cursos de cálculo e as referências utilizadas são (LEITHOLD, 1994; PINTO e MORGADO, 1997; STEWART, 2007) como também seus resultados são amplamente utilizados pela física, em especial no que tange a campos conservativos como pode ser visto em (FEYNMAN *et al.*, 2005; NUSSENZVEIG, 2002). Portanto, pelo bem da simplicidade, provas serão omitidas nesta seção. Assume-se familiaridade do leitor com o cálculo diferencial e integral e as notações mais difundidas.

Definição C.2.1 (Integral de Linha de Campo Vetorial). *O campo vetorial $\mathbf{f}: U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$, a integral de linha ao longo da curva suave $\mathcal{C} \subset U$, na direção de $\boldsymbol{\theta}$, é definida como:*

$$\int_{\mathcal{C}} \mathbf{f} \cdot d\boldsymbol{\theta} = \int_a^b \mathbf{f}(\boldsymbol{\theta}(t)) \cdot \boldsymbol{\theta}'(t) dt \quad (\text{C.2.1})$$

$$d\boldsymbol{\theta} = \boldsymbol{\theta}'(t) dt \quad (\text{C.2.2})$$

Onde \cdot é o operador de produto vetorial e $\boldsymbol{\theta}: [a, b] \rightarrow \mathcal{C}$ é uma parametrização bijetora da curva \mathcal{C} tal que $\boldsymbol{\theta}(a)$ e $\boldsymbol{\theta}(b)$ são, respectivamente, o início e o fim da curva \mathcal{C} .

Definição C.2.2 (Integral de Linha Fechada de Campo Vetorial). *O campo vetorial $\mathbf{f}: U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$, a integral de linha ao longo da curva suave $\mathcal{C} \subset U$, na direção de $\boldsymbol{\theta}$, é definida como:*

$$\oint_{\mathcal{C}} \mathbf{f} \cdot d\boldsymbol{\theta} = \int_a^b \mathbf{f}(\boldsymbol{\theta}(t)) \cdot \boldsymbol{\theta}'(t) dt \quad (\text{C.2.3})$$

$$d\boldsymbol{\theta} = \boldsymbol{\theta}'(t) dt \quad (\text{C.2.4})$$

Onde \cdot é o operador de produto vetorial, $\boldsymbol{\theta}: [a, b] \rightarrow \mathcal{C}$ é uma parametrização bijetora em $[a, b)$ e sobrejetora em $[a, b]$ da curva \mathcal{C} tal que $\boldsymbol{\theta}(a)$ e $\boldsymbol{\theta}(b)$ são, respectivamente, o início e o fim da curva \mathcal{C} e $\boldsymbol{\theta}(a) = \boldsymbol{\theta}(b)$.

Definição C.2.3 (Gradiente). *Em um sistema de coordenadas cartesianas, o gradiente é o campo vetorial cujas componentes são as derivadas parciais da função escalar \mathcal{F} :*

$$\nabla \mathcal{F} = \sum_{i=1}^n \frac{\partial \mathcal{F}}{\partial \theta_i} \hat{\mathbf{e}}_{\theta_i} \quad (\text{C.2.5})$$

Onde $\{\hat{\mathbf{e}}_{\theta_i} \mid 1 \leq i \leq N \text{ e } i \in \mathbb{N}_*\}$ é um conjunto de vetores ortonormais e cada $\hat{\mathbf{e}}_{\theta_i}$ diz respeito a uma direção e sentido na coordenada do parâmetro independente θ_i de \mathcal{F} .

Definição C.2.4 (Campo Gradiente e Campo Potencial). *O gradiente de um campo escalar é um campo vetorial. Se \mathcal{F} for um campo escalar e \mathbf{f} for o campo vetorial definido por $\mathbf{f} = \nabla\mathcal{F}$ então: \mathcal{F} é **uma** função potencial para \mathbf{f} ; e \mathbf{f} é campo vetorial gradiente de \mathcal{F} .*

Teorema C.2.1 (Campo Conservativo). *Para qualquer caminho seccionalmente suave \mathcal{C} , contido numa esfera aberta B em \mathbb{R}^n , do ponto \mathbf{s}_0 ao \mathbf{s}_1 . Seja $\boldsymbol{\theta}(t)$ a curva paramétrica de \mathcal{C} , tal que $\mathbf{s}_0 = \mathbf{s}(a)$ e $\mathbf{s}_1 = \mathbf{s}(b)$. Se \mathbf{f} for um campo vetorial conservativo contínuo em B e \mathcal{F} for **uma** função potencial possível para \mathbf{f} (isto é $\mathbf{f} = \nabla\mathcal{F}$) então*

$$\int_{\mathcal{C}} \mathbf{f} \cdot d\boldsymbol{\theta} = \mathcal{F}(\mathbf{s}_1) - \mathcal{F}(\mathbf{s}_0) \quad (\text{C.2.6})$$

será independente do caminho \mathcal{C} , dependendo somente dos pontos inicial e final. Logo, se \mathcal{C} for uma curva fechada então

$$\oint_{\mathcal{C}} \mathbf{f} \cdot d\boldsymbol{\theta} = 0 \quad (\text{C.2.7})$$

Demonstração. Teoremas 6.2 e 6.4. em (PINTO e MORGADO, 1997, p. 69, 218 e 235); ou ver (STEWART, 2007, p. 1032) □

Apêndice D

Espaço de Termômetros

Dave, stop. Stop, will you? Stop, Dave. Will you stop, Dave? Stop, Dave. I'm afraid. I'm afraid, Dave. Dave, my mind is going. I can feel it. I can feel it. My mind is going. There is no question about it. I can feel it. I can feel it. I can feel it.

HAL, 2001: A Space Odyssey

D.1 Introdução

Este capítulo começou a ser desenvolvido por ter sido inspirado no conceito de anisotropia, encontrado na mineralogia e física. A anisotropia é a característica que algumas substâncias possuem, na qual uma determinada propriedade física é dependente da direção. Assim, apesar do nome, a princípio não tem conexão direta com o conceito de entropia. A ideia inicial foi desenvolver um modelo matemático abstrato no qual um ponto poderia ter mais de uma “temperatura” associada a ele – para ser aplicado à qualidade de dados (uma vez que um dado poderia ter diferentes “temperaturas” dependendo do observador; ou dependendo da direção em que se olhasse para ele diferentes temperaturas poderiam ser objetivamente obtidas). Somando-se a isso, a definição das operações vetoriais e do produto interno vetorial poderiam fornecer uma maneira de manipular esse tipo de dado – e ao mesmo tempo se beneficiando do amplo conhecimento matemático existente sobre vetores na matemática.

D.2 Definições

Definição D.2.1 (Assinatura Térmica). *Seja $\mathcal{D} \subseteq \mathbb{Z}$. Se t é uma assinatura térmica em \mathcal{D} então $\forall d \in \mathcal{D} : [t_d] \in \mathbb{R}_*^+$. $[t_d]$ denota uma temperatura de t em d . $\mathcal{T}_{\mathcal{D}}$ é o universo de assinaturas térmicas em \mathcal{D} tal que $t \in \mathcal{T}_{\mathcal{D}}$ se, e somente se, t é uma assinatura térmica em \mathcal{D} .*

Definição D.2.2 (Similaridade de Assinaturas Térmicas). *Sejam t uma assinaturas térmicas em \mathcal{D} e t' uma assinaturas térmicas em \mathcal{D}' . $t \simeq t'$ se, somente se, $\forall d \in \mathcal{D} \cap \mathcal{D}': [t_d] = [t'_d]$.*

Definição D.2.3 (Igualdade de Assinaturas Térmicas). *Sejam t uma assinaturas térmicas em \mathcal{D} e t' uma assinaturas térmicas em \mathcal{D}' . $t = t'$ se, somente se, $t \simeq t'$ e $\mathcal{D} = \mathcal{D}'$.*

Definição D.2.4 (Soma Vetorial e Multiplicação por Escalar). *Seja $\mathcal{T}_{\mathcal{D}}$ universo de assinaturas térmicas em \mathcal{D} . Sejam $t' \in \mathcal{T}_{\mathcal{D}}$, $t'' \in \mathcal{T}_{\mathcal{D}}$ e $\alpha \in \mathbb{R}$ então:*

Soma Vetorial: $t' \oplus t''$, se, somente se, $\forall d \in \mathcal{D}: [(t' \oplus t'')_d] = [t'_d] \cdot [t''_d]$; e

Multiplicação por Escalar: $\alpha \odot t'$, se, somente se, $\forall d \in \mathcal{D}: [(\alpha \odot t')_d] = [t'_d]^\alpha$.

Corolário D.2.1 (Fechamento da Soma Vetorial e Multiplicação por Escalar). *Seja $\mathcal{T}_{\mathcal{D}}$ o universo de assinaturas térmicas em \mathcal{D} . Para todo $t' \in \mathcal{T}_{\mathcal{D}}$, $t'' \in \mathcal{T}_{\mathcal{D}}$ e $\alpha \in \mathbb{R}$ então:*

- $t' \oplus t'' \in \mathcal{T}_{\mathcal{D}}$; e
- $\alpha \odot t' \in \mathcal{T}_{\mathcal{D}}$

Demonstração D.2.2 (Corolário D.2.1). *A demonstração possui duas partes independentes:*

Fechamento da Soma Vetorial *Sejam $d \in \mathcal{D}$, $t' \in \mathcal{T}_{\mathcal{D}}$ e $t'' \in \mathcal{T}_{\mathcal{D}}$. Pela definição D.2.1, então $[t'_d] \in \mathbb{R}_*^+$ e $[t''_d] \in \mathbb{R}_*^+$. O que implica $[t'_d] \cdot [t''_d] \in \mathbb{R}_*^+$ para qualquer $d \in \mathcal{D}$ (propriedade do fechamento da multiplicação dos números não-negativos). Como $[(t' \oplus t'')_d] = [t'_d] \cdot [t''_d]$, pela definição D.2.4 (Soma Vetorial). No que resulta em $t' \oplus t'' \in \mathcal{T}_{\mathcal{D}}$, pela definição D.2.1, pois para qualquer $d \in \mathcal{D}$: $[(t' \oplus t'')_d] \in \mathbb{R}_*^+$. \square*

Fechamento do Multiplicação por Escalar *Sejam $\alpha \in \mathbb{R}$, $d \in \mathcal{D}$ e $t' \in \mathcal{T}_{\mathcal{D}}$. Pela definição D.2.1, então $[t'_d] \in \mathbb{R}_*^+$. O que implica $[t'_d]^\alpha \in \mathbb{R}_*^+$, para qualquer $d \in \mathcal{D}$ (propriedade do fechamento da potência real de números não-negativos). Como $[(\alpha \odot t')_d] = [t'_d]^\alpha$, pela definição D.2.4 (Multiplicação por Escalar). No que resulta em $\alpha \odot t' \in \mathcal{T}_{\mathcal{D}}$, pela definição D.2.1, pois para qualquer $d \in \mathcal{D}$: $[(\alpha \odot t')_d] \in \mathbb{R}_*^+$. \square*

Observação D.2.3. Assim, de agora em diante, sabemos que ao efetuar as operações de soma de termômetros e multiplicação de termômetros por escalar acima definidas sempre se obtém um para o mesmo dado. Isso é importante, pois ressalta a natureza vetorial do termômetro e indica a possibilidade da construção de termômetros mais complexos. Isso só foi possível devido à noção de um termômetro como uma função sobrejetora.

Definição D.2.5 (Elemento Neutro e Inverso Aditivo). *Seja \mathcal{D} um conjunto de dados e $t \in \mathcal{T}_{\mathcal{D}}$:*

- \mathcal{O} , $\forall d \in \mathcal{D}$: $[\mathcal{O}_d] = 1$; e
- $(-t) = -1 \odot t$;

Corolário D.2.4. $\mathcal{O} \in \mathcal{T}_{\mathcal{D}}$ e $-t \in \mathcal{T}_{\mathcal{D}}$

Demonstração D.2.5. *Como $[\mathcal{O}_d] = 1$ para todo $d \in \mathcal{D}$ e $1 \in \mathbb{R}_*^+$. Então $\mathcal{O} \in \mathcal{T}_{\mathcal{D}}$, por definição D.2.1. Como $t \in \mathcal{T}_{\mathcal{D}}$, então $(-1 \odot t)$ é uma assinatura térmica em \mathcal{D} (Fechamento, corolário D.2.1). Portanto, $(-t) \in \mathcal{T}_{\mathcal{D}}$, pois $(-t) = (-1 \odot t)$. \square*

Definição D.2.6 (Subtração de Vetores). $t, t' \in \mathcal{T}_{\mathcal{D}}$:

- $t \ominus t' = t \oplus (-t')$;

Conforme pode ser observado na tabela D.2, a precedência das operações vetoriais será análoga àquela que comumente utilizada para os operadores de soma vetorial e multiplicação por escalar no clássico espaço vetorial \mathbb{R}^n .

Tabela D.1: Precedência Geral de Operadores Vetoriais

Ordem	Operação	Símbolo
1°	Parênteses	()
2°	Multiplicação por Escalar	\odot
3°	Adição e Subtração Vetoriais	\oplus e \ominus

D.3 Propriedades da Soma Vetorial

1

Propriedade D.3.1 (Associativa).

$$(u \oplus v) \oplus w = u \oplus (v \oplus w), \forall u, v, w \in \mathcal{T}_{\mathcal{D}}$$

Demonstração D.3.2. *Sejam $u, v, w \in \mathcal{T}_{\mathcal{D}}$, para todo $d \in \mathcal{D}$*

$$\begin{aligned} [((u \oplus v) \oplus w)_d] &= [(u \oplus v)_d] \cdot [w_d] = [u_d] \cdot [v_d] \cdot [w_d] = \\ &= [u_d] \cdot ([v_d] \cdot [w_d]) = [u_d] \cdot [(v \oplus w)_d] = [u \oplus (v \oplus w)]_d \end{aligned}$$

\square

Então para todo $d \in \mathcal{D}$, $[((u \oplus v) \oplus w)_d] = [u \oplus (v \oplus w)]_d$ então $(w \oplus v) = (v \oplus w)$.

¹(STEINBRUCH e WINTERLE, 1987, p. 19)

Propriedade D.3.3 (Comutativa).

$$(w \oplus v) = (v \oplus w), \forall w, v \in \mathcal{T}_{\mathcal{D}}$$

Demonstração D.3.4. *Sejam $w, v \in \mathcal{T}_{\mathcal{D}}$, para todo $d \in \mathcal{D}$, por aplicação da definição D.2.1:*

$$[(w \oplus v)_d] = [w_d] \cdot [v_d] = [v_d] \cdot [w_d] = [(v \oplus w)_d]$$

Então para todo $d \in \mathcal{D}$, $[(w \oplus v)_d] = [(v \oplus w)_d]$ então $(w \oplus v) = (v \oplus w)$. \square

Propriedade D.3.5 (Existência do Elemento Neutro Aditivo).

$$\mathcal{O} \in \mathcal{T}_{\mathcal{D}}, \forall u \in \mathcal{T}_{\mathcal{D}}, \mathcal{O} \oplus u = u$$

Demonstração D.3.6. *Seja $u \in \mathcal{T}_{\mathcal{D}}$, para todo $d \in \mathcal{D}$, pela definição D.2.5 então:*

$$[\mathcal{O}_d] = 1$$

$$[(\mathcal{O} \oplus u)_d] = [\mathcal{O}_d] \cdot [u_d] = 1 \cdot [u_d] = [u_d]$$

Para todo $d \in \mathcal{D}$, $[(\mathcal{O} \oplus u)_d] = [u_d]$ então $(\mathcal{O} \oplus u) = u$.

Então \mathcal{O} é um elemento neutro para u . \square

Propriedade D.3.7 (Existência do Simétrico Aditivo).

$$\forall u \in \mathcal{T}_{\mathcal{D}}, \exists (-u) \in \mathcal{T}_{\mathcal{D}}, u \oplus (-u) = \mathcal{O}$$

Demonstração D.3.8. *Seja $u \in \mathcal{T}_{\mathcal{D}}$. Pela definição D.2.5 então $(-u) = (-1 \odot u)$. Para todo $d \in \mathcal{D}$*

$$[(u \oplus (-1 \odot u))_d] = [u_d] \cdot [(-1 \odot u)_d] = [u_d] \cdot [(u_d)^{-1}] = \frac{[u_d]}{[u_d]} = 1 = [\mathcal{O}_d]$$

Para todo $d \in \mathcal{D}$, $[(u \oplus (-1 \odot u))_d] = [\mathcal{O}_d]$ então $u \oplus (-1 \odot u) = \mathcal{O}$.

Então $(-1 \odot u)$ é um simétrico aditivo para u . \square

D.4 Propriedades da Multiplicação por Escalar

2

²(STEINBRUCH e WINTERLE, 1987, p. 19)

Propriedade D.4.1 (Associativa de Produtos por Escalares e Vetoriais). *Para todo $u \in \mathcal{T}_{\mathcal{D}}$ e para todo $\alpha, \beta \in \mathbb{R}$:*

$$(\alpha \cdot \beta) \odot u = \alpha \odot (\beta \odot u)$$

Demonstração D.4.2. *Sejam $u \in \mathcal{T}_{\mathcal{D}}$ e $\alpha, \beta \in \mathbb{R}$, para todo $d \in \mathcal{D}$*

$$\begin{aligned} [((\alpha \cdot \beta) \odot u)_d] &= [u_d]^{\alpha \cdot \beta} = ([u_d]^\beta)^\alpha = \\ &= [(\beta \odot u)_d]^\alpha = [(\alpha \odot (\beta \odot u))_d] \end{aligned}$$

Para todo $d \in \mathcal{D}$, $[((\alpha \cdot \beta) \odot u)_d] = [(\alpha \odot (\beta \odot u))_d]$ então $(\alpha \cdot \beta) \odot u = \alpha \odot (\beta \odot u)$. □

Propriedade D.4.3 (Distributiva da Multiplicação Vetorial em Relação à Soma de Escalares). *Para todo $u \in \mathcal{T}_{\mathcal{D}}$ e para todo $\alpha, \beta \in \mathbb{R}$:*

$$(\alpha + \beta) \odot u = (\alpha \odot u) \oplus (\beta \odot u)$$

Demonstração D.4.4. *Sejam $u \in \mathcal{T}_{\mathcal{D}}$ e $\alpha, \beta \in \mathbb{R}$, para todo $d \in \mathcal{D}$*

$$\begin{aligned} [((\alpha + \beta) \odot u)_d] &= [u_d]^{\alpha + \beta} = [u_d]^\alpha \cdot [u_d]^\beta = \\ &= [(\alpha \odot u)_d] \cdot [(\beta \odot u)_d] = [((\alpha \odot u) \oplus (\beta \odot u))_d] \end{aligned}$$

Para todo $d \in \mathcal{D}$, $[((\alpha + \beta) \odot u)_d] = [((\alpha \odot u) \oplus (\beta \odot u))_d]$ então $(\alpha + \beta) \odot u = (\alpha \odot u) \oplus (\beta \odot u)$. □

Propriedade D.4.5 (Distributiva da Multiplicação por Escalar em Relação à Adição Vetorial). *Para todo $u, v \in \mathcal{T}_{\mathcal{D}}$ e para todo $\alpha \in \mathbb{R}$:*

$$\alpha \odot (u \oplus v) = (\alpha \odot u) \oplus (\alpha \odot v)$$

Demonstração D.4.6. *Sejam $u, v \in \mathcal{T}_{\mathcal{D}}$ e $\alpha \in \mathbb{R}_*^+$, para todo $d \in \mathcal{D}$*

$$\begin{aligned} [(\alpha \odot (u \oplus v))_d] &= [(u \oplus v)_d]^\alpha = ([u_d] \cdot [v_d])^\alpha = \\ &= [u_d]^\alpha \cdot [v_d]^\alpha = [(\alpha \odot u)_d] \cdot [(\alpha \odot v)_d] = [((\alpha \odot u) \oplus (\alpha \odot v))_d] \end{aligned}$$

Para todo $d \in \mathcal{D}$, $[(\alpha \odot (u \oplus v))_d] = [((\alpha \odot u) \oplus (\alpha \odot v))_d]$ então $\alpha \odot (u \oplus v) = (\alpha \odot u) \oplus (\alpha \odot v)$. □

Propriedade D.4.7 (Elemento Neutro da Multiplicação por Escalar). *Para todo $u, v \in \mathcal{T}_{\mathcal{D}}$ e para todo $1 \in \mathbb{R}$:*

$$1 \odot u = u$$

Demonstração D.4.8. *Sejam $u \in \mathcal{T}_{\mathcal{D}}$ e $1 \in \mathbb{R}$, para todo $d \in \mathcal{D}$:*

$$[(1 \odot u)_d] = [u_d]^1 = [u_d]$$

Para todo $d \in \mathcal{D}$, $[(1 \odot u)_d] = [u_d]$ então $1 \odot u = u$. □

D.5 Espaço Vetorial

Análise D.5.1. Um conjunto, independente da sua natureza, que tenham uma soma e produto definidos que atendam atender as propriedades das seções D.3 e D.4 são um espaço vetorial (STEINBRUCH e WINTERLE, 1987, p. 18 e 19). A definição D.2.4 estabelece uma soma e um produto para o conjunto $\mathcal{T}_{\mathcal{D}}$ (definição D.2.1).

D.6 Produto Interno e Norma Euclidiana

(STEINBRUCH e WINTERLE, 1987, p. 109) define produto “interno em espaço vetorial uma $V \times V$ em \mathbb{R} que a todo para de vetores $(u, v) \in V \times V$ associa um número real, indicado por $\langle u, v \rangle$ (...)”

Definição D.6.1 (Produto Interno e Norma Euclidiana). *Seja $t: \mathcal{D} \rightarrow \mathbb{R}_+^* \in \mathcal{T}$ é um conjunto de termômetros tais que. t' e $t'' \in \mathcal{T}$.*

- $\langle t', t'' \rangle = \sum_{\forall d \in \mathcal{D}} \ln[t'_d] \cdot \ln[t''_d]$; e
- $\|t\| = \sqrt{\langle t, t \rangle}$

Propriedade D.6.1 (Comutativa).

$$\langle u, v \rangle = \langle v, u \rangle \forall u, v \in \mathcal{T}_{\mathcal{D}}$$

Demonstração D.6.2. *Sejam $u, v \in \mathcal{T}_{\mathcal{D}}$:*

$$\langle u, v \rangle = \sum_{\forall d \in \mathcal{D}} \ln[u_d] \cdot \ln[v_d] = \sum_{\forall d \in \mathcal{D}} \ln[v_d] \cdot \ln[u_d] = \langle v, u \rangle$$

□

Propriedade D.6.3 (Distributiva).

$$\langle u, v \oplus w \rangle = \langle u, v \rangle + \langle u, w \rangle \forall u, v, w \in \mathcal{T}_{\mathcal{D}}$$

Demonstração D.6.4. *Sejam $u, v, w \in \mathcal{T}_{\mathcal{D}}$:*

$$\begin{aligned}
\langle u, v \oplus w \rangle &= \sum_{\forall d \in D} \ln[u_d] \cdot \ln[(v \oplus w)_d] = \sum_{\forall d \in D} \ln[u_d] \cdot \ln([v_d] \cdot [w_d]) = \\
&= \sum_{\forall d \in D} \ln[u_d] \cdot (\ln[v_d] + \ln[w_d]) = \sum_{\forall d \in D} (\ln[u_d] \cdot \ln[v_d] + \ln[u_d] \cdot \ln[w_d]) = \\
&= \sum_{\forall d \in D} \ln[u_d] \cdot \ln[v_d] + \sum_{\forall d \in D} \ln[u_d] \cdot \ln[w_d] = \langle u, v \rangle + \langle u, w \rangle
\end{aligned}$$

□

Propriedade D.6.5 (Associativa).

$$\langle \alpha \odot u, v \rangle = \alpha \cdot \langle u, v \rangle \quad \forall \alpha \in \mathbb{R} \quad \forall u, w \in \mathcal{T}_{\mathcal{D}}$$

Demonstração D.6.6. *Sejam $u, v \in \mathcal{T}_{\mathcal{D}}$ e $\alpha \in \mathbb{R}$:*

$$\begin{aligned}
\langle \alpha \odot u, v \rangle &= \sum_{\forall d \in D} \ln[(\alpha \odot u)_d] \cdot \ln[v_d] = \sum_{\forall d \in D} \ln[u_d]^\alpha \cdot \ln[v_d] = \sum_{\forall d \in D} \alpha \cdot \ln[u_d] \cdot \ln[v_d] = \\
&= \alpha \cdot \sum_{\forall d \in D} \ln[u_d] \cdot \ln[v_d] = \alpha \cdot \langle u, v \rangle
\end{aligned}$$

□

Propriedade D.6.7 (Auto produto interno).

$$\langle u, u \rangle \geq 0 \quad \forall u \in \mathcal{T}_{\mathcal{D}} \quad \text{e} \quad \langle u, u \rangle = 0 \Leftrightarrow u = \mathcal{O}$$

Demonstração D.6.8. *A prova é realizada em duas partes, na primeira $\langle u, u \rangle \geq 0 \quad \forall u \in \mathcal{T}_{\mathcal{D}}$ e a segunda se subdivide na prova da implicação $\langle u, u \rangle = 0 \Rightarrow u = \mathcal{O}$ e na da sua recíproca. A prova da recíproca é feita por **reductio ad absurdum**. Sejam $u \in \mathcal{T}_{\mathcal{D}}$ e $\alpha \in \mathbb{R}$:*

Primeira Parte: $\langle u, u \rangle = \sum_{\forall d \in D} \ln[u_d] \cdot \ln[u_d] = \sum_{\forall d \in D} (\ln[u_d])^2 \geq 0$

Segunda Parte: Implicação: $\langle \mathcal{O}, \mathcal{O} \rangle = \sum_{\forall d \in D} (\ln[\mathcal{O}_d])^2 = \sum_{\forall d \in D} (\ln[1])^2 = \sum_{\forall d \in D} 0 = 0$

Recíproca: *Suponha que $\langle u, u \rangle = 0$ e $u \neq \mathcal{O}$. Então existe pelo menos $d' \in D$ tal que $[u_{d'}] \neq 1$. Então $\langle u, u \rangle = \sum_{\forall d \in D} (\ln[u_d])^2 = \sum_{\forall d \in D \setminus \{d'\}} (\ln[u_d])^2 + (\ln[u_{d'}])^2$. Por hipótese, $0 = \sum_{\forall d \in D \setminus \{d'\}} (\ln[u_d])^2 + (\ln[u_{d'}])^2$. Assim,*

$-(\ln[u_{d'}])^2 = \sum_{\forall d \in D \setminus \{d'\}} (\ln[u_d])^2$. Como \ln é uma função monotonicamente crescente então $\ln[u_{d'}] \neq 0$, uma vez que $\ln 1 = 0$ e, por hipótese, $[u_{d'}] \neq 1$. Então $0 > -(\ln[u_{d'}])^2 = \sum_{\forall d \in D \setminus \{d'\}} (\ln[u_d])^2$. Então $0 > \sum_{\forall d \in D \setminus \{d'\}} (\ln[u_d])^2$. Uma soma de quadrados de números reais não pode ser necessariamente não-positiva.

□

Apêndice E

Outros Estudos

$$S = k \cdot \log W$$

Ludwig Eduard Boltzmann, Gravação em sua lápide

“Begin at the beginning,” (...), “and go on till you come to the end: then stop.”

Lewis Carroll, Alice’s Adventures in Wonderland

E.1 Dados e Programas Programas de Computador

A palavra dados apareceu na língua portuguesa em 1813 e na língua inglesa em 1646 (*data*) – portanto antes de se pensar no conceito de uma máquina de computar de uso genérico (conceito atribuído à Charles Babbage na década de 1830 pelo projeto de sua máquina analítica). Em ambas as línguas, dados deriva da palavra latina *data* – o particípio passado do verbo *dare* – significando aquilo que é fornecido (ou, seja, é dado). Na língua inglesa, *data* significa: alguma coisa fornecida ou tida como certa; algo conhecido ou assumido como fato, e feito como base para *pensar racionalmente*¹ ou *calcular* (DA CUNHA, 2010; FLORIDI, 2010; LEWIS e SHORT, 1879; OXFORD, 1973).

Na matemática ocidental, tradução em latim para a obra *Dedomena* de Euclides, é *Data* – que em português é *Dados* (LIVIO, 2011, p. 94). Nessa obra, a entrada de um problema e sua solução seriam, ambos, dados (TAISBAK, 2003, p. 13): um dado seria algo que é fornecido ou deduzido.

A palavra dados (ou *data*) aparece em várias expressões, ou termos, na ciência da computação: compressão de dados, *data handling*, *data privacy*, *data search engine*,

¹Pensar racionalmente: reasoning

data security, estrutura de dados e algoritmos, banco de dados, *data warehouse*, *deductive database*, armazenamento de dados, metadados e – inclusive – podem ser considerados “crus”². Mas qual é o papel dos dados em uma computação?

(KNUTH, 1996a) abordou a diferença conceitual entre *dado* e *informação*:

The letter by Dr. Huber defines “algorithm” in terms of programming languages. I would like to take a slightly different point of view, in which algorithms are concepts that have existence apart from any programming language. To me the word *algorithm* denotes an abstract method for computing some output from some input, while a *program* is an embodiment of a computational method in some language. I can write several different programs for the same algorithm (e.g. in *Algol 60* and in *PL/I*, assuming that those languages are given an unambiguous interpretation).

Of course if I am pinned down asked to explain more precisely what I mean by these remarks, I am forced to admit that I don’t know any way to define any particular algorithm except in a programming language. Perhaps the set of all concepts should be regarded as a formal language of some sort. But I believe that algorithms were present long before *Turing et al.* formulated them, just as the concept of the number “two” was in existence long before the writers of first grade textbooks and other mathematical logicians gave it a certain precise definition.

(...)

By *computation* I mean essentially the same thing as what many people nowadays call “data processing”, “symbol manipulation”, or more generally “information processing”.

There seems to be confusion between the words *information* and *data* much like that between *algorithm* and *program*. When a scientist conducts an experiment in which some quantity is being measured, we have four things present, each of which is often called *information*: (a) the true value of the quantity; (b) the approximation to this true value that is actually obtained by the measuring device; (c) a representation of the value (b) in some formal language; and (d) the concepts learned by the scientist by a study of the measurements. The word *data* is most approximately applied to (c), and the word “information” when used in a technical sense should be further qualified by stating what kind of information is meant.

Em breve síntese, para (KNUTH, 1996a), a informação é algo mais intangível ou abstrato do que o dado. Ou melhor, o dado seria um valor de representação, numa linguagem formal, de uma informação. Tal representação pode ser aproximada ou exata dessa informação. Essa distinção é reafirmada em (KNUTH, 1996b):

²Dados crus: raw data

A particular representation of an algorithm is called program, just as we use the word “data” to stand for a particular representation of “information”.

Importante notar que há uma distinção implícita entre o conceito de valor e quantidade. Uma quantidade é um valor, mas o contrário não é necessariamente verdade. Um valor booleano não é uma quantidade, por exemplo.

Não será difícil perceber que a definição acima para dados não é muito diferente da para dados em (ISO/IEC/IEEE, 2010) (item 1.):

data. 1. a representation of facts, concepts, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automatic means (...) 2. collection of values assigned to base measures, derived measures and/or indicators. ISO/IEC 15939:2007, Systems and software engineering — Measurement process.3.4. 2. collection of values assigned to base measures, derived measures and/or indicators. ISO/IEC 25000:2005, Software Engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Guide to SQuaRE.4.9. 3. the representation forms of information dealt with by information systems and users thereof. ISO/IEC 10746-2:1996, Information technology — Open Distributed Processing — Reference Model: Foundations.3.2.6. 4. a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or communication, or processing. ISO/IEC 2382-1:1993, Information technology — Vocabulary — Part 1: Fundamental terms.01.01.02

Para (KNUTH, 1996b) a *ciência da computação* consistiria no campo de estudos dos “(...) phenomena surrounding computers (...), requiring descriptions and explanations; (...)”. Nesse sentido, como pode então ser estabelecida uma relação fenomenológica entre o conceito do dado? Na ciência da computação, o dado tem vários papéis — o mais primitivo deles é ser fornecido a programas de computadores para manipulação. (KNUTH, 1996b), por exemplo, percebeu que nomes em outras línguas para *Ciência da Computação* destacou que os nomes desse campo em francês, alemão e dinamarquês – respectivamente, *Informatique*, *Informatik* e *Data-logi* – eram orientados às entidades “manipuladas” pelos programas de computador, isto é: os dados, ou informações, em sentido amplo.

Em (SHANNON, 1948), a palavra dados (*data*) aparece somente três vezes: no sentido dos símbolos discretos que eram transmitidos em canal de comunicação – mesmo que os dados que chegassem ao receptor não fossem os mesmos enviados pelo emissor (devido a erros ou ruídos no canal de comunicação). Nesse contexto, os dados participam do processo de preservar a informação (dados de correção de

erros), mas sem necessariamente fazer parte da representação da informação em si. Os seja, os dados seriam entidades capazes de influir no processo de seleção de informações.

Na tentativa de procurar entender qual seria propriedade mais objetiva da informação, (NYQUIST, 1924), (HARTLEY, 1928) e (SHANNON, 1948) procuraram excluir os fatores relativos a dependência psicológicas.

No clássico *Go To Statement Considered Harmful* de (DIJKSTRA, 1968), há uma discussão sobre a diferença entre programa e processo (computacional):

My first remark is that, although the programmer's activity ends when he has constructed a correct program, the process taking place under control of his program is the true subject matter of his activity, for it is this process that has to accomplish the desired effect; it is this process that in its dynamic behavior has to satisfy the desired specifications. Yet, once the program has been made, the "making" of the corresponding process is delegated to the machine.

My second remark is that our intellectual powers are rather geared to master static relations and that our powers to visualize processes evolving in time are relatively poorly developed. For that reason we should do (as wise programmers aware of our limitations) our utmost to shorten the conceptual gap between the static program and the dynamic process, to make the correspondence between the program (spread out in text space) and the process (spread out in time) as trivial as possible.

No trecho acima, (DIJKSTRA, 1968) defendeu a abolição das instruções **GOTO** de códigos em alto nível para aumentar sua legibilidade e compreensão. O que melhoraria a qualidade do programa tanto por questões conceituais e formais como também uma maneira de se prevenir processos computacionais erráticos. DIJKSTRA reconhece que nem sempre o que se quer dizer com um programa de computador é o que de fato está escrito, pois seres humanos são sujeitos a não ter completa capacidade de previsão cognitiva da totalidade das sequência computacional, que é determinada pelo programa efetivamente escrito — ou seja, o programa efetivamente escrito pode não produzir “the desired effect”.

De acordo com (SHANNON, 1955), na década de 1950, um programa de computador poderia ser entendido da seguinte forma:

One of the most important technological advances of the last twenty years has been the development of large scale electronic computing machines. These computers are able of carrying out automatically and a speed of thousands of operations per second a long sequence of numerical operations. The series of instructions which tells the computer exactly what it should do is called the program.

Pode-se dizer que quando se está procurando um erro de programação existem técnicas para exercitar o código e testar os limites de abstração do programador ou do engenheiro de software — a escolha de dados de entrada é importante para essa tarefa, tem o poder de “modificar” a sequência de instruções e revelar uma falha³ (ou bug). Em (PRESSMAN, 2001, pp. 43–506), são ilustradas técnicas modernas e estratégias de testes de software. As técnicas de teste são voltadas para os caminhos possíveis de execução possíveis de um programa e erros de lógica de programação. Enquanto as estratégias de teste tratam o software como produto modularizado: um conjunto de programas interconectados. Uma anomalia⁴ é mais abrangente que o de falha, pois está relacionado a desvios ou erros da normalidade e dizem respeito à efetividade no domínio de aplicação – e não a erros de software ou hardware.

A seguir segue uma proposta de como lidar com a relação entre dados e programas tendo em vista a execução de programas de computador.

Definição E.1.1. *Os conceitos de programa de computador, processo computacional e dados se relacionam da seguinte forma: um programa define as possibilidades de processos computacionais enquanto os dados de entrada restringem tais possibilidades.*

Cabe dizer que as definições acima não dependem da eficácia ou eficiência do programa ou das intenções reais do programador – o que sugere que tais definições não possuem dependência psicológica. Uma vez concebido, um programa será executado e os dados definirão a sequência de execução (ou conjunto de execuções) – mesmo que, por exemplo, que o programa não descreva corretamente o algoritmo.

Em (ARAUJO, 2005; MARICONDA, 2006) e (HELFERICH, 2006c, p. 140 e 141) é possível perceber que Newton e Galileu não perdiam de vista o termo mecânica. É pelo título da obra de Galileu *Discursos e demonstrações matemáticas acerca de duas novas ciências* (a saber, a *mecânica* e os *movimentos locais*); e no prefácio do *Principia* de Newton, a abaixo (WESTFALL e COHEN, 2002, p. 275 e 276):

Posto que os antigos (como nos informa Papus) julgavam a ciência da mecânica de máxima importância na investigação das coisas naturais, e que os modernos, rejeitando as formas substanciais e as qualidades

³falha: **fault** 1. a manifestation of an error in software. 2. an incorrect step, process, or data definition in a computer program. 3. a defect in a hardware device or component. Syn: bug (ISO/IEC/IEEE, 2010)

⁴anomalia: **anomaly** 1. condition that deviates from expectations, based on requirements specifications, design documents, user documents, or standards, or from someone’s perceptions or experiences. IEEE Std 1028-2008 IEEE Standard for Software Reviews and Audits.3.1. 2. anything observed in the documentation or operation of software or system that deviates from expectations based on previously verified software products, reference documents, or other sources of indicative behavior. IEEE Std 829-2008 IEEE Standard for Software and System Test Documentation.3.1.4 (ISO/IEC/IEEE, 2010)

ocultas, esforçaram-se por submeter os fenômenos da natureza às leis da matemática, cultivei a matemática, neste tratado na medida em que ela se relaciona com a filosofia. Os antigos consideravam a mecânica por um vértice duplo: como racional, procedendo com exatidão pela demonstração, e prática. À mecânica prática pertencem todas as artes manuais de que a mecânica extraiu seu nome. Mas como os artífices não trabalham com a perfeita exatidão, sucede à mecânica distinguir-se a tal ponto da geometria que o que é perfeitamente recebe o nome de geométrico, e o menos preciso é chamado de mecânico. Os erros, contudo, não estão na arte, mas nos artífices. Aquele que trabalha com menos exatidão é um mecânico imperfeito; e se algum deles pudesse trabalhar com perfeita precisão, ser o mais perfeito de todos os mecânicos, pois a descrição das linhas retas e dos círculos em que se fundamenta a geometria pertence à mecânica. A geometria não nos ensina a traçar essas linhas, mas exige que sejam traçadas, pois requer que primeiro se ensine o aprendiz a descrevê-las com exatidão, para que ele possa penetrar na geometria, e depois mostra como se podem resolver problemas através dessas operações. Descrever retas e círculos são problemas, mas não problemas geométricos. A solução desses problemas é exigida da mecânica, e pela geometria se mostra o uso deles, quando assim solucionados; e a glória da geometria é poder, a partir desses poucos princípios trazidos de fora, produzir tantas coisas. A geometria, portanto, baseia-se na prática mecânica, e nada mais é do que a parte da mecânica universal que propõe e demonstra com exatidão a arte de medir. Contudo, visto que as artes manuais são predominantemente empregadas para mover corpos, sucede à geometria ser comumente referida à magnitude deles, e à mecânica, a seu movimento. Nesse sentido a mecânica racional é a ciência dos movimentos resultantes de toda e qualquer força e das forças exigidas para a produção de qualquer movimento, propostos e demonstrados com exatidão. (...)

Ou seja, Sir Isaac Newton percebia que sua teoria sobre o movimento dos corpos não deveria padecer das imperfeições humanas, ao mesmo tempo deveria ter a perfeição dos entes geométricos para que a exatidão fosse alcançada. O que poderia ser comparável⁵ analogamente a uma descrição independente de fatores psicológicos – como em (NYQUIST, 1924), (HARTLEY, 1928) e (SHANNON, 1948).

O mais importante observado até este momento é essa tática de se encontrar definições e relações esvaziadas de fatores psicológicos.

⁵comparável, mas com risco de incorrer em anacronismo.

E.2 Super-entropia

Definição E.2.1 (super-entropia). *Sejam $T = \sum_i t_i$, $f_i = t_i/T$ e $G = \sqrt[T]{\prod_i t_i^{t_i}}$. \mathcal{J} é a super-entropia em função de \dots, t_i, \dots então:*

$$\mathcal{J} = -\log_s G \quad (\text{E.2.1})$$

Teorema E.2.1 (super-entropia). *Seja \mathcal{J} uma super-entropia em função de \dots, t_i, \dots então:*

$$\mathcal{J} = -\log_s T - \sum_i f_i \log_s f_i \quad (\text{E.2.2})$$

Demonstração. $G = \sqrt[T]{\prod_i t_i^{t_i}}$

$$G = \prod_i t_i^{t_i/T}$$

$$G = \prod_i t_i^{f_i}$$

$$\log_s G = \log_s \prod_i t_i^{f_i}$$

$$\log_s G = \sum_i \log_s t_i^{f_i}$$

$$\log_s G = \sum_i f_i \log_s t_i$$

$$\log_s G = \sum_i f_i \log_s f_i T$$

$$\log_s G = \sum_i f_i (\log_s f_i + \log_s T)$$

$$\log_s G = \sum_i f_i (\log_s f_i + \log_s T)$$

$$\log_s G = \sum_i f_i \log_s f_i + f_i \log_s T$$

$$\log_s G = \sum_i f_i \log_s f_i + \sum_i f_i \log_s T$$

$$\log_s G = \sum_i f_i \log_s f_i + \log_s T$$

$$\log_s G = \log_s T + \sum_i f_i \log_s f_i$$

$$-\log_s G = -\log_s T - \sum_i f_i \log_s f_i$$

□

Teorema E.2.2 (super-entropia). *Seja \mathcal{J} uma super-entropia em função de \dots, t_i, \dots , $T = 1$ e $s > 1$ então \mathcal{J} é equivalente a entropia de Shannon.*

Demonstração. $\mathcal{J} = -\log_s T - \sum_i f_i \log_s f_i$

$$\mathcal{J} = -\log_s T - \frac{1}{\log_2 s} \sum_i f_i \log_2 f_i$$

$$\mathcal{J} = -\log_s T - k \sum_i f_i \log_2 f_i$$

$$\mathcal{J} = -\log_s 1 - k \sum_i f_i \log_2 f_i$$

$$\mathcal{J} = -k \sum_i f_i \log_2 f_i$$

□