Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia

AN AESTHETIC METRIC FOR MULTIPLAYER TURN-BASED GAMES

Eduardo Freitas Mangeli de Brito

Dissertação de Mestrado 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 Mestre em Engenharia de Sistemas e Computação.

Orientador: Geraldo Bonorino Xexéo

Rio de Janeiro
Fevereiro de 2016

AN AESTHETIC METRIC FOR MULTIPLAYER TURN-BASED GAMES

Eduardo Freitas Mangeli de Brito

DISSERTAÇÃO 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 MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

____
Prof. Geraldo Bonorino Xexéo, D.Sc.

____
Prof. Jano Moreira de Souza, Ph.D.

____
Profª. Renata Mendes de Araújo, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
FEVEREIRO DE 2016

Brito, Eduardo Freitas Mangeli de

An Aesthetic Metric for Multiplayer Turn-based Games/Eduardo Freitas Mangeli de Brito. – Rio de Janeiro: UFRJ/COPPE, 2016.

XI, 57 p.: il.; 29, 7cm.

Orientador: Geraldo Bonorino Xexéo

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2016.

Bibliography: p. 54 – 57.

1. games. 2. game metric. 3. quality criteria. 4. quality evaluation. I. Xexéo, Geraldo Bonorino. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

To my wife Janaina, the
lighthouse of my life.

Acknowledgments

I would like to start thanking the faculty members, the staff, and my colleagues from PESC, without whom this work would not be possible.

I worked with many brilliant people during my time in PESC. We shared ideas, joy and stress. They helped me with their criticism and encouragement, and I learned a lot in our time together. So, I want to thank everyone that gave some idea that enriched my rationalization of the problem addressed in this dissertation, or showed that I was wrong and saved me from lost my mind in the trials. Among all of them, I would like to especially thank Luís Fernando Oliveira, Bruno Adam Osiek, Ricardo Oliveira de Barros, and Luiz Fernando C. Tomaz.

My advisor always confirmed, through the course of my work in this dissertation, his ability to mentoring me. He fought my ideas while, at the same time, provided material to feed them. He drove my attention to subjects and analytical tools that I marginally knew, or never had heard about. This work was only feasible because of his help. Thanks, Geraldo Xexéo.

The examination board deserves all my gratitude for their questions and suggestions that made this work clearer and will support me in my future works.

I would also like to thank CAPES—Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—for the scholarship as well as the COPPE/UFRJ Business Incubator and SEBRAE for have kindly provided the data from Desafio Sebrae Business Game (Sebrae Challenge).

And finally, I need to say that I am really glad for the support of everyone who helped this work in any way.

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

UMA MÉTRICA ESTÉTICA PARA JOGOS MULTIJOGADOR BASEADOS EM TURNOS

Eduardo Freitas Mangeli de Brito

Fevereiro/2016

Orientador: Geraldo Bonorino Xexéo

Programa: Engenharia de Sistemas e Computação

Esse trabalho pretende apresentar uma métrica fuzzy para critérios estéticos de jogos multijogadores em turnos. As métricas previamente estabelecidas drama, mudança na liderança e incerteza foram generalizadas para sua adequação ao modelo de jogo genérico introduzido nesse trabalho. Então, novas interpretações matemáticas da análise comportamental inicial foram desenvolvidas. As métricas resultantes para cada um dos critérios foram agregadas em uma métrica fuzzy que foi aplicada a dados do Desafio Sebrae e do Campeonato Brasileiro de Futebol. Os resultados foram validados em comparação a avaliação de juízes humanos.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

AN AESTHETIC METRIC FOR MULTIPLAYER TURN-BASED GAMES

Eduardo Freitas Mangeli de Brito

February/2016

Advisor: Geraldo Bonorino Xexéo

Department: Systems Engineering and Computer Science

This work intends the development of a fuzzy aesthetic metric suitable for multiplayer turn-based games. The previously established aesthetic criteria drama, lead change and uncertainty were generalized to suit the generic game model introduced in this work. Therefore, new mathematical interpretations for the original behavior analysis were developed. The resulting metrics for each individual criterion were aggregated in a fuzzy aesthetic metric that measures the game aesthetic from data of the game execution. The novel metric was applied to data from the Desafio Sebrae and the Brazilian Football Championship. The results were validated against human judge's evaluation.

Contents

List of Figures **x**

List of Tables **xi**

1 Introduction **1**

1.1 Motivation . . . 1
1.2 Reaserch Problem . . . 2
1.3 Scope and key terms definition . . . 2
1.4 Methodology . . . 3
1.5 Document structure . . . 5

2 Background Knowledge **6**

2.1 Games as Systems . . . 6
2.1.1 MDA Framework . . . 7
2.1.2 Game Quality and Attractiveness . . . 9
2.2 Aesthetics Metrics by Dynamics Measures . . . 11
2.2.1 Move History Representation . . . 12
2.2.2 Drama . . . 13
2.2.3 Uncertainty . . . 14
2.2.4 Lead Change . . . 16
2.3 Fuzzy Logic . . . 17

3 Proposal **20**

3.1 Game model . . . 20
3.2 Dynamics representation . . . 21
3.3 Aesthetic Metrics . . . 22
3.3.1 Drama . . . 22
3.3.1.1 Drama by Points . . . 23
3.3.1.2 Drama by Positions . . . 23
3.3.1.3 Drama by Paths . . . 24
3.3.2 Lead Change . . . 25
3.3.3 Uncertainty . . . 26

3.3.3.1 Uncertainty by entropy 26
3.3.3.2 Uncertainty by probability distributions distance . . 27
3.3.3.3 Probabilities of winning 29
3.4 A Fuzzy Aesthetic Metric 31

4 Implementation 33
4.1 Evaluation Framework 33
4.2 Data 37
4.2.1 Brazilian Football Championship 37
4.2.2 SEBRAE Challenge 38
4.2.2.1 The Game 38
4.2.2.2 Dataset 39
4.2.2.3 Changes in the Game 40
4.3 Graphical User Interface for Sebrae Challenge 40
4.4 Metrics Validation 42
4.4.1 Drama Validation 43
4.4.2 Uncertainty Validation 46
4.4.3 Lead Change Validation 48
4.5 The Overall Aesthetic Metric 49
4.5.1 Overall Aesthetic Metric Validation 50

5 Conclusion and Future Work 52
5.1 Conclusion 52
5.2 Future Work 53

Bibliography 54

List of Figures

2.1 MDA diagram 8
2.2 Quality categories for serious business games by VARGAS *et al.* (2014). 11
2.3 Browne's Move History of a game between blue player and red player 13

3.1 Generic Multiplayer Game Model 21
3.2 Lead changes throughout game history by scores examples 22
3.3 Maximum Drama Path (MDP) examples 25
3.4 Comparison between the behavior of uncertainty metrics. 29

4.1 Evaluation Framework 34
4.2 Model implementation classes 35
4.3 Measures implementation classes 36
4.4 Match Viewer 41
4.5 Group Viewer 42
4.6 Drama Measures by Championship Edition 44
4.7 MDP versus winner path comparison - Brazilian football championship editions 2005 and 2009 44
4.8 Uncertainty Measures by Championship Edition 47
4.9 Lead Change Measures by Championship Edition 49
4.10 Overall Aesthetic Evolution in Sebrae Challenge 51

List of Tables

1.1 Preliminary considerations of the research questions 4

2.1 Järvinen's game elements taxonomy 7
2.2 MDA components description 7
2.3 Thompson's game "qualities" 10
2.4 Malaby's sources of contingency 15
2.5 Costikyan's sources of uncertainty 15

4.1 Optional parameters for constructor methods of the measure classes . 35
4.2 Summary of Brazilian National Football Major League Championship Recent History 37
4.3 Matches' distribution in sample 39
4.4 Subset 'Others' Detail 39
4.5 Drama Measures Values by Championship Edition 43
4.6 Judges Evaluation of Drama 45
4.7 Kendall tau ranks correlation between judge's rank and Drama measures 45
4.8 Uncertainty Measures Values by Championship Edition 46
4.9 Kendall tau ranks correlation between judges' rank and Uncertainty measures 47
4.10 Lead Change Values by Championship Edition 48
4.11 Kendall tau ranks correlation between judges' rank and Lead Change measures 49
4.12 Overall Aesthetic Values by Brazilian Football Championship Edition 50
4.13 Kendall tau ranks correlation between judge's rank and Overall Aesthetic Metric 50

Chapter 1

Introduction

This chapter exposes the research motivation and states its research problem. The scope definition is presented and the methodology used is explained, followed by the document structure.

1.1 Motivation

We live surrounded by games. More than a childish thing, games are an artifact present in our adult and elderly life. Today, housewives play in their cell phones, or tablets; stakeholders are former, or even current, video game players; and gamification plays a significant part in many companies business process. The predicted global video game market total revenue in 2015 is $91.5 billion dollars (WARMAN, 2015). Based on these examples, even being non-exhaustive, one cannot deny current ubiquity of the game culture.

In that scenario, industry technicians, independent game developers, scholars and professionals of different science fields cope with the challenge of understanding and improve the game development process. This effort includes a quest for a better comprehension of the human interaction with games as well as the formalization of various aspects of those artifacts. Because of that, "game scholars [...] are now being asked to illuminate policy makers, interested media, and the public at large on what games mean", as denoted by MALABY (2007).

Furthermore, the relation between the development of computational techniques based on games, especially on chess, and the progress in solving other classes of problems like medical diagnosis programs, automated reasoning, automated theorem proving and oil searching can be perceived through the course of history (NEWBORN, 2012). Since the SHANNON (1950) proposal of building a computer program capable of playing a reasonable chess match, taking into account the development of machines specially designed to play (CAMPBELL *et al.*, 2002), games have been acting as tools to increase the computer's ability of automatic problem-solving.

This dissertation explores games' aesthetic concepts established by previous works and investigates their ability to address multiplayer turn-based games. Thus, it hopes to contribute towards the generalization of some game aesthetic metrics besides to present a novel fuzzy game aesthetic metric.

1.2 Reaserch Problem

The problem addressed by the research is if there are aesthetic criteria, previously established to more strict classes of games and measurable by game history analysis, that could compose an aesthetic metric for multiplayer turn-base games. So, the research problem is defined as follow:

- If there exist aesthetic criteria, measurable by the game history analysis, that can be combined in a fuzzy aesthetic metric for multiplayer turn-based games.

1.3 Scope and key terms definition

In this work, numerous terms from games' literature are used with a meaning other than the common one in daily use. Hence, to avoid semantic misunderstood, some key terms definition are presented, besides the scope circumscription.

The word **game** is used, in this dissertation, as a type of "ludic artifact", as defined by KOSTER (2013) so in the sense of the conceptual game, and its infinite set of possible **matches**. Thus, players engage directly in matches of a certain type of game.

Turn-based games are those in which the matches are composed of turns, and in each turn all players make their moves. Furthermore, the concept of a **move** concerns the players' choices of actions, among those inherent to the game and possible at the time. The players' moves, within a turn, can be made simultaneously or one at a time. Therefore, a turn can be seen as a game segment that allow players' states evaluation at its end.

SHANNON (1950) defined a position P as a game state that can be evaluated by a hypothetical **evaluation function** $f(P)$ in order to determine if it leads to a win, draw or lose for a certain player. With chess as a subject, he also showed that in practical situations, and considering an interesting game, an evaluation function shouldn't be capable of determining the final result after each move or turn, but it should be able of comparing players' campaigns evolution through the match up to the game goal.

The **game history** (BROWNE, 2008) presents the players' campaign evolution, after each game turn, based in an evaluation function's output. Thus, it directly shows a facet of the game dynamics and is a source to other facets extraction.

The term **aesthetic** is used, in this work, in the sense of an emotional response from the player with positive valence. So, with a meaning similar to that stated by BATEMAN (2015) as "a specific value judgment (or set of value judgments) that valorizes some forms of play and that may or may not denigrate other forms of play", and consistent with the MDA framework (HUNICKE *et al.*, 2004).

The term **fuzzy logic** is used in its wider sense that embraces, besides the narrow meaning of the fuzzy logic as an extension and generalization of multivalued logics, the theory of fuzzy sets as pointed by KLIR & YUAN (1995).

The scope of this dissertation comprises the development of metrics, from previously established game's aesthetic criteria, suitable to a generic model of multiplayer turn-based games. The developed metrics are intended to measure the game aesthetic from game history. Thus, this work proposes extract aesthetic evaluation from game dynamics. The ideas exposed have game designers and developers as their primary audience. Notwithstanding, those who are interested in game quality or automatic aesthetic evaluation could take advantage of the concepts presented.

1.4 Methodology

The research problem specification led to the definition of the following research questions:

- Which theoretical framework should be used?
- Which aesthetic criteria should be chosen?
- How the chosen criteria must be adapted, targeting multiplayer games?
- How those adapted criteria should be combined into a fuzzy aesthetic metric?
- How to test the new metric?

These questions bring some preliminary considerations that are presented in Table 1.1.

Therefore, the research was developed seeking to address these questions and considerations. Thus, this dissertation accepts the view of games as information systems. Starting from that rationalization, a quality model that comprises aesthetic criteria could be built.

The aesthetic criteria **Drama**, **Leader Change** and **Uncertainty** were selected among others candidates due their relevance in the literature about games, their foundational nature to the concept of what a game is, and other particular features that are detailed in chapter 2. So, new mathematical interpretations of the original

Table 1.1: Preliminary considerations of the research questions

theoretical framework	The chosen theoretical framework should be capable of sustaining the view of games as structured complex artifacts. It should also comprise tools to understand and to represent the game model used in the proposals implementation.
choosing criteria	The capacity of aesthetic evaluation from game history is crucial to the criteria chosen, in face of intended aesthetic evaluation extraction from game execution data.
criteria generalization	The adaptation of the chosen criteria must regard their applicability to games that fit the multiplayer game model used in this work, besides the maintenance of the originally proposed behavior analysis.
fuzzy agregation	The criteria analysis must lead toward specific metrics that should be combined in a fuzzy aesthetic metric able to quantify a facet of the game quality.
metric validation	The resulting fuzzy aesthetic metric must be validated against human judges' evaluations.

behavior analysis were developed in order to adapt these criteria to turn-based multiplayer games.

The developed metrics were combined by a fuzzy aggregation operation to build an aesthetic metric for the generic game model proposed in this work.

Data from the Sebrae Challenge and the Brazilian Football Championship were used to evaluate and validate the proposals. The datasets were prepared to fit the multiplayer game model developed targeting generalization. The Sebrae Challenge dataset was stored in a PostgreSQL[1] database system in accordance to the relational model shown in the section 3.1. The Brazilian Football Championship data was stored in text files using JSON (JavaScript Object Notation[2]), one file for each annual edition.

An Evaluation Framework (section 4.1), developed in the branch version 2.7 of the Python[3] programming language, was used to work with the data and run the computation of the developed metrics.

Two Graphical User Interfaces (section 4.3) were built to deal with various options of data visualization from Sebrae Challenge, and to verify the implemented code of the developed metrics.

The metrics were applied to the data of Brazilian Football Championship. Pictorial representations of the games behavior were presented to human judges so they

[1]http://www.postgresql.org/
[2]http://www.json.org/
[3]https://www.python.org/

could analyze the facets of the game dynamics that the metrics intend to measure (section 4.4).

The judges' ranks were aggregated in accordance with the Schulze Method (SCHULZE, 2003) and, after that, were compared to the computed measures in order to verify the correlation between the judges' preferences and the rank stated by the proposed metrics. This method enables the election of metrics that are combined, by a fuzzy aggregation operation, in an overall aesthetic metric that reflects a facet of the game quality.

The resulting overall aesthetic metric was applied to data of Sebrae Challenge in order to analyse the evolution of aesthetic levels, measured by it, through different series in the same championship and to verify if the changes introduced in the game led to an increase in those levels as intended by the game designers.

1.5 Document structure

This work comprises four chapters other than this.

In chapter 2 formal concepts needed to the development of the ideas of this dissertation are presented. So, games are formally described as systems, a quality model that comprises aesthetic criteria is presented, previous interpretations and representations of desirable game behaviors are shown as well as a brief explanation of some fuzzy logic concepts.

The chapter 3 brings the contributions developed in order to address the research problem. Thereby, the generic game model for multiplayer turn-based games, a pictorial representation of game dynamics, novel aesthetic metrics, and the definition of an overall aesthetic metric are introduced.

The chapter 4 details the implementation of the ideas developed in this work. It shows a description of the evaluation framework and graphical user interfaces that were built. Also, it describes the data used in the metrics evaluation and present the validation of all developed metrics.

Finally, chapter 5 comprises the conclusion and future work indications.

Chapter 2

Background Knowledge

This chapter introduces some background knowledge needed to concepts development that take place at chapter 3. Hence, it formally defines games as systems and, from that rationale, establishes a quality model comprising aesthetic criteria. Furthermore, the previously established criteria Drama, Uncertainty and Lead Change are presented, likewise a brief exposition of the fuzzy logic.

2.1 Games as Systems

As pointed by SALEN & ZIMMERMAN (2004), games are systems and, even further, about this there is little disagreement. Nevertheless, a system is a broad concept and, for instance, can be seen as a set, like "a set of connected items or devices that operate together", or as a method, like "a way of doing things"[1]. Therefore, this concept must be refined to allow a practical use. Thus, there are typical tools used to achieve this refinement like an application domain, the types of the components, the types of relations among components, or the purpose of the system *per se* (HIRSCHHEIM *et al.*, 1995).

JÄRVINEN (2008) introduced a classification of game elements that can be seen as an illustration of the prior assertions. His work classifies all game elements in three classes and nine categories as shown in Table 2.1. So, games can be seen as a system that comprises a set of components, including the players and information.

Other very common type of system is information system. These systems have been interesting scholars and researchers for a long time and have a vast set of tools to analyze and describe it. Formally, they can be defined as a "set of interrelated components that collect (or retrieve), process, store, and distribute information" (LAUDON *et al.*, 2011), or understood as an "interactive web of people and machines which together constitute a social system" (GRUNDY, 1997). The terms used in

[1] Cambridge English Dictionary at http://dictionary.cambridge.org/us/dictionary/english/

Table 2.1: Järvinen's game elements taxonomy

Systemic Elements	components environment
Compound Elements	ruleset game mechanics theme interface information
Behavioral Elements	players contexts

these definitions do not have narrow meanings. Thus, the term "machines", for instance, should not be associated only with physical machines, such as computers, instead, it also should be associated with abstract or mechanical machines, such as algorithms or game rules.

From those grounds, one can clearly make an association between games and information systems, as SALEN & ZIMMERMAN (2004) already did. However, seems to be a valid point remembering the fact that information systems can be built without computers, solely using pen and paper, for instance. In the same way, the association between information systems and games is not restricted only to video games or computer games, but it comprises games in general, like board games, card games, and even sports or gambling.

This work assumes that premise, so it views games as systems and makes use of a quality model and other tools suitable for information systems and, as a consequence, for games.

2.1.1 MDA Framework

MDA is an acronym that stands for Mechanics, Dynamics and Aesthetic besides identifying a framework proposed by HUNICKE *et al.* (2004). This framework, in adition to establishing a classification of the game components, sets up a causal relation between these classes that is foundational to this work. The MDA framework was built intending to comprise concepts that help designers, researchers and scholars perform the decomposition of games into coherent and understandable parts.

Table 2.2: MDA components description

Mechanics	The particular components of the games, as tokens, theme, characters, rules and so on.
Dynamics	The run-time behavior of the mechanics.
Aesthetics	Emotional responses evoked in the players.

The game elements are classified into three distinct components that are briefly describes in Table 2.2.

As one can see, the mechanics are the only components directly accessible to game designers or developers. This concept comprises the items created, changed or manipulated by who build the game. Hence, an ontology of game mechanics can be seen as a boilerplate to build games. On the other hand, dynamics and aesthetics components are not accessible to designers. Game dynamics originate from the player's interaction with mechanics. Thereby, if the rules allow the exchanging of resources by players in order to reach a common objective, this can lead towards a collaborative dynamic. As for aesthetics, they emerge from dynamics, as another layer of components, and are the emotional responses of players.

The view of the classes as superposed layers of components in which mechanics, the support layer, act as a base for dynamics that, in turn, originate aesthetics is proper to the designer's perspective. On the contrary, aesthetics play the major part in the player's perspective, it "is born out in observable dynamics and eventually, operable mechanics" (HUNICKE *et al.*, 2004). Figure 2.1 shows the components of the MDA framework as well as their relations under designer's and player's perspectives.

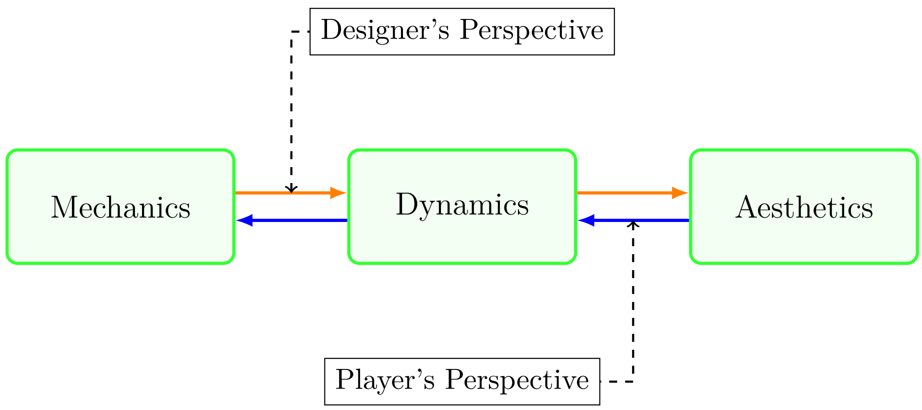


Figure 2.1: MDA diagram

Each one of the MDA framework components comprises a model that describe it. Therefore, as noted earlier, a sub ontology of game mechanics can describe the mechanics component for a particular game. In the same way, models can be built to describe dynamics and aesthetics components. Targeting a common terminology that could be shared by who is interested in development or analysis of games, HUNICKE *et al.* proposed the non-exhaustive taxonomy of aesthetics listed below.

Sensation Game as sense-pleasure.
Fantasy Game as make-believe.
Narrative Game as drama.
Challenge Game as obstacle course.

Fellowship Game as social framework.
Discovery Game as uncharted territory.
Expression Games as self-discovery.
Submission Game as pastime. (HUNICKE *et al.*, 2004)

This work uses the propositions of the MDA framework in the sense of the segmentation of game elements into subsets. In particular, this work uses the aesthetic meaning embedded in the aesthetic component of the MDA framework. However, as was already stated, this meaning is more clearly pointed by BATEMAN (2015) as "a specific value judgment (or set of value judgments) that valorizes some forms of play and that may or may not denigrate other forms of play".

2.1.2 Game Quality and Attractiveness

This section presents a formalization of the quality model that compound the theoretical framework of this work. As a result, the logical connection between the quality of a game and its attractiveness is established.

The American Society for Quality defines *Quality* as "the characteristics of a product or service that bear on its ability to satisfy stated of implied needs" and as "free of deficiencies" (AMERICAN SOCIETY FOR QUALITY, 2015). This is in accordance with the definition of quality provided by the most cited authors in this field, Juran and De Feo, that define quality as "features of products which meet customer needs and thereby provide customer sotisfaction" and "freedom from deficiencies". A simplier, yet to general, definition is also "fitness for use" (JURAN & DE FEO, 2010), which broadly describes both definitions.

In game literature, many works clearly lengthily discuss game attractiveness aspects without directly fully discussing the technical concept of quality. However, some of them are cornerstones of the contemporary field of game quality analysis.

THOMPSON (2000) pointed four "qualities" that good "abstract strategy games" must have. Table 2.3 shows these desirable game characteristics as well as a short description of them. The kind of game that he was dealing with have perfect information, so chance does not play a part, and are theme independent. Although coping with a very specific type of game, and despite his lack of formality, his ideas became a foundation for future analysis of game attractiveness.

The automatic game evaluation was addressed by ALTHOFER (2003). He proposed a set of criteria to evaluate the "interestingness" of a game using a computer program, aiming the computer-aided invention of games. All criteria in the proposed set were chosen heaving in mind that they should be evaluated automatically from statistic analysis of game records but some of them were related to computational performance, since computer games are his object of analysis.

Table 2.3: Thompson's game "qualities"

Depth	The game must allow different levels of expertise.
Clarity	The game must be understandable without so much effort.
Drama	It should be possible for a player to recover from a weaker position and still win.
Decisiveness	It should be possible for one player to achieve an advantage from which the other player cannot recover.

SCHELL (2014) discuss 113 lenses that can be used to guide game development. Each of such lenses provide a definition of explanation of use and some questions that can be answered, with different levels of subjectiveness. The first lens, for example, is "The Lens of Emotion", which aims to "make sure you create the right emotions", and one of the questions is "What emotions would I like my player to experience".

BROWNE (2008) also discuss various different quality characteristics, with very objective metrics that can be measured, since he aims the automatic generation of games. However, his vision of quality as a sub-set of aesthetic measurements invert the concept of quality with one of is its perspectives in a hierarchical model, as it would be accepted by most quality researches.

Many other authors also discuss quality, including the use of other names such as heuristics for game evaluation or assessment. Some authors discuss *game analysis*, which include quality but also refers to social impacts an other Game Studies concerns. FERNÁNDEZ-VARA (2014) is a Game Studies text that discuss "Formal Qualities" of a game. Also, ZAGAL & BRUCKMAN (2008) and CHAN & YUEN (2008) propose antologies, such as the Game Ontology Project[2] that can be used to guide a game quality study.

Quality is a subject for standardization by ISO, IEC and other standardization organizations, such as IEEE. Currently, the most generic quality related to games is ISO/IEC (2011), which describes a quality model for systems and software that can be applied directed to games, although it hardly describes common game evaluation heuristics such as "enjoyment".

VARGAS *et al.* (2014) provide a systemic review of Serious Game Quality that found 112 papers on the subject. They matched all quality concepts to the ISO/IEC 25010 standard, which resulted in a the model that is shown in Figure 2.2.

This work already stated that games can be viewed as systems and, now, a more comprehensive definition can be used. BRUNNSTRÖM *et al.* (2013), for instance, defined an application, as "a software and/or hardware that enables usage and interaction by a user for a given purpose". Their definition also pointed that

[2] http://www.gameontology.com

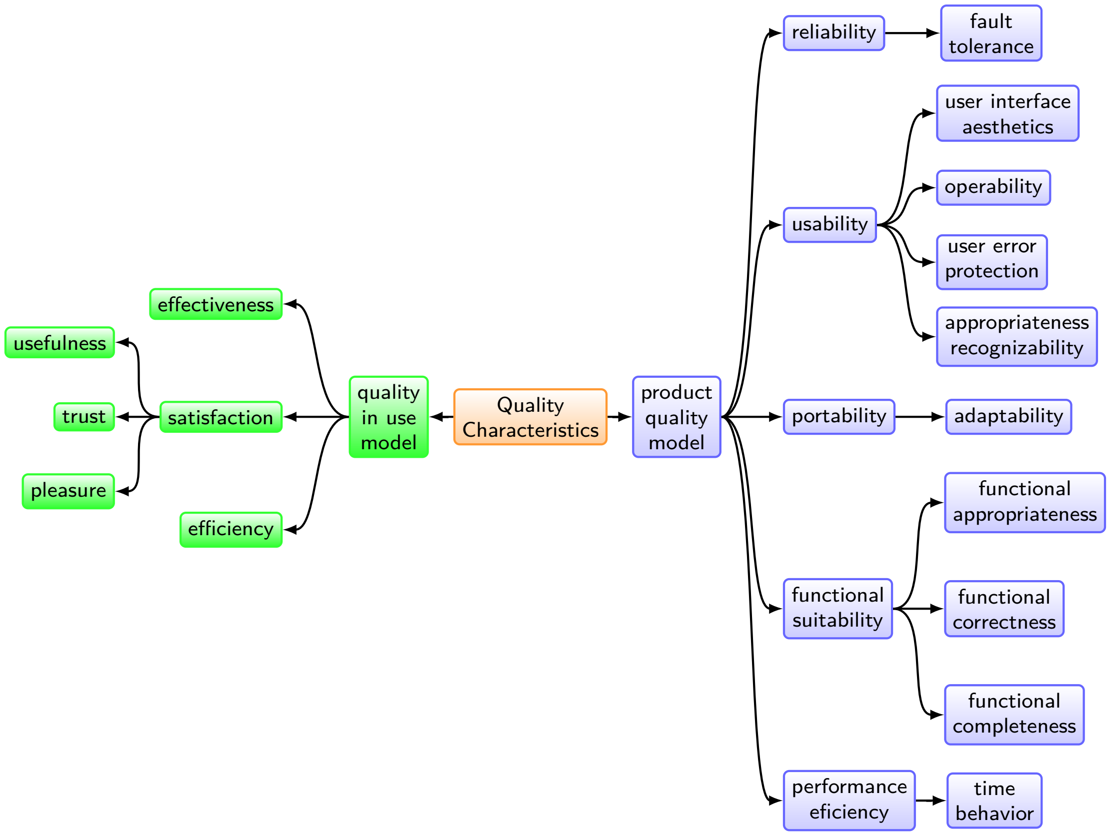


Figure 2.2: Quality categories for serious business games by VARGAS *et al.* (2014).

"such purpose may include entertainment or information retrieval, or other". So, clearly, games also can be viewed as an application. BRUNNSTRÖM *et al.* (2013), in addition, set the concept of Quality of Experience (QoE) as "the degree of delight or annoyance of the user of an application or service" that emerges "from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state".

MOLLER *et al.* (2013) use the QoE concept as a foundation to build a taxonomy of factors, aspects, and features that are relevant to games. This taxonomy set an organized way of view games and analyze their quality. They consider two aesthetic aspects—system personality and appeal—as quality features.

The concept of aesthetic used in this work fits in the characteristic *satisfaction* as well as in the sub-characteristic *pleasure* of VARGAS *et al.* (2014) classification and it is also in accordance with the aesthetic aspects categorization of MOLLER *et al.* (2013) taxonomy.

2.2 Aesthetics Metrics by Dynamics Measures

YANNAKAKIS *et al.* (2013) present a classification for player modeling approaches into model-based and model-free approaches. A model-based, or top-down, approach starts from some premises about the human behavior and may be inspired by theories of emotions or by "a general theoretical framework of behavioral analysis and/or cognitive modeling", like KOSTER's theory of 'fun'. A model-free, or bottom-up, approach makes no assumptions about the player behavior and, generally, uses methods from machine learning to extract behavioral playing patterns from game data. Moreover, the authors pointed that the majority of the existing works is a hybrid of the two types of approaches and present characteristics from both.

This work does not intend to propose or use a model that entirely explain the aesthetic component of the MDA framework or completely translate the quality aspects of a game. Instead, it intends to use some of the quality criteria previously established in the literature, adapting them for the proposed game model and interpreting their behavior in game data. Thus, this work uses a player model that comprises player behavior regardless their cultural context or other restrictions. In order to do that, a set of quality criteria linked with foundational aspects of games was chosen to overpass the boundaries of age, social class *et cetera*. Moreover, the chosen aesthetic metrics, besides their expected universal adequacy, should also be able to extract desirable game behavior from game data.

This section describes the criteria chosen to compose the quality metric—Drama, Lead Change and Uncertainty—besides detailing previous approaches to measuring them. It also presents a pictorial representation of player's score evolution—Move History—that directly shows a facet of the game dynamics and is a source for extraction of other facets associated with the game aesthetics.

2.2.1 Move History Representation

BROWNE (2008) has presented a graphic representation of player's score evolution, the *Move History*. These graphs reflect the score's progression during a match, move by move, and concern combinatorial games that were defined by him as:

- *Discrete*: Players alternate making moves (no simultaneous moves).
- *Finite*: Produces a well-defined outcome after a finite number of moves.
- *Deterministic*: Chance plays no part.
- *Perfect information*: No hidden information.(BROWNE, 2008)

Figure 2.3 shows a *Move History* of a game between two players, blue and red. The vertical axis presents an estimation of the evaluation function output associated to the game. Negative values in this axis are an original feature of Browne's *Move History* to cope with the representation of multiple information in the same graph, like the score gap to the leader. The dots symbolize players' moves. So, the moves

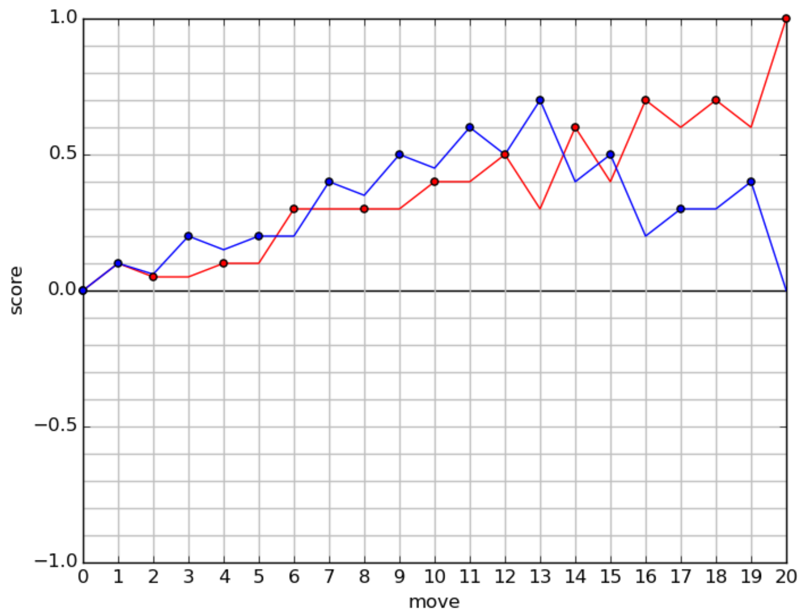


Figure 2.3: Browne's Move History of a game between blue player and red player

identified by even numbers are red player's moves, while the odd moves are from the blue player. The player's score decreases or stays unchanged after a move of the other player.

2.2.2 Drama

THOMPSON (2000) proposes that a game has *Drama* if "it should be possible for a player to recover from a weaker position and still win the game". He uses chess as an evidence of this rationale, arguing that a chess match between good players rarely unfolds till a checkmate, one of them resigns when it becomes clear that he can not win so when the *Drama* ceases. BROWNE & MAIRE (2010), rephrasing him, affirmed that "there should be at least the hope of recovery from bad positions".

For two players, BROWNE (2008) calculated *Drama* A_{drav} as:

$$A_{drav} = \sum_{g=1}^{G} \frac{\sum_{n=M_{gr}+1}^{M_g-1} E_w(m_n) < E_l(m_n) \begin{cases} \sqrt{E_l(m_n) - E_w(m_n)} \\ 0 \end{cases}}{\underset{[M_{gr}+1 \leq n \leq M_g - 1]}{count} \quad (E_w(m_n) < E_l(m_n))} / G \tag{2.1}$$

This average *Drama* formula takes the summation of all observed game instances *Drama* values divided by the observation quantity in sample G. Therefore, looking at the equation, one can describe *Drama* for each match as the summation of differences between the winner E_w and an eventual leading player E_l scores in those turns m_n when the winner player didn't lead, leaving out the turns with random moves M_{gr}. Random moves are a particular feature Browne's automatic analysis process, used in the initial game configuration, that is disregarded in this work since it is an artifact from his experiment setup.

Therefore, the *Drama* for a single match g is:

$$Drama_g = \frac{\sum_{n=1}^{M_g-1} E_w(m_n) < E_l(m_n) \begin{cases} \sqrt{E_l(m_n) - E_w(m_n)} \\ 0 \end{cases}}{\underset{[1 \leq n \leq M_g - 1]}{count} \quad (E_w(m_n) < E_l(m_n))} \tag{2.2}$$

Since when E_w is the leader in a move $E_l(m_n) - E_w(m_n) = 0$, one can use a simpler notation:

$$Drama_g = \frac{\sum_{n=1}^{M_g-1} \sqrt{E_l(m_n) - E_w(m_n)}}{|\{m|(E_w(m) < E_l(m))\}|} \tag{2.3}$$

2.2.3 Uncertainty

Regardless what formal game definition was the chosen one, the uncertainty is a fundamental aspect of the activity of play a game sustained by various authors. SHANNON (1950), for instance, recognized that if there was a method that could determine the outcome of a chess match analyzing any general position, this game would lose most of its interest. CAILLOIS (1961), in a more general way, stated that play "is also uncertain activity. Doubt must remain until the end, and hinges upon the denouement".

SALEN & ZIMMERMAN (2004) are quite straight when recognizing uncertainty as "a key component of meaningful play". They also argue that the outcome of a game is uncertain even without a "die roll or random algorithm", or "even though the game is a game of skill, not chance".

In spite of the fact that he personally prefers the term "contingency" describing the unpredictability of game outcomes, MALABY (2007) proposed a classification concerning the sources of contingency (Table 2.4) that was expanded by COSTIKYAN (2013) as a taxonomy of sources of uncertainty in games shown in Table 2.5.

DE KOVEN (2013), in the same rationale line of the foundational aspect of uncertainty in games, has asserted: "Imagine how incomplete you would feel if, before the game, you were already declared the winner. Imagine how purposeless the game would feel".

Table 2.4: Malaby's sources of contingency

Stochastic contingency	Randomness produced.
Social contingency	Uncertainty about the another's point of view.
Performative contingency	Related to the execution of actions that may fail or succeed.
Semiotic contingency	Unpredictability of the game outcomes meaning.

More recently, ABUHAMDEH *et al.* (2014) noticed that game outcome uncertainty "appeared to promote feelings of suspense, which lead to greater enjoyment of the process of engagement".

Table 2.5: Costikyan's sources of uncertainty

Performative Uncertainty	The uncertainty related to physical performance, like in first-person shooters and driving games.
Solver's Uncertainty	Uncertainty linked to puzzles solving.
Player Unpredictability	The source of uncertainty concerns actions or choices of the other players, like in combat games and most multiplayer games.
Randomness	The uncertainty arises from random elements of a game, like dice rolling or card draws.
Analytic Complexity	The outcome of the chosen action is uncertain due to the complexity of the game tree, like in the chess.
Hidden Information	The players know only part of the information about the game state, like poker.
Narrative Anticipation	The players can not know how the game story, or play arc, will unfold.
Development Anticipation	The uncertainty results from new features and expansions provided by the game's developers after the game launching.
Schedule Uncertainty	The game imposes a limit on the available resources in each game session. Thus, to re-engage in the game, the player must wait until new resources are available, leading him towards juggling with their use and the time needed to refilling their stocks.
Uncertainty of Perception	The difficulty of the player in perceiving the whole state of the game, even if there are not hidden information, produces the uncertainty.
Semiotic Uncertainty	The cultural meaning that emerges from the game outcome generates uncertainty. Some games intentionally create cultural meaning and others are made to manipulate it.

BATEMAN (2015) argues that the uncertainty aesthetic is applicable to the point of view of the game's attractiveness judgment. In fact, this author goes beyond and says that the uncertainty is the most suitable aesthetic value for being placed as a boundary around what is play, among the various aesthetic values described in his analysis.

But still, IIDA *et al.* (2004) were definitive setting up the relation between the uncertainty and the attractiveness of games with the assertion: "Interesting games are always uncertain until the last end of games". Therefore, the measure of the uncertainty $U_p(H)$ of a group of similar positions H, for a player p can be defined as the entropy of the probability of the possible p_i outcomes, as follow:

$$U_p(H) = -\sum_{i=1}^{k} p_i \cdot \ln(p_i) \tag{2.4}$$

This work does not disregard the importance of all types of uncertainty in games. However, it centers its analysis to the uncertainty concerning the final outcome of the match. Hence, the scope of this work comprises the uncertainty about who will win the match and how long this doubt is sustained. Hence, in a very similar sense of BROWNE's assertion about what uncertainty is: "the tendency for the outcome of the game to remain uncertain for as long as possible, [...]. The sooner a game's outcome is known the less interesting it becomes (especially for the losing player)".

2.2.4 Lead Change

Lead Change is a behaviour closely related to the uncertainty of the match outcome. ABUHAMDEH *et al.* (2014) argue that in a "closer" match, in which the score gap is tight, the uncertainty about the final outcome is higher. In some two player game, like chess, this behaviour can be perceived as a lead change after each ply—a half move or a movement of only one player—in accordance with the chosen evaluation function.

Sports are a type of game in which lead changes are frequently observed, especially in competitive sports for teams or individual players. Competitive teams sports, as pointed by CLAUSET *et al.* (2015), are those in which two teams compete to achieve intermediary goals in order to increase their score, and the one with the higher score at the end of the match is the winner. This definition was stated to contrast with individual sports like running, and judged sports like figure skating, but the rationale can be applied not only to sports teams but also to individual players that play against each other in sports like tennis, for instance, without losing any characteristic. One must note that these can be viewed as two player games, even when teams of players are involved.

CLAUSET *et al.* (2015), when dealing with competitive teams sports, as defined by them, also formally defined lead changes as "the times in a game when the lead changes" that "occurs whenever the score difference $X(t)$ returns to 0", and pointed that they are "often the most exciting" moments in a match.

BROWNE (2008) uses, in his analysis of aesthetic criteria, the Lead Change criterion that as defined as the number of times the lead changes in the intelligent moves—not random—of a match. The Average Lead Change metric defined by him is showed in Equation 2.5. The notation and the motivation of the random moves used by BROWNE are explained in details in subsection 2.2.2. However, for simplicity, the notation is briefly described again: M_{gi} is the number of intelligent moves in a match g, M_{gr} is the number of random moves in a match g, m_n is a move and m_{n-1} is the previous one, and G is the total number of game trials or matches in sample.

$$A_{lead} = \sum_{g=1}^{G} \frac{\sum_{n=M_{gr}+1}^{M_g} leader(m_n) \neq leader(m_{n-1}) \begin{cases} 1 \\ 0 \end{cases}}{M_{gi} - 1} / G \tag{2.5}$$

Hence, the *Lead Change* metric for a single match g, as defined by BROWNE for combinatorial games, can be described by Equation 2.6.

$$Lead\ Change_g = \frac{\sum_{n=M_{gr}+1}^{M_g} leader(m_n) \neq leader(m_{n-1}) \begin{cases} 1 \\ 0 \end{cases}}{M_{gi} - 1} \tag{2.6}$$

2.3 Fuzzy Logic

This work does not introduce any contribution to the fuzzy logic field. The concepts that are briefly explained here intend to provide a basic structure for producing an aesthetic metric that comprises different metrics developed addressing each aesthetic criterion. To accomplish this purpose, the concepts of fuzzy set, membership function, and an aggregation operation will be presented.

As KLIR & YUAN (1995), let's call "crispy" the traditional, or non-fuzzy, sets. Crispy sets can have their elements defined in two ways: by an *extension definition*, enumerating their elements, or by an *intensional definition*, where its elements share a common property. Another way to describe a crispy set is by its *characteristic function* that declares which element is in the set. Equation 2.7 shows the characteristic function for the set A that maps the elements of the universe of discourse X for a value of either zero or one, so $\mathcal{X}_A : X \mapsto \{0, 1\}$.

$$\mathcal{X}_A(x) = \begin{cases} 1 \; for \; x \in A \\ 0 \; for \; x \notin A \end{cases} \tag{2.7}$$

A *membership function* is a generalization of the *characteristic function* concept. Instead of returning a value equal to 0 or 1, a *membership function* assigns values within a range that denote the membership grade of the element in the set. The most common range for returned values is the interval of real numbers between zero and one, large values point to higher degrees of membership. The set defined by a *membership function* is a *fuzzy set*.

The Equation 2.8 presents an example of membership function, that defines the fuzzy set A. As one can see, μ_A maps the elements of the universe of discourse X to the range of real values $[0, 1]$. One must note that "each fuzzy set is completely and uniquely defined by one particular membership function" (KLIR & YUAN, 1995).

$$\mu_A : X \mapsto [0, 1] \tag{2.8}$$

Aggregation operations on fuzzy sets are defined as those operations that use several fuzzy sets as input, combining them in a desirable way, to produce a single fuzzy set (ROSS, 2009). These operations have the ability to produce meaningful expressions from a combination of distinct concepts (BĚLOHLÁVEK *et al.*, 2002). Let's take as an example the results of a personal computer's evaluation in three categories named as memory speed, graphics quality, and processor speed; each of them graded in the range $[0, 1]$. A proper *aggregation operation* would lead to a meaningful expression of the overall computer's performance by a single fuzzy set.

Formally, an *aggregation operation* on fuzzy sets as defined as follow:

$$h : [0, 1]^n \mapsto [0, 1] \; for \; n \geq 2 \tag{2.9}$$

Therefore, when an *aggregation operation* is applied to fuzzy sets $B, C, \ldots, D$ defined by the membership functions $\mu_B, \mu_C, \ldots, \mu_D$, in order to produce a single fuzzy set A, it operates over the membership grades of each one of theses sets for the elements x in the universe of discourse X as shown in Equation 2.10.

$$\mu_A(x) = h(\mu_B(x), \mu_C(x), \cdots, \mu_D(x)) \tag{2.10}$$

The function h must satisfies, at least, the following three axiomatic requirements to be qualified as an *aggregation operation* (KLIR & YUAN, 1995):

- Axiom h_1: $h(0, 0, \ldots, 0) = 0$ and $h(1, 1, \ldots, 1) = 1$ (boundary condition)

- Axiom h_2: For any pair $\langle a_1, a_2, \ldots, a_3 \rangle$ and $\langle b_1, b_2, \ldots, b_3 \rangle$ of n-tuples such that $a_i, b_i \in [0,1]\ \forall i \in \mathbb{N}_n,$ if $a_i \leq b_i\ \forall i \in \mathbb{N}_n$, then $h(a_1, a_2, \ldots, a_n) \leq h(b_1, b_2, \ldots, b_n)$ (monotonic increasing)

- Axiom h_3: h is a continuous function

Besides these basic requirements, *aggregation operations* are usually expected to additionally satisfy the two following axiomatic requirements:

- Axiom h_4: h is a symmetric function, so
  $h(a_1, a_2, \ldots, a_n) = h(a_{p(1)}, a_{p(2)}, \ldots, a_{p(n)})$
  for any permutation p on $\mathbb{N}_n$

- Axiom h_5: h is idempotent, so $h(a, a, \ldots, a) = a,\ \forall a \in [0,1]$

This work is interested in two well known operations that can be classified as *aggregation operations*: the *arithmetic mean*, and the *weighted average* with the sum of weights equals to 1.

Chapter 3

Proposal

This chapter presents the novel ideas proposed in this dissertation. Thus, it introduces the generic game model used as a base to implementation and data fit, the game dynamics graphical representation, the original aesthetic metrics developed, and the overall fuzzy aesthetic metric. Therefore, it addresses the research questions about the chosen theoretical framework, the generalization of the chosen metrics and the fuzzy aggregation of the metric values.

3.1 Game model

This dissertation uses a formal generic model of multiplayer games. This model is intended to enrich the theoretical framework shown in the chapter 2 and to allow the use of the developed aesthetic metrics.

This work uses, as a base for the new aesthetic metrics, metrics developed addressing the aesthetic criteria chosen but suitable to strict classes of games like games that involve only two players or combinatorial games. In order to generalize the evaluation of the same game behavior analysis made by these metrics, a more general game model was developed. This model can comprise a large set of games, in where one can fit games with multiple players, a non-arbitrary range of players' scores, without a common final score besides establishing a turn as a segment of the game that can be evaluated to analyze the players progress and other game facets. In addition, championships of sports or other classes of games can be viewed as an instance of the proposed multiplayer turn-based game model showed in Figure 3.1.

The Entity-Relationship model shown in Figure 3.1 emphasizes the distinct nature of its entities using colors. The light blue entities are not part of the game model *per se*. They are aggregation entities and were important in the process of fitting the data from Sebrae Challenge (subsection 4.2.2). As these entities seem to be suitable to represent other kinds of games, they are presented as a complement of the game model's kernel. The model's core is the Match entity, represented in light green. The other entities of the model's kernel are presented in yellow.

Thus, a tournament could be composed of various series, and a series would comprise matches. The use of series is a common expedient to split a tournament into subsets of matches in order to decrease the number of players as the tournament unfolds, implementing sub-goals that force the players to achieve a good rank position for progressing.

A player can enroll in various matches, but each enrollment is related to only one match. The Enrollment entity allows this behavior and stores the final score of the player in a match.

The TurnResult entity implements a NxN relation between the Enrollment and the Turn entities besides storing the turn score, the total score, and the player status. These attributes can generalize the model in order to suit it to games with or without a cumulative score, and also for those with a premature elimination of players or penalty situations in which a player must wait a certain number of turns until play again be allowed.

This game model was crucial for the storing and organization of the data from Sebrae Challenge. However, it was intended to be a generic model convenient to represent multiplayer turn-based games.

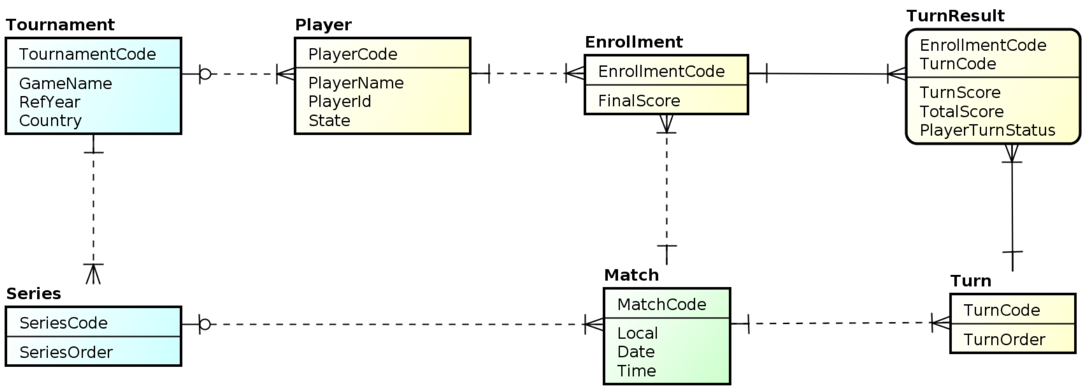
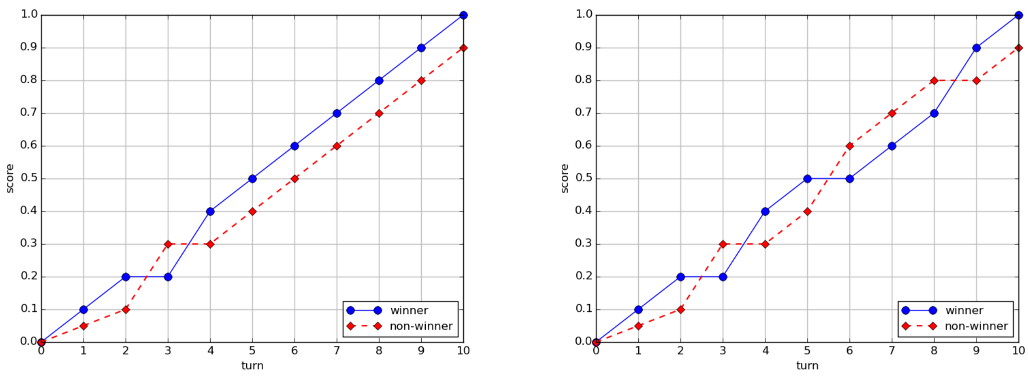


Figure 3.1: Generic Multiplayer Game Model

3.2 Dynamics representation

BROWNE (2008) has presented a graphic representation of players' score evolution, the *Move History* (subsection 2.2.1). These graphs reflect the score progression during a match, move by move. However, a better representation to the intent of this work is shown in Figure 3.2 and was named *Match History by Scores*. It plots player's scores in each game turn, after all players have made their moves, instead after each individual move. Thereby, the graph in Figure 3.2 (a) shows a match example with a single lead change while the one in Figure 3.2 (b) shows a match with multiple lead changes. Both hypothetical games presented in these graphs are two player games, but both representations can be applied to multiplayer games.



(a) Single lead change

(b) Multiple lead changes

Figure 3.2: Lead changes throughout game history by scores examples

3.3 Aesthetic Metrics

This section presents the novel metrics developed to address the generalization of the game aesthetic concepts showed in section 2.2. Hence, it introduces metrics for the criteria *Drama—Drama by Points*, *Drama by Positions* and *Drama by Paths—*, *Lead Change*, and *Uncertainty—Uncertainty by Entropy* and *Uncertainty by PDD*. The problem about how to assign probabilities of winning to players is also discussed.

3.3.1 Drama

Targeting the generalization of *Drama*, one need to bypass combinatorial game restrictions other than the number of players. First of all, the player's score should vary in non-arbitrary ranges, within the particular limitations of each game's rule. The direct implication of this assertion is that there is not a final score that ends the match. In other words, unlike in a combinatorial game, if a player achieves one or any other score value it is not a sign of winning. This work is interested in discussing full matches, even including turns that are played with the winner already defined, since no other player can achieve her score.

In order to implement that rationale, players' scores can be normalized to emulate the same underlying progress presented by combinatorial game, in those games with cumulative points characteristics. So, the ending score of the winner will be the higher one, and all other scores must be normalized in relation to it.

3.3.1.1 Drama by Points

Those little modifications bring the opportunity to introduce some values' representations other than those used by Browne, aiming at a better result. Let us call the winner higher, and last score, as S_H. The lower limit of the points range, to those games with a non-zero minimum, will be named S_L. Also, we'll call the scores after some move m as $S_l(m)$ for the eventual leader, and $S_w(m)$ for the winner player. For any p player in the game, $S(p, m)$ is the score of p in turn m, and $S(p, m_n) \leq S(p, m_{n+1})$. Browne's random moves will be disregarded within our study context, so we'll call M the set of all m moves in a match mch of a game g. After that, we can redefine the *Drama* equation, named now *Drama by Points*.

$$Drama\,by\,Points_{mch} = \frac{\sum_{n=1}^{|M|-1} \sqrt{\frac{S_l(m_n)-S_w(m_n)}{(S_H-S_L)}}}{|\{m|(S_w(m) < S_l(m))\}|} \tag{3.1}$$

3.3.1.2 Drama by Positions

When analysing multiplayer games, one finds the underlying implication of changes in position and distance to the leader associated to total players' scores. Considering that the generic main goal of a game is winning it and, in some situations, there is a secondary goal related to achieve a good position among the first ones leading to a progression in a tournament or a secondary prize, the current player position and the number of players up to the leader are strongly related to a disadvantage perception. Thus, analyzing players ranked at each turn, as well the comparison between each of these and the final ranking, can lead toward a measure that translate the *Drama* as perceived by players in a multiplayer game.

Furthermore, when dealing with games, we are coping with systems that largely use feedback loops to handle point gaps between players (SALEN & ZIMMERMAN, 2004). In this context, neither a distance measured in points nor changes in its amplitude is necessarily related to the players ability. Thereby, this work now introduces some aspects of player position change and its formalization.

The *Players Rank Vector* after move m, with a set of players P, can be defined as an ordered vector PRV_m that holds $|P|$ values denoting each player, from the first position x_1 until the last position $x_{|P|}$. This vector should be built with an appropriate, in game basis, evaluation function for players position.

$$PRV_m = (x_1, x_2, \cdots, x_{|P|}) \tag{3.2}$$

The position function $P_f : P \mapsto \{\mathbf{N} - 0\}$ returns the ranked position for each player given a *Players Rank Vector*.

$$P_f(x_i, m) = i, \forall i \in \{1, 2, \cdots, |P|\} \tag{3.3}$$

Thus, with the P_f defined in Equation 3.3, one can define the normalized distance between two players positions $d_p(x_i, x_j)$ as follow:

$$d_p(x_i, x_j) = \frac{|P_f(x_i) - P_f(x_j)|}{|P| - 1} \tag{3.4}$$

As a consequence of previously presented formal concepts, and with P_w as the match winner, we can define the *Drama by Positions* measured in a match *mch* like:

$$Drama\,by\,Positions_{mch} = \frac{\sum_{n=1}^{|M|-1} \sqrt{\frac{P_f(P_w, m_n) - 1}{|P| - 1}}}{|\{m|(P_f(P_w, m) > 1)\}|} \tag{3.5}$$

3.3.1.3 Drama by Paths

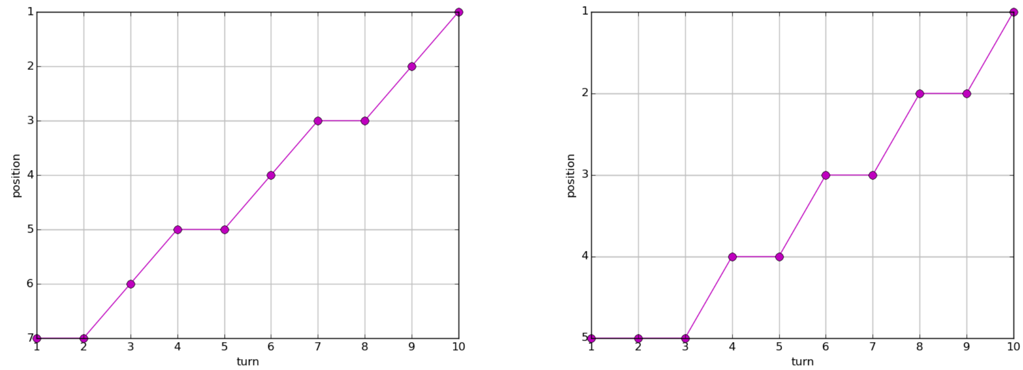
The seminal *Drama* definition used in this work includes an observation about one desirable winning campaign. THOMPSON (2000) claims that "the suspense should continue through an extended campaign" while BROWNE (2008), inspired by that, says that "a player's recovery should not occur in a single killer move, but that the suspense should build up over an extended campaign". As a result, the *Maximum Drama Path* (MDP) for a multiplayer game can be defined as the longest path that a player should traverse from the last position, after the initial turn, up to the first at the end of the game.

$$MDP(m) = \left\lceil |P| + \frac{(1 - |P|)(m - 1)}{|M| - 1} \right\rceil \tag{3.6}$$

In Figure 3.3 one can see two examples of MDPs from matches with a different number of players but the same number of turns. Figure 3.3 (a) shows an MDP for a match with seven players while Figure 3.3 (b) shows an MDP for a five players match. Both figures show paths from matches with 10 turns.

Therefore, *Drama* in a match can be seen as how similar are the path traversed by winner P_w and the MDP. However, when the winner is the better-positioned player after any move other than the last one, the *Drama* measure has to show the earlier goal achievement. After all, when the main goal is fulfilled in advance there is less Drama or, in the case of the winner be always in the first place, no *Drama* at all. So, the *Drama by Path* in a match *mch* is defined as:

$$Drama\,by\,Path_{mch} = \frac{|\{m|P_f(P_w, m) > 1)\}|}{|M| - 1} \left(1 - \sum_{m=1}^{M} \frac{|P_f(P_w, m) - MDP(m)|}{(|P| - 1)(|M| - 1)}\right) \tag{3.7}$$



(a) MDP - 7 players, 10 turns

(b) MDP - 5 players, 10 turns

Figure 3.3: Maximum Drama Path (MDP) examples

3.3.2 Lead Change

This work focuses on two different aspects of lead changes. First, one must think in the implication of the quantity of players that lead the match, after some move. Moreover, the number of times that the match's leader changes must also be considered. With this rationale in mind, an approach of building two components that translate these aspects appears to be appropriate. Hence, the first assertion, about the number of leaders, can be mathematically modelled as the Equation 3.8; and the later one, about the number of changes, as the Equation 3.9.

$$\sqrt{\frac{|\{p \in P | P_f(p, m) = 1, \forall m \in M\}| - 1}{|P| - 1}} \tag{3.8}$$

$$\sqrt{\frac{|\{m \in M | P_f(p_\alpha, m_i) = 1 \wedge P_f(p_\alpha, m_{i-1}) \neq 1, \forall i \in \{2, 3, \cdots, |M|\}\}|}{|M| - 1}} \tag{3.9}$$

P is the set of players that joined the match, and M is the set of turns in a match, alike in the previous equations. The function P_f returns the ranked position of a player after a given move (turn) and is defined in Equation 3.3.

The existence of changes in leadership is a desirable characteristic to a game. However, each lead change is less important to the game's appeal. The same occurs with the number of leaders. In other words, if a match had only one leader, it probably was a boring match. But, in a match with 15 players, if 13 or 14 of them have been leaders, the game's appeal should vary little. Thus, the square root was applied to the equations in order to decrease the marginal contributions for each element added to the sets, since the radicands of each of them result in a value between 0 and 1. Hence, the individual contribution of the elements, to the final value in each equation, is higher in sets with fewer elements, considering the same number of players (Equation 3.8) or moves(Equation 3.9).

In order to simplify the notation, let's call L the set of players that lead during a match. So, it is the set of players with the highest score per turn, after the players' moves, as shown in Equation 3.10. In the same way, let's call $LChange$ the set of turns in which there was a change in the leadership, as presented in Equation 3.11.

$$L = \{p \in P | P_f(p, m) = 1, \forall m \in M\} \tag{3.10}$$

$$LChange = \{m \in M | P_f(p_\alpha, m_i) = 1 \wedge P_f(p_\alpha, m_{i-1}) \neq 1, \forall i \in \{2, 3, \cdots, |M|\}\} \tag{3.11}$$

Thence, the *Lead Change* metric, for a single match, can be defined as the arithmetic mean of the two components showed in Equation 3.8 and Equation 3.9; and it can be represented using the notation showed in Equation 3.10 and Equation 3.11 as follow:

$$LeadChange_{mch} = \frac{\sqrt{\frac{|L|-1}{|P|-1}} + \sqrt{\frac{|LChange|}{|M|-1}}}{2} \tag{3.12}$$

3.3.3 Uncertainty

In subsection 2.2.3, this work already showed that Uncertainty is a fundamental aspect of all type of games as sustained by a pleiad of authors. In addition, the equation Equation 2.4 registered that the entropy concept already had been used to describe the uncertainty related to games.

This work is interested in measuring the uncertainty related to the final game outcome so in the uncertainty about who will win the match and how long this doubt is sustained.

So, let's call $\mathbb{P}(p, m)$ the probability of the player p win the match evaluated after the turn m and, as a consequence for coping with a discrete probability distribution, we have: $\sum_{p \in P} \mathbb{P}(p, m) = 1 \ \forall m \in M$.

3.3.3.1 Uncertainty by entropy

The Shannon entropy is the best know measure of uncertainty for systems (PAL, 1999), and as defined as Equation 3.13. In the entropy calculations, let's define $0 \log_2(0) = 0$.

$$\text{Shannon Entropy} = -\sum_{n=1}^{N} p_i \log_2(p_i) \tag{3.13}$$

By definition, the maximum Shannon entropy in a discrete probability distribution occurs when it is uniform, so all n events have the same probability of $\frac{1}{n}$. The Equation 3.14 shows how the maximum of the Shannon entropy can be expressed by $\log_2(n)$.

$$\begin{aligned} \max(\text{Shannon Entropy}) &= -\sum_{n=1}^{N} \left(\frac{1}{n}\right) \log_2\left(\frac{1}{n}\right) \\ &= -\log_2\left(\frac{1}{n}\right) \\ &= \log_2(n) \end{aligned} \tag{3.14}$$

Hence, for each turn m in the set M for a match mch, with a set P of players p, the uncertainty about the winner can be defined as in Equation 3.15.

$$Uncertainty(m) = \frac{-\sum_{p \in P} \mathbb{P}(p, m) \log_2(\mathbb{P}(p, m))}{\log_2(|P|)} \tag{3.15}$$

And, as a consequence, the *Uncertainty by Entropy* metric, for a single match, can be defined as the average of the uncertainty in each of its turns apart from the last, as in Equation 3.16.

$$Uncertainty\ by\ Entropy_{mch} = -\sum_{n=1}^{|M|-1} \sum_{p \in P} \frac{\mathbb{P}(p, m_n) \log_2(\mathbb{P}(p, m_n))}{\log_2(|P|) \cdot (|M| - 1)} \tag{3.16}$$

3.3.3.2 Uncertainty by probability distributions distance

The maximum uncertainty occurs when all players have the same probability of winning the match. Using the entropy as a measure of uncertainty, one can easily see that the minimum uncertainty—entropy equals to 0—occurs when the probability of winning for one of the players is equal to 1 and, as a consequence, for all the others is equal to zero. So, the uniform distribution of probabilities of winning lead to the maximum uncertainty and the probability distribution D of the form $D = (1, 0, 0, \ldots, 0)$ result in the absence of uncertainty.

From that rationale, one can use another approach, other than the one based on the entropy concept, to measure the degree of uncertainty related to the final game outcome. As each turn presents a different probability distribution, the distance from the current distribution to a hypothetical uniform distribution can guide to a measure of uncertainty.

The Hellinger distance H between two discrete probability distributions $W = (w_1, \ldots, w_k)$ and $Q = (q_1, \ldots, q_k)$ is defined as follow:

$$H(W, Q) = \sqrt{\frac{1}{2}\sum_{i=1}^{k}(\sqrt{w_i} - \sqrt{q_i})^2} \tag{3.17}$$

For the purpose of this dissertation, let's assume the distribution W as an unknown probability of winning distribution related to players after a turn m of a match, and Q as the uniform probability distribution with the same number of events, so players, with the form $Q = \left(\frac{1}{|P|}, \frac{1}{|P|}, \ldots, \frac{1}{|P|}\right),\ for\ |P| = number\ of\ players$. Thus, $H(W, Q)$, as shown in Equation 3.17, can be viewed as the degree of certainty associated with the distribution W. From that, the maximum degree of certainty (Equation 3.18), related to a probability of winning distribution, can be defined as the Hellinger Distance $H(W, Q)$ when $W = (1, 0, 0, \cdots, 0)$ and Q is the uniform distribution described.

$$\begin{aligned} \max(Degree\ of\ Certainty) &= \sqrt{\frac{(\sqrt{1} - \frac{1}{\sqrt{|P|}})^2 + (|P| - 1)\left(-\frac{1}{\sqrt{|P|}}\right)^2}{2}} \\ &= \sqrt{\frac{1 - \frac{2}{\sqrt{|P|}} + \frac{1}{|P|} + (|P| - 1)\left(\frac{1}{|P|}\right)}{2}} \\ &= \sqrt{\frac{2 - \frac{2}{\sqrt{|P|}}}{2}} \end{aligned} \tag{3.18}$$

Hence, the Uncertainty after a turn m, varying in a range between 0 and 1, can be defined as in Equation 3.19, for a set of players P. As expected, $\mathbb{P}(p, m)$ is the probability of winning for the player p evaluated after the turn m.

$$Uncertainty(m) = 1 - \sqrt{\sum_{p \in P} \frac{\left(\sqrt{\mathbb{P}(p, m)} - \frac{1}{\sqrt{|P|}}\right)^2}{2 - \frac{2}{\sqrt{|P|}}}} \tag{3.19}$$

As a consequence, the Uncertainty for a match mch computed by the probability distributions distances—*Uncertainty by PDD*—is defined as equal to the average of uncertainty after each turn of the match, apart from the last, as shown in Equation 3.20.

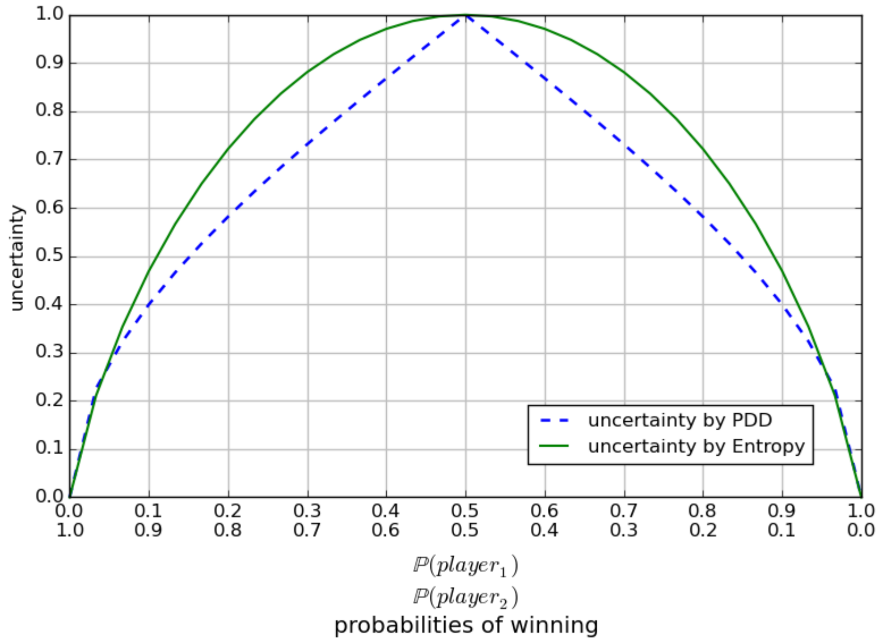


Figure 3.4: Comparison between the behavior of uncertainty metrics.

$$Uncertainty\ by\ PDD_{mch} = \frac{\sum_{n=1}^{|M|-1} 1 - \sqrt{\sum_{p \in P} \frac{\left(\sqrt{\mathbb{P}(p,m_n)} - \frac{1}{\sqrt{|P|}}\right)^2}{2 - \frac{2}{\sqrt{|P|}}}}}{|M| - 1} \tag{3.20}$$

Figure 3.4 presents the behavior of both proposed metrics for uncertainty in all possible values of probabilities of winning for a two player game. The blue dashed line shows the computed values for uncertainty by the metric $Uncertainty\ by\ PDD$ while the continuous green line shows the values evaluated by $Uncertainty\ by\ Entropy$ metric. As one can see, in the great majority of the cases the $Uncertainty\ by\ Entropy$ metric returns a value higher than the one returned by the $Uncertainty\ by\ PDD$ metric.

3.3.3.3 Probabilities of winning

The game score is the key to assign the probability of a player win the match after a given turn as it can be seen as the result of an evaluation function that shows the player progress until the game goal. Hence, the player with the highest score would be the closest to winning. Thus, the probabilities of the players win the match, after a given turn, depends on their scores.

However, apart other distinctions on player's score nature, there are games with a limit in player's score variation after each turn, while there are others without it. In those games with this limit, the score gap between a player and the leader could be unachievable, so the probability of winning for this player should be zero.

Two slightly distinct methods are used in this work to assign the probabilities of winning to players, $\mathbb{P}_a(p, m)$ and $\mathbb{P}_b(p, m)$. Although the both starts from the same principle to assign the probability of winning $\mathbb{P}$ to player p evaluated after the turn m, one method simply uses the player's score and the total score at the turn to accomplish its task while the other also take into account the score gap to the leader and the number of remaining turns.

When no assumptions can be made about the limit of score variation, the probabilities of winning were assigned to players in accordance with the $\mathbb{P}_a(p, m)$ formula, showed in Equation 3.21. Starting from the premise that the initial probability distribution is uniform, the *a priori* probability $\frac{1}{|P|}$ is modified by the normalized score. $S_l(m)$ is the score of the eventual leader as already shown in the section Drama by Points. The new values must be normalized to make their sum equal to one. Gratefully, this equation can be conveniently simplified to the ratio between the player's score and the total score of the players in the turn.

$$\begin{aligned} \mathbb{P}_a(p, m) &= \frac{\frac{1}{|P|} \cdot \frac{S(p,m)}{S_l(m)}}{\sum\limits_{k \in P} \left(\frac{1}{|P|} \cdot \frac{S(k,m)}{S_l(m)} \right)} \\ &= \frac{\frac{1}{|P| \cdot S_l(m)} \cdot S(p, m)}{\frac{1}{|P| \cdot S_l(m)} \sum\limits_{k \in P} S(k, m)} \\ &= \frac{S(p, m)}{\sum\limits_{k \in P} S(k, m)} \end{aligned} \tag{3.21}$$

Some concepts must be defined prior to the description of the probability formula to those games with a limit in score increasing per turn. Initially, that limit to increasing in score can change in each turn so let's call it $MS(m)$. Furthermore, if some player can not reach the score of the leader in the remaining turns, its probability of winning the match must be equal to zero. Thus, the achievement test function A_f (Equation 3.22) computes the difference between the score of the player $S(p, m)$ and the score of the eventual leader at the turn $S_l(m)$, returning zero in the case of the leader's score be unachievable by the player or an adjustment factor inversely proportional to the size of the score's gap.

$$A_f(p, m_n) = \begin{cases} 0, \text{ if } S_l(m_n) - S(p, m_n) > \sum\limits_{k=n+1}^{|M|} MS(m_k) \\ \frac{1}{1+S_l(m_n)-S(p,m_n)} \text{ , otherwise} \end{cases} \tag{3.22}$$

Therefore, the probability of winning $\mathbb{P}_b(p, m_n)$ in those games with a limit to increasing in score per turn can be inferred from Equation 3.21 and Equation 3.22, and defined as follow.

$$\mathbb{P}_b(p, m_n) = \frac{S(p, m_n) \cdot A_f(p, m_n)}{\sum_{k \in P} S(k, m_n) \cdot A_f(k, m_n)} \tag{3.23}$$

3.4 A Fuzzy Aesthetic Metric

All the developed metrics define fuzzy sets. The aesthetic facet of the game measure by each metric is the property shared by the elements of the set in the degree returned by the metric. Thus, they are *membership functions* that uniquely and completely define fuzzy sets. Hence, their output can be used in a fuzzy aggregation operation to produce an overall aesthetic evaluation for matches and, as a consequence, for games.

Addressing the formalization of the previous assertions, let's call the infinite "crispy" set of all possible matches *mch*, of all games that fit in the model proposed in this work, as the universe of discourse MCH. In addition, let's assume Am as a generic representation of any metric among the developed aesthetic metrics. Hence, one can note by the rationale presented in section 3.3 that the functions Am are defined as in Equation 3.24 and they denote the membership grade of a match to the fuzzy set uniquely defined by each one of them.

$$Am : MCH \mapsto [0, 1] \tag{3.24}$$

Six metrics were presented, two for the *Uncertainty* criterion, three for *Drama*, and one for *Lead Change*. Although one can use all of them into an aggregation operation, this work intends to use one metric for each criterion. So, to clarify the notation, let's call the membership functions $\mu_{Uncertainty}$, μ_{Drama}, and $\mu_{Lead\ Change}$, respectively. The definition of which metric will be used to measure the criteria, when there is a choice to make, is done in chapter 4.

The fuzzy set specifies by the aggregation function h will be called **AM**. Thus, for each element *mch* for the set MCH described above, $\mu_{\mathbf{AM}}$ can be defined as follow:

$$\mu_{\mathbf{AM}}(mch) = h(\mu_{Uncertainty}(mch), \mu_{Drama}(mch), \mu_{Lead\ Change}(mch)) \tag{3.25}$$

The aggregation function h will be initially defined as the arithmetic mean. However, it should evolve to a weighted average mean, with the sum of its weights equals to one, to improve predictions regarding human perception about the games aesthetic criteria. Accordingly, the fuzzy aesthetic metric, that expresses the overall aesthetic evaluation of a match, will be called *Overall Aesthetic Metric* and is defined as follow:

$$Overall\ Aesthetic\ Metric(mch) = \frac{\mu_{Uncertainty}(mch) + \mu_{Drama}(mch) + \mu_{Lead\ Change}(mch)}{3} \tag{3.26}$$

Chapter 4

Implementation

This chapter presents the implementation of the ideas proposed in the last one. Therefore, it shows the computational artifacts developed in order to evaluate the proposed metrics, exposes and explains the datasets used in that evaluation as well as the validation of the metrics. Thus, it addresses the research question about the metrics validation.

The developed metrics were applied to the data of Brazilian Football Championship, the results were compared with the evaluation of the criteria made by human judges. In a first round, this method led to the election of the metrics that would compose the overall aesthetic metric. Later, the same method was used to evaluate the resultant measure. Finally, the overall aesthetic metric was applied to Sebrae Challenge dataset to analyze the changes in the aesthetic levels in this game.

4.1 Evaluation Framework

An *Evaluation Framework* was developed in order to implement the exposed proposals. That framework comprises a template for metrics and a generic game structure. The `MeasureTemplate` class offers the basic architecture for the metrics, while the `GenericGame` class implements operations and attributes of those games that fit into the model presented in this work.

Figure 4.1 shows a UML class diagram of those classes as well as the data structures needed to manipulate player's scores and cope with metric's identification. The base classes of the framework were specialized to implement the metrics and to retrieve and store data from two distinct datasets.

To deal with the different data organization of the datasets used in the analysis of the metrics, two separate structures were built. The `BrasileiroGame` class—that cope with the dataset of Brazilian Football Championship—and the `DesafioGame` class—that cope whit data of Sebrae Challenge—are specialized classes of `GenericGame` that construct the attribute `_gameData` through the method `_-`

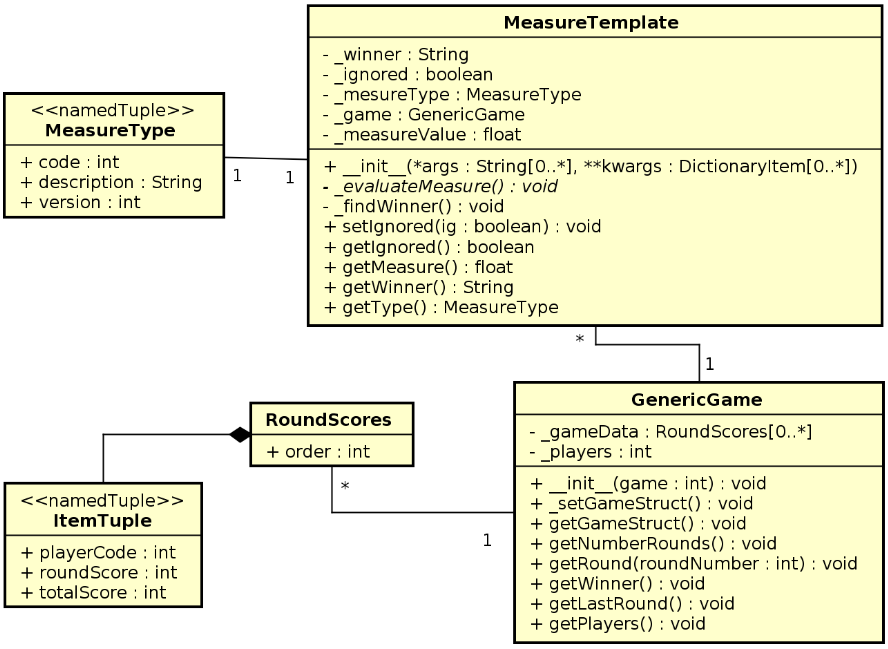


Figure 4.1: Evaluation Framework

`setGameStruct` using different inputs to do so. The former uses a `Match` class from the package *brasileiro.model* while the latter uses one from the package *desafio.model*. The `Match` class of the `desafio.model` package operates on data in a PostgreSQL database system, it retrieves and store game data and computed metrics, while the `Match` class of the `brasileiro.model` operates on files stored in JSON format only retrieving information.

All the metric for the matches in the dataset of Brazilian Football Championship, operated by the `brasileiro.model`, are not stored due to the reduced number of records, so they are computed at runtime. Figure 4.2 shows a UML class diagram with these classes besides the classes for cope with specific characteristics of the data of the Brazilian football championship: `MatchRound` and `ItemBRTuple`.

Figure 4.3 shows the classes that implement the developed metrics. Each metric was implemented as a subclass of the `MeasureTemplate` and have a different version of `_evaluateMeasure` method, that is the responsible for computing the measures. The `DramaByPaths` class, additionally, have a method that returns the MDP point, `_maxDrama`, and other that returns the ranked position of a player, `_playerPosition`, to do its calculations. The classes that implement the *Uncertainty* measures, `UncertaintyPDD` and `UncertaintyEntropy`, also have an additional method to evaluate the probability of winning for a certain player, `_probPlayer`. The implementation of the `_probPlayer` method is selected from the decorator module `probWinning`. The constructor method of these classes selected the version of the evaluation probability method. If there is a score limit for

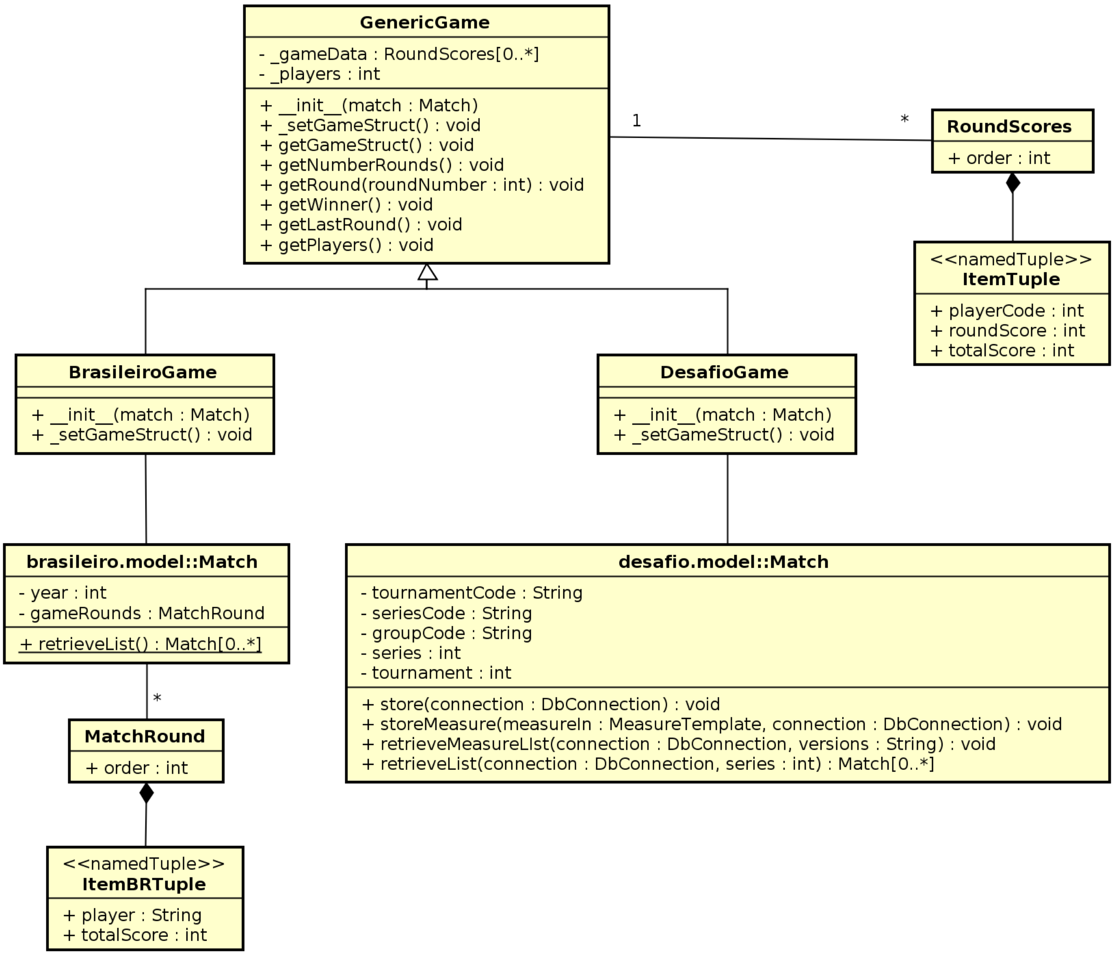


Figure 4.2: Model implementation classes

each turn among the parameters received by the constructor, `_probPlayer` implements `probB` version that was developed in accordance with the $\mathbb{P}_b(p, m)$ proposed in 3.3.3.3, otherwise it implements `probA` that complies with $\mathbb{P}_a(p, m)$.

The constructor method of the classes that implement the developed metrics can adapt the object's behavior built by them to the game subject of the analysis. The parameters optionally received by the constructors guide this process and are listed in Table 4.1.

Table 4.1: Optional parameters for constructor methods of the measure classes

Parameter	Description	Default value
`ignored`	Number of initial turns that must be ignored in game analysis.	0
`normScores`	Boolean value denoting if the scores must be normalized.	False
`minScore`	Value of the minimum score for a player.	0

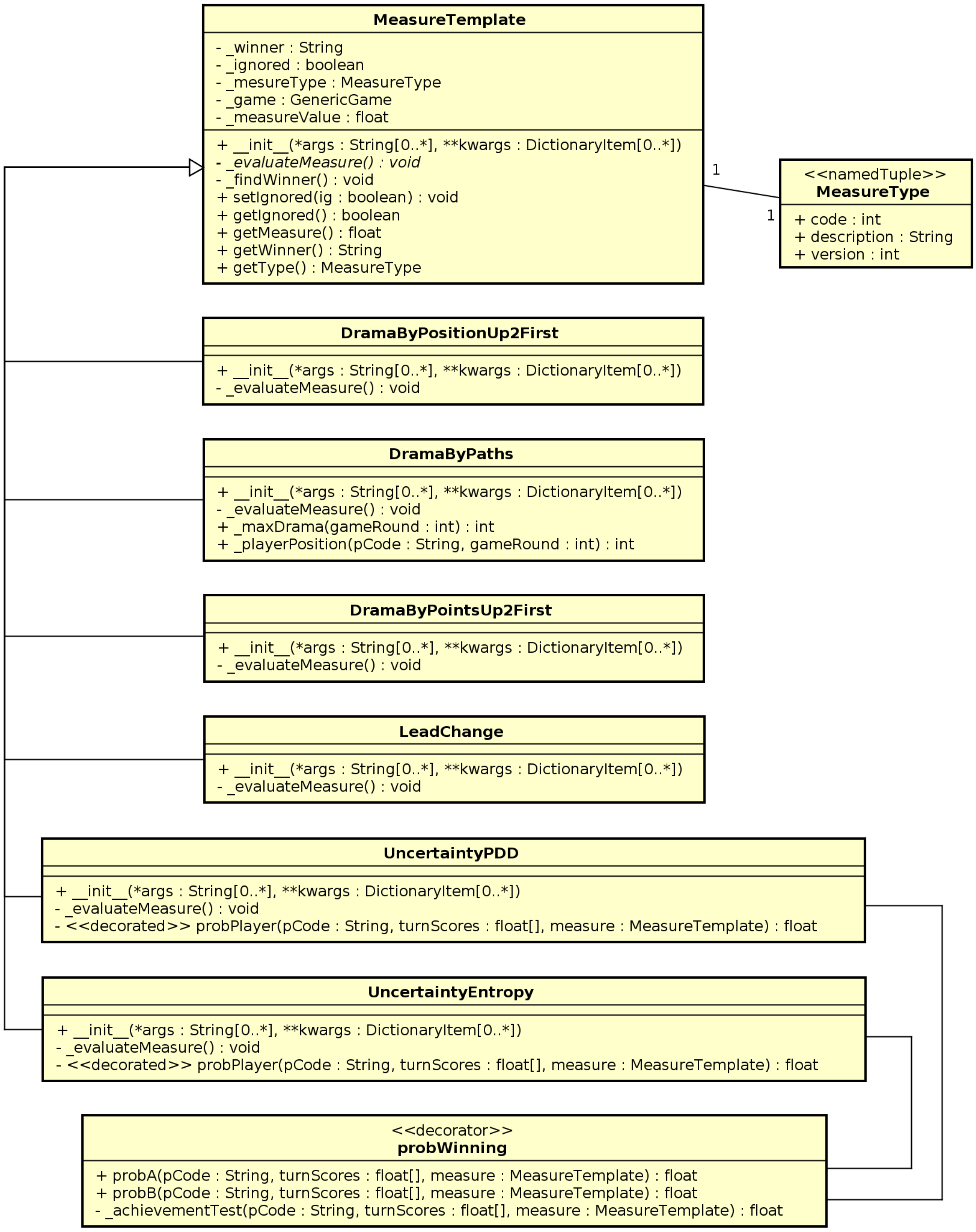


Figure 4.3: Measures implementation classes

The classes that implement *Uncertainty* measures additionally can receive the `scoreLimit` parameter with the possible value for the maximum score increase per turn.

The evaluation framework was developed targeting the expansion of the number of metrics and games implemented. The base classes were thought to be easily used, presenting a basic interface for the common operations needed to deal with the generic multiplayer game model proposed in section 3.1.

4.2 Data

In this section, the datasets used in the implementation and in the outcome validation of the metrics are presented. Two distinct datasets were used, one of them comprises the recent history of the Brazilian national football championship and the other consist of matches from the Sebrae Challenge. The datasets were managed, in different ways, in order to fit its data into the generic game model presented at section 3.1.

The Sebrae Challenge dataset was stored in a PostgreSQL[1] database system in accordance to the Entity-Relationship model shown in the section 3.1. The Brazilian Football Championship data was stored in text files using JSON (JavaScript Object Notation[2]), one file for each annual edition.

4.2.1 Brazilian Football Championship

The various measures above-mentioned were applied to the recent history of the Brazilian national football major league championship, know as *Brasileirão – Série A*. The dataset comprises data from annual editions from 2003 to 2014. Since 2003, the tournament adopted the double round-robin system. Because of that, the full championship can be viewed as a single match in which the teams, acting as players, make their moves at each turn, in this case, a championship round. Therefore, in each move, a team can score zero points, in the case of loss, one point in a draw, or three points when it wins.

Table 4.2: Summary of Brazilian National Football Major League Championship Recent History

Edition	Teams	Rounds	Winner	Winner final score	Winner effectiveness
2003	24	46	Cruzeiro	100	72.5
2004	24	46	Santos	89	64.5
2005	22	42	Corinthians	81	64.3
2006	20	38	São Paulo	78	68.4
2007	20	38	São Paulo	77	67.5
2008	20	38	São Paulo	75	65.8
2009	20	38	Flamengo	67	58.8
2010	20	38	Fluminense	71	62.3
2011	20	38	Corinthians	71	62.3
2012	20	38	Fluminense	77	67.5
2013	20	38	Cruzeiro	76	66.7
2014	20	38	Cruzeiro	80	70.2

[1] http://www.postgresql.org/
[2] http://www.json.org/

Table 4.2 shows the number of teams and rounds, as well as the winner, its final score and effectiveness for each championship edition in the dataset. Winner effectiveness is the ratio of total points achieved by the winner and total possible points in the tournament. There were championships with 20, 22, and 24 teams. As each team plays against all the others twice, therefore there were championships with 38, 42, and 46 turns[3]. Moreover, in twelve editions there were only six distinct winners.

This set of data was obtained from the Internet and parsed with the help of the BeautifulSoup[4] python library. Each championship was stored as a JSON object in a single file. All data operations – acquisition, parsing, storing and retrieving – were made using the python programming language.

4.2.2 SEBRAE Challenge

The Sebrae Challenge was a business game applied to undergraduate students in different countries. The game was intended to encourage entrepreneurship and simulates several facets of a company like inventory control, marketing strategies, human and financial resources management, and so on.

4.2.2.1 The Game

The game was developed by the COPPE/UFRJ Business Incubator in 2000 to comply with a SEBRAE's demand. SEBRAE is a Brazilian institution with the goal to encourage entrepreneurship. That game was applied during 13 years to more than one million undergraduate students in Brazil and other countries, like Argentina, Chile, Colombia, Ecuador, Panama, Paraguay, Peru and Uruguay.

DE BAKKER *et al.* (2011) brings a short description of that game: "A representative business game example is the *Desafio Sebrae* (Sebrae Challenge), in which college students work as a team managing a virtual enterprise and need to make periodic business decisions. Thus, the game seeks to disseminate the culture and experience of entrepreneurial management for college students even before they start their careers."

The tournaments were organized in series composed by matches with the same number of turns. The number of turns was a parameter fixed in advance, before the series start, and was unknown by the participants. The teams with better performance among all the participants, in each series, were promoted to the next

[3]The different number of teams and rounds in the editions was caused by early adjustments in the championship format. The old championship format used to have more teams. The current number of teams was gradually achieved.

[4]http://www.crummy.com/software/BeautifulSoup/

Table 4.3: Matches' distribution in sample

Country	Series	Number of Matches	Percentage
Brazil	S1	4531	58.17%
	S2	2017	25.90%
Peru	S1	374	4.80%
	S2	187	2,40%
Others		680	8.73%

one. Thus, the series in a tournament had a decreasing number of participants and the final series had a single match.

The players teams engage in a half-blind contest in which each team can't directly see the current scores of others. Although teams don't know the other teams' scores during the match, they can follow the development of other players through the market information that can be acquired in the game. During the matches, teams should take and register their decisions at each turn, otherwise they were eliminated.

4.2.2.2 Dataset

Table 4.4: Subset 'Others' Detail

Country	Series	Number of Matches	Percentage
Argentina	S1	109	1.40%
	S2	57	0.73%
	S3	2	0.03%
Brazil	S3	64	0.82%
	S4	16	0.21%
	S5	2	0.03%
Colombia	S1	49	0.63%
	S2	30	0.39%
	S3	1	0.01%
Ecuador	S1	92	1.18%
	S2	54	0.69%
	S3	1	0.01%
Panama	S1	69	0.89%
	S2	30	0.39%
	S3	2	0.03%
Paraguay	S1	63	0.81%
	S2	36	0.46%
	S3	1	0.01%
Peru	S3	1	0.01%

The Sebrae Challenge dataset comprises a total of 7983 matches that occurred in seven countries in 2011 and 2012, from which this work took 7789 as valid matches.

As the teams could be eliminated by didn't register their decisions in each turn, those matches that ended with only one active player, as well as those ones in which all teams were eliminated before the ending turn, were discarded.

Table 4.3 shows the distribution of matches in the final sample. The subset named 'others' comprises the data from the series listed in Table 4.4 table. The number of matches in the represented series is the sum of those ones from the tournaments in 2011 and 2012.

4.2.2.3 Changes in the Game

At each year the Sebrae Challenge had its graphical user interface changed targeting a better user experience. Besides that, there were always improvements in the game core like adjustments in the mathematical model, security upgrades and so on. As the dataset used in this work comprises matches of two distinct years, the changes that happened are relevant to the analysis.

In 2012, the most significant modifications in Sebrae Challenge were the changes in the content communication and the introduction of services decisions in the game. The former was a redesign of the content structure implementing an organization more didactic than the previous one. This new organization included features like tips in the screen about decision-making, learning tutorials, new help files, an exhibition of the consequences that may arise from decisions, a business sectorial guide, consulting simulations, business concepts guides, and explanatory links. The latter was caused by a sponsor's request and introduce the implementation of decision-making tasks for services, including new evaluation methods and contents for that module, besides adapting the whole mathematical model.

To avoid undesirable effects in the game balance, the inclusion of new decisions led to an interface adaptation, and to changes to the game metaphor.

4.3 Graphical User Interface for Sebrae Challenge

Two Graphical User Interfaces (GUIs) were built to deal with several options of data visualization from Sebrae Challenge dataset, and to verify the implemented code of the developed metrics. They retrieve original data from matches, as well as the computed metrics in their newer version, from a PostgreSQL database system.

Figure 4.4 shows the *Match Viewer*, a GUI developed to allow the data visualization of a single match. The three list boxes at the upper-left corner show, respectively, a list of all tournaments in the dataset, a list of the series of the selected tournament, and a list of matches of the selected series. Just below them,

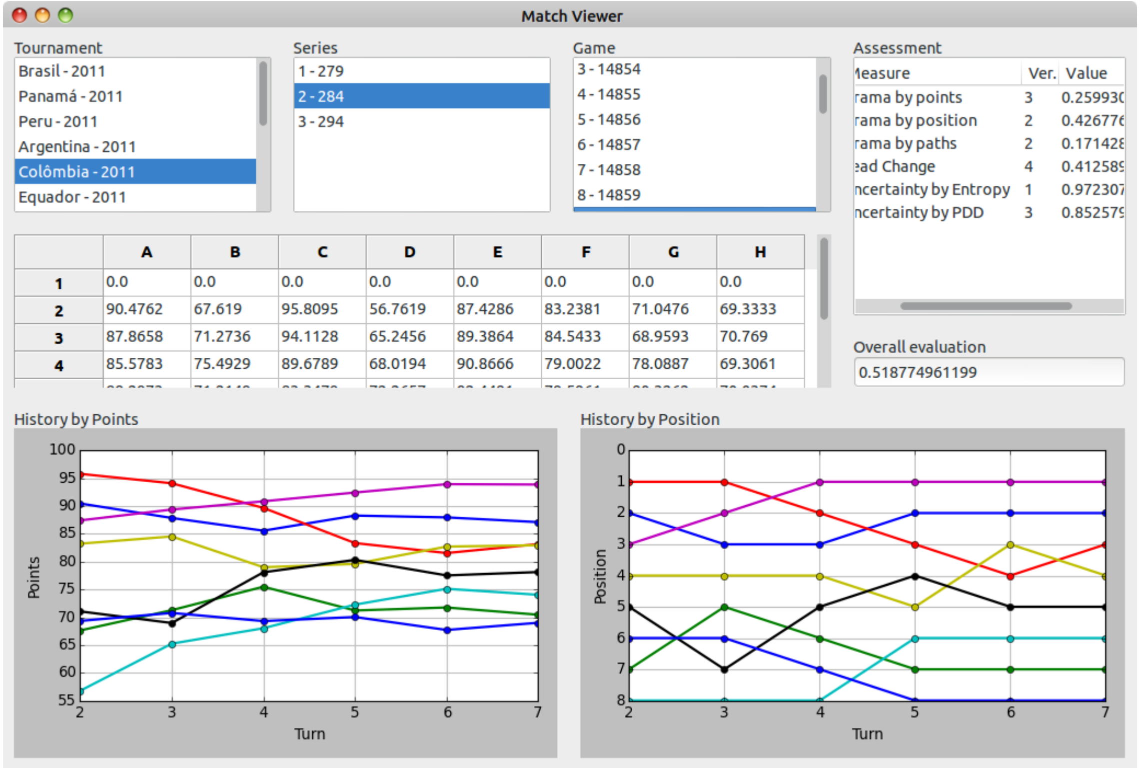


Figure 4.4: Match Viewer

there is a table view of players' scores within the match, one column for each player and one row for each turn. On their right, there is a list box with the values of all computed measures in their later version, right on top of the overall evaluation value. In the lower section of the window, there are two 2D graphics presenting the player's campaigns through the match. The left one presents the evolution of the campaigns using the score of players, so the vertical axis represents the scores and the horizontal the turns. The graphic on the right also presents the campaigns evolution but using the position of players after each turn. One must note that the vertical axis of the latter one is reversed so the first position is on top of the axis. Both graphics have the horizontal axis starting from 2, the first turn of matches from Sebrae Challenge was ignored because they always present all scores equal to zero.

Figure 4.5 presents other GUI, the *Group Viewer*, that was developed to view the distribution of the metrics' values in data samples. The tree view in the window's upper-left corner supports the navigation through the dataset. On its right, one can see information about the group of matches selected and a list of the measures' values as well as the overall evaluation. The presented values are averages of the values for the metric in the sample unless there is only one selected match. The lower section of the window presents a graphic representation of the information about the sample or match selected in accordance to the chosen metric in the combo box

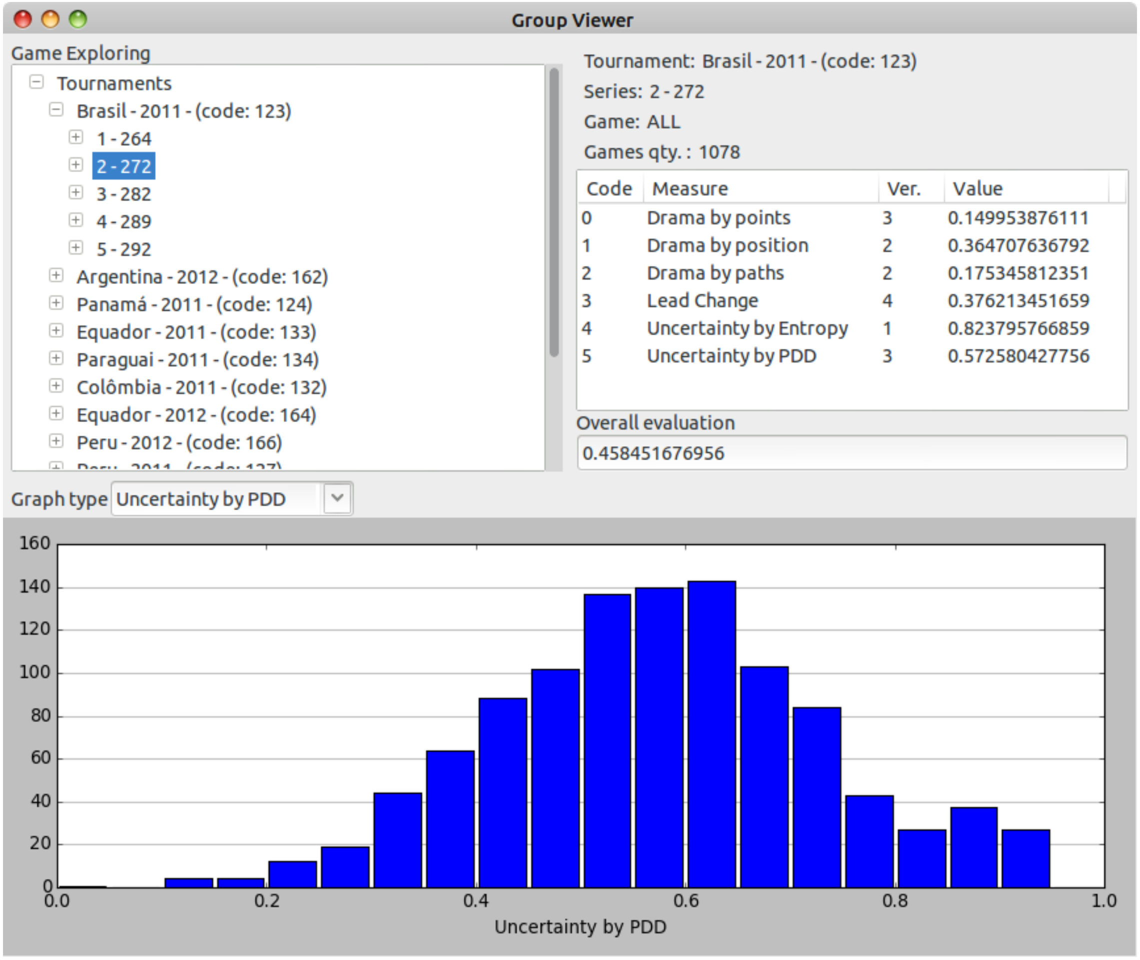


Figure 4.5: Group Viewer

just above it. If the selection in the tree view is a set of matches the graphic always will be a histogram, otherwise it may be a history by points or a history by position, the most suitable to the chosen metric.

4.4 Metrics Validation

In this section, the measures of the aesthetic criteria are validated. First, the metrics for the *Drama* criterion are compared for choose one of them as a component of the fuzzy aesthetic metric. The Brazilian Football Championship was used to select the metrics that better emulate the human choices. After that, the same method of choice is used to select the *Uncertainty* criterion metric. And finally, a similar validation process is used to evaluate the fuzzy aggregated metric. Thus, this section addresses the research question about the metrics validation and considers the implication related to human validation.

4.4.1 Drama Validation

The three novel metrics for the *Drama* criterion—*Drama by Points*, *Drama by Position* and *Drama by Path*—were validated using the dataset of the Brazilian football championship. Table 4.5 shows the values of the three different *Drama* measures presented, using data from the 12 last Brazilian national football championship editions, as well as the rank of each edition as stated by each measure.

Table 4.5: Drama Measures Values by Championship Edition

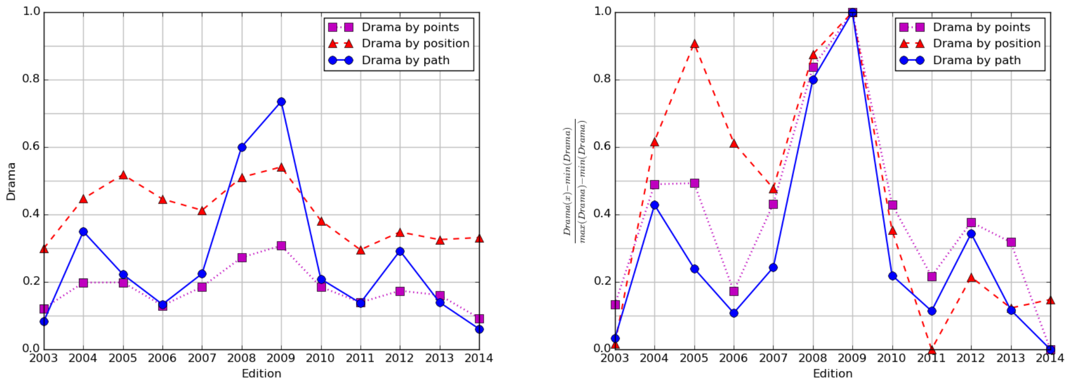
Edition	Drama by Points		Drama by Position		Drama by Path	
	value	rank	value	rank	value	rank
2003	0.1213	11	0.2993	11	0.0828	11
2004	0.1979	4	0.4470	4	0.3495	3
2005	0.1986	3	0.5183	2	0.2225	6
2006	0.1298	10	0.4458	5	0.1334	10
2007	0.1854	5	0.4125	6	0.2245	5
2008	0.2725	2	0.5103	3	0.6004	2
2009	0.3076	1	0.5409	1	0.7349	1
2010	0.1851	6	0.3821	7	0.2088	7
2011	0.1392	9	0.2953	12	0.1374	9
2012	0.1738	7	0.3480	8	0.2923	4
2013	0.1610	8	0.3254	10	0.1396	8
2014	0.0927	12	0.3316	9	0.0607	12

Figure 4.6 shows two graphical representations of the *Drama* measure values. In Figure 4.6 (a), one can see how *Drama by Path* presents a variance higher than others measures so is more responsive to changes in *Drama* levels among the editions analyzed as matches. In Figure 4.6 (b), the *Drama* values are normalized, for each *Drama* measure, according to the formula: $y = (Drama(x) - min(Drama))/(max(Drama) - min(Drama))$. Therefore, one can clearly see the points of disagreement.

Human judges[5] with different views of games—designers and players—were invited to also evaluate the *Drama* level in the data. We presented to them printed graphs similar to those in Figure 4.7 but not showing the MDP neither the measures values. They were introduced to the concept of *Drama* used in this work to avoid semantic misunderstood.

The judges' ranks do not present a consensus, even in the very first or last positions. According to MESKANEN & NURMI (2006), in this condition, the Schulze Method (SCHULZE, 2003) has the better appliance to represent their aggregate

[5]Three game designers, 2 undergraduate students, 1 graduate student and 1 project manager in the game development field.



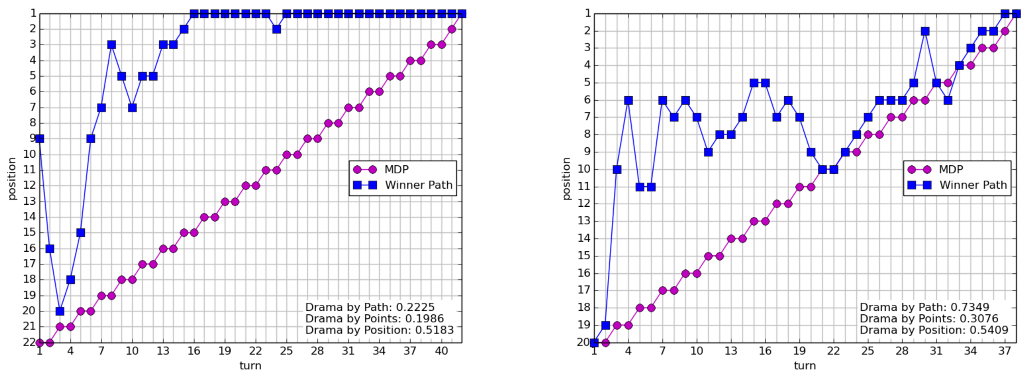
(a) Drama measure values (b) Normalized drama measure values

Figure 4.6: Drama Measures by Championship Edition

preferences, exhibiting the strongest path. Table 4.6 shows the judges' ranks and the final rank in accordance with the aggregation method used.

In order to verify the better association of rankings between judges and dramas measures, Table 4.7 shows the Kendall tau[6] coefficient and the two-sided *p-value* for a hypothesis test where H_0 is $\tau = 0$ denoting no correlation between ranks (ABDI, 2007).

One must pay attention for the fact that the Browne's measure of *Drama* is a simple measure, with some characteristic effects. For example, a game with an unique movement where the second player becomes the first player, overcoming a difference of δ points, has the same *Drama* as a game where this happens twice, or even more times, if the average difference is δ points. Thus, the same drama value is assigned to games with quite distinct behaviour, like the ones in Figure 3.2. However,



(a) 2005 Edition (b) 2009 Edition

Figure 4.7: MDP versus winner path comparison - Brazilian football championship editions 2005 and 2009

[6]The τ_B version of the function in python scipy library (http://scipy.org/) was used for the computation.

Table 4.6: Judges Evaluation of Drama

Edition	Judges							Final Rank
	A	B	C	D	E	F	G	
2003	11	11	11	11	11	11	8	11
2004	3	1	3	3	3	3	2	3
2005	5	5	5	8	7	6	6	6
2006	8	6	10	10	10	8	11	10
2007	6	7	8	7	8	5	9	8
2008	2	2	2	2	2	2	4	2
2009	1	3	1	1	1	1	1	1
2010	9	4	4	4	5	4	3	4
2011	10	12	7	5	4	9	5	7
2012	4	8	6	6	6	10	7	5
2013	7	9	9	9	9	7	10	9
2014	12	10	12	12	12	12	12	12

since Browne measures 57 aesthetic criteria, he can afford some shortcoming in one of them, since this can be compensated by another. For example, it measures other quality criteria related to *Drama* such as *Leaded Change*, *Permanence*, *Killer Moves*, and *Uncertainty*.

The same weakness above mentioned is inherited by the measures *Drama by Points* and *Drama by Positions*. However, the cumulative points nature of the game used as the object of this study has the property of reducing the possibility of a killer move. As the maximum addition to a player's score after a move is limited to three, the game progress can impose score gaps that are impossible to reverse with a single move. Furthermore, this characteristic, allied to the great number of players and turns, appears to be responsible for bring closer the behaviour of the *Drama by Points* and *Drama by Path* measures.

The great correlation presented between the judges' preferences and the ranks stated by *Drama by Points* and *Drama by Path* must be seen as a sign of strength of the assumption that those measures are able to evaluate the *Drama* in a multiplayer game. In addition, the *Drama by Path* does not present the same limitations of the other measures shown due to its penalty factor and the MDP use, in the sense it penalizes the early goal achievement and is independent of the cumulative points

Table 4.7: Kendall tau ranks correlation between judge's rank and Drama measures

Measure	Judges' Rank	
	τ	p-value
Drama by Points	0.7878	0.0004
Drama by Positions	0.5454	0.0136
Drama by Path	0.8182	0.0002

nature of the game. Also, it is the measure with the better correlation with judges' choices combined by the Schulze Method.

Drama by Path appears to be the better among the shown measures related to the *Drama* criterion, because it proved to be more responsive to changes in *Drama* level, is independent of the cumulative points nature of the game, is the closest to the seminal *Drama* concept definition, and presented great correlation to choices made by humans. Because of that, it will be used in the following development of this work.

4.4.2 Uncertainty Validation

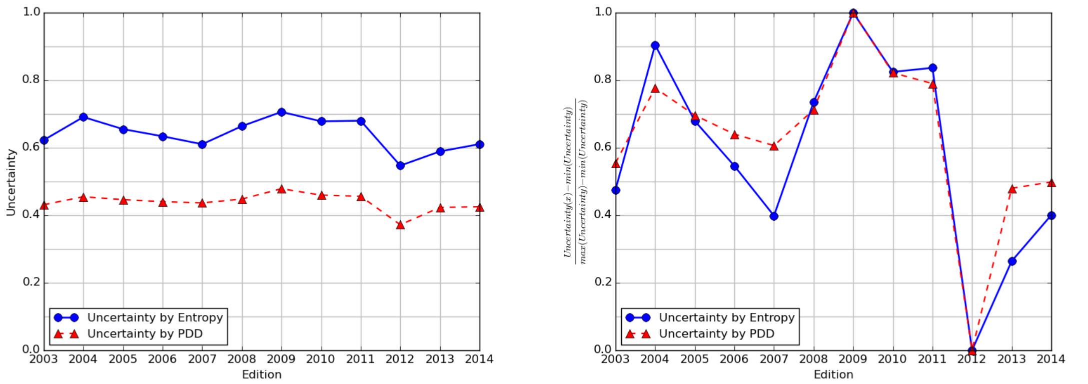
The same method used in validation of *Drama* metrics was used to validate the novel metrics for *Uncertainty*—*Uncertainty by Entropy* and *Uncertainty by PDD*. Therefore, the two measures were applied to data from the Brazilian football championship dataset. The score limit was set up to 3, in accord with the game characteristic, so $\mathbb{P}_b(p, m)$ (propose in 3.3.3.3) was used to evaluate the probabilities of winning. Table 4.8 presents the values computed by each metric and the rank of each edition as stated by them.

Table 4.8: Uncertainty Measures Values by Championship Edition

Edition	Uncertainty by Entropy		Uncertainty by PDD	
	value	rank	value	rank
2003	0.6227	8	0.4311	9
2004	0.6909	2	0.4548	4
2005	0.6551	6	0.4462	6
2006	0.6339	7	0.4402	7
2007	0.6104	10	0.4366	8
2008	0.6639	5	0.4479	5
2009	0.7061	1	0.4786	1
2010	0.6782	4	0.4596	2
2011	0.6801	3	0.4562	3
2012	0.547	12	0.3719	12
2013	0.589	11	0.4231	11
2014	0.6107	9	0.4251	10

Figure 4.8 shows two plots of the *Uncertainty* measure values. Figure 4.8 (a) shows the raw values of the metrics while Figure 4.8 (b) presents their computed values normalized, according to the formula: $y = (Uncertainty(x) - min(Uncertainty))/(max(Uncertainty) - min(Uncertainty))$. In the former, one can see that *Uncertainty by Entropy* assign higher values to the championship edi-

tions and appears to present a higher variance. In the latter, the points of disagreement, although they are few, are made clear.



(a) Uncertainty measure values (b) Normalized uncertainty measure values

Figure 4.8: Uncertainty Measures by Championship Edition

As was made for *Drama* metrics, the Kendall tau coefficient was used to verify the association between the judges' rank and the *Uncertainty* metrics. Table 4.9 presents the resulting coefficient and the two-sided *p-value* for a hypothesis test where H_0 is $\tau = 0$ denoting no correlation between ranks.

Table 4.9: Kendall tau ranks correlation between judges' rank and Uncertainty measures

Measure	Judge's Rank	
	τ	p-value
Uncertainty by Entropy	0.4545	0.0397
Uncertainty by PDD	0.4848	0.0282

Although *Uncertainty by PDD* outperforms *Uncertainty by Entropy* in the test, the size of the sample and the similarity between the results do not lead to a conclusive decision. Both metrics appears to present a very similar behaviour, varying in a small range of values. *Uncertainty by Entropy* seems to be more responsive to changes in the criterion level even in the subject game that is a cumulative points game, with a maximum score per turn. These characteristics of the game, allied to the number of players and turns, are responsible for sustaining high levels of uncertainty in the initial turns. Thus, even with a drastic decrease in the last turns of some championships, when the winner is already defined, the overall level of *Uncertainty* varies little in the sample of the subject game.

The game nature is not the only factor to consider. The rank of the judges was made with little information about the facet of the game dynamic examined by these measures. This potential lack of information may have negatively contributed to the coefficient values. Hence, in a probably very arbitrary decision, the *Uncertainty by*

PDD metric will be the chosen one to the following development of the overall aesthetic metric. This decision does not discard the use of *Uncertainty by Entropy* in a future analysis.

4.4.3 Lead Change Validation

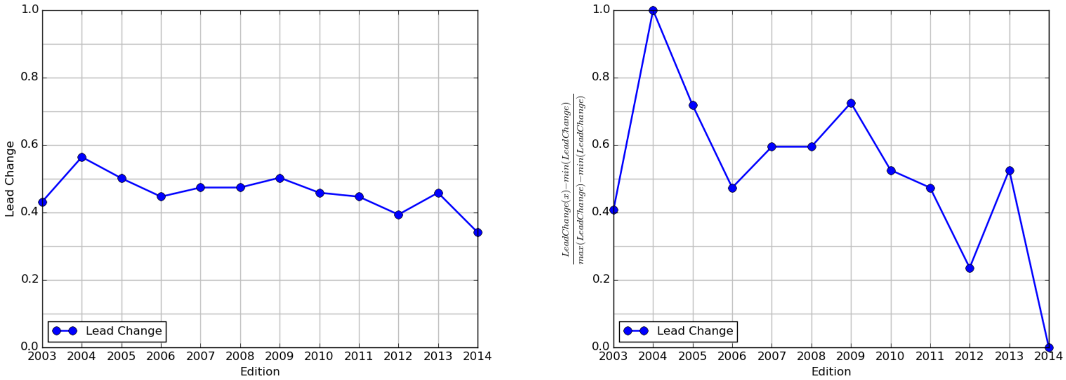
Although there is not a choice concerning which *Lead Change* measure should be chosen, the values computed for this aesthetic criterion are presented below, with a concise discussion.

Table 4.10: Lead Change Values by Championship Edition

Edition	Lead Change	
	value	rank
2003	0.4321	10
2004	0.5645	1
2005	0.5015	3
2006	0.4469	8
2007	0.4740	4
2008	0.4740	5
2009	0.5031	2
2010	0.4586	6
2011	0.4469	8
2012	0.3938	11
2013	0.4586	6
2014	0.3411	12

Table 4.10 shows the values computed by the measure as well as the rank stated by it. The measure assigned the same value to two championship editions, 2007 and 2008. The values are equal even with more precision digits. In situations like that, the higher rank position was assigned for the older edition.

Figure 4.9 shows two plots of the *Lead Change* measure values. Figure 4.9 (a) shows the raw values of the metrics while Figure 4.9 (b) presents their computed values normalized, according to the formula: $y = (Lead\ Change(x) - min(Lead\ Change))/(max(Lead\ Change) - min(Lead\ Change))$. In the former, one can see that Lead Change varies in a small range of values. The latter allows observing the decreasing tendency of the values through time.



(a) Lead Change measure values

(b) Normalized Lead Change measure values

Figure 4.9: Lead Change Measures by Championship Edition

Table 4.11 shows the values for the test of ranks correlation between the rank defined by the Lead Change measure and the one stated by the judges (Table 4.6). This comparison between the ranks suffers the same weakness than the previous one, with the *Uncertainty* measures. The judges received little information about the facet of the aesthetic measured by this metric.

Table 4.11: Kendall tau ranks correlation between judges' rank and Lead Change measures

Measure	Judges' Rank	
	τ	p-value
Lead change	0.6358	0.0040

The small range of variation for this metric can be explained by the cumulative points nature of the game, the limit for scoring per turn, the number of player and the number of turns. Thereby, the majority of players never reach the first place and the number of players that can do it reduces as the game unfolds.

4.5 The Overall Aesthetic Metric

As proposed in section 3, the Overall Aesthetic Metric is an aggregation operation on the fuzzy sets defined by the metrics for the aesthetic criteria *Uncertainty*, *Drama*, and *Lead Change*. In section 4, the metrics for the *Uncertainty* and *Drama* were determined and, now, the membership functions for these criteria can be specified. Thereby, the membership function for the three aesthetic criteria is defined as follow, with the functions redefined for single letters:

Drama $\mu_{Drama}(x) = Drama\ by\ Path(x) = \mathcal{D}(x)$

Uncertainty $\mu_{Uncertainty}(x) = Uncertainty\ by\ PDD(x) = \mathcal{U}(x)$
Lead Change $\mu_{Lead\ Change}(x) = Lead\ Change(x) = \mathcal{L}(x)$

And, as a consequence, the Overall Aesthetic Metric can be specified as in Equation 4.1.

$$Overall\ Aesthetic\ Metric(x) = \frac{\mathcal{D}(x) + \mathcal{U}(x) + \mathcal{L}(x)}{3} \quad (4.1)$$

4.5.1 Overall Aesthetic Metric Validation

The Overall Aesthetic Metric was applied to the data from the Brazilian Football Championship dataset in order to verify its adequacy for evaluating game aesthetics and emulating human choices. The results, likewise the aesthetic metric for each criterion, was compared with the rank stated by the human judges. In addition, the metric was applied to data from Sebrae Challenge dataset to analyse the resulting values' variation through the series and editions.

Table 4.12: Overall Aesthetic Values by Brazilian Football Championship Edition

Edition	Overall Aesthetic Value	
	value	rank
2003	0.3153	11
2004	0.4563	3
2005	0.3901	4
2006	0.3402	10
2007	0.3784	5
2008	0.5074	2
2009	0.5722	1
2010	0.3757	6
2011	0.3468	8
2012	0.3526	7
2013	0.3404	9
2014	0.2756	12

Table 4.13: Kendall tau ranks correlation between judge's rank and Overall Aesthetic Metric

Measure	Judges' Rank	
	τ	p-value
Overall Aesthetic Metric	0.9090	4×10^{-5}

Table 4.12 presents the values computed by the Overall Aesthetic Metric and the rank set by it.

Table 4.13 shows the Kendall tau coefficient of the correlation between the rank built by the Overall Aesthetic Metric and the one defined by the judges, shown in Table 4.6. The great correlation denoted by the Kendall tau value, allied to the very small p-value, may be caused by the small size of the sample. However, all the metrics were developed without targeting any specific game. Moreover, as exposed in section section 3.4, the fuzzy aggregation operation chosen to combine the metrics does not set any differentiation between them.

In Figure 4.10, one can see the evolution of *Overall Aesthetic* levels in distinct series of Sebrae Challenge. The values presented in the graph are the average of the *Overral Aesthetic* evaluated using the developed metric in those matches.

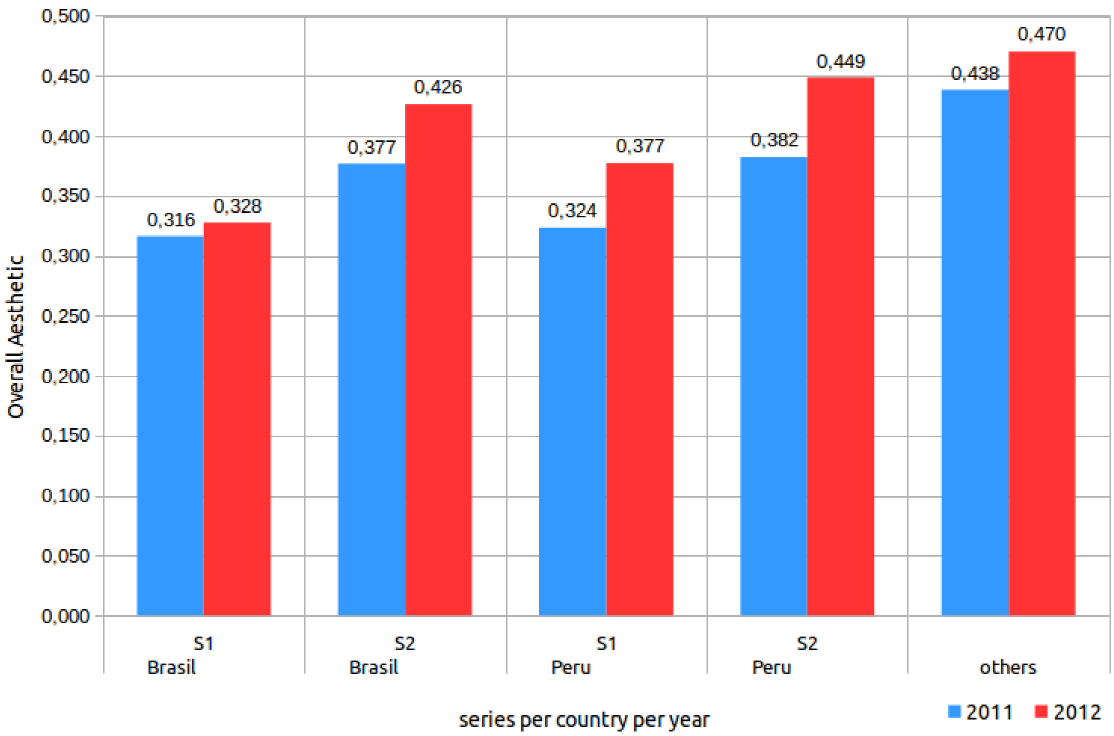


Figure 4.10: Overall Aesthetic Evolution in Sebrae Challenge

There is an increase in the average Overall Aesthetic while the game unfolds. It is an expect game behaviour since better players are engaged in the matches. Also, the number of matches where the winner has the best score in all the turns decreases because the game balance is improved. Therefore, one can see the Overall Aesthetic values in the Brazilian and Peruvian second series higher than in the first ones.

The Overall Aesthetic level evolution through the years can also be seen in the Figure 4.10. For each considered series, even for the subset 'others', one can see the Overall Aesthetic rising. This behaviour is consistent with the game designer's intentions on implementing the changes detailed in 4.2.2.3.

Chapter 5

Conclusion and Future Work

This chapter presents the final considerations of this work. So, it makes a review of the approach used to address the research problem and the research questions, enumerates its contributions, and propounds future work.

5.1 Conclusion

The research problem addressed in this work is if there are aesthetic criteria, previously established to more strict classes of games and measurable by game history analysis, that could compose an aesthetic metric for multiplayer turn-base game.

The problem investigation started defining a theoretical framework that sustains the rationale of aesthetic criteria as a facet of game quality. This formal base also comprises a revision of desirable game behaviors previously established in the literature that reflect fundamental characteristics of games. The theoretical framework was enriched with a generic model of turn-based multiplayer games. This model can comprise a large set of games, in where one can fit games with multiple players, a non-arbitrary range of players' scores, without a common final score, besides establishing a turn as a segment of the game that can be evaluated to analyze the players progress.

The selected desirable game behaviors and their respective metrics were generalized and new metrics suitable to the proposed generic game model were developed.

An *Overall Aesthetic Metric* was described by an aggregation operation in the fuzzy sets defined by the developed metrics. Thereby, the distinct aesthetic concepts of *Drama*, *Lead Change*, and *Uncertainty* could be expressed by a single and meaningful value.

An evaluation framework was developed to deal with two distinct datasets of matches and allow the implementation of the novel metrics. To support data visualization e simple information extraction, two graphical user interfaces were built.

Six novel aesthetic metrics were developed. They were mathematically modeled to extract information about desirable game behaviors from data. The new metrics were analyzed and validated against choices made by humans. Three of them were selected to compose the Overall Aesthetic Metric.

The *Overall Aesthetic Metric* was applied to matches of the both datasets used in this work in order to validate its ability to evaluate game aesthetics.

Thence, the main contributions of this dissertation are:

i The aesthetic metrics defined for multiplayer turn-based games:
- Drama by Points.
- Drama by Positions.
- Drama by Path.
- Lead Change.
- Uncertainty by Entropy.
- Uncertainty by Probability Distributions Distance.

ii The concept of Maximum Drama Path (MDP).

iii The Overall Aesthetic Metric.

iv The formal model for multiplayer turn-based games.

5.2 Future Work

The ideas exposed in this work can lead to other developments using the same design approach. New metrics can be built to reflect other aesthetic criteria or even measures for other facets of game quality.

The evaluation method used to validate the developed metrics can be replicated with a larger set of matches already evaluated by humans. Hence, the ability of the *Overall Aesthetic Metric* for emulating player preferences can be improved .

The information extraction realized by the metrics can be used to automatically evaluate games still under development even discarding gameplay sessions if those games can be formally described and automatically played by computer. This approach is already in use, but the metrics developed in this work, and the generic game model for multiplayer turn-based games, can lead towards an improvement in this field.

Bibliography

ABDI, H., 2007, "The Kendall rank correlation coefficient", *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pp. 508–510.

ABUHAMDEH, S., CSIKSZENTMIHALYI, M., JALAL, B., 2014, "Enjoying the possibility of defeat: Outcome uncertainty, suspense, and intrinsic motivation", *Motivation and Emotion*, (sep.), pp. 1–10. ISSN: 0146-7239, 1573-6644. doi: 10.1007/s11031-014-9425-2.

ALTHOFER, I., 2003, *Computer-aided game inventing.*

AMERICAN SOCIETY FOR QUALITY, 2015. "Quality Glossary - Q - ASQ". `http://asq.org/glossary/q.html` .

BATEMAN, C., 2015, "Implicit Game Aesthetics", *Games and Culture*, v. 10, n. 4 (jul.), pp. 389–411. ISSN: 1555-4120, 1555-4139. doi: 10.1177/1555412014560607.

BROWNE, C., 2012, "Elegance in Game Design", *IEEE Transactions on Computational Intelligence and AI in Games*, v. 4, n. 3 (sep.), pp. 229–240. ISSN: 1943-068X. doi: 10.1109/TCIAIG.2012.2197621.

BROWNE, C., MAIRE, F., 2010, "Evolutionary Game Design", *IEEE Transactions on Computational Intelligence and AI in Games*, v. 2, n. 1 (mar.), pp. 1–16. ISSN: 1943-068X. doi: 10.1109/TCIAIG.2010.2041928.

BROWNE, C., 2008, *Automatic generation and evaluation of recombination games.* Ph.D. Thesis, Queensland University of Technology. `http://core.ac.uk/download/pdf/10885627.pdf` .

BRUNNSTRÖM, K., BEKER, S. A., DE MOOR, K., et al., 2013, "Qualinet white paper on definitions of quality of experience", .

BĚLOHLÁVEK, R., KLIR, G. J., LEWIS, H. W., et al., 2002, "On the capability of fuzzy set theory to represent concepts", *International Journal of General Systems*, v. 31, n. 6, pp. 569–585.

CAILLOIS, R., 1961, *Man, play, and games*. University of Illinois Press.

CAMPBELL, M., HOANE JR., A. J., HSU, F.-H., 2002, "Deep Blue", *Artificial Intelligence*, v. 134, n. 1–2 (jan.), pp. 57–83. ISSN: 0004-3702. doi: 10.1016/S0004-3702(01)00129-1.

CHAN, J. T., YUEN, W. Y., 2008, "Digital game ontology: Semantic web approach on enhancing game studies". In: *9th International Conference on Computer-Aided Industrial Design and Conceptual Design*.

CLAUSET, A., KOGAN, M., REDNER, S., 2015, "Safe leads and lead changes in competitive team sports", *Physical Review E*, v. 91, n. 6, pp. 062815.

COSTIKYAN, G., 2013, *Uncertainty in games*. Mit Press.

DE BAKKER, A. R., DE M PACHECO, B., DIPOLITTO, C., et al., 2011, "Emotions in Business Game: Case Study in Desafio Sebrae". In: *Games and Digital Entertainment (SBGAMES), 2011 Brazilian Symposium on*, pp. 141–147. IEEE.

DE KOVEN, B., 2013, *The well-played game: A player's philosophy*. MIT Press.

FERNÁNDEZ-VARA, C., 2014, *Introduction to Game Analysis*. Routledge.

GRUNDY, A. F., 1997, *Women, work and computerization: Spinning a web from past to future*, v. 6. Springer Science & Business Media.

HIRSCHHEIM, R., KLEIN, H. K., LYYTINEN, K., 1995, *Information systems development and data modeling: conceptual and philosophical foundations*. Cambridge University Press.

HUIZINGA, J., 1949, *Homo Ludens : A Study of the Play-Element in Culture*. Routledge & Kengan Paul.

HUNICKE, R., LEBLANC, M., ZUBEK, R., 2004, "MDA: A formal approach to game design and game research". In: *Proceedings of the AAAI Workshop on Challenges in Game AI*, v. 4.

IIDA, H., TAKAHARA, K., NAGASHIMA, J., et al., 2004, "An Application of Game-Refinement Theory to Mah Jong". In: Hutchison, D., Kanade, T., Kittler, J., et al. (Eds.), *Entertainment Computing – ICEC 2004*, v. 3166, Springer Berlin Heidelberg, pp. 333–338, Berlin, Heidelberg. ISBN: 978-3-540-22947-6 978-3-540-28643-1.

ISO/IEC, 2011. "ISO/IEC 25010:2011 - Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models". http://www.iso.org/iso/catalogue_detail.htm?csnumber=35733 .

JURAN, J. M., DE FEO, J. A., 2010, *Juran's Quality Handbook: The Complete Guide to Performance Excellence 6/e*. 6 edition ed. New York, McGraw-Hill Education. ISBN: 978-0-07-162973-7.

JÄRVINEN, A., 2008, *Games without Frontiers: Theories and Methods for Game Studies and Design*. Tampere University Press. ISBN: 978-951-44-7252-7. http://tampub.uta.fi/handle/10024/67820 .

KLIR, G., YUAN, B., 1995, *Fuzzy sets and fuzzy logic*, v. 4. Prentice hall New Jersey.

KOSTER, R., 2013, *Theory of fun for game design*. " O'Reilly Media, Inc.".

LAUDON, K. C., LAUDON, J. P., OTHERS, 2011, *Essentials of management information systems*. Pearson Upper Saddle River.

MALABY, T. M., 2007, "Beyond play a new approach to games", *Games and culture*, v. 2, n. 2, pp. 95–113.

MESKANEN, T., NURMI, H., 2006, "Distance from consensus: A theme and variations". In: *Mathematics and Democracy*, Springer, pp. 117–132.

MOLLER, S., SCHMIDT, S., BEYER, J., 2013, "Gaming taxonomy: An overview of concepts and evaluation methods for computer gaming QoE". In: *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 236–241, jul. doi: 10.1109/QoMEX.2013.6603243.

NEWBORN, M., 2012, *Kasparov versus Deep Blue: Computer chess comes of age*. Springer Science & Business Media.

PAL, N. R., 1999, "On quantification of different facets of uncertainty", *Fuzzy Sets and Systems*, v. 107, n. 1, pp. 81–91.

ROSS, T. J., 2009, *Fuzzy logic with engineering applications*. John Wiley & Sons.

SALEN, K., ZIMMERMAN, E., 2004, *Rules of play: Game design fundamentals*. MIT press.

SCHELL, J., 2014, *The Art of Game Design: A book of lenses*. CRC Press.

SCHULZE, M., 2003, “A new monotonic and clone-independent single-winner election method”, *Voting matters*, v. 17, n. 1, pp. 9–19.

SHANNON, C. E., 1950, “Programming a Computer for Playing Chess”, *Philosopical Magazine*, v. 41, n. 314 (mar.).

SHANNON, C. E., 2001, “A mathematical theory of communication”, *ACM SIGMOBILE Mobile Computing and Communications Review*, v. 5, n. 1, pp. 3–55.

THOMPSON, J. M., 2000. “Defining the Abstract”. jul. http://www.thegamesjournal.com/articles/DefiningtheAbstract.shtml .

VARGAS, J. A., GARCÍA-MUNDO, L., GENERO, M., et al., 2014, “A systematic mapping study on serious game quality”. In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, p. 15. ACM.

WARMAN, P., 2015. “The Global Games Market: Trends, Market Data and Opportunities”. T11 Global Mobile Data Summit Presentation, sep. http://www.newzoo.com/wp-content/uploads/2011/06/Newzoo_T11_Beijing_FINAL_Public.pdf .

YANNAKAKIS, G. N., SPRONCK, P., LOIACONO, D., et al., 2013, “Player modeling”, *Dagstuhl Follow-Ups*, v. 6.

ZAGAL, J. P., BRUCKMAN, A., 2008, “The game ontology project: Supporting learning while contributing authentically to game studies”. In: *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 2*, pp. 499–506. International Society of the Learning Sciences.