Player Behavior Modeling In Video Games

Norouzzadeh Ravari, Yaser

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

PLAYER BEHAVIOR MODELING

IN VIDEO GAMES

YASER NOROUZZADEH RAVARI

Player Behavior Modeling in Video Games

Yaser Norouzzadeh Ravari

Player Behavior Modeling in Video Games

Yaser Norouzzadeh Ravari
PhD Thesis
Tilburg University, 2021

ISBN: 978-94-6421-344-7
Cover design: Bregje Jaspers
Print: IPSKAMP

Player Behavior Modeling in Video Games

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. W.B.H.J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op dinsdag 22 juni 2021 om 16.00 uur

door

Yaser Norouzzadeh Ravari

geboren te Tehran, Iran.

Promotores
Prof.dr. P.H.M. Spronck, Tilburg University
Prof.dr. E.O. Postma, Tilburg University

Promotiecommissie
Prof.dr. R.C. Veltcamp, Utrecht University
Prof.dr. W.J.A.M. van den Heuvel, Tilburg University
Prof.dr. P.L. Lanzi, Politecnico di Milano
Dr. M. Preuss, Leiden University
Dr. M. Theune, University of Twente

Abstract

In this research, we study players' interactions in video games to understand player behavior.

The first part of the research concerns predicting the winner of a game, which we apply to *StarCraft* and *Destiny*. We manage to build models for these games which have reasonable to high accuracy. We also investigate which features of a game comprise strong predictors, which are economic features and micro commands for *StarCraft*, and key shooter performance metrics for *Destiny*, though features differ between different match types.

The second part of the research concerns distinguishing playing styles of players of *StarCraft* and *Destiny*. We find that we can indeed recognize different styles of playing in these games, related to different match types. We relate these different playing styles to chance of winning, but find that there are no significant differences between the effects of different playing styles on winning. However, they do have an effect on the length of matches. In *Destiny*, we also investigate what player types are distinguished when we use Archetype Analysis on playing style features related to change in performance, and find that the archetypes correspond to different ways of learning.

In the final part of the research, we investigate to what extent playing styles are related to different demographics, in particular to national cultures. We investigate this for four popular Massively multiplayer online games, namely *Battlefield 4*, *Counter-Strike*, *Dota* 2, and *Destiny*. We found that playing styles have relationship with nationality and cultural dimensions, and that there are clear similarities between the playing styles of similar cultures. In particular, the Hofstede dimension Individualism explained most of the variance in playing styles between national cultures for the games that we examined.

Contents

Listing of figures

To my parents,
who are the reason for all the successes in my life!

Acknowledgments

A PhD program normally is challenging in that it requires skills in both doing research and project management. For me, it was unusually challenging. I needed more than one year, filled with a lot of uncertainty, to move to the Netherlands; my research topic was different from my background; and I struggled with some cultural differences. My passion for travelling and understanding the cultural differences in an international environment motivated me to overcome these challenges. The support and love that I have received from my family, friends, and colleagues helped me reach this point.

I am thankful to Eric Postma and Pieter Spronck as my promoters for giving me the opportunity to continue my PhD at Tilburg University and guiding me during these years by providing comments and feedback. Pieter, you always have been supportive and took my personal life matters into account. You agreed that I work remotely when I moved to Amsterdam after my marriage. You gave me the freedom to choose my research direction. As head of department, you had a tight schedule, but you have made time for our meetings even in the evenings. I would like to thank Sander Bakkes as my daily supervisor for some period of time, for his comments and feedback.

I'd like to thank Anders Drachen and Rafet Sifa for our collaborations during my research. It was a fantastic opportunity to work with you. I'd like to thank Drew Hendrickson for the interesting conversations that we had about our research.

I'd like to thank Eva Verschoor and Grace Goossens for facilitating the administrative tasks for me. I'd like to thank my colleagues, Adriana, Angelique, Charlotte, Chris, Chrissy, Debby, Emmelyn, Ingrid, Jan, Akos, Katja, Lieke, Lisa, Maira, Mariana, Mirjam, Nadine, Paris, Parisa, Renske, Rianne, Shoshannah, Tess, Tycho, Yevgen, and Yueqiao, for all of the good memories that we share. I enjoyed our chats and the department events that we participated in. Alexandra, Julie, Moinuddin, Thiago, Wilma, and Yu I thank for all the good times that we spent together.

I'd like to thank Maarten de Rijke for providing me the opportunity to join the top researchers in ILPS in the first year of my PhD at University of Amsterdam. I gained a lot of experience in a short time under his supervision. I'd like to thank Ilya Markov, as my daily supervisor in some period of time, for giving feedback.

I'd like to thank my colleagues and the department staff at University of Amsterdam for their energy and being open to answer my questions: Alexey, Anne, Arianna, Artem, Christof, Christophe, Cristina, Daan, Dan, David, Evgeny, Fei, Hamid Reza, Harrie, Hendra, Hosein, Isaac, Ivan, Ke, Maartje, Manos, Marlies, Marzieh, Mostafa, Nikos, Petra, Praveen, Richard, Ridho, Shangsong, Tom, Xinyi, and Zhaochun.

I would like to thank my friends for their support, especially at the times that the family was far away, and for the joy that we had together. Before moving to the Netherlands, they provided the required information and later helped me by sharing their experiences about living abroad, adapting to a new environment, moving to a new place, and progressing on my career path. Their love and support when I became a father was incredible: Ali and Azadeh, Ali and Bita, Ali and Fatemeh, Ali and Zeinab, Ali Z., Amin and Parisa, Amin and Fatemeh, Amir and Fatemeh, Amirhosein and Irene, Babak and Haleh, Behrouz and Shima, Behrouz and Hedwig, Esmaeal and Elham, Erfan, Giti, Hadi, Hamed and Fatemeh, Hamid and Mina, Hanyeh and Farshid, Hoda and Sina, Hojat and Marzieh, Hosein and Zahra, Hosein and Shaghayegh, Jafar and Leila, Keyvan, Mahdi and Mahdieh, Mahdi and Samira, Mahnaz, Marzieh and Abbas, Masoud and Nasrin, Masoud and Parisa, Masoud and Rozita, Martijn and Niloufar, Mehran, Miad, Mina, Mohammad and Fahimeh, Mohammad L., Mohammad M., Mohammad and Mahsa, Mojtaba, Mostafa and Samira, Naser and Aylar, Nazila, Neda, Negar, Kamal and Vida, Reza and Melisa, Roham and Nasibeh, Rokhsare, Saied and Pardis, Shayan and Narges, Sima, Vahid and Sara, Yasaman, Zahra, and Zeinab.

My biggest thanks go to my family for all the incredible support and love during my life. I'd like to thank my parents who always did their best and encouraged me to follow my dreams. I am always grateful for their endless love and support. I'd like to thank Dr. Rahimi and Khale Maryam for their love, support and especially for their help and advice in difficult situations. I'd like to thank Maryam and Mohsen, and Ali and Zoha for their love, support and especially for arranging matters on behalf of me in my absence.

I'd like to thank my family in-law for their love, help, and warm company that allow me to focus on doing my PhD. I'd like to thank my wife, Fereshteh, who has been extremely supportive. From the bottom of my heart, I would like to say big thank you for all moments that we shared and will share together and being by my side in all good times and hard times. Last but not the least, to my lovely son, Liam: you brought more love and happiness to our life and showed us how beautiful life is. I love to see your progress and I'm proud of you.

1 Introduction

The internet comprises a communication environment for devices and users. Game consoles, GPS-enabled devices, Personal Computers (PCs), and other devices support and entertain people over the internet. To improve user experiences, providers of services over the internet desire to understand the different types of users they facilitate and to predict their interactions. For this they often employ the concept of user modeling.

This research concerns user modeling in games that are played over the internet; we refer to this as player modeling. The fact that video games have access to a large number of details on player interactions, may seem to make the task of constructing of player models relatively easy. However, it is actually quite a challenging task, as one has to select from the large number of features available which to focus on and to interpret what they mean.

In the following sections, we introduce the main concepts in this thesis. In section 1.1, we describe the concept of user modeling. In section 1.2, we describe the concept of behavior data that we will use in this research. In section 1.3, we explain video games. In section 1.4, research questions are proposed. In section 1.5, we explain the outline of this research.

1.1 User Modeling

Fischer [30] describes user models as "models that systems have of users that reside inside a computational environment. They should be differentiated from mental models that users have of systems and tasks that reside in the heads of users, in interactions with others and with artifacts."

In this study, we define a user model as a representation of a user of a system, based on user behavior data. Such data may encompass the user's skills, characteristics, behavioral patterns, and preferences. User models may be used to predict user interactions. A valuable application of user modeling is the improvement of systems and applications by having them adapt automatically to user needs.

User modeling received considerable attention in recent decades [30, 102]. With the growth in usage of internet-enabled devices, gathering user data has gotten increasingly easier over time. Currently, for many systems large amounts of user behavior data are available. Artificial Intelligence (AI) techniques can be used to construct user models based on this data [102].

In this research, we examine user modeling in video games in particular. This domain is discussed next. When user modeling is studied in *video games* it is called player modeling. Player modeling concerns understanding players' behavior through their interactions in a game environment [28, 106].

Yannakakis and Togelius [107] divided player modeling approaches into three types: model-based [45], model-free [103], and hybrid-model [70]. In model-based (top-down) approaches, researchers assume a hypothesis and use observations to test the hypothesis. In model-free approaches, a machine learning model or a statistical model is used to fit the player's behavior data into a model without any initial assumptions. Hybrid models combine features of model-based and model-free approaches. In the present study, we use the model-free player modeling approach. We discuss player modeling in more detail in section 1.3.

1.2 User Behavior Data

User behavior data shows how users interact while using a device. In user modeling, different types of behavior data are studied for different purposes. We divide user behavior data into two categories:

- Physiological/External behavior data: this data concerns measurements such as heart rate and skin conductance which are collected during the time that users interact with the device that they are using. This kind of data is collected by external measurement devices.

- Natural/Internal behavior data: this data concerns measurements such as clicks and scrolling which are employed by the user during their interaction with a device. This kind of data is logged by the device itself; no external devices are needed to collect them.

In this study, we use *natural behavior data* because the models that are developed based on these data can be built without relying on external devices to provide measurements, i.e., they can be built in any environment. Therefore, from hereon, whenever we use the term “user behavior data,” we refer to “natural/internal behavior data.”

User behavior data that can be gathered may be depended on both the application and the device type. Since in this research the application is always an online video game, and the device type is either a PC or a game console connected to the internet, we focus specifically on data that can be gathered using these devices. Hereby we focus purely on behavioral data, and not data that potentially is depended on the device that is used.

1.3 Video Games

Video games are games which employ electronics to create an interactive environment which the player interacts with. Feedback is provided via a video device such as a TV screen or a computer monitor. Video games can be played on a variety of devices, such as smartphones, PCs, laptops, tablets, or consoles. Players use game controller devices including mouse, keyboard, joystick, and touchscreen to interact with the game environment.

In this research, we mainly study two online multiplayer video games: *StarCraft* and *Destiny* (although in one chapter also the games *Dota 2* and *Battlefield 4* are used). *StarCraft* is a Real-Time Strategy (RTS) game; *Destiny* is a hybrid of Role-Playing Game (RPG) and an Action game.

In many video games (and definitely in the ones used in this research), the goal of a player is to win a match. The result of a play session may vary from simply showing who wins and who loses, to a ranking in which players are sorted by their scores.

Players can play differently to win a match. Players make different decisions before starting the match and during the match. They may be able to choose one of multiple character types, and their in-game behaviors vary according to their character abilities, the construction of their team, and their chosen intermediate goals. As a result, different players may show different playing styles in the same match. Therefore, player modeling must encompass a wide variety of information, including the different choices made by players before and during the game. The goal which one wants to achieve using player modeling must be used to restrict which information is used in the player model.

1.4 Problem Statement and Research Questions

As argued above, video game analysis may benefit from player modeling. In particular, player modeling may be employed to make predictions on player behavior, which allows video games to adapt to the players' needs. In our research, we aim to use only natural interaction data, so that the results can be applied widely. Consequently, our problem statement reads as follows:

Problem Statement: *To what extent can we use natural interaction data to create predictive player models in video games?*

To answer the problem statement, we study various aspects of player modeling, in relation to particular goals. We formulated the following Research Questions (RQs):

1. To what extent is it possible to predict the winner of matches using in-game interactions?

2. To what extent is it possible to predict the winner of matches using post-game data?

3. To what extent can different playing styles in a game be distinguished from each other?

4. To what extent can a player model relate a player's profile to playing style?

5. To what extent can a player model relate national culture to playing style?

1.5 Thesis Overview

In this thesis, we use data science methods to analyze big data of player behavior to create player models in the domains video games. We use multiple datasets to answer the different research questions. Details on these datasets follow in the corresponding chapters.

The structure of this thesis is as follows.

In chapter 2, we use the game *StarCraft* to answer RQ 1. *StarCraft* is a Real-Time Strategy (RTS) game with high complexity. In *StarCraft*, the environment is partially observable, and dynamic. Additionally, the player can choose one of three different races that have different functionalities. Beside that, players have many possible actions and units that they can choose. As a result, the complexity of *StarCraft* motivates AI researchers to study different approaches and techniques to create strong AI, analyse the games, and model players [66, 87, 90]. We investigate to what extent for *StarCraft* such models can be used for winner prediction.

In chapter 3, we study winner prediction in *Destiny* and we answer RQ 2. When comparing *Destiny* to *StarCraft*, we note that *Destiny* is a hybrid of a First Person Shooter and a Role Playing Game, while *StarCraft* is a RTS game. Moreover, *Destiny* is a multi-player game with teams of 3 to 6 players, while *StarCraft* is a one-vs-one game. Additionally, *Destiny* includes both win-loss and ranking game modes, while *StarCraft* only include win-loss game modes. These differences motivated us to study winner prediction in *Destiny*.

In chapter 4, we analyze playing styles in *StarCraft* and answer RQ 3. This work extends previous research of other researchers [23, 37] by developing player models based on playing styles in *StarCraft*. To answer RQ 3, we use in-game behavior data of *StarCraft* expert players. Then we develop models to discover variations between playing styles across matchups in *StarCraft*. In this chapter we investigate the relation between playing styles, match-length and win-rate.

In chapter 5, we analyze playing styles in *Destiny* and answer RQ 4. To answer RQ 4, we use post-game behavior data of *Destiny* players. We propose features to represent players' learning behavior across time. Then, we develop models to investigate the variations between playing styles across matchups in *Destiny*.

In chapter 6, we study the relation between national culture and playing styles, and address RQ 5. We use datasets of player behavior in *Destiny*, *Dota 2*, *Counter-Strike*, and *Battlefield 4*. We develop models to predict nationality based on playing style. Moreover, we investigate the relationship between national culture and playing style, to determine whether there are significant differences between the playing styles of different countries and regions in the world.

In chapter 7, our work is discussed briefly. We reach conclusions based on our studies in chapter 8, where we combine the answers of the research questions to answer the problem statement.

This chapter tackles research question 1 on winner prediction in *StarCraft*. It is based on the original work listed below.

Research Question

1. To what extent is it possible to predict the winner of matches using in-game interactions?

Original Work:

Norouzzadeh Ravari, Y., Bakkes, S., and Spronck, P. (2016). Star-Craft winner prediction. In Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE) [58].

2

Winner Prediction in *StarCraft*

In this chapter, we develop models to predict the winner in *StarCraft* matches by employing user behavior data. Next, we analyze the most important features that influence the probability of winning.

Among AI researchers, RTS games have been a popular research domain in the past decade. In particular, the complex, partially observable, and dynamic environments of RTS games motivate AI researchers to study different approaches and techniques to create strong AI, analyzing the games and modeling players. Winner prediction is a highly relevant topic of AI research. In *StarCraft*, winner prediction is challenging because players have many action choices, in a discrete environment where players manage their units concurrently. Moreover, the strategy of players depends on the match type. This increases the complexity of winner prediction.

Our work investigates to what extent it is possible to predict the winner of a match, regardless of the match type and the character types that are involved. Two groups of models are presented for predicting match results: One group predicts match results for each individual match type and the other group predicts match results in general (mixed match types), without considering specific match types. We compare the prediction performance of developed models to figure out to what extent different match types need different models. We also analyze performance metrics and their influence on each model.

In the following sections, we explain our study in winner prediction. In section 2.1, we explain player modeling in RTS games. In section 2.2, we introduce a dataset of *StarCraft* that we used in this study. In section 2.3, we explain used features for winner prediction in *StarCraft*. In section 2.4, we explain our winner models. In section 2.5, we discuss top features in *StarCraft* winner prediction. In section 2.6, we discuss the results of our models. In section 2.7, we present our conclusion.

2.1 Related Work

In our research, we build a model of *StarCraft* players. This is a challenging task, as RTS games have a very large state space [71] and are only partially observable [63].

Player modeling encompasses a player's in-game behavior [41, 64, 71, 106] including actions, skills, and strategies. Player modeling in RTS games has been studied from different perspectives. Gagné et al. [35] used telemetry and visualization to understand how players learn and play a basic RTS game. They reported that their approach does not suffice to understand players. Since RTS games are partially observable, not all behaviors of an opponent can be known at all times. To model the opponent, different techniques have been used. Schadd et al. [75] classified an opponent's playing style and strategy in the RTS game

Spring. They found it difficult to determine the opponent strategy in the early game. Dereszynski et al. [21] successfully used a statistical model for predicting opponent behavior and strategy in *StarCraft*.

Multiple researchers have investigated detection of player skills in RTS games. Avontuur et al. [4] built a model to determine a player's *StarCraft* league based on observations of player features during the early game stages. Thompson et al. [94] examined the differences between player skills across the leagues. They reported that experts have automated many behaviors, i.e., the higher a player's skill, the less control they need to spend on basic game tasks, and thus have room to develop other skills.

Park et al. [66] and Hsieh and Sun [43] predicted opponent strategy by analyzing build orders. Synnaeve and Bessiere [90] presented a Bayesian model to predict the first strategy of the opponent in RTS games. Hsieh and Sun [43] used case-based reasoning for this purpose. They managed to model different strategies that could then be recognized. They did this for all three playable races in *StarCraft* (Protoss, Terran, and Zerg). On a limited winner-prediction scale, Stanescu et al. [87] showed that the winner of a small battle in *StarCraft* can be predicted with high accuracy. Bakkes et al. [6] predicted the outcome of the RTS game *Spring* using the phase of the game. Hsu et al. [44] utilized an evolutionary method to predict the winning rate between EISBot and human players for ZvZ (Zerg vs. Zerg), ZvT (Zerg vs. Terran), and ZvP (Zerg vs. Protoss) match types. They formulated the winner prediction as an optimization task. Their approach achieved 61% accuracy on average for ZvZ and less than 2% for ZvT and ZvP.

Predicting match up outcome is more challenging than combat outcome. During the match up, players lose their units or buildings during combats that affect the match up outcome. Meanwhile, the number of units and their locations changes and thus the player has to adjust their strategy. If a suitable prediction model can be built, an interesting application would be the possibility of game personalization. Moreover, it can be used as an evaluation function to design AI bots that behave like human players.

Closest to what we intend to do with our research is the work by Erickson and Buro [27], who used state evaluation to predict the winner of a *StarCraft* match in human vs. human play. They limited themselves to matches between Protoss players in games of a particular length. In contrast, in our work, we investigate all races, in all possible match ups, with less limitation on game length.

2.2 StarCraft Dataset

StarCraft has been a popular RTS game since 1998. The game play is explained in B.1. *StarCraft* includes three different playable races: Terran, Zerg, and Protoss. The player chooses one of the races to play at the start of a match. Figure 2.2.1 shows a battle between Terran and Zerg.

We used the dataset that was provided by Robertson and Watson [72] that is publicly available at `https://github.com/phoglenix/ScExtractor`. This dataset has been created based on human vs. human replays from professional players that were collected by Synnaeve and Bessiere [91]. The database includes replay data and state information provided by the Brood War API (BWAPI).

Table 2.2.1 shows the number of replays for each match type. We filtered the replays to exclude replays with a length of less than 10 minutes to have reasonable data for feature extraction. Also, we removed replays with a length of more than 50 minutes, in order to limit the diversity of the replays' length.

The dataset also included a race indicator. After the filtration, we collected 24k, 9k, and 9k samples for PvT, PvZ, and TvZ respectively.

Figure 2.2.1: Terran vs. Zerg

For symmetric matches, we have 3K, 1K, and 4K samples for PvP, ZvZ, and TvT respectively.

We computed the fractions of victories in non-symmetric matches in our dataset. The results show Protoss won a fraction of 0.55 of the matches vs. Terran and 0.51 vs. Zerg. The winning rate of Terran vs. Zerg was 0.56. This implies that the winner/loser classes are balanced in our dataset with respect to the percentage of winning in different match types.

2.3 Used Features

In this section, we explain how features are extracted from the dataset. The features are time-dependent or time-independent. The list of pro-

Table 2.2.1: Number of replays in the used database [72].

Race	PvT	PvZ	TvZ	PvP	ZvZ	TvT
Number of replays	2017	840	812	392	199	395
Number of replays(After filtering)	1490	579	612	263	115	298

posed features is summarized in table 2.3.1. The time-dependent features are extracted for each player in 10-second intervals. For instance, we extracted *unspent resources* and *income* as follows [27]: R_t is the total of resources (minerals and vespene gas) at time t (increments in intervals of 1 second), and T is the passed time in seconds (T always being a multiple of 180 seconds). The *unspent resources* U (i.e., how many resources are available on average at any given time) are calculated as:

$$U = (\sum_{t=1,2,...,T} R_t)/T$$

The *income* I is computed as the total resources R_{tot} collected over time T, averaged per second:

$$I = R_{tot}/T$$

For each time-dependent feature, over the last 3 minutes we calculated the mean, the variance, and the difference between the two players. For instance, let b_t denote the number of build commands during t, t being a multiple of 10 seconds. Then, B_T is an array of b_t during last 180 seconds: $B_T = [b_{t_1}, b_{t_2}, ..., b_{t_{18}}]$. We computed $mean(B_T)$ and $var(B_T)$. In addition, if b_{A_t}, and b_{B_t} are number of build commands for player A and B during 10-second interval t, the difference between players A and B in the number of build commands for the past 180 seconds is calculated as:

Table 2.3.1: Proposed features

Time-dependent	Time-independent
move	number of regions
build	buildable ratio tiles
tech	walkable ratio tiles
hold	average of choke distances
siege	height levels ratio
burrow	map dimension
micro	
macro	
control	
strategy	
tactic	
unique regions	
region value	
commands diversity	

$$d_T = \sum_{t=T-180}^{T} (b_{A_t} - b_{B_t})$$

Time-Dependent Features Expert players use time more efficiently when they play *StarCraft* [94]. To capture the skills of players in this regard, we used the following features. We counted the frequency of commands for each match type, and we found that the most frequent commands include: *move, build, tech, hold, siege, and burrow*. The order of command frequencies differs across the match types.

We categorized the commands into *micro* and *macro* commands. A command is considered *micro* if it does not cost minerals or gas; otherwise, it is considered *macro*. Then, we computed the number of *micro* and *macro* commands during every 10 seconds for each player.

Inspired by the work done by Ontanón et al. [63], we put the commands in one of three categories: *control, strategy*, and *tactic* commands. We computed the number of commands in each category for 10 seconds intervals per player.

Regions are extracted by the method that authors in [68] proposed. A region includes adjacent walk-able tiles that do not include choke points. We counted the number of *unique regions* that have a building for a player during every 10 seconds. The game assigns buildings different values. For a player, we also stored the sum of the building values minus the sum of the opponent's building values as *region value*.

Time-Independent Features To study the effect of maps on the winner prediction, we recorded some features that reflect the static characteristics of the map. The size of the map is indicated by the total number of regions.

Maps contain different areas, including; *build-able areas, walk-able areas*, and *the average of choke distances*. The height of an area is one of six different *height* levels. For each map, we counted the number of *build-able* tiles, and we computed the ratio of the total number of build-able areas to the total number of tiles.

We did the same for the other types of areas. Since maps have different dimensions, we included the dimension of the map in terms of length and width as number of tiles.

2.4 Winner Prediction Model

In this section, we explain our winner prediction models across the *StarCraft* races. *StarCraft* is a zero-sum game, but in some matches there is no winner in our replays. Therefore, we filtered the matches that do not have a winner, and we represent the winner prediction as a binary classification problem: win(1), and lose(0).

We follow two approaches for winner prediction: individual models for each match type, and mixed models. The individual models include six binary classifiers for PvT, PvZ, TvZ, PvP, ZvZ, and TvT matches. We used P, T, and Z for Protoss, Terran, and Zerg races respectively. The mixed models include the following tree binary classifiers: a model for non-symmetric matches (PvT, PvZ, and TvZ), a model for symmetric matches (PvP, ZvZ, and TvT), and a general model for all matches.

We employed two classification methods: Gradient Boosting Regression Trees (GBRT) [33] and Random Forest (RF) [13]. GBRT uses an ensemble of weak learners, such as regression trees, and optimizes a loss function to generalize them. GBRT is robust to outliers, can handle combined type features, does not need to normalize the inputs, and can handle non-linear dependencies between the feature values and the outputs. RF also is an ensemble learning method, which uses decision trees for prediction. GBRT and RF have been used successfully for prediction tasks in video games [25, 82, 83].

We did 10-fold cross validation on the samples. To avoid bias, for any match the samples are either in the training set or in the test set, but not in both. We used the Scikit-learn package in Python [67] for developing our models.

In the following sections, we present the results of our approaches for winner prediction in *StarCraft*. In section 2.4.1 individual models are presented. In section 2.4.2 mixed models are presented.

2.4.1 Prediction Per Match Type

In this section, we explain our models that are developed for each individual match type. The winner prediction results across the match types are summarized in table 2.4.1. The table also includes the baseline victory fractions. The baseline represents the majority winning rate in all match types according to our dataset. The performance of the models is presented in terms of accuracy. The features are grouped into three categories: Category A contains actions per minute (APM), income, and unspent resources, category B contains time-dependent features (features measured during a particular time slice), and category C contains time-independent features.

We compared the performances in two cases: modeling using all mentioned features (A, B, C), and modeling excluding time-independent features (A, B). The reason to exclude the time-independent features from the second modeling approach is that player's strength, and therefore chances at victory, tend not to be influenced by static map features, which are the core of category C.

We attempted to improve the results for both approaches by employing RF for feature selection, but we did not observe a significant improvement in the predictions. Therefore these results are left out.

From table 2.4.1 it can be observed that with the (A, B, C) modeling approach, a small improvement to winner prediction over the baseline can be achieved for PvT and PvZ matches (for the PvT matches, a *very* small improvement). No improvement is achieved for the other matches. However, for the (A, B) modeling approach, a considerable improvement of winner prediction over the baseline is achieved for all

Table 2.4.1: Winner prediction performance across non-symmetric matches in terms of accuracy. A=APM and economy features, B=time-dependent features, C=time-independent features

Model	Features	PvT	PvZ	TvZ	PvP	ZvZ	TvT
baseline		0.55	0.51	0.56	0.50	0.50	0.50
RF	A,B,C	0.59	0.61	0.50	0.50	0.51	0.49
GBRT	A,B,C	0.59	0.62	0.50	0.50	0.50	0.48
RF	A,B	**0.64**	**0.63**	**0.62**	**0.64**	**0.58**	**0.64**
GBRT	A,B	**0.64**	**0.63**	**0.62**	**0.64**	**0.58**	0.63

match types.

From these results, we see that time-independent features seem to have a negative effect on most predictions. Thus, we may assume that the inclusion of map properties in the feature set leads to detrimental results of the classification. Since our dataset contains mainly replays of expert players, it seems that they are capable of incorporating map properties in their playing style, regardless of match type.

2.4.2 Prediction For Mixed Match Types

As we mentioned earlier, the winner prediction is possible across the match types by individual models. In the next step, we are interested to see how accurately we can predict the match results when we mix the races. Therefore, we employed three mixed models: one for non-symmetric match types, one for symmetric match type, and one for all match type (general model).

The prediction performance of the mixed models is shown in table 2.4.2. The first two rows represent the performance of the models that use all features, while the last two rows show the performance for the models without time-independent features. The table shows a similar result as found for the models for the individual match types: when all features are included, the models do not perform well, while when time-dependent features are removed from the data, all models perform reasonably well with an accuracy of more than 63%, even for the generalized model that predicts the results for all match types.

2.5 Top Features

In this section, we explain the top features in *StarCraft* winner prediction. GBRT provides relative importance of features expressed as values in the range [0, 1]. The bigger number shows higher importance.

Table 2.4.2: Winner prediction performance across mixed match types in terms of accuracy. A=APM and economy features, B=time-dependent features, C=time-independent features

Model	Features	NonSym	Sym	General
RF	A,B,C	0.57	0.50	0.59
GBRT	A,B,C	0.58	0.50	0.59
RF	A,B	**0.64**	**0.64**	**0.64**
GBRT	A,B	0.63	0.63	0.63

Table 2.5.1 presents the relative importance of the top 10 time-depended features for individual models, of which the results are given in table 2.4.1 as the models for feature sets (A, B). The importance rates are given between parentheses.

Our feature set includes three variations of features (mean, variance, and the difference between players). For the top feature list, we ignored variations of the features. For instance, if mean and variance of income are amongst the top features, we only included 'income' on the list once; however, we summed the importance rates for the different variations of a feature and ranked them by these sums.

Generally, most features have some predictive value for each of the match types, and when examining the rankings, we see that they tend to be ordered similarly across the match types, with some notable exceptions. Income and unspent resources are always amongst the top three features for all match types. This shows that having a strong economy is an important element to win a match for any of the match types.

The biggest exceptions are found for the ZvZ matches. In ZvZ, *micro* commands have a stronger predictive value compared to the other match types. According to the table 2.5.1, while the importance rate of *micro* commands (0.233) in ZvZ is close to the importance rates of

Table 2.5.1: Top time-depended features per match type

PvT	PvZ	TvZ
Income (0.203)	Income (0.189)	Income (0.198)
Unspent (0.141)	Unspent (0.157)	Unspent (0.140)
Micro (0.094)	Micro (0.129)	Micro (0.140)
Control (0.091)	Control (0.096)	Control (0.095)
Region value (0.076)	Region value (0.080)	Region value (0.067)
Unique regions (0.052)	Unique regions (0.035)	Unique regions(0.044)
Builds (0.020)	Slice (0.027)	Race (0.027)
Slice (0.020)	Race (0.024)	Slice (0.027)
APM (0.017)	Unique commands (0.017)	Burrow (0.018)
Unique commands (0.016)	Burrow (0.016)	Unique commands(0.012)
PvP	**ZvZ**	**TvT**
Income (0.219)	Micro (0.233)	Income (0.206)
Unspent (0.201)	Unspent (0.229)	Unspent(0.192)
Micro (0.174)	Income (0.217)	Micro (0.161)
Control (0.140)	Control (0.092)	Control (0.134)
Region value (0.033)	Region value (0.031)	Region value(0.032)
Unique regions (0.030)	Slice (0.022)	Unique regions(0.027)
Slice (0.017)	Unique commands (0.012)	Slice (0.025)
Builds (0.013)	Tech (0.008)	player distance (0.019)
Unique commands (0.009)	Unique regions (0.008)	Unique commands (0.015)
APM (0.009)	Tactics (0.007)	APM (0.011)

income (0.229) and *unspent* (0.229), in the other match types *micro* commands are placed in the third rank of the top features, and have a considerably lower importance rate. This shows that ZvZ matches have to be approached by the players in a different way than they approach the other match types.

Control and *region value* are strong predictive features across all match types. *Control* commands are issued on a unit, and include move, gather, build, and repair; i.e., they are a combination of micro and macro commands. They reflect the general process of enriching the economy and spending resources on buildings. *Region value* is the difference between the values of the players' buildings during the specified time interval. I.e., it reflects how the resources are spent to construct

Table 2.5.2: Top time-depended features for mixed models

Non-Sym	Sym	General
Income (0.181)	Income(0.184)	Income (0.177)
Unspent (0.118)	Unspent(0.150)	Region value (0.112)
Region value (0.107)	Micro(0.138)	Unspent (0.104)
Control (0.074)	Control(0.118)	Control (0.079)
Micro (0.074)	Region value (0.044)	Micro (0.071)
Unique regions (0.062)	Unique regions (0.043)	Unique regions (0.066)
Race (0.028)	APM (0.019)	Slice (0.023)
Slice (0.025)	Slice (0.019)	Race (0.023)
APM (0.017)	Unique commands (0.016)	APM (0.020)
Unique commands (0.017)	Builds (0.012)	Unique commands (0.018)

buildings.

The top 10 features, with their importance rates, for each of the mixed models that do not include time-independent features, are given in table 2.5.2. The importance rates are presented in parenthesis.

Income is the most predictive feature for all of the mixed models. For the non-symmetric and symmetric match types, again *income* and *unspent* are the most predictive features. For the mixed models, *unspent* is moved to the third place in the ranking, while *region value* is in second place – however, the importance of *unspent* is still very close to the importance of *region value*. This means that for all match types, economic features play a decisive role in determining the match outcome.

From the table, we can see that the top six features are the same for each of the combined match types, though they sometimes appear in a slightly different order. We also see that of these six features, for symmetric matches, there is a considerable gap between the importance of the top-4 features, and the features on the fifth and sixth place. For the other two combined match types, that gap is found between the sixth and seventh ranked features. From this we conclude that *income*, *unspent*, *micro*, and *control* are the most important features overall, while in non-symmetric matches *region values* and *unique regions* also play a role in determining the match outcomes.

2.6 Discussion of the Results

From the results of winner prediction in *StarCraft*, we conclude that including time-independent features in the dataset actually has a detrimental effect on the classification algorithms, creating classifiers that perform worse than those created using a dataset without these time-independent features. We offer the following explanation for this observation: Each match is divided into multiple time-slices (180 seconds); each slice from a match has the same winner, and also exactly the same time-independent features and thus, there are correlations among several samples in the training set. Therefore, a classification algorithm may uncover a strong relationship between these time-independent features and the ultimate winner. However, since the time-slices of each match are stored only in one specific fold for the evaluation, in the fold that is used as test the relationships found in the folds used for training are non-existent. Therefore, the inclusion of time-independent features creates classifiers that work well on a training set but not as well on a test set.

We surmise that there still might be an interesting relationship between time-independent features and the ultimate winner of a match, but such a relationship cannot be found using our approach with match slices. A separate classification run using a dataset that only stores features of complete matches may uncover such relationships.

As for the individual features, we see that the general class of *micro* features ranks fairly high in victory prediction, but that the two most important features (*income* and *unspent*) for winner prediction are both *macro* features. Therefore, we conclude that while micro commands are important for winning *StarCraft* matches, the strategic and tactical aspects of *StarCraft*, which are exemplified by macro actions, have more importance overall.

2.7 Chapter Conclusion

In this chapter, we studied the winner prediction of a matches across *StarCraft* races using individual and mixed models for match types.

The individual models for match types show that winner prediction is possible for all of the match types, with an accuracy of 63% or higher for all match types except ZvZ, as long as only time-dependent features are included in the dataset. Moreover, we designed more general models that contain non-symmetric match types, symmetric match types, and all match types. The results show that these mixed models manage to predict the match winner, also with an accuracy of 63% or higher.

Our work is the first work in comparing the performance of winner prediction across the races and analyzing the relative importance of the features in this task. For all classifiers, the top-10 features used for prediction are more or less the same, with economic features having the highest predictive value in all cases, followed by micro commands.

Our results improve considerably on previous work done in this area, where only symmetric matches were used, and where accuracies achieved were much lower than we managed to find. Further improvements might still be possible, if more detailed features of matches are incorporated.

This chapter tackles research question 2 on winner prediction in *Destiny*, and it is based on the following original work.

Research Questions.

2. To what extent is it possible to predict the winner of matches using post-game data?

Original Work: Norouzzadeh Ravari, Y., Spronck, P., Sifa, R., and Drachen, A.(2017). Predicting victory in a hybrid online competitive game: The case of destiny. In Proceedings of the Thirteenth Artificial Intelligence and Interactive Digital Entertainment International Conference (AIIDE), pages 207–214. York [61].

3

Winner Prediction in *Destiny*

In this chapter, we develop models to predict a winner of a match in *Destiny* by employing user behavior data. Next, we analyze the most important features that influence the probability of winning.

In chapter 2, we discussed winner prediction in *StarCraft* [58]. Winner prediction in *Destiny* differs from winner prediction in *StarCraft* in multiple aspects: first, the game genres are different; *Destiny* is a hybrid of a Massively Multi-player Online Role-Playing Game (MMORPG) and a First-Person Shooter (FPS) game, while *StarCraft* is a RTS game. Second, *Destiny* is a multi-player game, which means that players function in a team, while *StarCraft* is a one-vs-one player game. Third, *Destiny* includes both ranking by score and win-loss game modes, while *StarCraft* only includes win-loss game modes.

Our work investigates to what extent it is possible to predict the winner of a match in *Destiny*, regardless of the game mode and the character types that are involved. Two groups of models are presented for predicting match results: One group predicts match results for each individual match type and the other group predicts match results in general (combined models), without considering specific match types. We also analyze performance metrics and their influence on each model.

In the following sections, we discuss our study in winner prediction for *Destiny*. In section 3.1, we discuss winner prediction in video games. In section 3.2, we introduce the dataset of a *Destiny* that we used in this study. In section 3.3, we explain the features used for winner prediction in *Destiny*. In section 3.4, we build our winner prediction models and discuss our results. In section 3.5, we present our conclusion.

3.1 Related Work

This work deals with match result prediction in the game *Destiny*. Winner prediction (sometimes referred to as victory prediction or match result prediction) concerns analyses in the domain of electronic sports (esports). In esports, the most popular game genre to which winner prediction is applied, is the Multiplayer Online Battle Arena (MOBA), which features such games as *DotA2*, *League of Legends*, *StarCraft* and *Destiny*.

Winner prediction in competitive games has been studied primarily from two perspectives:

1. AI-driven work in multi-player video games for the purpose of developing AI players. For example, Cole et al. [17] developed AI bots at expert level for the FPS *Counter-Strike*.

2. Behaviorally driven work in esports for the purpose of providing knowledge to players and teams. For example, Schubert et al. [78] developed encounter-based models for evaluating MOBA matches

and predicting match results. Additionally, player behavior has been investigated from a broad array of perspectives across scientific disciplines. For example, Schatten et al. [76] proposed an agent-based model to study the organizational behavior of players.

The prediction of match- and combat outcome or match results (winners) has been the focus of research across different genres of games, notably RTS games. For example, Bakkes et al. [6] utilized match status in different phases to predict the match result in *Spring*. Yang et al. [105] investigated common patterns of winning teams in combat tactics. Erickson and Buro [27] used players' features and battle information to predict match results in *StarCraft*. Our own work discussed in chapter 2 concerns winner prediction in *StarCraft* as well. In contrast to our work in chapter 2, the focus in the present chapter is on the prediction of match results within and across multiple different game modes of *Destiny* using post-game scores.

In esports, winner prediction forms a key focus in the limited literature that is available, summarized by Schubert et al. [78]. While there has been almost no work on winner prediction in FPS games outside of the broader esports community, analytics for MOBAs has been the focus of more than a dozen publications. The consensus is that winner prediction is possible but there is as yet no substantial body of publicly available work to compare performance results [78, 105].

3.2 Destiny Dataset

The *Destiny* dataset that we used includes players' end-match performances from September 2014 to January 2016. In the game, each player can have up to three characters, of three character class types, namely Titan, Hunter, and Warlock. The game play is explained in B.2.

In total, the dataset includes performances of 15,249 characters. The number of characters in Titan, Hunter, and Warlock class types are 4,847, 5,445, and 4,957, respectively. The numbers of samples for Titan, Hunter, and Warlock players are 627,191, 833,251, and 707,141 respectively. Each sample shows a summary of the performance of a character of a player at the end of a match. This information includes the player ID, character ID, class type, date of activity, and more than 1,000 features that represent the player's performance metrics such as the number of kills, deaths, and assists.

In Player-vs-Player (PvP) game modes, two teams of players play against each other. The number of players in each team can vary. Players team up before the match. However, in the dataset team information and match ID are missing, which entails that we are unable to determine which players were in a match together.

In the dataset, the result of a match is denoted by a variable named 'standing.' Its value is an integer in the range 0 to 5. In win-loss matches, the match has a winning and a losing team; the standing value is 0 if the player was part of the winning team, and 1 if the player was part of the losing team. In free-for-all matches, the standing value is in the range 0 to 5, indicating the player's rank, with 0 going to the best player. In this dataset, we have 2 million samples for win-loss matches, and 145,000 samples for free-for-all matches.

3.3 Proposed Features

We used 34 features that were tracked in the game. These represent typical FPS metrics as well as metrics that try to capture the unique elements of *Destiny*, e.g., stats and assists. Table 3.3.1 shows the list of used features, and mean and standard deviation of them in our dataset. Most of the features have a self-explanatory name, but for some, this is less clear. These are explained below:

Stat-agility: affects movement speed and jumps height.
Stat-armor: the higher Armor, the less damage the player will take.
Stat-discipline: affects grenade cooldown time.
Stat-intellect: influences the super cooldown time.
Stat-light: it is the second leveling method for players that reached maximum level and increases output damage.
Stat-optics: influences the zoom capability of the weapon while aiming.
Stat-recovery: shows how fast the player's heath and shields regenerated after taking damage.
Stat-strength: influences the cooldown of melee ability.
Completion reason: multiple possibilities, a.o.: killing all of the opponents, earning specified points, reaching the match time limit, or achieving the objective.
Current progress: earned points.
Leave remaining seconds: remaining seconds of activity, if a player leaves the activity before it ends.
Next level at: required points to reach the next level.
Player count: the number of players in the match.

3.4 Winner Prediction Model

In this section, we explain our prediction models for winner prediction in *Destiny*. For our classification efforts, since win-loss matches have only two possible standing values, we formulated the prediction for win-loss matches as binary classification. In contrast, for free-for-all matches, we used multiclass classification. The player's features are considered inputs, and the standing value the classifier's output. We use the one-vs-all strategy for multiclass classification, because this strategy is computationally efficient and interpretable. In this strategy, one classifier is fitting each class. Thus, the number of classifiers

Table 3.3.1: Mean and standard deviation of the features in the PvP dataset from *Destiny*.

Feature	mean	stdev
stat-agility	2.58	5.21
stat-armor	2.22	6.33
stat-discipline	90.96	149.51
stat-intellect	103.00	168.07
stat-light	76.93	249.50
stat-optics	21.87	64.28
stat-recovery	2.23	5.06
stat-strength	72.96	115.46
activity duration seconds	1504.63	601.82
activity length	668.21	638.78
assists	2.71	3.53
average score per kill	195.85	189.35
average score per life	257.73	186.96
class type	0.78	1.03
completed	0.27	0.91
completion reason	6.32	0.27
current progress	51189.37	316352.80
daily progress	1399.75	50.05
deaths	4.95	10.93
gender type	0.40	0.20
kills	6.63	10.76
kills deaths assists	1.29	1.32
kills deaths ratio	1.18	1.14
leave remaining seconds	25.79	0.39
level	5.03	37.20
minutes played this session	71.83	62.37
minutes played total	22459.69	28466.16
mode	4.34	13.13
next level at	4278.25	2726.71
percent to next level	21.54	8.60
player count	3.30	11.03
progress to next level	2075.96	807.86
weekly progress	5122.14	297.93

is equal to the number of classes. We employed GBRT and RF classification methods. We used the Scikit-learn package in Python [67] for developing our models.

Two groups of models were developed: *combined models* and *individual models*. Combined models predict the match result by ignoring the game modes, while individual models take the game mode into account. We distinguish the following three types of combined models:

- Combined models:
  - Win-loss model: this model predicts the result of win-loss game modes (0 or 1).
  - Ranking model: this model predicts the rank of player for free-for-all game modes (range 0 to 5).
  - Binary-ranking model: this model is a binary version of the ranking model. In this model, we divided the ranks into two groups: the first group includes ranks 0, 1, and 2, while the second group includes ranks 3, 4, and 5.
- Individual models: 13 models predict the match results for each game mode (binary or multi-class, depending on the game mode).

To train the models, we divided data into randomized training (70%) and test (30%) sets, ensuring that a player who is in the training set is not in the test set. For each model, we trained the model on the training set, and we evaluated the model on the test set.

The performance of the models is represented by Area Under Curve (AUC) and average precision. These are two common metrics to show the performance of classifiers in machine learning [31, 32]. AUC is the area under the Receiver Operating Characteristic (ROC) curve. ROC curve represents the true positive rate (recall) against the false positive rate (FPR) for different classification thresholds. Generally, AUC is in

[0.5, 1]. AUC=1 means ideal performance, while AUC=0.5 represents the worst performance.

Average precision is also a common performance measurement in machine learning where there is a ranking. Therefore, we used average precision to show how accurate the ranking model can predict the rank of a player in free-for-all matches.

3.4.1 Combined Models

Table 3.4.1 shows the performance for the combined models. For the binary models, AUC and average precision tend to be close, but for multiclass models average precision is much lower than AUC. As Davis and Goadrich (2006) states, a precision-recall curve provides more insight into the accuracy of ranking problems. Thus, we compare the performances in terms of average precision, and later we discuss the precision-recall curves of our multiclass classification models.

As table 3.4.1 shows, RF models outperform GBRT models in most classification tasks. Thus, we focus on RF models. The comparison of RF models shows that the win-loss model, ranking model, and binary-ranking model achieved 84%, 68%, and 94% average precision respectively. It is not unexpected that a win-loss model outperforms a ranking model, as the ranking model has more classes.

Table 3.4.1: Performance of combined models.

Model	Classification	AUC	avg precision
win-loss	GBRT-Binary	82%	81%
win-loss	RF-Binary	84%	84%
ranking	GBRT-Multiclass	88%	63%
ranking	RF-Multiclass	90%	68%
binary-ranking	GBRT-Binary	**95**%	**94**%
binary-ranking	RF-Binary	94%	**94**%

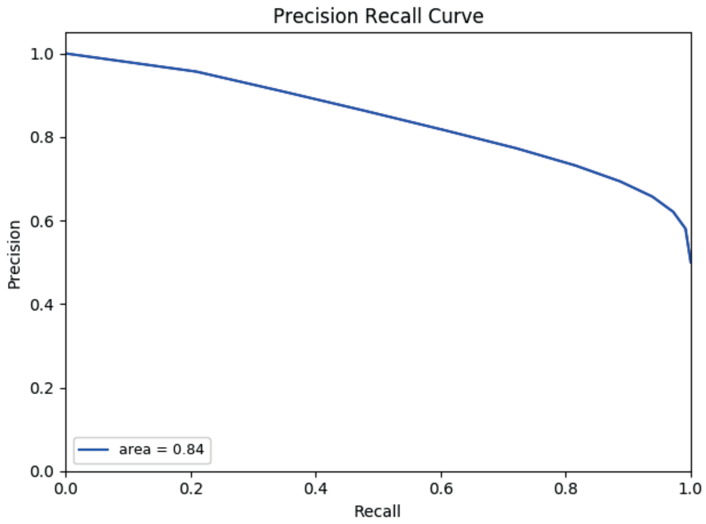

Figure 3.4.1: Precision-recall curve for RF win-loss model.

Figure 3.4.1 shows precision-recall curve for RF win-loss model. As the plot shows, precision starts at 1 for recall 0, and steadily decreases to precision 0.5 at recall 1. In total, the model achieved 84% accuracy in terms of AUC.

Figure 3.4.2 shows the precision-recall curve for the RF ranking model. In this figure, classes 0, 1, and 5 have higher performances compared to the other classes. In other words, the model predicts higher ranks (0 and 1) and the lowest rank (5) more accurately than the mid-ranks 2, 3, and 4.

Figure 3.4.3 shows the performance of the RF ranking models for each class by a normalized confusion matrix. The columns show predicted class labels and rows represent the true class labels [88]. For instance, the value in column 0 and row 1 is 0.05, which indicates that 5% of class 1 samples are assigned to class 0. As we saw before, ranks 0, 1, and 5 are

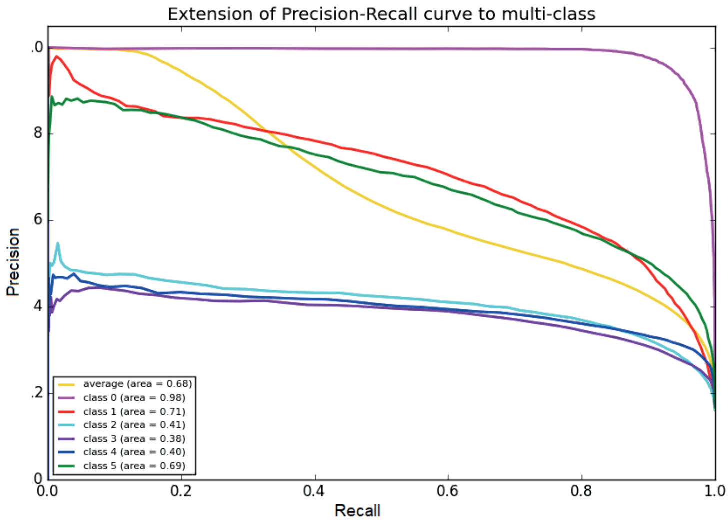

Figure 3.4.2: Precision-Recall curve in ranking model of RF classifier.

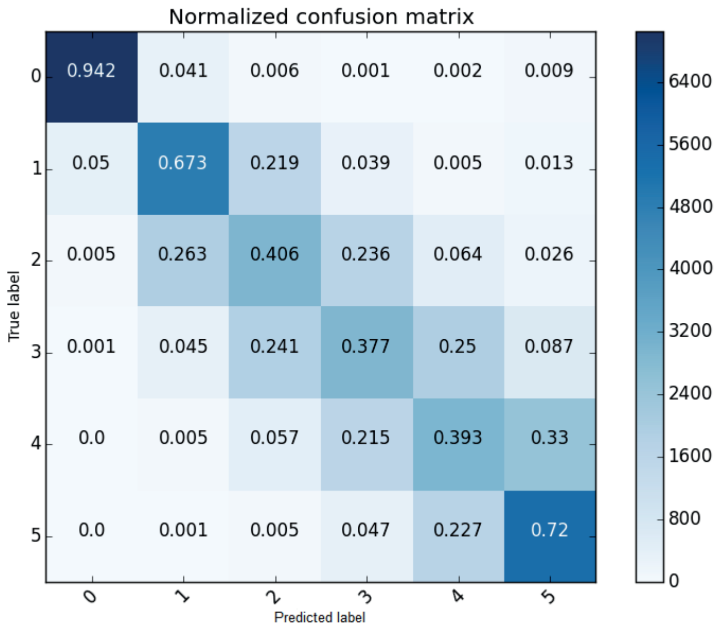

Figure 3.4.3: Normalized confusion matrix in ranking model of RF classifier.

predicted with high accuracy, while ranks 2, 3, and 4 are predicted with lower accuracy. Avontuur et al. [4] predicted league levels of *StarCraft* players with 44% accuracy on average, but with close to 90% accuracy for the best players. Similar to our ranking model, their model has higher performance in prediction of top and low levels, but it has lower performance in prediction of mid-levels.

Feature importance in RF model is calculated according to Gini importance [56]. The top-5 features in the win-loss model and ranking model are similar, but differently ordered. Score-Per-Life (SPL) is of high importance in both models. Kills (K) causes the biggest differences in the top-5 features in these models. While Kills has the highest order in the ranking model, it is not among the top-5 of the win-loss model. In free-for-all game modes (which the ranking model encompasses) getting kills and avoiding being killed are of utmost importance. The third column of table 3.4.2 shows top-5 features for the binary-ranking model. In comparison with the ranking model, Kill-Death ratio (KD) is at the top of the binary-ranking model, SPL is missing, and Team ID is added to the top-5 features. A possible explanation for SPL being of lower importance for the binary-ranking model is that it is mainly used to distinguish between ranks which are close together.

Table 3.4.2: Top-5 features per RF combined models with their relative importance rate. D:Deaths, K:Kills, KD: Kill death ratio, KDA: kills deaths assists, SPL: average score per life, ADS: activity duration seconds, SPK: average score per kill.

win-loss	ranking	binary-ranking
SPL (0.06)	K (0.12)	KD (0.6)
SPK (0.05)	SPL (0.11)	K (0.3)
D (0.04)	KD (0.10)	ADS (0.05)
KDA (0.03)	KDA (0.09)	KDA (0.04)
KD (0.03)	ADS (0.06)	TeamID (0.04)

3.4.2 Individual Models

The dataset includes 13 game modes. Supremacy and Rumble are ranking modes, the others are win-loss modes. Thus, we employed a binary classifier for each win-loss mode and multiclass classifiers for the ranking modes.

Table 3.4.3 shows the performance of individual models in terms of AUC and average precision. Since the performance of the RF classifiers was similar to the GBRT classifiers, we removed the RF models in this table. The comparison of individual win-loss models with the combined win-loss model in table 3.4.1 in terms of average precision shows most of the individual models have higher performance except for the Clash, Rift, and Mayhem Clash models. The low performance of these models may be due to the fact that these three game modes represent 6v6 matches, while all the others represent 3v3 matches. Among the individual models, Trials Of Osiris, Elimination, and Supremacy models have a very high performance, around 99%. This may be explained by the fact that for these game modes the outcome is almost exclusively determined by the number of kills, which is one of the features in our dataset.

While most PvP modes are cooperative, our models were trained on the features of individual players. And yet the models can still predict match outcomes with high accuracy. This may be due to some features representing cooperative performance metrics, such as SPL and KDA. SPL is the sum of scores that a player earned during his life that includes cooperative actions such as assist, revive, and capture a zone. KDA also includes assists. As table 3.4.4 shows, SPL and KDA have an important role in all PvP game modes.

Table 3.4.4 shows top-5 features for individual models. In all of the win-loss models, SPL is the strongest predictive feature, while in ranking models kills is the strongest predictive feature. SPK and deaths

Table 3.4.3: Performance of models per game mode.

Game-mode	Classification	AUC	avg precision
Skirmish	GBRT-Binary	89%	89%
Control	GBRT-Binary	83%	82%
Salvage	GBRT-Binary	94%	94%
Clash	GBRT-Binary	77%	75%
Trials Of Osiris	GBRT-Binary	99%	99%
Doubles	GBRT-Binary	94%	94%
Iron Banner	GBRT-Binary	83%	82%
Elimination	GBRT-Binary	99%	99%
Rift	GBRT-Binary	80%	78%
Mayhem Clash	GBRT-Binary	75%	73%
Zone Control	GBRT-Binary	95%	94%
Rumble	GBRT-Multiclass	86%	84%
Supremacy	GBRT-Multiclass	98%	98%

are also strong predictive features in win-loss models.

Top-5 features in win-loss models are very similar, with different orderings. For the two ranking models, the list of top-5 features is also quite similar. The weight of SPL is especially high for Elimination (0.44), Trials Of Osiris (0.40), and Zone Control (0.40). In these game modes, players must capture a zone or kill all of the opponents. KDA is the most frequent feature in top-5 features after SPL. KDA is found in most of the individual models, except Zone Control and Supremacy models.

In Skirmish, Salvage, Clash, Doubles, and Mayhem Clash models, deaths is the second strongest predictive feature. In these game modes, keeping teammates alive is critical. In general, kills, deaths, KD, KDA, SPL, ADS, and SPK are the most important player's performance metrics in different PvP game modes. Kills, deaths, and KD show how much a player is involved in fighting other players. KDA also reflects the cooperation between team members in addition to kills and deaths.

SPL represents how much the player earned points during his life. Players can earn points from activities other than kills, deaths, and assists, namely actions such as capturing, neutralizing, or defending a zone, and reviving a teammate. Thus, SPL includes scores that are related to cooperation. ADS shows how long players spend time in a match. SPK shows the points that a player gets for kills. A high value may entail that a player often manages to pull off complex kills such as headshots, or kills using melee weapons or grenades. Most of the player's performance metrics that are evaluated in this study are available in the other combined MMO games and FPS games. We expected that a similar approach would work for these games.

To sum up, the results show that winner prediction is possible in *Destiny*. In win-loss matches, the models predict the winner with an accuracy higher than 80%. In ranking matches, where six outcome classes exist, the models' prediction accuracy is at least 68% in terms of average precision.

In ranking matches, top and bottom ranks can be predicted with higher accuracy than the mid-ranks. As expected, the individual models have higher performance compared to the combined models. Interestingly, some of the individual models predicted the match results by 99%, i.e., almost perfectly. The comparison of top performance metrics in win-loss models and ranking models shows that in the ranking game modes, kills is the most important player performance metric to get the best result, while in win-loss modes avoiding to die is more important.

In individual models, the top-5 performance metrics are almost the same, but with different orderings in different game modes. SPL is the strongest predictive feature in both win-loss matches and ranking matches. KDA is the second strongest predictive feature across game modes. Both of these metrics integrate elements of cooperation. Generally, players seem to focus on the actions that earn more points in different game modes (which comes at no surprise).

Table 3.4.4: Top-5 features per RF individual models with their relative importance rate. D:Deaths, K:Kills, KD: Kill death ratio, KDA: kills deaths assists, SPL: average score per life, ADS: activity duration seconds, SPK: average score per kill.

Skirmish	Control	Salvage	Clash	Trials Of Osiris	Doubles	Iron Banner
SPL(0.14)	SPL(0.15)	SPL(0.26)	SPL(0.10)	SPL(0.40)	SPL(0.17)	SPL(0.14)
D(0.14)	SPK(0.13)	D(0.10)	D(0.08)	SPK(0.19)	D(0.15)	SPK(0.13)
KDA(0.90)	KDA(0.07)	SPK(0.09)	KDA(0.07)	D(0.17)	KDA(0.13)	KD(0.07)
KD(0.06)	ADS(0.06)	KD(0.07)	ADS(0.06)	KD(0.07)	KD(0.09)	KDA(0.07)
ADS(0.06)	D(0.06)	KDA(0.06)	SPK(0.06)	KDA(0.04)	ADS(0.06)	ADS(0.06)
Elimination	Rift	Mayhem Clash	Zone Control	Rumble	Supremacy	
SPL(0.44)	SPL(0.12)	SPL(0.08)	SPL(0.40)	K(0.12)	K(0.39)	
SPK(0.17)	KDA(0.09)	D(0.08)	SPK(0.10)	KDA(0.10)	SPL(0.12)	
D(0.10)	D(0.07)	KDA(0.08)	D(0.07)	SPL(0.08)	KD(0.09)	
KDA(0.06)	SPK(0.07)	SPK(0.07)	ADS(0.05)	SPK(0.08)	SPK(0.09)	
KD(0.05)	ADS(0.07)	ADS(0.06)	KD(0.05)	KD(0.07)	D(0.07)	

We point out that with these results we only predict a winner based on features at the end of the game, when this winner is already known. One may think that tracking these features during the game might indicate already early in the game who is going to be the winner. However, in general these games are designed in such a way that these features are easy to influence by the players during the game, and thus they generally cannot be used to predict a winner until very late in the game.

3.5 Chapter Conclusion

In this chapter, we presented our work on winner prediction in *Destiny* using post-game player behavior data. We show that winner prediction in *Destiny* is possible using post-game player behavior data (in the previous chapter we showed that it is also possible using in-game data in *StarCraft*). Moreover, we developed models for individual match types and for mixed match types. We found that both approaches can predict the winner with reasonable accuracy, but individual models outperform mixed models considerably.

We also noticed that for different game modes, predictive features tended to differ, e.g., the number of kills proved to be highly predictive for ranking modes, while not among the important features for win-loss game modes. A possible explanation for this finding is that different game modes require different styles of playing. In the following chapters, we investigate this in more depth, by building models that attempt to distinguish players based on playing style.

This chapter tackles research question 3 on player profiling in *StarCraft*, and it is based on the following original work.

Research Questions.

3. To what extent can different playing styles in a game be distinguished from each other?

Original Work: Norouzzadeh Ravari, Y., Bakkes, S., and Spronck, P. (2018). Playing styles in StarCraft. In European GAME-ON Conference on Simulation and AI in Computer Games [59].

4

Playing Styles in StarCraft

In this chapter, we study player profiling in *StarCraft*. Understanding playing styles in video games may help game designers to create entertaining game content for a variety of players. In Real-Time Strategy (RTS) games, players play differently based on their preferences and their strategy. In our research, we use *StarCraft*, as it is one of the most commonly used RTS games in research. In this study, we present human playing styles across the match types in *StarCraft*, in particular where related to the different playable races in the game. We propose features that reflect playing styles and discover the variations in playing styles by PCA (Principal Component Analysis). We employ K-means clustering to cluster the *StarCraft* players according to their playing styles.

In section 4.1, we discuss related work. In section 4.2, we explain our proposed features. In section 4.3, we explain playing style analysis by PCA. In section 4.4, we explain our clustering model to cluster playing styles. In section 4.5, we explore whether playing styles influence winning rate and/or game-length. In section 4.6, we present our conclusion.

4.1 Related work

Different players employ different playing styles. To distinguish different players, several attempts have been made to define player typologies [98]. Some researchers focused on analyzing the players' in-game behavior and personality. One of the earliest player typologies in game contexts is the Bartle types [7]. Bartle analyzed players of online text-based worlds, called *Multi-User Dungeons* (MUDs). According to his study, a playing style has two dimensions: action vs. interaction and player-orientation vs. world-orientation. Thus, according to the position of the playing style in this two-dimensional structure, the player type would be one of four different possibilities, which Bartle named Achiever, Explorer, Socializer, and Killer.

Yee [108] reported that Bartle types are not general typologies, and that they suffer from biases. He studied motivational factors of the users of Massively Multi-User Online Role Playing Games (MMORPGs), and he presented player typologies in the categories Achievement, Social aspects and Immersion. Zackariasson et al. [110] extended Yee's types for players' buying behavior.

The BrainHex method [55] is a common questionnaire-based approach to characterize playing styles. Toker et al. [96] predicted players' types based on their musical interests. They supposed that playing style is related to preference for music styles, and they classified users in seven types that are measured by the BrainHex method [55]. Si et al. [79] conducted a study on map exploration style in *StarCraft*.

They asked volunteer players to do different missions, answer a questionnaire, and participate in an interview. Players were required to finish a mission in a limited time. The first mission was to explore the whole map in three minutes, the second was to discover and kill some units in limited time, and the last task was to find the enemy base in limited time. They reported four exploring styles that are related to the BrainHex models [55].

The connection between player typology and in-game behavior was made by multiple researchers. Drachen et al. [23] modeled players' behavior in *Tomb Raider Underworld*. They extracted causes of death, the total number of deaths, completion time, and the number of requests for help, to model the players. They observed four playing styles that are labeled as Veterans, Solvers, Pacifists, and Runners.

Other researchers found different playing styles for different games [5, 37, 55, 69, 79, 96]. Ferro et al. [29] posed that one will have to take into account game elements and game mechanics if one wants to determine player typologies and personalities. Thus, they argued that it depends on the game in which playing styles can be distinguished.

That a player can exhibit multiple different playing styles simultaneously, was shown by a number of researchers. Gow et al. [37] studied playing styles in two different games. First, they investigated the playing style in *Snakeotron*, an arcade game with a highly limited state space, and they identified some variety in playing styles. In a second study, they investigated player's behavior in *Rogue Trooper*, a third-person shooter game. Their investigation of combat data shows the variations of playing styles in this game. In RTS games, Bakkes et al. [5] modeled opponents by using a case base of game observations to improve adaptive game AI. Ponsen et al. [69] discovered the players' strategic behavior in online poker. They studied players' behaviors in a series of games, and they showed that players switch between styles according to the match-up conditions. Normoyle and Jensen [57] studied playing styles in *Battlefield 3*, and they found that a player can exhibit multiple playing styles simultaneously.

In this chapter, we use an approach to determine player typologies based on match types. In our research in chapter 2, we found that a general model can predict the winner across all of the match types in *StarCraft* [58]. However, each match type has different unit types and a different possible command set. We propose different feature sets based on match types that exceed 300 features for each player. Features are reduced and analyzed by PCA. We use the top principal components to show how playing styles vary across the match types. We then employ k-means to cluster playing styles across match types.

4.2 Proposed Features

We use the same dataset that was introduced in section 2.2. The game play is explained in B.1. Our analysis of playing styles is done for different match types, as styles are influenced by the match types. The six different match types are distinguished by the races involved in the match. The three races are Protoss (P), Terran (T), and Zerg (Z). There are three non-symmetric match types, namely PvT, PvZ, and TvZ, and three symmetric match types, namely PvP, TvT, and ZvZ. The non-symmetric match types are sometimes listed at TvP, ZvP, and ZvT, respectively when the analysis is done from the perspective of the first-mentioned race.

We distinguish playing styles based on features that are derived from player actions. For each action that a player performs, the dataset includes data for the following fields: player ID, frame number, action type, group ID, and action target position.

Some action types are limited to a specific race, others are available to all races. When a player selects action A, he also selects a unit or multiple units to perform the action. Players select multiple units as a

Table 4.2.1: Feature set dimensions.

	PvT	PvZ	TvZ	PvP	ZvZ	TvT
Number of players	4032	1680	1624	748	398	790
Features	551	479	533	371	380	416

group to manage the tasks that they want to accomplish in the game. Generally, a group can include one or more units that can be selected from different unit types.

We define features per race type. Since possible actions vary based on race type, the number of proposed features vary based on the race type. Table 4.2.1 shows, besides the number of players, also the number of features per match type. Proposed features represent the type, frequency, distance to base, and the number of units that are involved in an action that the player used.

We use a subset of the features for our analyses. Specifically, for each action, the following features are extracted:

- *Action frequency*: For each player, we counted how many times an action is repeated during the match up.
- *Group size*: To see how many units the player associate to an action, we computed the average group size.
- *Group size diversity*: The variance of group size is computed to represent the style of player in creating groups of units that vary during the match up.
- *Number of unique groups*: Players can use a group many times, or they create different groups for different tasks. Thus, we computed the diversity of groups by variance value.

- *Mean and variance of distance between base and target*: The location of units that perform an action gives useful information. For instance, if a player places mines near his base he is probably using his action defensively, while if he places them close to the opponent's base he is using it offensively. For each action, we computed the mean and variance of distance between the primary base and group target. In our dataset, matches are played on varying map sizes. We normalize the distances based on the size of the map. Note that later we found that normalized distance is highly correlated to non-normalized distance, and thus we could have forgone this normalization step.

4.3 Analysis of Playing Styles Using PCA

We did not want to make assumptions on particular player typologies, therefore we used PCA to analyse the features. PCA is widely used for dimension reduction and for discovering discriminative features [99]. We analyzed the features we extracted for all match types by PCA (4 components) to discover player typologies across match types. PCA components with a coefficient below 0.1 we did not consider in our analysis as we wanted to focus on only the strongest features. We found that, together, the components that we did consider cover more than 25% of the variance of features. Top PCA components are the most discriminative features for distinguishing playing styles. In the following sections, we explain the most discriminative features in non-symmetric (different races playing against each other) and symmetric (races playing against themselves) match types.

4.3.1 Analysis of Non-symmetric Match Types

In this subsection, we show the top principal components of playing styles in non-symmetric match types. The principal components that

Table 4.3.1: Derived players' typology in non-symmetric match types.

PCA	PvT	PvZ	TvZ
1	No scan	Burrow	No defensive matrix
2	Mind control	Mind control	No plague
3	Nuke	No carrier attack	No Yamato gun
4	Freeze	Spawn broodling	No cloak

distinguish the two players in a match have been determined. Table 4.3.1 shows the top four principal components for non-symmetric match types. The labels are chosen based on the name of the commands that a player selected during a match up. If a component is negatively related, we precede it with 'No'.

In Table 4.3.1, PvT playing styles can be distinguished by using or not using 'scan', 'mind control', 'nuke', and 'freeze' commands, while in PvZ 'burrow', 'mind control', 'carrier attack', and 'spawn broodling' are the top principal components. PvZ and PvT only have 'mind control' in common. In addition, the top principal components for TvZ do not have anything in common with the top principal components for PvZ and PvT. This raises the question to what extent playing style depends on the race type of the opponent. In section 4.4.2, we will answer this question.

4.3.2 Analysis of Symmetric Match Types

In Table 4.3.2, the top four principal components for symmetric match types are presented. Playing styles in PvP are determined by 'freezing', 'carrier attack', and 'move' commands. In TvT, 'scan', 'restore', and 'move' reflect playing styles. Playing styles in ZvZ are characterized by 'burrow', 'rally', 'move', and 'guard' commands. One of the interesting observations in this table is the 'move' command which affects the playing style in all of the symmetric races.

Table 4.3.2: Derived players' typology in symmetric match types.

PCA	PvP	TvT	ZvZ
1	No freezing	No scan	Burrow
2	No carrier attack	No restore	Rally
3	Carrier attack	Move	No move
4	Move	No move	Guard

4.3.3 Analysis by Race Type

In section 4.3.1, we observed that playing styles in non-symmetric match types vary. In this section, we look at whether playing style for a race type differs based on the opponent. So, for each race type, we determined the top four principal components for when their opponent is Terran, Protoss, or Zerg. Our observations are summarized in Tables 4.3.3 (for Protoss), 4.3.4 (for Terran), and 4.3.5 (for Zerg).

Protoss Player's Typology Table 4.3.3 represents playing styles of Protoss players when they play against Protoss, Terran, or Zerg players. Comparing a Protoss player's playing style in PvP with PvT and PvZ shows that Protoss players in PvP and PvT use different techniques, and the only common style in both of them is using 'carrier attack'. Comparing PvP with PvZ, also the used commands are mostly different, and the exception is 'No freezing', that means some P players in PvP and PvZ do not use 'freezing' commands.

Table 4.3.3: Derived Protoss Player Typology.

PCA	PvP	PvT	PvZ
1	No freezing	Freezing	Freezing
2	No carrier attack	Carrier attack	No freezing
3	Carrier attack	No freezing	No harvest
4	Move	Harvest	Harvest

Table 4.3.4: Derived Terran Player Typology.

PCA	TvT	TvP	TvZ
1	No scan	No scan	No freeze
2	No restore	No heal	Shield
3	Move	Blind	Freeze
4	No move	Blind, no heal	Bunker

Terran Player's Typology Table 4.3.4 represents the T players' playing style across different match types. The comparison of principal components in TvT, TvP, and TvZ shows they do not have any playing style in common. Thus, it appears that T players have more diverse playing styles compared to the P players.

Zerg Player's Typology Z player's typology is summarized in Table 4.3.5. Z players have diverse playing styles in different match types. The comparison of playing styles in ZvZ, ZvP, and ZvT shows they only have one style in common, namely the 'guard' technique. In comparison with P and T players, we observe that the T players have the most diverse playing style, and P players have less diversity in playing styles compared to T and Z players.

To sum up, we found that players who play against different race types have different playing styles in *StarCraft*. Moreover, the playing style is not the same in different match types. For instance, when P

Table 4.3.5: Derived Zerg Player Typology.

PCA	ZvZ	ZvP	ZvT
1	Burrow	No burrow	No plague/scarify
2	Rally	Guard	No reducing speed
3	No move	No move	Guard
4	Guard	Transport	No guard

players play against P players, they use ‘carrier attack’, while P players in PvZ use the more effective ‘freezing‘ commands.

4.4 Clustering Playing Styles

In the cases that a designer wants to customize the game to a group of players, it helps if he can determine discrete groups of players, each group fitting a particular player typology. To automatically determine such groups for *StarCraft* players, we decided to use k-means clustering. We used the Scikit-learn package in Python [67] for developing our model.

We cluster on the features listed in section 4.2. These features use different unit measures, such as pixels for distance and count for group size. Therefore we decided to normalize the feature values.

To get an idea about the number of clusters we can expect to get, we utilized the Calinski-Harabasz criterion [14]. The CH criterion has been used successfully for cluster analysis [53]. With this method, we change the number of clusters and we look at the CH index value. When the value reaches its maximum, we can select the optimum number of clusters. We observed that the CH value has a peak at four clusters for each match type.

The extracted feature set is high-dimensional. We employ PCA for dimension reduction. We found that PCA with two components will cover more than 25% of the variance of features. Therefore we used the top two principal components with k-means from the scikit-learn package in Python for clustering. Figures 4.4.1 to 4.4.6 show the clusters of playing styles across the match types. The component names are those of the most influential feature in the corresponding component (though it is by no means the only feature that determines this component).

For the non-symmetric match types, we examine player typologies based on the principal components using two approaches:

- Principal component analysis of playing styles without considering the race type of the opponent (opponent independent).
- Principal component analysis of playing styles by considering the race type of the opponent (opponent dependent).

4.4.1 Opponent-Independent Playing Styles in Non-Symmetric Match Types

Player typologies in non-symmetric match types are presented in figures 4.4.1 to 4.4.3. The top two principal components for each of the match types are shown on the axes.

We observe that the players from different race types are generally placed in different clusters. For instance, in figure 4.4.1 Protoss players are mostly located in clusters 0 and 1, while Terran players belong to clusters 2 and 3. We see the same in figures 4.4.2 and 4.4.3. Therefore, in Section 4.4.2 we will analyze each race in a non-symmetric match type to discover playing styles in each race.

Although in figure 4.4.1, P players and T players are clustered in different groups, we observe that the members of cluster 2 (Terran, top) and cluster 0 (Protoss, right) are closer to each other compared to the other clusters. The same kind of observation can be made in figures 4.4.2 and 4.4.3.

A comparison between dispersions in the different match types shows that dispersion in PvT is lowest, which means that in a PvT match similarity in playing styles within a race is highest.

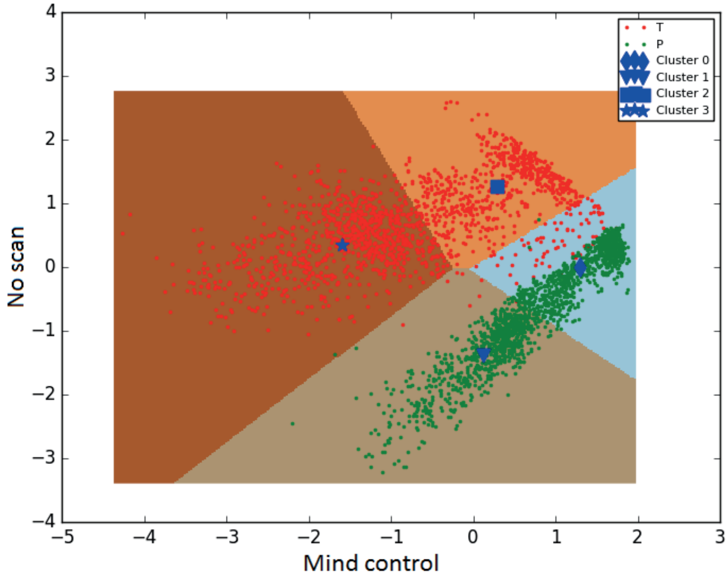

Figure 4.4.1: Observed playing styles in PvT match type by clustering top two principal components.

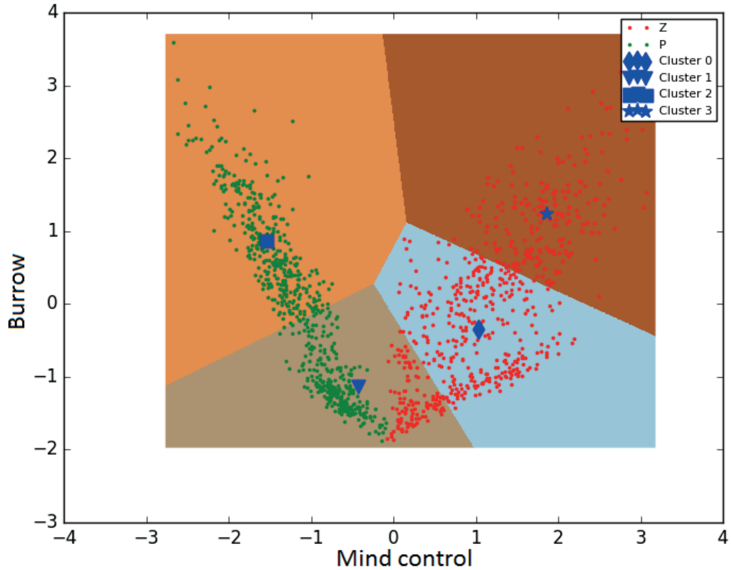

Figure 4.4.2: Observed playing styles in PvZ match type by clustering top two principal components.

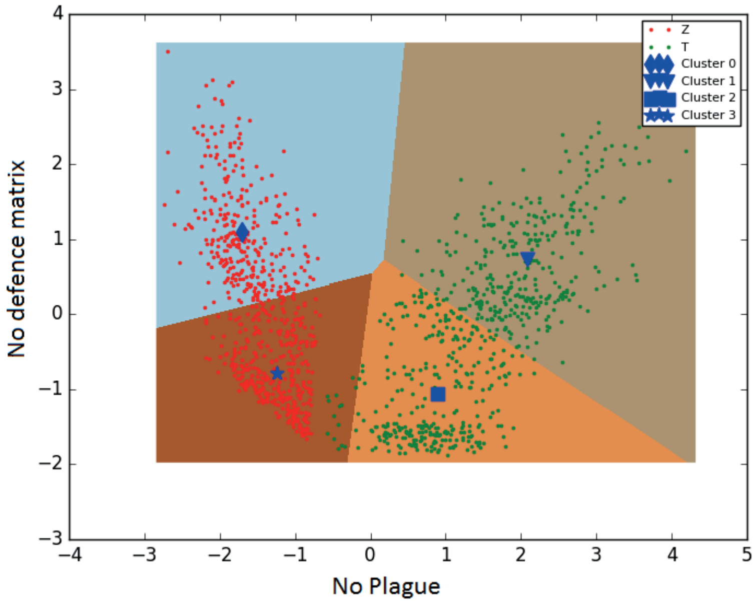

Figure 4.4.3: Observed playing styles in TvZ match type by clustering top two principal components.

4.4.2 Opponent-Dependent Playing Styles in Non-Symmetric Match Types

In the previous section, we observed that different races have different playing styles. To discover playing styles in a race, firstly we separate players of two races in a match type. Then, we utilize PCA to find more informative features for each race. We keep the principal components above 0.1. These coefficients cover more than 37% of the variance of features. Table 4.4.1 shows the top two principal components for each race. We select top-two principal components because they provide good insight into the playing style variations and they cover the variance of features sufficiently.

As can be observed in Table 4.4.1, P and T players change their strategy based on the opponent race type, but Z players use the same strategy regardless of the opponent race type. P players use 'upgrading'

Table 4.4.1: Player typology for each races in non-symmetric match types.

PCA	P in PvT	T in PvT
1	train and upgrade	siege tank
2	freezing	lift
	P in PvZ	**Z in PvZ**
1	research and upgrade	burrow and research
2	psionic storm	no morphing
	T in TvZ	**Z in TvZ**
1	research and train	burrow
2	irradiate attack (surround by radioactive)	no morphing or upgrading

against both T and Z players. T players use 'radiate attack' versus Z players to kill more army units at once, but they use 'siege tank' versus P players that has the greatest attack range. Z players do not use 'morphing' when they playing against P or T players.

In PvZ, P players play differently compared to P players in PvT. Here, the first principal component shows the P players that use 'research and upgrading'. The second group uses 'psionic storm' to kill the opponent's army in a target area. The first group of Z players uses 'burrow' technique to hide underneath the surface, and they use 'research' command. The second group does not use 'morphing'.

The two top principal components of T and Z players in TvZ are as follows. The first principal component of T players shows the players that employ 'research' and 'training' commands, while the second group uses 'radioactive attack' that affects the target area. Interestingly, both P and T players use similar attack types versus Z players. Both 'radioactive attack' and 'psionic storm' target an area that helps to kill multiple Z armies simultaneously. The first group of Z players uses 'burrow' techniques as in PvZ, and the second group does not use the 'morphing' command.

Based on Table 4.4.1, P playing styles differ in using 'research',

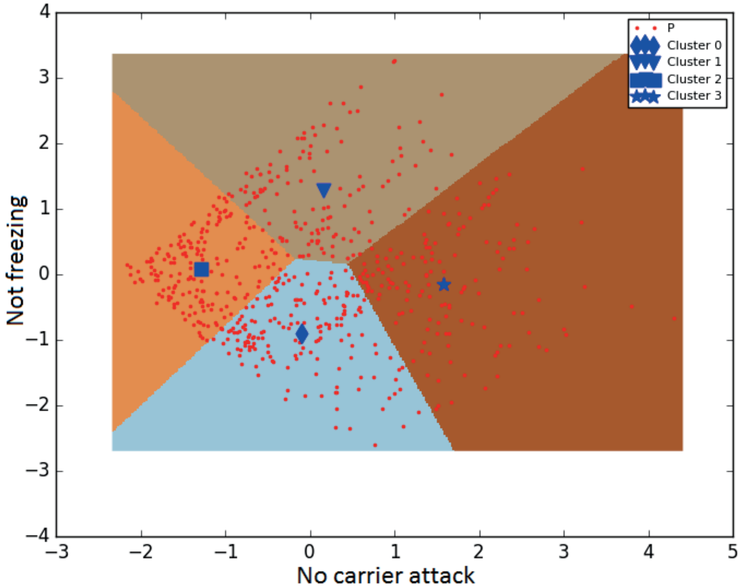

Figure 4.4.4: Observed playing styles in PvP match type by clustering top two principal components.

'train', 'upgrade', and 'freezing' commands. P players use 'freezing' techniques versus T players and 'psionic storm' against Z players. T playing styles are more varied than P playing styles. T players are different in using 'siege tank,' 'irradiate attack,' 'research,' and 'train'. They use 'siege' techniques versus T players, while they do more 'research' and 'train' against Z players, and also use the 'irradiate attack' versus Z players. Z players are different in using 'burrow,' 'morph,' and 'research' commands and they have relatively the same playing style versus P and T players.

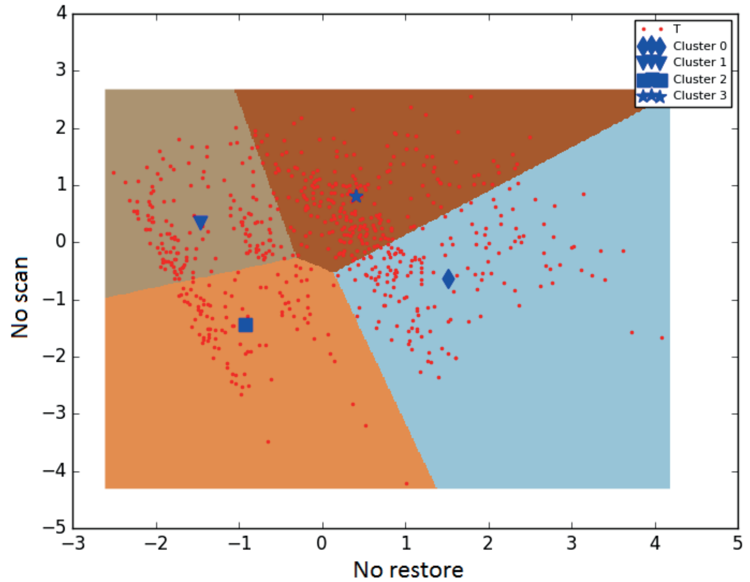

Figure 4.4.5: Observed playing styles in TvT match type by clustering principal components.

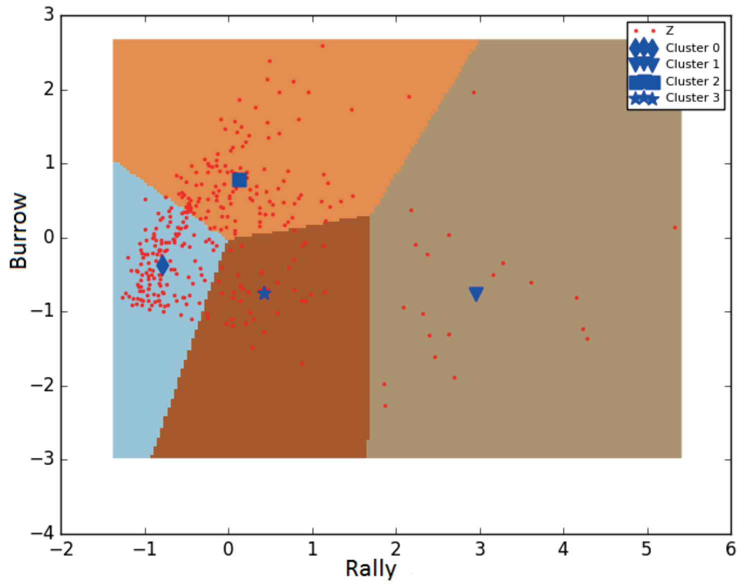

Figure 4.4.6: Observed playing styles in ZvZ match type by clustering principal components.

4.4.3 Symmetric Match Types

In symmetric match types, the players are from the same race. Thus, they can build the same building, and select the same commands. We are interested in seeing whether playing styles vary by match type. To determine this, we clustered playing styles by utilizing two top principal components.

Figures 4.4.4 to 4.4.6 show four types of playing styles across symmetric match types. Figure 4.4.4 shows player typology in PvP. The clusters show to what extent a player uses ‘freezing’ and ‘carrier attack’ commands. The TvT playing styles reflect how much a player uses ‘scan’ or ‘restore’ techniques. In ZvZ, player typology is based on ‘burrow’ and ‘rally’ commands. In this match type, the playing styles are not well distributed as in PvP or TvT, which can be explained by the fact that we had a much smaller sample size in ZvZ.

In both non-symmetric and symmetric match types, we discovered distinct playing styles. We expect that in general, players in RTS games exhibit different styles of playing based on the available commands in the game.

4.5 Winning Rate and Game-Length

Playing style can influence different aspects of the game, so they may influence the final results of the game. In this section, we investigate the relationship between players’ typologies and win-loss ratio. Moreover, we compare the game length in different playing styles to determine if there is a relationship between the playing style and the game-length?

In the previous section, we observed that for non-symmetric match types, players from different races are clustered in different clusters. In the other words, the playing style depends on the race type. To clarify player typologies for a race in non-symmetric match types, we separate players by their race type. Then we repeat PCA analysis and k-means clustering, so that we distinguish typologies per race/match type, and can relate these typologies to win-loss ratio and game length. The results are shown in table 4.5.1. We do the same for symmetric matches, but here we do not need to separate players by race. The results are shown in table 4.5.2.

In tables 4.5.1 and 4.5.2 we show, besides match type, race, and cluster number, the size of the cluster, the win-loss ratio (this is the number of wins for the race divided by the number of losses in the cluster, i.e., if it is higher than 1, the cluster causes more wins than losses), the mean game-length (mean-gl) and the variance in game length (var-gl).

In table 4.5.1, player's typologies are presented for each race. In PvT, the most interesting observation is that for P players the win-loss values are always higher than for T players. Clusters P2 and T0 show the lowest mean of game-lengths in PvT match types with 18 and 20 minutes respectively, and the variance of game-lengths is also low.

In PvZ, the win-loss values for P players and Z players are relatively balanced (no race seems to have an upper hand), though some clusters do much better than others.

In TvZ, T players always have a higher win-loss rate compared to Z players. T2 and Z2 have the lowest mean of game lengths in this match type with 15 and 16 minutes respectively.

A casual observation of table 4.5.1 gives the impression that the Protoss players have an advantage over Terran players, while Terran players have an advantage over Zerg players (and, strangely, Zerg and Protoss players are reasonably matched). However, we found that none of the win-loss ratios differ at a statistically significant rate ($p<0.05$), so these casual observations can be ignored. Interestingly, the game lengths for most match types differ significantly, except for T1 and T2

Table 4.5.1: Comparison of win-loss and game-length(minutes) in non-symmetric match types. mean-gl=mean of game-lengths for each cluster, var-gl=variance of game-lengths for each cluster. p-value= 0.05.

Match Type (number of players)	Race(Cluster)	Size	win-loss	mean-gl	var-gl
PvT (4032)	P 0	616	1.20	26	86.23
	P 1	301	1.11	30	154.35
	P 2	490	1.01	18	39.95
	P 3	609	1.20	32	91.28
	T 0	558	0.94	20	38.43
	T 1	667	0.88	25	84.26
	T 2	328	0.82	27	72.20
	T 3	463	0.86	37	131.82
PvZ (1680)	P 0	188	1.07	26	164.24
	P 1	281	0.85	30	44.08
	P 2	180	1.17	18	127.32
	P 3	191	0.71	32	41.66
	Z 0	199	0.90	37	171.62
	Z 1	225	1.14	16	43.83
	Z 2	244	1.10	21	37.14
	Z 3	172	1.26	33	110.01
TvZ (1624)	T 0	204	1.17	29	68.67
	T 1	239	1.13	24	67.63
	T 2	196	1.31	15	28.20
	T 3	173	1.16	26	60.41
	Z 0	206	0.84	20	28.91
	Z 1	178	0.94	29	58.55
	Z 2	206	0.69	16	39.68
	Z 3	222	0.93	30	58.71

Table 4.5.2: Comparison of win-loss and game-length(minutes) in symmetric match types. mean-gl=mean of game-lengths for each cluster, var-gl=variance of game-lengths for each cluster.

Match Type (number of players)	Cluster	Size	win-loss	mean-gl	var-gl
PvP (784)	0	157	0.99	32	178.32
	1	184	1.17	31	174.25
	2	228	0.90	14	44.29
	3	215	0.99	16	27.48
TvT (790)	0	189	0.87	40	183.99
	1	281	1.05	20	43.11
	2	194	0.94	18	39.97
	3	124	1.26	40	123.24
ZvZ (398)	0	164	0.99	10	18.74
	1	31	1.07	43	103.12
	2	48	1.18	26	36.19
	3	154	0.95	15	43.15

in PvT, P0 and P3 in PvZ, and Z0 and Z3 in TvZ.

In symmetric match types, cluster 1 in PvP has the highest win-loss value. In terms of game-length, clusters 0 and 1 are longer than clusters 2 and 3. In TvT, clusters 1 and 2 have win-loss ratio above 1. In this match type, clusters 0 and 3 include very long-lasting matches with 40 minutes of mean game-length. In ZvZ, clusters 1 and 2 have higher win-loss ratios and game-lengths.

As we found for non-symmetric match types, in symmetric match types the win-loss ratios are not significantly different ($p<0.05$), but game-lengths are significantly different, except for P0 and P1 in PvP, and T0 and T3 in TvT.

4.6 Conclusion

In our investigation of playing styles in *StarCraft*, we found that playing styles differ based on race and match type, i.e., the different races use different styles, and these styles also depend on their opponents. Only the Zerg players seemed to use a similar playing style regardless of the opponent race. When examining whether playing styles influence win-loss ratios (i.e., whether certain styles are clearly better or worse than others), we found no significant differences between the win-loss ratios of the playing styles. However, playing styles clearly influence the length of the game.

We see that the design of *StarCraft* allows for considerably different playing styles for all races and that none of these playing styles provides a dominating position in the game. Such design can be considered 'strong,' as it provides players with variety without enforcing dominating tactics.

This chapter tackles research question 4 on learning processes in *Destiny*. The work in this chapter has not been published previously.

Research Question

4. To what extent can a player model relate a player's profile to playing style?

5
Learning Processes in Destiny

In this chapter, we analyze the players learning processes in *Destiny*. We use a dataset of the historical behavior of players to find the learning rate across different performance metrics. The used dataset captures team-based behaviors from Player-vs-Player (PvP) and Player-vs-Environment (PvE) aspects of *Destiny*. In this investigation, we use a performance rating as a function of the number of played matches.

In the following sections, we discuss our study into the learning processes in *Destiny*. In section 5.1, we discuss skill acquisition in video games. In section 5.2, we explain our proposed features. In section 5.3, we discuss the learning behavior of players. In section 5.4, we explain how we cluster players by their learning behavior. In section 5.5, we present our conclusion.

5.1 Related Work

Research into player skill acquisition has been performed with varying objectives, some having to do with improving physical skills and some with mental skills. Page et al. [65] studied the relationship between motor skill development and playing active video games. They found that active video games improve motor skill development in children. Green and Bavelier [38] studied the influence of action video games on cognition and their application in training for job-related skills. Boot [11] summarized previous research into video games used by psychologists for understanding skill acquisition and cognitive processes.

Studies on players' skill acquisition resulted in different findings for different games. Thompson et al. [95] studied action sequences in StarCraft in different levels and they found that the first action in a group of actions has a latency. Boot et al. [12] studied players' skill acquisition in the *Space Fortress* game and they found that some game rules were not understood by several players after training.

Different studies show the relationship between social skills and learning processes. Gutwin et al. [39] studied the effect of the assistance of team members on improving skill development in FPS games. Later, Stafford and Dewar [86] studied several factors that influence the learning rate of players and they found that a focus on social play decreases this learning rate. Schrader and McCreery [77] studied learning processes in MMOGs and their influence on learning in educational systems. They stated that the social skills that learners developed in MMOGs help them to achieve their goals.

In our work, we aim to determine whether players are learning, by linking performance rating to the ordering of the matches played. Then, we use archetype analysis for investigating learning profiles. Archetype analysis has been used for research in video games before by Sifa et al. [81], Sifa et al. [80], and by Drachen et al. [24].

Table 5.2.1: PCA component coefficients of learning rates for all players

Performance Metric	PCA Coefficient (var=0.51)
assists	-0.40
average score per kill	-0.36
average score per life	-0.34
deaths	-0.13
kills-deaths-assists	-0.36
kills-deaths ratio	-0.38
kills	-0.41
score	-0.36
team score	-0.08

5.2 Experimental Setup

For this study, we used the same dataset of *Destiny* that is discussed in section 3.2. We selected the following game modes: ‘skirmish’, ‘control’, ‘ironbanner’, ‘clash’, and ‘trials’. The game play is explained in B.2. To study the learning processes of the players, we applied PCA analysis to the features in the dataset, and we determined the features which ended up with the highest PCA coefficients. We selected the components with the highest coefficients (see Table 5.2.1) to study the learning processes.

In the following sections, we investigate to what extent we can observe the learning process in *Destiny*. In section 5.3, we investigate the learning process across performance metrics using average performance ratings. In section 5.4 we use archetype analysis to discover learning profiles.

5.3 Players’ Learning Process

In this section, we analyze the learning processes of *Destiny* players across performance metrics in the selected game modes (‘skirmish’, ‘control’, ‘ironbanner’, ‘clash’, and ‘trials’). We employed the average performance rating proposed by Stafford and Dewar [86] and used by Stafford et al. [85]. In this approach, players that have competed in at least 50 matches are selected; this includes 2,021,764 players. Players are ranked by the average of their top-three scores in a performance metric. The ranking is then divided into three groups that show the average score per performance metric in the top third, middle third, and the lowest third of the metrics.

Figures 5.3.1 to 5.3.7 show the average performance rating across players’ performance metrics. Each of these figures shows three lines, which represent the top third, middle third, and the lowest third of the metrics. In the figures, the x-axis shows the index of the game, and thus represents the order of matches that a player has played, in the range 0 to 50. The y-axis shows the average of the corresponding performance metric.

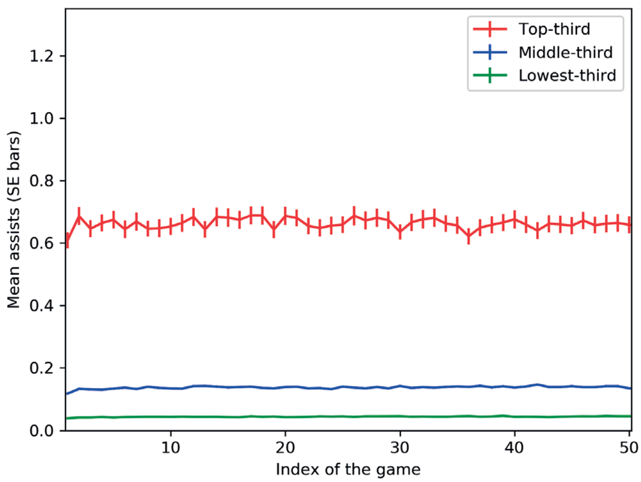

Figure 5.3.1: Average assist based on players’ highest three ratings

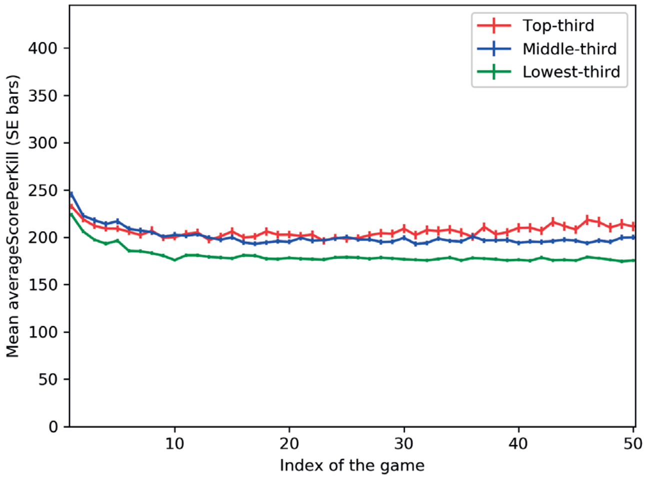

Figure 5.3.2: Average Score Per Kill based on players' highest three ratings

Figure 5.3.1 shows the average of assists based on players' highest three ratings. The assists are clearly different in players' highest three ratings, but do not change over time. The figure shows learning does not occur for the assist, but it shows the players start with different levels of assist.

In figure 5.3.2, the average score per kill for players' highest three ratings is shown. At the beginning, this metric decreases; then it remains static for the rest of the games. The decrease at the start may simply indicate that players become more effective at killing, but their scores remain overall the same, therefore the score per kill goes down. However, over the time for the best players (the red line) the score per kill again increases somewhat.

Similar learning processes can be observed in figures 5.3.3, 5.3.4, and 5.3.5. They show a learning curve which increases over time. While a

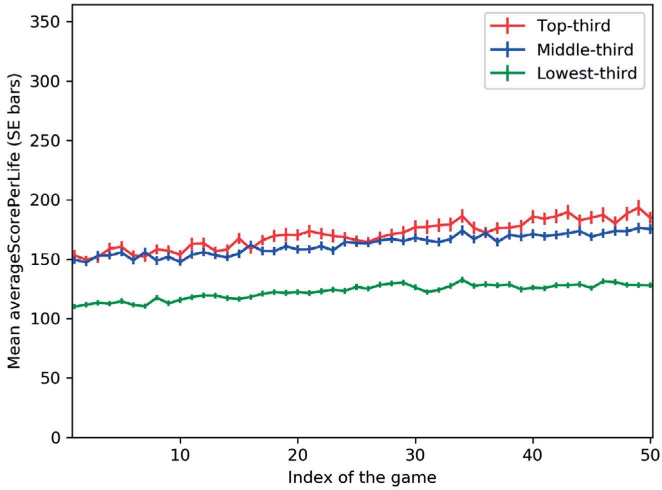

Figure 5.3.3: Average Score Per Life based on players' highest three ratings

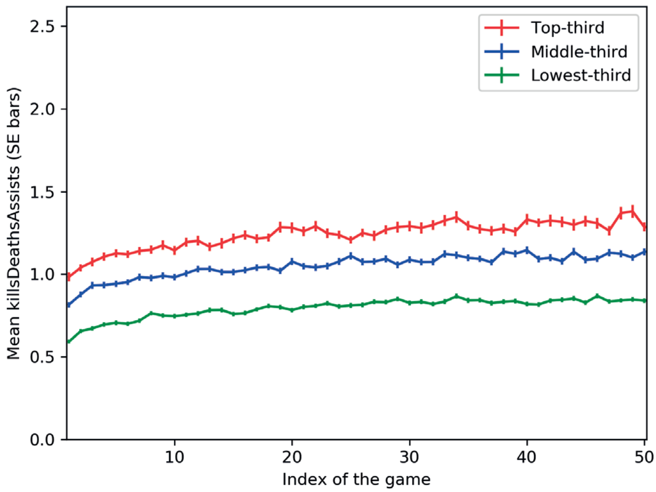

Figure 5.3.4: Average kills Deaths Assists based on players' highest three ratings

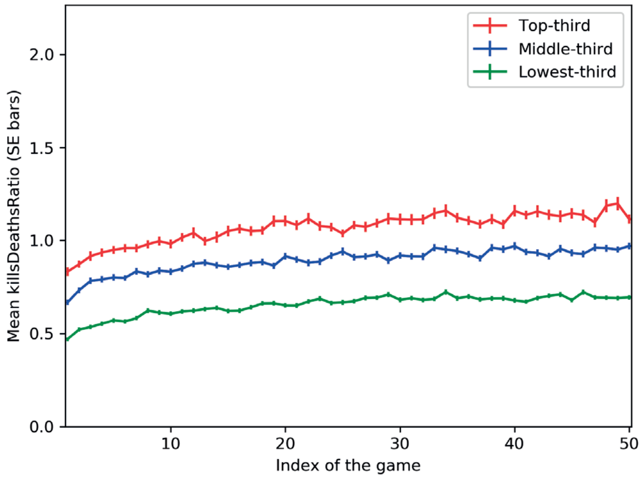

Figure 5.3.5: Average kills Deaths Ratio based on players' highest three ratings

difference can be observed between the best and the worst players, all players manage to improve their skills over time.

Figure 5.3.6 shows the progression of the average score for the players. There is a considerable difference between the score of top-third players compared to middle-third and lowest-third players. Moreover, the top-third of players manage to increase their average score over time.

Figure 5.3.7 shows the average team score over time. In this figure, we notice a striking difference between the progression of the best players, compared to the other players: the best players manage to increase their average team score considerably, while the other players hardly improve. Learning can be observed for all three groups, but learning for the top-third players is much faster than for the other players. This confirms the results that we explained in the previous chapter, and in Norouzzadeh Ravari et al. [61].

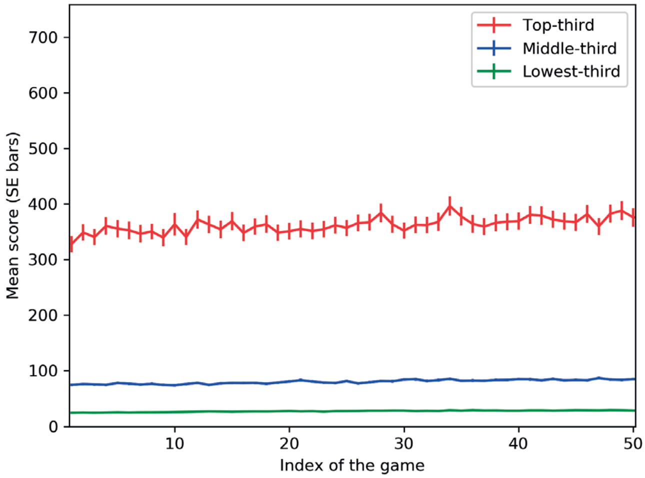

Figure 5.3.6: Average Score based on players' highest three ratings

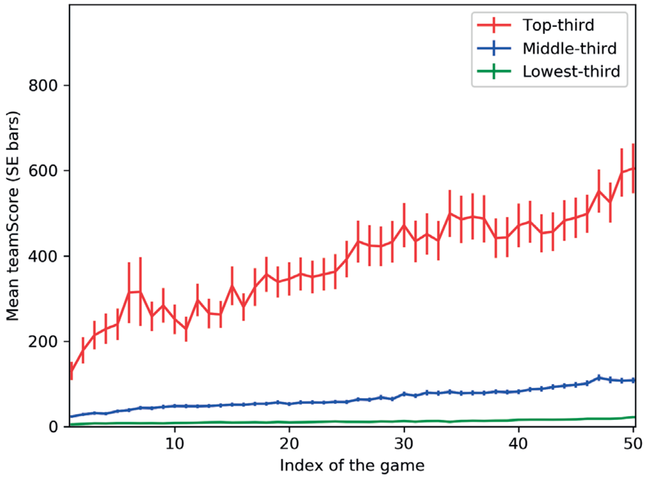

Figure 5.3.7: Average Team Score based on players' highest three ratings

Overall, we observed that learning occurs for most of the players' performance metrics over time. The learning process is not the same for all performance metrics, and not for all groups of players. To answer the question of whether we can distinguish different classes of players based on their learning capabilities, we investigate the variations in players' learning processes using clustering techniques in the next section.

5.4 Clustering Based on Learning Process

In section 5.3, we observed that different performance metrics indicate different learning speeds. In this section, we want to discover how learning profiles relate to performance metrics.

We use the same data that is used in section 5.3. The features that are used in this section are extracted as follows. First, performance metrics are normalized by a min-max method. Then, for each performance metric, we fitted a model to relate it to the match index (running from 1 to 50). For instance, for the assists performance metric, we select the first 50 scores in assists for a player. Then we fit a linear model to find the relation between these scores and the match index. Finally, we compute the slope of the model as a feature. For this purpose we used different fitting methods. Based on MSE and the Akaike Information Criterion (AIC), we selected linear regression and isotonic regression models. Next, we used the slope of the fitted models as a feature for clustering learning profiles.

We used Silhoute score Rousseeuw [73] to find the optimum number of clusters/archetypes, which proved to be 3 clusters. We tried different clustering models such as k-means, DBSCAN, and mixture models, but the results were not satisfying. Then, we used archetype analysis to investigate the learning profiles.

Archetype analysis is a soft clustering method that is proposed by Cutler and Breiman [18]. Archetype analysis extracts archetypes that

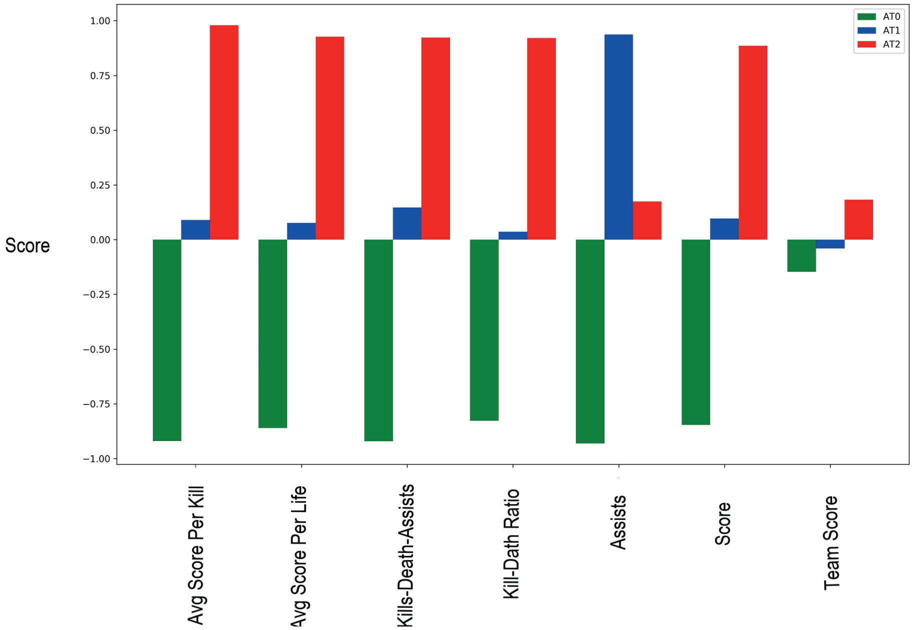

Figure 5.4.1: Archetype analysis on learning rates

represent profiles. Archetype analysis uses matrix decomposition to find a combination of archetypes that approximates the convex hull of the original data. We used py-cha package in Python for archetype analysis Aslak [3] and Scikit-learn package Pedregosa et al. [67] for developing our models.

Figure 5.4.1 shows the result of the archetype analysis on learning rates of player performance metrics. Three archetypes are distinguished, which are represented by blue, green, and red; the bars show the role of different performance metrics in the clusters. The green cluster is clearly different from the blue and red clusters in all performance metrics. Both the blue and red clusters have archetypes that score positively on (almost) all of the features, while green scores negative on all of them. All clusters are different in their levels of performance metrics. Overall, the levels in the red cluster are higher than the levels of performance metrics in the blue cluster, except for the assists. In the other words, in the blue cluster, the assists are the performance metric that this group of players excels at. Therefore, we may call the red archetype the “fast-learners,” the blue archetype the “supportive-learners” and the green archetype the “slow-learners.”

5.5 Chapter Conclusion

In this study, we investigated the learning processes related to different performance metrics in *Destiny*. We studied the performance metrics that show the most variation in players’ profiles. We observed that there are differences in performance between players even at the start of playing the game, i.e., expertise levels already differ for novice players. We also observed differences in learning progress between players. Learning improves player performance on several metrics (but not all of them). For some learning metrics considerable difference in learning rates between groups of players can be observed (e.g., for average team score).

We investigated the variations in players’ learning processes using archetype analysis. We observed three groups of players: fast-learners, supportive-learners, and slow-learners. Fast-learners obtain the skills more efficiently than the other two groups. The supportive-learners, however, show progress in most of the metrics, but excel in particular on their progress in assists. For slow-learners, the performance over the different metrics actually decreases. Note that these three groups are archetypes, i.e., they are extreme examples of players. It is not the case that a big group of players actually decrease in performance over time, as all players may exhibit some of the features of the slow-learner archetype.

Our findings in this study can be used for personalization of the game based on the players’ learning progress. For instance, the game environment could be adopted based on players learning and figure out the weaknesses of the players to improve it. Moreover, the result of this study can be used for team building such that a combination of different learning groups could be chosen to build a team.

This chapter tackles research question 5 on the relation between playing styles and national culture.

Research Question

5. To what extent can a player model relate national culture to playing style?

Original Work: Norouzzadeh Ravari, Y., Strijbos, L., and Spronck, P. (2020). Investigating the Relation between Playing Style and National Culture. IEEE Transaction on Games. pages 1-10 [62].

Acknowledgement: The work on *Battlefield 4* discussed in this chapter was originally done by de Vries [20], and after that expanded upon by the author and collaborators with investigations of *Counter-Strike*, *Dota 2*, and *Destiny*. The work by De Vries is mainly reported on in sections 6.4.1 and the *Battlefield 4* parts of section 6.6.

6
Playing Style and National Culture

In this chapter, we examine playing styles in four popular Massively Multiplayer Online Games (MMOGs), namely *Battlefield 4*, *Counter-Strike*, *Dota 2*, and *Destiny*. We investigate to what extent national culture has relationship with these playing styles, and whether players from countries with similar cultures exhibit similar playing styles as well. In section 6.1, we review related work. In section 6.2, we explain the concept of playing style. In section 6.3, we explain the concept of national culture. In section 6.4, we explain the datasets we generated for the four games used in the study. In section 6.5, we explain our study implementation. In section 6.6, we present our results. In section 6.7, we discuss our findings, and in section 6.8 present our conclusion.

6.1 Related Work

Studies into playing styles approach the topic from different directions. Below we discuss three of these directions, namely (1) the differences between playing styles; (2) the relation between a player's personality and their playing style; and (3) the relationship between playing style and national culture.

Among others, differences in playing styles were studied by Eggert et al. [26]. They labeled playing style data from 708 players of the game *Dota 2*, and predicted nine player roles based on playing style. Liu et al. [50] showed that it was possible to identify a player based on their playing style in *StarCraft II*. They also predicted the player's next action based on their in-game behavior.

The relation between personality and playing style was studied by Yee and Ducheneaut [109], who used an online survey about personality and motivation to investigate players' in-game behavior. Van Lankveld et al. [100] measured the relation between players' behavior and the Five Factor Model of personality using questionnaires. Canossa et al. [15] investigated the correlation between players' motivational factors and their in-game behavior by analyzing *Minecraft* data. Bean and Groth-Marnat [8] studied the relation between player behavior in *World of Warcraft* and their personality via Big Five Factor Model.

The fact that relations exist between culture and emotional expressions [112] is used by Desai et al. [22], who show that players from different cultures understand emotions in video games differently.

The relation between national culture and playing style has been researched very little. That such a relationship exists is to be expected, as differences between members of different cultures are widely observed. The relationship between national culture and playing style is, however, has been studied only sparingly [20, 89]. In this study, we investigate the relationship between national culture and playing style in more depth, and in particular aim to determine whether such relationships can be generalized.

6.2 Playing Style

We define playing style as a set of features generated by a video game player's gaming behavior. Collections of features may be interpreted to reflect a more general playing style, e.g., "aggressive" or "exploratory." Depending on the game and the goals of the research, few or many playing styles may be distinguished. In principle, one could argue that every player exhibits their own playing style, but for the purposes of research and application, it makes more sense to distinguish playing styles in a more general way.

We are aware that the particular game that is used to study playing style has a strong relationship with the kind of playing styles that can be observed. Different games offer different types of actions, different environments, different characters, and different tools. For instance, in *Destiny* players choose one of the main characters, who all function differently. What can be observed of a player is for a large part determined by the character that they play. As such, the character selection is part of the playing style.

6.3 National Culture

National culture concerns a set of cultural values, such as norms, behaviors, beliefs, and customs, which are common to the population of a nation. Naturally, nations may be close in their culture to each other, or quite distant from each other. A well-known framework to define national culture was given by Hofstede [40, 97]. A major competing framework is GLOBE (Global Leadership and Organizational Behavior Effectiveness) [42], but in our study we use the Hofstede dimensions as

GLOBE is focused on the leadership in industries [101].

By means of replicated surveys concerning the values of people all over the world, Hofstede created six cultural dimensions based on statistical relationships. Each cultural dimension indicates a preference towards one state of affairs over another. Scores for each Hofstede dimension are available for more than 50 countries. Hofstede distinguishes the following six dimensions [97]:

- *Power Distance (PD).* A high value for Power Distance indicates acceptance of a hierarchical order, a low value indicates that a society seeks an equal power distribution.
- *Individualism (IDV) vs. Collectivism.* The Individualism dimension is concerned with self-image, which is defined in terms of "I" (Individualism) or "we" (Collectivism).
- *Masculinity (MA) vs. Femininity.* The Masculinity dimension is concerned with a society's preference for values which are traditionally associated with males or females, such as achievement (masculine) or modesty (feminine).
- *Uncertainty Avoidance (UA).* A high value for the Uncertainty Avoidance dimension indicates feeling threatened by uncertainty and trying to cope by avoiding unorthodox behavior. A low value indicates a relaxed attitude towards uncertainty.
- *Long-Term Orientation (LTO) vs. Short-Term Orientation.* Long-term orientation concentrates on thrift and future rewards, in contrast to short-term orientation, which appreciates traditions and fulfilling social obligations.
- *Indulgence (IDL) vs. Restraint.* An indulgent society puts great importance on personal happiness, whereas restrained societies are more prone to regulating positive emotions by strict norms.

The Hofstede dimensions have been used for game research before. They were, for instance, used to customize online gaming web sites [46], and to investigate the relation between national culture and players' experiences in MMOGs [111]. As in the previous research, in our research we assign to a player the Hofstede dimensions of their home country. Naturally, individual players may diverge from that, but we assume that the assignment will be approximately correct for most of them.

In our study, we did not pre-select the features which we surmise relate to the Hofstede dimensions. Instead, we built models based on all available features, and then examined the relationship between the strongest features and the Hofstede dimensions. Since all games that we study are team games, it is to be expected that many features relate in particular to the Individualism dimension, as players have the opportunity to express team-directed behavior vs. individualistic behavior.

6.4 Data

In this section, we describe the datasets created for each of the games used in the study: *Battlefield 4*, *Counter-Strike*, *Dota 2*, and *Destiny*.

6.4.1 Battlefield 4

Battlefield 4 is a First Person Shooter (FPS) game. The game enables many options for different strategies and tactics, both related to personal achievements and team support. The game play is explained in B.3.

Statistics of *Battlefield 4* players can be acquired from the BF4stats website (`bf4db.com/stats`). It provides an Application Programming Interface (API) which can be called with a player's username and platform to retrieve statistics of the player. A web crawler was used to retrieve player names from the "general score" leaderboard on the website [20]. Subsequently, a second web crawler was built to retrieve playing style statistics for every player name. The crawler retrieved information on 158 features, such as kill/death ratio (KDR), objective scores, scores for different roles within the game, scores for different game modes, and the use of different weapons. Playing style statistics were retrieved for 119,834 players.

Only players who had played more than 24 hours total were taken into account, and only countries with 500 or more players were selected. To keep the dataset balanced, a random selection of 514 players from each country was taken (the maximum possible). The dataset used for the remainder of this study therefore consisted of 14,906 players from 29 countries.

6.4.2 Counter-Strike

Counter-Strike is a FPS game, played with two competing teams. In this study we used *Counter-Strike: Global Offensive*. The game support different scenarios. The game play is explained in B.4.

We used the Steam *Counter-Strike* API to retrieve *Counter-Strike* for 125,127 accounts data [89]. The data was retrieved in two parts: stats and achievements. The stats data contains 218 playing style features per player such as 'total kills,' 'total deaths,' and 'total planted bombs.' Each achievement feature represents one of the 167 possible in-game achievements, and indicates whether the achievement was acquired. We grouped the achievement data into five categories (team tactics, combat skills, weapon specialist, global expertise, and arms race and demolition) and transformed them into a ratio (the number of completed achievements by a player per category divided by the total number of achievements per category). We also grouped the weapon features into five categories.

We excluded players with less than 12 hours of total playing time. We also removed all countries with less than 500 players. Players from Belarus and Kazakhstan were removed from the dataset, because Hofstede cultural dimensions are not available for these countries. This left players from 26 different countries. To keep the dataset balanced, from each country a random sample of 537 players was taken (the maximum possible). Therefore, the final *Counter-Strike* dataset used for further analysis consists of 13,962 players.

6.4.3 Dota 2

Dota 2 (Defense of the Ancients) is a Multiplayer Online Battle Arena (MOBA) game, in which the player controls a single hero character (selected from 115 possibilities) in one of two teams, each team consisting of five players. The game play is explained in B.5.

The *Dota 2* data was collected from Opendota (`www.opendota.com/`) via the Opendota API [89]. We selected a subset of the 232,326 *Dota 2* Steam accounts to gather data from, namely players which were active on Opendota and have a matchmaking rating (MMR) between 1,500 and 6,000 (thereby excluding very low and very high outliers). A total of 117,514 accounts met the criteria; we collected 78 *Dota 2* playing style features per player.

For the *Dota 2* dataset, only countries with at least 500 players were selected. Belarus, Kazakhstan, Myanmar, and Mongolia were excluded because their Hofstede cultural dimensions are not available. To keep the dataset balanced, from each of the remaining 30 countries we took a random sample of 512 players (the maximum possible).

6.4.4 Destiny

Destiny is a FPS game with strong influences of Massively Multi-player Online Role-Playing Games (MMORPG). Players play in a small team against another small team, in one of 13 different game modes. The game play is explained in B.2.

We collected players' information from the DestinyTracker website (`destinytracker.com`) in September 2017. The dataset includes 94 playing style features of 11,637 players from 41 countries. The features include player skills such as Kill-Death ratio (KD), Kill-Death-Assist ratio (KDA), and kills per game; weapon features that show percentage of kills by different weapons; teamwork features such as number of revives that the player performed or received; and performance features such defensive kills and offensive, and average distance of kills.

Players with less than 24 hours total playing time were removed from the dataset. Countries with less than 250 players were also removed reducing the number of countries to 27. We decided not to balance the resulting dataset, considering that 250 players is relatively low, so we wanted to include more players for countries which have them. In the end, the dataset includes 8,240 players from 27 countries. For most countries the dataset contains between 250 and 300 players.

6.5 Study Implementation

This study is divided into four parts, of which the results are reported in section 6.6. First, in our datasets we look at variations in playing styles based on nationalities, using Analysis of Variance (ANOVA) tests. Second, we look at variations in playing styles based on Hofstede dimensions, again using ANOVA tests. Third, we look at how the playing styles of nationalities are related to each other using t-distributed Stochastic Neighbor Embedding (t-SNE) dimension reduction. Finally, we use machine learning models to predict nationality and national culture based on playing style. We used Scikit-learn package Pedregosa et al. [67] and Keras for developing our models Chollet et al. [16].

In pre-processing the data, all time-dependent features (such as number of kills) were divided by the total playing time in minutes. Normalization was done by centering and scaling, i.e., by removing the mean of every feature and dividing every feature by its standard deviation.

All countries in the dataset were categorized into different groups based on their scores for each Hofstede dimension. Hofstede scores are specified in the range $[0, 100]$. We created four categories of scores for each dimension. Category 1 represents 'low' scores $[0, 24]$, category 2 represents 'medium low' scores $[25, 49]$, category 3 represents 'medium high' scores $[50, 74]$, and category 4 represents 'high' scores $[75, 100]$. An overview of these categorizations for the countries in our datasets is given in table 6.5.1. Naturally the numbers do not represent a full, inclusive view on a country's national culture, but merely serve as a pragmatic basis for research. The table also includes for each country for which games we have data.

6.6 Results

We used a one-way ANOVA to determine whether cross-cultural differences exist in playing styles. We performed ANOVA tests to determine for playing style variables the proportion variance explained by nationalities (6.6.1) and by Hofstede dimensions (6.6.2). Individual players get assigned to the Hofstede group to which their country is assigned. To determine the relation between national culture and playing style, we used t-SNE (6.6.3). We used classification to predict nationality and categories of cultural dimensions based on playing style (6.6.4).

6.6.1 Nationality and Playing Style

We analyzed the effect of nationality on the playing styles in our datasets. Top features with the highest proportion variance explained (η^2) are shown in tables 6.6.1 to 6.6.4 in descending order of variance explained.

Table 6.5.1: Categorization of countries based on their values for each Hofstede dimension.

Country	code	PD	IDV	MA	UA	LTO	IDL	Dataset
Argentina	ar	2	2	3	4	1	3	CS, Dota
Austria	at	1	3	4	3	3	3	B4, CS, Dota, Destiny
Australia	au	2	4	3	3	1	3	B4, CS, Dota, Destiny
Belgium	be	3	4	3	4	4	3	B4
Brazil	br	3	2	2	4	2	3	B4, CS, Dota, Destiny
Canada	ca	2	4	3	2	2	3	B4, CS, Dota, Destiny
China	cn	4	1	3	2	4	1	B4, CS, Dota
Czech Republic	cz	3	3	3	3	3	2	B4, CS, Dota
Germany	de	2	3	3	3	4	2	B4, CS, Dota
Denmark	dk	1	3	1	1	2	3	B4
Finland	fi	2	3	2	3	2	3	B4, CS, Dota, Destiny
France	fr	3	3	2	4	3	2	B4, CS, Dota, Destiny
Hungary	hu	2	4	4	4	3	2	B4
India	in	4	2	3	2	3	2	CS, Dota
Indonesia	id	4	1	2	2	3	2	CS, Dota
Italy	it	3	4	3	4	3	2	B4, CS, Dota, Destiny
Japan	jp	1	2	4	4	4	2	B4, CS, Dota, Destiny
Malaysia	my	4	2	3	2	2	3	CS, Dota
Mexico	mx	4	2	3	4	1	4	B4
Netherlands	nl	2	4	1	3	3	3	B4, Dota
Norway	no	2	3	1	3	2	3	B4, CS, Dota, Destiny
Peru	pe	3	1	2	4	2	2	CS, Dota
Philippines	ph	4	2	3	2	2	2	CS, Dota
Poland	pl	3	3	3	4	2	2	B4, CS, Dota, Destiny
Portugal	pt	3	2	2	4	2	2	B4
Romania	ro	4	2	2	4	3	1	CS, Dota
Russia	ru	4	2	2	4	4	1	B, CS, Dota, Destiny
Serbia	rs	4	2	2	4	3	2	Dota
Singapore	sg	3	1	2	1	3	2	Dota
South Africa	za	2	3	3	2	1	3	B4, CS
South Korea	kr	3	1	2	4	4	2	B4
Spain	es	3	3	2	4	2	2	B4
Sweden	se	2	3	1	2	3	4	B4, CS, Dota, Destiny
Switzerland	ch	2	3	3	3	3	3	B4
Thailand	th	3	1	2	3	2	2	Dota
Turkey	tr	3	2	2	4	2	2	B4, CS, Dota, Destiny
Ukraine	ua	4	2	2	4	3	1	B4, CS, Dota
United Kingdom	gb	2	4	3	2	3	3	B4, CS, Dota, Destiny
United States	us	2	4	3	2	2	3	B4, CS, Dota, Destiny
Vietnam	vn	3	1	2	2	3	2	Dota

Table 6.6.1: ANOVA test on the effect of nationality on 15 features with the highest proportion variance explained in *Battlefield 4* with p=.001 and df=(28, 14877).

Feature	F	η^2
Rounds played	124.74	.190
Medals	59.40	.101
Ribbons per round	55.32	.094
Light Machine Gun shots fired	4.15	.070
Rounds finished/Rounds played	37.31	.066
Assault Rifle shots fired	37.26	.066
Suppression assists	35.37	.062
Gadget accuracy	32.37	.057
Assignments	28.26	.051
Sniper Rifle shots fired	28.21	.050
Team score	27.30	.049
Kill assists	24.06	.043
Vehicle damage	23.86	.043
Shotgun shots fired	23.68	.043
Savior kills	23.37	.042

Battlefield 4 In table 6.6.1, it is shown from the 15 features with most variance explained that in *Battlefield 4* nationality affects playtime (rounds played), achievements (medals, ribbons per round, and assignments), and social behavior (suppression assists, team score, kill assists, and savior kills). Experience related features such as 'rounds played,' 'medals,' and 'ribbons per round' are at the top of these features. This can be interpreted as nationality has relationship with the players' engagement and skills de Vries [20].

Counter-Strike Table 6.6.2 shows the 15 features with highest proportion variance explained in *Counter-Strike*. The results show that countries may have preferences for particular battle maps. A considerable amount of variance is explained by nationality in all five 'achieve-

Table 6.6.2: ANOVA on the effect of nationality on the 15 features with highest proportion variance explained in *Counter-Strike* with p=.001 and df=(25, 13936).

Feature	F	η^2
Total Rounds Map Nuke	58.597	.095
Team Tactics Achievements Ratio	49.015	.081
Total Wins Map Nuke	48.790	.080
Combat Skills Achievements Ratio	47.762	.078
Global Experience Achievements Ratio	39.383	.066
Weapon Specialist Achievements Ratio	39.283	.066
Total Rounds Map Dust2	37.398	.063
Total Wins Map Dust2	34.986	.059
Arms Demolition Achievements Ratio	34.766	.059
Total Rounds Map Inferno	31.538	.053
Total Wins Map Inferno	3.819	.052
Total Wins Pistolround	28.591	.049
Total Rounds Map Train	26.449	.045
Total Kills Enemy Weapon	26.061	.045
Total Rounds Played	23.639	.041

ment ratio' features. This means that players from particular countries are more achievement-oriented than players from other countries.

Dota 2 Table 6.6.3 shows the effect of nationality on the 15 features with highest proportion variance explained in *Dota 2*. The 'total matches' feature is at the top, meaning that players from certain countries play more matches than players from other countries. The second feature is 'ping.' By using a ping, a player alerts team members about an event at a specific location. The high proportion variance of 'ping' and 'total words said' shows that players from different countries may have different styles of communication. A considerable amount of variance is explained by several 'hero' features. A player's choice for a hero depends on their preferred in-game role. This may therefore indicate

Table 6.6.3: ANOVA on the effect of nationality on the 15 features with highest proportion variance explained in *Dota 2* with p=.001 and df=(29, 15330).

Feature	F	η^2
Total Matches	64.527	.109
Ping	55.628	.095
Total Words Said	5.871	.088
Hero Damage	29.065	.052
Deaths	25.370	.046
Level	21.299	.039
XP per Min Avg per Match	2.513	.037
Hero Ranged	18.644	.034
Assists	18.369	.034
Purchase Force Staff	17.453	.032
Hero Disabler	17.129	.031
Purchase Blink	16.950	.031
Tower Damage	16.864	.031
Hero Support	16.799	.031
Hero Jungler	16.401	.030

differences in in-game roles between different countries.

Destiny Table 6.6.4 shows the top-15 features with the highest proportion variance explained in *Destiny*. 'Kill distance' is the top feature; it is influenced by the players' shooting behavior and types of weapons used. The second feature is 'DTR score,' which is a combined rating, calculated by the Destiny Tracker site, for all skill-based features of playing style, such as kills, medals, assists and deaths. This entails that there is a considerable difference in playing skills between certain countries. Several features related to 'kills' also explain a high amount of variance, as do features that relate to cooperative behavior such as 'assists' and 'revives.'

Table 6.6.4: ANOVA test on the effect of nationality on 15 features with the highest proportion variance explained in *Destiny* with p=.001 and df=(40, 10440).

Feature	F	η^2
Kill Distance	209.6	.445
Player DTR score	191.4	.423
Life Span	186.5	.416
Assists Per Game	168.5	.392
Precision Kills	145.5	.358
Assists	133.3	.338
Ability Kills	131.3	.334
Player kills	128.8	.330
Orbs Dropped	91.20	.330
Deaths	119.8	.314
Player Games	119.5	.314
Player time Played	112.7	.301
Revives Performed	104.0	.284
Revives Received	98.90	.274
Orbs Gathered	94.11	.265

6.6.2 Hofstede Dimensions and Playing Style

In a separate series of ANOVA tests, we analyzed the effect of Hofstede dimensions on the playing styles in our datasets (tables 6.6.5 to 6.6.8). From our discussion we left out the playing style features for which we found the interpretation is ambiguous – for instance, for *Counter-Strike* we found that particular maps were preferred in particular countries, for which many interpretations are possible, but without more research it is not possible to say which interpretations are likely to be correct. The values in bold indicate the Hofstede dimension which explains most variance for each feature.

Battlefield 4 Table 6.6.5 shows that for *Battlefield 4* in particular Individualism (IDV) explains most variance in the playing style features, namely for 9 of the 11 features. For the resulting two features, 'medals' and 'ribbons per round,' Individualism explains almost as much variance as Indulgence, the dimension which explains most variance for those two features. The explanatory power of Individualism is not surprising, as *Battlefield 4* is a multi-player game in which players co-operate in teams and are rewarded for team-play, but can choose to play alone de Vries [20].

The first three features, 'medals,' 'ribbons per round,' and 'assignments,' reflect award-collecting behavior, and are most explained by Indulgence (IDL). Award-collecting behavior is related to restrained societies, because awards are clear status symbols that represent a player's performance in the game.

The variance of 'light machine gun shots fired,' 'assault rifle shots fired,' 'sniper rifle shots fired,' and 'shotgun shots fired', is most explained by the Individualism (IDV) dimension, for which we see no obvious explanation.

The variance of the last four features, 'suppression assists,' 'team

Table 6.6.5: Proportion variance explained by each Hofstede dimension category in *Battlefield 4* with df=(3, 14902).

Feature	PD	IDV	MA	UA	LTO	IDL
Medals	.011	.037	.009	.002	.023	**.040**
Ribbons per round	.010	.035	.009	.003	.023	**.037**
Assignments	.006	**.015**	.003	.003	.010	**.015**
LM Gun shots fired	.014	**.058**	.004	.003	.023	.021
AR shots fired	.020	**.050**	.013	.010	.016	.028
SR shots fired	.011	**.034**	.012	.004	.016	.020
Shotgun shots fired	.011	**.028**	.010	.005	.010	.021
Suppression assists	.008	**.037**	.005	.000	.024	.020
Team score	.004	**.028**	.004	.000	.014	.012
Kill assists	.008	**.027**	.008	.003	.013	.015
Savior kills	.008	**.030**	.006	.003	.012	.014
Sum of variance	.111	**.379**	.083	.036	.184	.243

score,' 'kill assists,' and 'savior kills,' is again most explained by the Individualism (IDV) dimension. These four features express social and cooperative behavior. This result is not unexpected, as collectivist countries tend to consider in-group goals as more important than personal goals.

Counter-Strike Table 6.6.6 shows that for *Counter-Strike* the Individualism dimension explains most variance in 14 of the 15 features. All 'map' features are most explained by Individualism, which seems to indicate that the gameplay features of particular maps are appreciated more by individualistic countries than collectivistic countries, and vice versa. It may also indicate a preference for strategic gameplay of certain cultures, as the Nuke map (ranking high among the features) is strategically among the most balanced maps. The 'achievement' features are all most explained by Individualism. The mean value of all these fea-

Table 6.6.6: Proportion variance explained by each Hofstede category in *Counter-Strike* where df=(3, 13958). (TTA: Team Tactics Achievements, CSA:Combat Skills Achievements, TK:Total Kills, TR:Total Rounds, TW:Total Wins)

Feature	PD	IDV	MA	UA	LTO	IDL
TR Map Nuke	.038	**.061**	.015	.025	.002	.014
TTA Ratio	.020	**.043**	.017	.018	.003	.006
TW Map Nuke	.031	**.050**	.013	.019	.002	.012
CSA Ratio	.023	**.043**	.016	.018	.001	.006
TR Map Dust2	.013	**.025**	.006	.008	.001	.011
TW Map Dust2	.012	**.023**	.006	.007	.001	.010
TR Map Inferno	.009	**.031**	.012	.015	.006	.008
TW Map Inferno	.009	**.030**	.012	.015	.006	.008
TW Pistolround	.004	**.015**	.011	.008	.006	.005
TR Map Train	.013	**.025**	.006	.011	.001	.011
TK Enemy Weapon	.002	**.013**	.012	.008	.003	.006
Weapons Donated	.001	.009	**.015**	.003	.004	.009
TK Enemy Blinded	.001	**.012**	.009	.006	.002	.006
TK Zoomed Sniper	.011	**.020**	.005	.006	<.001	.005
TW Map Train	.008	**.016**	.003	.006	<.001	.006
Sum of variance	.275	**.518**	.199	.230	.050	.135

tures increases as the value for Individualism increases, meaning that players from individualistic countries collect more achievements and thus show more achievement-oriented behavior than players from collectivistic countries.

All remaining features are, again, most explained by Individualism, except ‘total weapons donated.’ The Masculinity dimension explains the highest proportion variance in this feature. The mean value of ‘total weapons donated’ increases as the value for Masculinity decreases. This is not surprising, as a low value for Masculinity represents a society with preference for cooperation and caring for others.

DOTA 2 Table 6.6.7 shows that also for *Dota 2*, the Individualism dimension explains most of the proportion variance in playing styles. The mean value of 'total matches' increases as the value for Individualism increases. There are several possible explanations for this finding. One is that players from individualistic countries are more likely to leave a match before the match is over; another is that players from individualistic countries simply play a higher variety of games. The cooperative behavior features such as 'ping' and 'assists' are most explained by Individualism. This is not unexpected, as collectivistic countries emphasize cooperative behavior. The feature 'hero jungler' is a lone-ranger role, and unsurprisingly associated with Individualism. The features 'hero disabler' and 'hero support' resemble supportive in-game roles. These features are most explained by Individualism and Uncertainty Avoidance, for which we see no obvious explanation.

DESTINY Table 6.6.8 summarizes the proportion variance explained by Hofstede dimensions in *Destiny*. Similar to the other video games in our study, the Individualism dimension explains the most of proportion variance, but the Masculinity dimension and Power Distance dimension also explain a considerable proportion variance. The 'kills' features such as 'kill distance,' 'DTR score,' and 'precision kills' are all explained by Individualism. These features relate to particular tactics used by players, and indicate that players from individualistic countries may prefer different tactics than players from collectivistic countries.

The two 'revives' features are most explained by the Masculinity dimension. A possible explanation is that revives are a sign of cooperation, which is more common for cultures which are more feminine.

Table 6.6.7: Proportion variance explained by each Hofstede dimension category in *Dota 2* with df=(3, 15356).

Feature	PD	IDV	MA	UA	LTO	IDL
Total Matches	.017	.032	.009	**.021**	.004	.009
Ping	.008	**.028**	.004	.012	.014	.021
Total Words Said	.001	.009	.004	.006	**.027**	.034
Hero Damage	.009	**.015**	.004	.009	.002	.006
Deaths	.008	.013	.007	**.018**	.003	.013
Level	.006	**.007**	.004	.006	.002	.005
XP per Min Avg	.008	**.009**	.003	.006	.001	.005
Hero Ranged	.007	**.018**	.003	.004	.001	.004
Assists	.006	**.014**	.006	<.001	.005	.009
Purch. Force Staff	<.001	.007	.001	**.015**	<.001	.004
Hero Disabler	.004	**.011**	.002	.010	<.001	.002
Purchase Blink	<.001	.004	.002	**.008**	.001	.006
Tower Damage	.001	**.006**	.003	.005	.001	.003
Hero Support	.008	**.015**	.005	.007	.001	.004
Hero Jungler	.011	**.013**	.002	**.013**	.001	.003
Sum of variance	.122	**.253**	.080	.165	.075	.147

Table 6.6.8: Proportion variance explained by each Hofstede dimension category in *Destiny* with df=(3,10477).

Feature	PD	IDV	MA	UA	LTO	IDL
Kill Distance	.040	**.080**	.048	.026	.045	.072
Player DTR score	.074	**.081**	.059	.038	.008	.011
Life Span	.039	**.093**	.064	.023	.047	.064
Assists Per Game	.023	.032	.026	.028	.029	**.046**
Precision Kills	.053	**.104**	.062	.028	.018	.028
Assists	.048	**.103**	.060	.023	.019	.023
Ability Kills	.045	**.102**	.065	.024	.021	.022
Player kills	.056	**.070**	.045	.038	.006	.008
Orbs Dropped	.048	**.104**	.056	.025	.019	.027
Deaths	.038	**.097**	.065	.018	.024	.016
Player Games	**.052**	.050	.046	.037	.005	.004
Player Time Played	**.051**	.047	.042	.034	.005	.004
Revives Performed	.037	.057	**.063**	.023	.018	.011
Revives Received	.035	.055	**.061**	.023	.018	.009
Orbs Gathered	.036	**.090**	.045	.025	.016	.020
Sum of variance	.673	**1.167**	.805	.413	.297	.363

6.6.3 National Culture and Playing Style

To investigate the relation between playing style and nationality we employed t-SNE. t-SNE visualizes high dimensional data in two dimensions that represent structures at different scales Maaten and Hinton [52]. Linderman and Steinerberger [49] proved that t-SNE has the capability to find relevant clusters. t-SNE has been applied to different domains including player modeling. For instance, Alves et al. [1] used t-SNE to distinguish paying players from non-paying players. Ryan et al. [74] used t-SNE to find to what extent video game descriptions reflect the similarity between video games.

We calculated the average over all playing style variables for all the players of a country. We projected the resulting vectors into 2-dimensional space, using the t-SNE algorithm. Figure 6.6.1 compares playing styles and nationality in different games. The subfigures indicate differences and similarities in playing styles between countries by the distance between the countries, i.e., countries which show similar playing styles are closer together.

Battlefield 4 Figure 6.6.1-a shows the comparison between playing styles of countries in *Battlefield 4*. Note that Russia, Ukraine, Turkey, South Korea, and China are in the lower right corner of the image and are dissociated from the other countries. China and South Korea are the only countries in this analysis in the ‘low’ group on Individualism, and the other three countries are in the ‘medium low’ group on Individualism. Moreover, Russia, Ukraine, and China are the only countries in the ‘low’ group on Indulgence, and the other countries are in the ‘medium low’ group. Thus, we consider it unsurprising that these countries differentiate from the other countries on average playing style.

Note also that the Scandinavian countries (Denmark, Finland, Norway and Sweden) are close together at the top of the graph. Moreover,

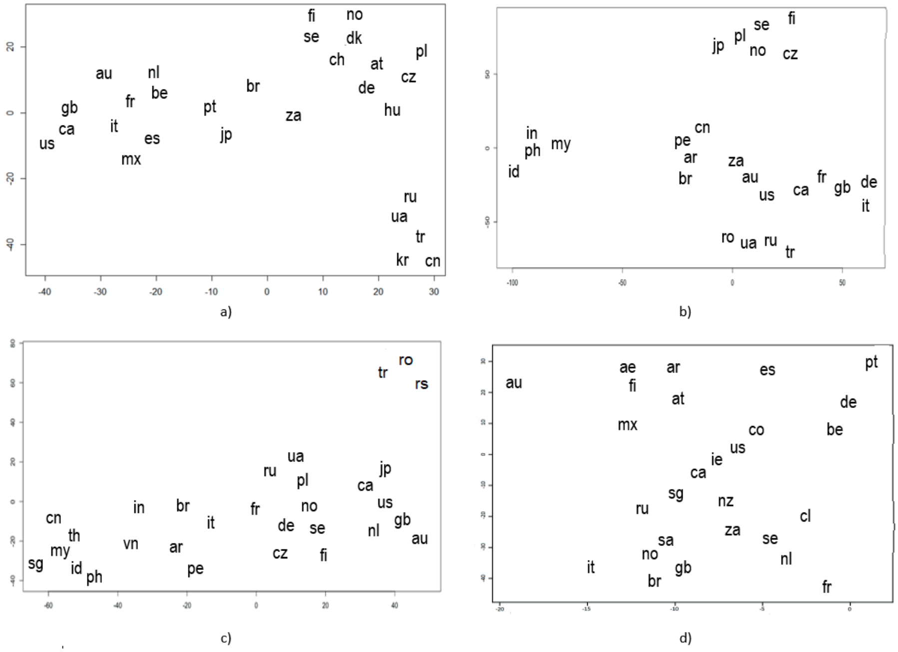

Figure 6.6.1: Comparing playing styles and nationality in different games. a) *Battlefield 4*. b) *Counter-Strike*. c) *Dota 2*. d) *Destiny*.

we observe that the Anglo-Saxon countries (US, Canada, Great Britain, and Australia), which are culturally very similar, are clustered at the left of the graph. Next to them, we find a cluster of typical Western European countries: Netherlands, Belgium, France, Italy, and Spain, with Mexico very close to Spain de Vries [20].

Counter-Strike As figure 6.6.1-b shows, Romania, Ukraine, Russia, and Turkey are grouped at the bottom of the figure. These countries are all score ‘high’ on Uncertainty Avoidance, and ‘medium low’ on Masculinity. Furthermore, at the left of the figure four Asian countries (India, Malaysia, Philippines, and Indonesia) are separated from the other countries. These countries have comparable scores for each of the Hofstede dimensions.

Several other groups of countries can be recognized as clusters. The South American countries (Peru, Argentina, and Brazil) are close together in the middle (with China being very close to them). As with the *Battlefield 4* figure, the Anglo-Saxon countries are again close to each other, as are the Scandinavian countries.

Dota 2 Figure 6.6.1-c shows the results of the t-SNE visualization of the country vectors for *Dota 2*. Although the differences between the countries are less clear, one can still recognize clusters of countries with comparable national cultures. Most Asian countries are grouped at the lower left and the Anglo-Saxon countries are clustered at the right. Moreover, Serbia, Romania, and Turkey are separated from the other countries at the top right of the figure. All these countries have comparable scores for most of the Hofstede dimensions.

Destiny The comparison between playing styles of countries in *Destiny* is shown in figure 6.6.1-d. For *Destiny* clusters are not clear. A potential explanation is that *Destiny*, which has only three different player classes, offers fewer possibilities to express cultural differences. Another potential explanation is that the game appeals to a more constricted type of player, and thus there can be found less variability in playing styles. The fact that Asian countries are hardly represented in our *Destiny* dataset is a sign of that.

Comparing countries To demonstrate how different playing styles for players in particular countries can be, in figure 6.6.2 we have clustered for each dataset all the players from the US versus all the players from China (except for *Destiny*, where we have no players from China, and thus we picked players from Brazil, which is far away from the US in figure 6.6.1-d).

Figure 6.6.2-a shows a t-SNE on playing style for Chinese and US

players in *Battlefield 4*. The figure shows that, in general, there is a clear difference in playing style of players from these countries. According to the Hofstede dimensions, China and the US are indeed quite different, in particular where Individualism, Indulgence, and Long-Term Orientation are concerned.

Figure 6.6.2-b shows Chinese players vs. US players in *Counter-Strike*. Most Chinese players are clustered at the left side of the figure, while most US players are at the right. We examined many other pairs of countries for *Counter-Strike*, and many show similar differences; some of these even more clear than the China/US plot.

Figure 6.6.2-c indicates some differences in playing style between Chinese and US players in *Dota 2*. US players are centered in the middle, while Chinese players are overall more spread out. Other pairs of countries show even clearer differences for *Dota 2* than the China/US plot.

For *Destiny*, we compared US players with Brazilian players. In figure 6.6.2-d we can see that there is relatively little overlap between the players from US and the players from Brazil, a picture similar to what we could see for the US and China for *Battlefield 4*. What is notable about this figure is that there seem to be many little clusters of players with playing styles that are very closely related. This may be the result of the different game modes of *Destiny*, combined with the limited selection of character classes.

What is striking in all these figures, is the fact that the US players tend to overlap a lot with the players they are compared to, but that the other country (China or Brazil) has more ‘unique’ players. A possible explanation is that the US is more a ‘melting pot’ of players where playing styles which are typical for other countries are also represented.

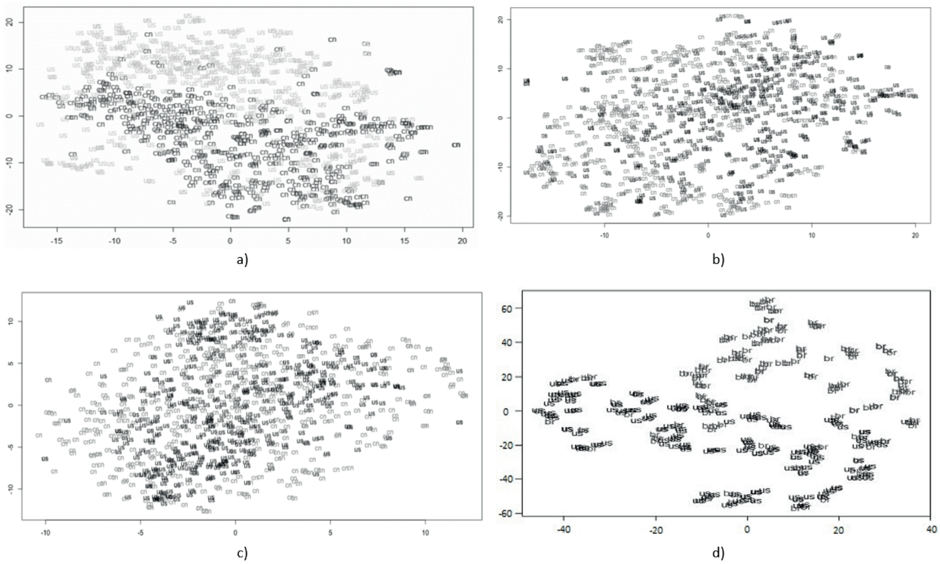

Figure 6.6.2: Comparing playing styles in different video games. a) China vs. US in *Battlefield 4*. b) China vs. US in *Counter-Strike*. c) China vs. US in *Dota 2*. d) Brazil vs. US in *Destiny*.

6.6.4 Predicting Nationality and Cultural Dimensions

We formulated the prediction of Hofstede cultural dimensions as a multi-class classification. For each cultural dimension a classification model was developed. Playing style features represented the input of the model. The output was the score value for cultural dimension, varying between 1 and 4. As classifiers we used Linear Regression, Random Forest, Support-Vector Machines (SVM), and Neural Networks. We applied these classifiers for all cultural dimensions in all four video games. We split each dataset into a training set consisting of 70% of the data, and a test set with the remainder of the data.

Classification accuracy, i.e., the percentage of correctly classified samples, was used to evaluate the different models. Five-fold cross-validation was performed on the training set to estimate the performance of the models using a particular set of parameters, and the

Table 6.6.9: Predicting nationality and cultural dimension from playing styles in *Battlefield 4*

Target	Baseline	RF	LR	SVM	NN
Nationality	3.48	15.49	**18.00**	17.45	16.79
PD	41.33	47.35	48.11	**48.62**	47.30
IDV	41.29	46.94	49.97	**50.34**	47.66
MA	44.87	47.32	**47.63**	47.56	46.42
UA	48.25	50.24	51.31	**51.93**	31.07
LTO	34.51	42.43	44.25	**44.46**	38.89
IDL	44.94	49.83	52.00	**52.72**	52.09

stratified option ensured that the training samples were equally distributed. For each model, the parameter set which had the best cross-validation score on the training set was selected. Finally, the accuracy scores on the test set were compared to the accuracy of the ZeroR baseline model (the majority class) to see if the model was able to determine the appropriate Hofstede dimension on the basis of the input parameters. The results are shown in tables 6.6.9 to 6.6.12. The values in bold indicate models with higher accuracy.

The results show that nationality can be predicted considerably better than the baseline. Usually this is because some countries can be predicted quite well – we found that in particular Asian countries can be distinguished from Western countries. The performance of nationality prediction in *Destiny* is lower than the other games in the study. This also might be due to including fewer Asian countries in the *Destiny* dataset.

The prediction of cultural dimensions also improves upon the base-

Table 6.6.10: Predicting nationality and cultural dimension from playing styles in *Counter-Strike*

Target	Baseline	RF	LR	SVM	NN
Nationality	3.87	16.79	18.08	**19.51**	16.65
PD	38.45	49.77	49.80	**50.13**	48.98
IDV	38.49	48.30	47.58	**48.48**	48.37
MA	53.81	**55.71**	53.71	54.57	55.25
UA	42.28	**54.57**	53.42	53.85	54.53
LTO	42.36	44.36	**44.65**	44.11	43.61
IDL	42.32	**48.55**	46.72	48.37	47.19

Table 6.6.11: Predicting nationality and cultural dimension from playing styles in *Dota 2*

Target	Baseline	RF	LR	SVM	NN
Nationality	3.35	15.63	17.48	**18.13**	16.70
PD	36.69	**48.27**	45.61	47.49	47.66
IDV	36.56	44.60	44.47	**45.21**	45.05
MA	43.29	**52.05**	48.86	49.15	49.54
UA	39.94	50.85	50.55	51.17	**51.33**
LTO	43.29	**47.40**	45.50	46.74	46.35
IDL	49.97	**55.66**	54.39	53.84	54.23

Table 6.6.12: Predicting nationality and cultural dimension from playing styles in *Destiny*

Target	Baseline	RF	LR	SVM	NN
Nationality	3.70	2.37	**7.70**	6.65	4.63
PD	31.43	22.85	**34.44**	31.54	33.60
IDV	25.81	13.03	24.38	17.67	**28.26**
MA	35.90	48.92	**61.64**	61.02	54.60
UA	30.56	**50.24**	32.66	37.33	44.22
LTO	31.43	36.60	19.71	28.60	**39.43**
IDL	38.10	3.00	46.09	38.76	**51.42**

line for almost every dataset (also in tables 6.6.9 to 6.6.12). The performance of the prediction depends on the chosen video game. In *Battlefield 4*, IDV, LTO, and IDL are predicted more accurately (compared to the baseline) than the other cultural dimensions. For *Counter-Strike*, this holds for PD, IDV, and UA. For *Dota 2*, PD, IDV, and UA are predicted better than the other cultural dimensions. In *Destiny*, MA, UA, and IDL are predicted better than the other cultural dimensions. We observe that Individualism (IDV) is predicted with considerable accuracy in all studied video games, except *Destiny*; a possible explanation is that we had less access to players from Asian countries in our *Destiny* dataset. These observations are in line with what we found in section 6.6.2, where we saw that Individualism explains the highest proportional variance in most cases.

6.7 Discussion

We analyzed four popular Massively Multiplayer Online Games across Western and non-Western countries and we looked at the relation between nationality, national culture, and playing style. Our research demonstrated clear differences in the playing styles of players from different countries. Moreover, we found that in many cases, the playing styles of people in countries which can be considered similar in culture (Anglo-Saxon, Scandinavian, Asian, Western-European) have clear similarities, as the average playing styles of their citizens seem to form clusters when using t-SNE dimensionality reduction. This demonstrates that cultural differences in playing styles can be recognized.

In part of our analysis, we did not consider countries as separate entities, but characterized them according to the Hofstede dimensions. Such a characterization can be called a "national culture." We found clear indications that national culture has relationship with playing styles, and thus on gameplay preferences. This entails that game development companies which aim to have their games played world-wide, need to take into account the fact that different countries may have different preferences for gameplay elements. The least that game designers could do would be to run early test sessions with a game in countries with widely different cultures, to see whether relatively simple enhancements to a game could make the game more supportive of playing styles that are prevalent in particular countries.

Naturally, while differences between countries as a whole can be recognized, different players will have wildly different styles, so when comparing players from two countries, their styles are likely to overlap. We observed this in the t-SNE visualizations for individual players from different countries. However, we could still observe that the overlap is only for part of the player base, while large groups of players could be labeled as "typical for their country."

6.8 Conclusion

We studied playing styles in four popular MMOGs across different countries. We analyzed the relation between nationality, cultural dimensions, and playing styles. Our findings show that not only nationality and cultural dimensions have relationship with playing styles, but also that players from countries with similar national cultures seem to have similar playing styles. It should be noted, however, that we cannot conclude that there exists a direct causal relationship between culture and playing style. Third-variable explanations for the relationships might play a role, for instance, explanations related to the economy of the countries. However, considering that a relationship between culture and playing style exists, it may be prudent to take national cultures into account when designing games that should have an international appeal. In particular, the Hofstede dimension "Individualism" explained most of the variance in playing styles between national cultures for the games that we examined.

In conclusion, we have demonstrated that a player's culture is related to their playing style, and that insights into culture can be gained from examining playing style. In this research we were restricted by our limited access to player information, which we mostly got from public websites. Thus, in our examination of player culture we were restricted to nationality. However, cultural values are expressed by more player features, such as their religion, their birth place, and their family values. Game publishers often have access to information of this kind for individual players. Expanding the current research using such information could increase its impact considerably.

7
Discussion

In this thesis, we analyzed players' interactions in video games. We found that the interactions provide relevant information about the players' behavior and their intent. This information provides an opportunity for player modeling and personalization. As a result, the system which the player interacts with can be adjusted to provide a better experience. Below we discuss several topics that bear further exploration after the research from the previous chapters.

We found that player modeling has some striking differences with user modeling that is generally applied in more 'serious' applications. We base this statement in particular because of early research that we did in user modeling for location search services. This research did not fit well with the rest of the thesis, and thus we moved it to appendix A. We discuss the similarities and dissimilarities between user modeling in location search and player modeling in video games in section 7.1.

Moreover, we discuss the usefulness of player modeling in section 7.2. Possible improvements that can make the research more valuable are explored in section 7.3. Finally, we look at possible future research in section 7.4.

7.1 Similarities and dissimilarities

In this section, we explain similarities and dissimilarities between user modeling in location search and player modeling in video games. We use location search, which we studied in appendix A, as a typical example of an industrial application.

As far as similarities go, the systems which run the software that the user interacts with can be mobiles, tablets, PCs, or any other type of computer system, and the applications run on the system either locally or via the internet. The major similarity between user modeling in location search and player modeling in video games is that in both cases the models are derived only from what the users provide to the systems. In our research, this is limited to the users' interactions.

Because the data derived from the users' interactions are comparable in both cases, one might think that the best approach to user modeling for location search is also the best approach for player modeling in video games. However, because the applications are rather different, this is not necessarily the case, which is exemplified by the dissimilarities between the applications. The four main dissimilarities are the following:

First, interaction types in location search are limited to selection and navigation, while in video games players' interactions have more variations, such as moving in the game environment, shooting, jumping, and collaborating with teammates.

Second, the intent of interactions is different between the cases. In location search, users are looking for information on a location in order to navigate to that place. In video games, players interact with the game environment in order to experience entertainment.

Third, in location search users only interact with the system, while in video games (at least in multi-player games) players interact with both the system and with the other players.

Fourth, in location search, the learning process of users is much shorter than in video games. Users can learn how to use location search in a short time, but in video games a user's learning process might continue until the end of the game. The reason for this is that in location search the intent of the system is to allow users to learn how to use it in the shortest possible time, while in video games the learning process usually is a major part of the gameplay and increases players' engagement while it is going on.

Regarding the dissimilarities between user modeling in location search (and other typical industrial applications of user modeling) and player modeling in video games, we might say that user modeling tends to be considerably simpler, because in industrial applications there are usually fewer interaction types, users only interact with the system, and users do not have to spend much time to learn how to use the system. The consequence is in general there is a much higher variety in player models for a particular game than in user models for industrial applications.

7.2 Usefulness

In video games, player modeling may improve players' experiences in several ways. In multiplayer games that use a matching system, player models may be used to form matches that are particularly engaging for the players involved. When player modeling is combined with knowledge on team construction, it can be used to form well-functioning player teams. Another possibility is to use the player model to adapt the game itself, for instance, the game environment. This last possibility is specifically relevant for adapting a game to players' cultural backgrounds, as discussed in chapter 6.

In general, we might say that the aim of player modeling is found in the term 'matching': the model tries to find the best match between a player profile and the game or to find the best match with other players as teammates. We note that 'matching' is also often the goal of user modeling in industrial applications, where usability of the application is matched to the user model.

7.3 Improvements

To improve upon the user models constructed in this study, probably the best approach is to incorporate more information on the players. In our studies, we limited ourselves to players' interaction, but the interpretation of these interactions can be improved by adding knowledge of the players.

The user's intent is a crucial factor in using player models in video games. For instance, if a player's intent is to learn and not just to get entertained, then the player model should focus on helping the players to improve their skills based on their learning profile. In addition to players' intent, other information about the player such as emotions, education levels, age, and gender could improve the player model's effectiveness.

7.4 Future

The natural interactions of a player with a game could be enhanced by incorporating data from a variety of extra inputs. Cheap sensor data is often available nowadays, such as camera data and sound data. Beyond that, many players have devices available that can give information on their emotional state, such as data delivered by wireless wearable technology. This data could be used to construct more accurate player models.

However, it should be noted that while there are many promising areas for using player modeling in video games, video games are so complex that it is hard to create models that work for all players, as there exist a high variety of different players for any game. Moreover, players themselves are not static, e.g., their moods change by influences from outside the game.

8 Conclusion

In this study, we researched to what extent we can use natural interaction data to create predictive player models in video games. In this research, we focused on five research questions, which should provide an answer to the problem statement. The research questions and problem statement are answered below.

RQ 1: *To what extent is it possible to predict the winner of matches using in-game interactions?*

To answer this question, in chapter 2 we used a dataset of expert players in *StarCraft*. The dataset includes players' interactions per frame. We proposed several time-independent and time-dependent features for winner prediction across different match types in *StarCraft*. We found that the winner prediction using time-dependent features is possible for the majority of match types with an accuracy of 63% or higher. Among the features, *income* is the most predictive feature in non-symmetric match types. In symmetric match types, *income* and *unspent* are the most predictive features. Thus, it appears that economic features are overall the most predictive.

RQ 2: *To what extent is it possible to predict the winner of matches using post-game data?*

To answer this question, in chapter 3 we used a dataset of the game *Destiny*. The dataset includes the post-game data of players across 13 different match types over two years. First, we developed two models: a winner prediction model in win-lose matches, and a winner prediction model in ranking-by-score matches. Then we developed prediction models for each of the 13 match types. The models developed for the second approach had much higher accuracy compared to the models for the first approach. We analyzed the top-10 most predictive features in each match type. In ranking matches, the number of kills is the most predictive feature. In win-lose matches, the importance of features varies based on the match type. We conclude that winner prediction is possible with reasonable accuracy.

RQ 3: *To what extent can different playing styles in a game be distinguished from each other?*

To answer this question, in chapter 4 we used the same dataset of *StarCraft* that we used in chapter 2. By analyzing *StarCraft* interaction data across six different match types, we see different clusters appear. These clusters are identified by the actions that the players do, but mostly depend on what is the in-game race of the opponent; except for Zerg players who tend to choose the same playing style against all opponent races. We thus found that we were able to distinguish different playing styles, which are mostly related to the actions taken by players. As possible actions differ between the races, the playing style clusters were based on in-game races, though even within each race different playing styles could be distinguished.

RQ 4: *To what extent can a player model relate a player's profile to playing style?*

To answer this question, in chapter 5 we used the same dataset as we used in chapter 3. We used a performance rating as a function of the number of played matches. Then we investigated the variations in learning profiles and we found three groups of learning profiles: fast-learners, supportive-learners, and slow-learners. Fast-learners acquire skills in a shorter time compared with two other groups. The supportive-learners improve most of their skills and are strong in assisting others. The performance of slow-learners decreased over time. As we managed to derive a player's learning profile from their playing style, we can conclude that at least aspects of a player's profile can be related to playing style.

RQ 5: *To what extent can a player model relate national culture to playing style?*

To answer this question, in chapter 6 we investigated the relationship between the playing styles of players and their nationalities in four MMOGs including *Battlefield 4*, *Counter-Strike*, *Dota 2*, and *Destiny*. We found that there are relationships between nationality and playing style. Moreover, we found that players from countries with similar national cultures show similar playing styles. The Individualism dimension of the Hofstede model seems to be most determining in these relationships.

Problem Statement: *To what extent can we use natural interaction data to create predictive player models in video games?*

From our research and our findings in this thesis, we conclude that player models can be constructed from interaction data for video games. While the accuracy and effectiveness of the models depends on parameters such as the provided information, variations of interactions, and the complexity of the environment, we found that the models have predictive capabilities with regards to a variety of player features in video games, such as playing style, ability to be victorious, learning behavior, and cultural background.

References

[1] Alves, J., Lange, S., Lenz, M., and Riedmiller, M. (2014). Case study: behavioral prediction of future revenues in freemium games. In *Workshop New Challenges in Neural Computation*, pages 26–33. Citeseer.

[2] Amin, A., Townsend, S., Van Ossenbruggen, J., and Hardman, L. (2009). Fancy a drink in canary wharf?: A user study on location-based mobile search. In *IFIP Conference on Human-Computer Interaction*, pages 736–749. Springer.

[3] Aslak, U. (2020). Pcha algorithm for archetypal analysis. [Online]. Available: `https://github.com/ulfaslak/py_pcha`.

[4] Avontuur, T., Spronck, P., and Van Zaanen, M. (2013). Player skill modeling in StarCraft II. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE)*.

[5] Bakkes, S., Spronck, P., and Van Den Herik, H. J. (2009). Opponent modelling for case-based adaptive game ai. *Entertainment Computing*, 1(1):27–37.

[6] Bakkes, S., Spronck, P., and van den Herik, J. (2007). Phase-dependent evaluation in RTS games. In *Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence*, pages 3–10.

[7] Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research*, 1(1):19.

[8] Bean, A. and Groth-Marnat, G. (2016). Video gamers and personality: A five-factor model to understand game playing style. *Psychology of Popular Media Culture*, 5(1):27.

[9] Berberich, K., König, A. C., Lymberopoulos, D., and Zhao, P. (2011). Improving local search ranking through external logs. In *the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 785–794.

[10] Bishop, C. M. (2007). *Pattern recognition and machine learning*. Springer.

[11] Boot, W. R. (2015). Video games as tools to achieve insight into cognitive processes. *Frontiers in Psychology*, 6:3.

[12] Boot, W. R., Sumner, A., Towne, T. J., Rodriguez, P., and Anders Ericsson, K. (2017). Applying aspects of the expert performance approach to better understand the structure of skill and mechanisms of skill acquisition in video games. *Topics in cognitive science*, 9(2):413–436.

[13] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[14] Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.

[15] Canossa, A., Martinez, J. B., and Togelius, J. (2013). Give me a reason to dig minecraft and psychology of motivation. In *Computational Intelligence in Games (CIG)*, pages 1–8. IEEE.

[16] Chollet, F. et al. (2015). Keras. [Online]. Available: `https://github.com/fchollet/keras`.

[17] Cole, N., Louis, S. J., and Miles, C. (2004). Using a genetic algorithm to tune first-person shooter bots. In *Evolutionary Computation Congress*, volume 1, pages 139–145. IEEE.

[18] Cutler, A. and Breiman, L. (1994). Archetypal analysis. *Technometrics*, 36(4):338–347.

[19] Davis, J. and Goadrich, M. (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240. ACM.

[20] de Vries, M. (2017). Mining the relationship between play style and culture. Master's thesis, Tilburg school of humanities and digital science.

[21] Dereszynski, E. W., Hostetler, J., Fern, A., Dietterich, T. G., Hoang, T.-T., and Udarbe, M. (2011). Learning probabilistic behavior models in real-time strategy games. In *Seventh Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE).*

[22] Desai, N., Zhao, R., and Szafron, D. (2016). Effects of gender on perception and interpretation of video game character behavior and emotion. *IEEE Transactions on Computational Intelligence and AI in Games*, 9(4):333–341.

[23] Drachen, A., Canossa, A., and Yannakakis, G. N. (2009). Player modeling using self-organization in tomb raider: Underworld. In *IEEE symposium on computational intelligence and games*, pages 1–8. IEEE.

[24] Drachen, A., Green, J., Gray, C., Harik, E., Lu, P., Sifa, R., and Klabjan, D. (2016a). Guns and Guardians: Comparative Cluster Analysis and Behavioral Profiling in Destiny. In *Proc. of IEEE Computational Intelligence and Games (CIG).*

[25] Drachen, A., Lundquist, E., Kung, Y., Rao, P., Sifa, R., Julian, R., and Klabjan, D. (2016b). Rapid Prediction of Player Retention in Free-to-Play Mobile Games. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE).*

[26] Eggert, C., Herrlich, M., Smeddinck, J., and Malaka, R. (2015). Classification of player roles in the team-based multi-player game dota 2. In *International Conference on Entertainment Computing*, pages 112–125. Springer.

[27] Erickson, G. K. S. and Buro, M. (2014). Global state evaluation in StarCraft. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE).*

[28] Farooq, S. S. and Kim, K.-J. (2016). Game player modeling. pages 1–15. Springer.

[29] Ferro, L. S., Walz, S. P., and Greuter, S. (2013). Towards personalised, gamified systems: an investigation into game design, personality and player typologies. In *9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, page 7. ACM.

[30] Fischer, G. (2001). User modeling in human–computer interaction. *User modeling and user-adapted interaction*, 11(1–2):65–86.

[31] Fogarty, J., Baker, R. S., and Hudson, S. E. (2005). Case studies in the use of roc curve analysis for sensor-based estimates in human computer interaction. In *Proceedings of Graphics Interface 2005*, pages 129–136. Canadian Human-Computer Communications Society.

[32] Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.

[33] Friedman, J. H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 38(4):367–378.

[34] Froehlich, J., Chen, M. Y., Smith, I. E., and Potter, F. (2006). Voting with your feet: An investigative study of the relationship between place visit behavior and preference. In *Proceedings of the 8th International Conference on Ubiquitous Computing*, UbiComp'06, pages 333–350, Berlin, Heidelberg. Springer-Verlag.

[35] Gagné, A. R., El-Nasr, M. S., and Shaw, C. D. (2011). A deeper look at the use of telemetry for analysis of player behavior in RTS games. In *International Conference on Entertainment Computing (ICEC 2011)*, pages 247–257. Springer.

[36] Gan, Q., Attenberg, J., Markowetz, A., and Suel, T. (2008). Analysis of geographic queries in a search engine log. In *Proceedings of the first international workshop on Location and the web (LOCWEB '08)*, pages 49–56.

[37] Gow, J., Baumgarten, R., Cairns, P., Colton, S., and Miller, P. (2012). Unsupervised modeling of player style with LDA. volume 4, pages 152–166. IEEE.

[38] Green, C. S. and Bavelier, D. (2015). Action video game training for cognitive enhancement. *Current Opinion in Behavioral Sciences*, 4:103–108.

[39] Gutwin, C., Vicencio-Moreira, R., and Mandryk, R. L. (2016). Does helping hurt?: Aiming assistance and skill development in a first-person shooter game. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, pages 338–349. ACM.

[40] Hofstede, G. H., Hofstede, G. J., and Minkov, M. (2005). *Cultures and organizations: Software of the mind*, volume 2. Mcgraw-hill New York.

[41] Holmgård, C., Liapis, A., Togelius, J., and Yannakakis, G. N. (2014). Personas versus clones for player decision modeling. In *International Conference on Entertainment Computin (ICEC 2014)*, pages 159–166. Springer.

[42] House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., and Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage publications.

[43] Hsieh, J.-L. and Sun, C.-T. (2008). Building a player strategy model by analyzing replays of real-time strategy games. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 3106–3111. IEEE.

[44] Hsu, C.-J., Hung, S.-S., and Tsay, J.-J. (2013). An efficient framework for winning prediction in real-time strategy game competitions. *Innovation, Communication and Engineering*, page 239.

[45] Isbister, K. and Schaffer, N. (2008). *Game usability: Advancing the player experience*. CRC press.

[46] Kale, S. H. (2006). Designing culturally compatible internet gaming sites. *UNLV Gaming Research & Review Journal*, 10(1):5.

[47] Kamvar, M., Kellar, M., Patel, R., and Xu, Y. (2009). Computers and iphones and mobile phones, oh my! a logs-based comparison of

search users on different devices. In *Proceedings of the 18th international conference on World wide web*, pages 801–810.

[48] Lane, N. D., Lymberopoulos, D., Zhao, F., and Campbell, A. T. (2010). Hapori: context-based local search for mobile phones using community behavioral modeling and similarity. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 109–118.

[49] Linderman, G. C. and Steinerberger, S. (2019). Clustering with t-sne, provably. *SIAM Journal on Mathematics of Data Science*, 1(2):313–332.

[50] Liu, S., Ballinger, C., and Louis, S. J. (2013). Player identification from rts game replays. In *28th Int. Conf. on Computers and their Applications (CATA)*.

[51] Lymberopoulos, D., Zhao, P., Konig, C., Berberich, K., and Liu, J. (2011). Location-aware click prediction in mobile local search. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 413–422.

[52] Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.

[53] Maulik, U. and Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1650–1654.

[54] Montanez, G. D., White, R. W., and Huang, X. (2014). Cross-device search. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 1669–1678. ACM.

[55] Nacke, L. E., Bateman, C., and Mandryk, R. L. (2011). Brainhex: preliminary results from a neurobiological gamer typology survey. In *International Conference on Entertainment Computing*, pages 288–293. Springer.

[56] Nembrini, S., König, I. R., and Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics*, 34(21):3711–3718.

[57] Normoyle, A. and Jensen, S. T. (2015). Bayesian clustering of player styles for multiplayer games. In *11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-15) pg*, pages 163–169.

[58] Norouzzadeh Ravari, Y., Bakkes, S., and Spronck, P. (2016). StarCraft winner prediction. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE)*.

[59] Norouzzadeh Ravari, Y., Bakkes, S., and Spronck, P. (2018). Playing styles in StarCraft. In *European GAME-ON Conference on Simulation and AI in Computer Games*.

[60] Norouzzadeh Ravari, Y., Markov, I., Grotov, A., Clements, M., and De Rijke, M. (2015). User behavior in location search on mobile devices. In *European Conference on Information Retrieval (ECIR)*, pages 728–733. Springer.

[61] Norouzzadeh Ravari, Y., Spronck, P., Sifa, R., and Drachen, A. (2017). Predicting victory in a hybrid online competitive game: The case of destiny. In *Proceedings of the Thirteenth Artificial Intelligence and Interactive Digital Entertainment International Conference (AIIDE)*, pages 207–214. York.

[62] Norouzzadeh Ravari, Y., Strijbos, L., and Spronck, P. (2020). Investigating the relation between playing style and national culture. *IEEE Transaction on Games*, pages 1–10.

[63] Ontanón, S., Synnaeve, G., Uriarte, A., Richoux, F., Churchill, D., and Preuss, M. (2013). A survey of real-time strategy game ai research and competition in StarCraft. *Computational Intelligence and AI in Games, IEEE Transactions on*, 5(4):293–311.

[64] Ortega, J., Shaker, N., Togelius, J., and Yannakakis, G. N. (2013). Imitating human playing styles in Super Mario Bros. *Entertainment Computing*, 4(2):93–104.

[65] Page, Z. E., Barrington, S., Edwards, J., and Barnett, L. M. (2017). Do active video games benefit the motor skill development of non-typically developing children and adolescents: A systematic review. *Journal of science and medicine in sport*, 20(12):1087–1100.

[66] Park, H., Cho, H.-C., Lee, K., and Kim, K.-J. (2012). Prediction of early stage opponents strategy for StarCraft ai using scouting and machine learning. In *Proceedings of the Workshop at SIGGRAPH Asia*, pages 7–12. ACM.

[67] Pedregosa, F., Varoquaux, G., Gramfort, A., and et al (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825—2830.

[68] Perkins, L. (2010). Terrain analysis in real-time strategy games: An integrated approach to choke point detection and region decomposition. *Sixth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE)*, 10:168–173.

[69] Ponsen, M., Tuyls, K., Dejong, S., Ramon, J., Croonenborghs, T., and Driessens, K. (2008). The dynamics of human behaviour in poker. In *20th Belgian-Dutch Conference on Artificial Intelligence*, pages 225–232. Universiteit Twente.

[70] Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., and Kivikangas, M. (2006). Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

[71] Robertson, G. and Watson, I. (2015). Building behavior trees from observations in real-time strategy games. In *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, pages 1–7. IEEE.

[72] Robertson, G. and Watson, I. D. (2014). An improved dataset and extraction process for StarCraft ai. In *The Twenty-Seventh International Flairs Conference*.

[73] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

[74] Ryan, J., Kaltman, E., Fisher, A. M., Owen-milner, T., Mateas, M., and Wardrip-fruin, N. (2017). GameSpace: An Explorable Visualization of the Video game Medium.

[75] Schadd, F., Bakkes, S., and Spronck, P. (2007). Opponent modeling in real-time strategy games. In *GAMEON*, pages 61–70.

[76] Schatten, M., Tomičić, I., and Durić, B. O. (2015). Multi-agent modeling methods for massivley multi-player on-line role-playing games. In *International Information and Communication Technology, Electronics and Microelectronics Convention on*, pages 1256–1261. IEEE.

[77] Schrader, P. G. and McCreery, M. (2008). The acquisition of skill and expertise in massively multiplayer online games. *Educational Technology Research and Development*, 56(5-6):557–574.

[78] Schubert, M., Drachen, A., and Mahlmann, T. (2016). Esports analytics through encounter detection. In *Proceedings of the MIT Sloan Sports Analytics Conference*, volume 1, page 2016.

[79] Si, C., Pisan, Y., and Tan, C. T. (2016). Understanding players' map exploration styles. In *Australasian Computer Science Week Multiconference*, page 69. ACM.

[80] Sifa, R., Bauckhage, C., and Drachen, A. (2014). Archetypal game recommender systems. *CEUR Workshop Proceedings*, 1226:45–56.

[81] Sifa, R., Drachen, A., and Bauckhage, C. (2018). Profiling in Games: Understanding Behavior from Telemetry. *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective*.

[82] Sifa, R., Hadiji, F., Runge, J., Drachen, A., Kersting, K., and Bauckhage, C. (2015). Predicting Purchase Decisions in Mobile Free-to-Play Games. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment (AIIDE) Conference*. AAAI.

[83] Sifa, R., Srikanth, S., Drachen, A., Ojeda, C., and Bauckhage, C. (2016). Predicting retention in sandbox games with tensor factorization-based representation learning. *Computational Intelligence in Games*.

[84] Song, Y., Ma, H., Wang, H., and Wang, K. (2013). Exploring and exploiting user search behavior on mobile and tablet devices to improve search relevance. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1201–1212.

[85] Stafford, T., Devlin, S., Sifa, R., and Drachen, A. (2017). Exploration and skill acquisition in a major online game. *The 39th Annual Meeting of the Cognitive Science Society (CogSci)*.

[86] Stafford, T. and Dewar, M. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological Science*, 25(2):511–518.

[87] Stanescu, M., Hernandez, S. P., Erickson, G., Greiner, R., and Buro, M. (2013). Predicting army combat outcomes in StarCraft. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE)*. Citeseer.

[88] Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment*, 62(1):77–89.

[89] Strijbos, L. (2018). Towards a further understanding of the relationship between culture and play sytle. Master's thesis, Tilburg school of humanities and digital science.

[90] Synnaeve, G. and Bessiere, P. (2011). A bayesian model for opening prediction in RTS games with application to StarCraft. In *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)*, pages 281–288. IEEE.

[91] Synnaeve, G. and Bessiere, P. (2012). A dataset for StarCraft AI & an example of armies clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 8, pages 25–30.

[92] Teevan, J., Karlson, A., Amini, S., Brush, A. B., and Krumm, J. (2011). Understanding the importance of location, time, and people in mobile local search behavior. In *the 13th international conference on human computer interaction with mobile devices and services*, pages 77–80.

[93] Tekofsky, S., Spronck, P., Goudbeek, M., Plaat, A., and van den Herik, J. (2015). Past our prime: A study of age and play style development in battlefield 3. *IEEE Transactions on Computational Intelligence and AI in Games*, 7(3):292–303.

[94] Thompson, J. J., Blair, M. R., Chen, L., and Henrey, A. J. (2013). Video game telemetry as a critical tool in the study of complex skill learning. *PloS one*, 8(9):e75129.

[95] Thompson, J. J., McColeman, C., Stepanova, E. R., and Blair, M. R. (2017). Using video game telemetry data to research motor chunking, action latencies, and complex cognitive-motor skill learning. *Topics in cognitive science*, 9(2):467–484.

[96] Toker, D., Steichen, B., and Birk, M. (2014). Predicting player type in social network games. In *User Modeling and Adaptive Hypermedia (UMAP) Workshops*.

[97] Triandis, H. C. (1993). Cultures and organizations: Software of the mind.

[98] Tuunanen, J. and Hamari, J. (2012). Meta-synthesis of player typologies. In *Nordic Digra 2012 Conference: Games in Culture and Society, Tampere, Finland*.

[99] Van Der Maaten, L., Postma, E., and Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, 10:66–71.

[100] Van Lankveld, G., Spronck, P., Van den Herik, J., and Arntz, A. (2011). Games as personality profiling tools. In *Computational Intelligence and Games (CIG)*, pages 197–202. IEEE.

[101] Venaik, S. and Brewer, P. (2008). Contradictions in national culture: Hofstede vs globe. In *Academy of International Business 2008 Annual Meeting, Milan, Italy*, volume 30.

[102] Webb, G. I., Pazzani, M. J., and Billsus, D. (2001). Machine learning for user modeling. *User modeling and user-adapted interaction*, 11(1-2):19–29.

[103] Weber, B. G. and Mateas, M. (2009). A data mining approach to strategy prediction. In *2009 IEEE Symposium on Computational Intelligence and Games*, pages 140–147. IEEE.

[104] West, R., White, R. W., and Horvitz, E. (2013). Here and there: Goals, activities, and predictions about location from geotagged queries. In *the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 817–820.

[105] Yang, P., Harrison, B. E., and Roberts, D. L. (2014). Identifying patterns in combat that are predictive of success in MOBA games. In *Foundational of Digital Games Conference*.

[106] Yannakakis, G. N., Spronck, P., Loiacono, D., and André, E. (2013). Player modeling. *Dagstuhl Follow-Ups*, 6.

[107] Yannakakis, G. N. and Togelius, J. (2011). Experience-driven procedural content generation. *IEEE Transactions on Affective Computing*, 2(3):147–161.

[108] Yee, N. (2006). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. volume 15, pages 309–329. MIT Press.

[109] Yee, N. and Ducheneaut, N. (2020). The gamer motivation model in handy reference chart and slides. [Online]. Available:`https://quanticfoundry.com/2015/12/15/handy-reference/`.

[110] Zackariasson, P., Wåhlin, N., and Wilson, T. L. (2010). Virtual identities and market segmentation in marketing in and through massively multiplayer online games (mmogs). *Services Marketing Quarterly*, 31(3):275–295.

[111] Zaharias, P. and Papargyris, A. (2009). The gamer experience: Investigating relationships between culture and usability in massively multiplayer online games. *Computers in Entertainment (CIE)*, 7(2):26.

[112] Zajonc, R. B. (1998). Emotions. In *In D.T. Gilbert, S.T. Fiske, G. Lindzey (Eds.), 4th ed. Handbook of social psychology*, volume 1, pages 591—632. New York: Oxford University Press.

[113] Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. (2009). Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800.

As a bonus, during early research we investigated the prediction of user behavior in location search, using online location search services. As this research did not fit well with the research in video games, we moved it to appendix A.

Original Work: Norouzzadeh Ravari, Y., Markov, I., Grotov, A., Clements, M., and De Rijke, M. (2015). User behavior in location search on mobile devices. In *European Conference on Information Retrieval (ECIR)*, pages 728–733. Springer [60].

A

User Modeling in Location Search

In this appendix, we study how users behave when they interact with a location search service by analyzing user behavior data from a popular GPS-navigation service. We aim to find out whether user behavior in location search varies across different device types. Next, we propose user models for interaction prediction in location search.

We analyze query- and session-based characteristics and the temporal distribution of location searches performed on smartphones, tablets, and GPS-devices. Then, we develop models to predict users' click and route interactions in location search. Our findings may be used to improve the design of search interfaces to help users perform location search more effectively and improve the overall experience on GPS-enabled mobile devices.

In section A.1, we explain user modeling in search services. In section A.2, we introduce a dataset of a location search services that we used in this study. In section A.3, we compare user behavior in search services on tablets versus mobile devices. In section A.4, we show how our models can predict user interactions. In section A.5, we discuss our findings in user modeling in location search. In section A.6, we summarize the chapter.

A.1 Related Work

Studies in user modeling in search services have been done with various objectives. Some studies focused on user modeling across different devices, while other studies focused on using contextual information such as user location to adapt the search results to user needs.

Several studies have analyzed the effect of device types on user behavior in search services. Kamvar et al. [47] analyzed mobile, tablet and desktop user interactions and they suggested that no single interface can fit all user needs, and that search experience should be changed based on the type of device. Song et al. [84] also compared different device types and they concluded that a single ranking algorithm cannot be used for all of them. They suggested using the characteristics of user behavior on tablets and mobile devices to improve the ranking algorithm. Berberich et al. [9] used external data resources to improve mobile local search. Kamvar et al. [47] observed that user behavior varies in computer and mobile devices and they suggested to use different search interfaces for different devices. Montanez et al. [54] studied transitions between devices to use for search, as they observed that people use a variety of devices for search during a day. They found that analyzing a user's log on a previous device helps to predict their behavior on the next device, which may help a search service to retrieve more relevant content based on logs of other devices.

Researchers have also used the actual location of the user to improve local search results. Lane et al. [48] proposed the Hapori framework that utilizes location, time, and weather for local search. Teevan et al. [92] conducted a user study into search, asking participants about their location, desired destination, plans about visiting a place, etc. The authors report that participants mostly search "on the go" and plan to visit destinations soon after querying.

Location-related queries have also been analyzed in web search services. Gan et al. [36] studied geographic searches using queries from AOL. The authors classify queries into Geo and non-Geo queries and report that non-Geo queries are related to Geo ones. Geo-queries mostly searching for the name of a location or a service in an area. West et al. [104] studied web search logs to explore the relation between mobile queries and their locations. The authors proposed a statistical model to predict whether a user is soon observed at the searched location.

The studies above are mostly concerned with user behavior in local search and are based on logs of a general web search service on desktops, mobiles or tablets. Our work differs from this previous work, as we study user interaction with *location search* within a GPS-navigation system. We first compare user behavior in location search to user behavior in web search. Then we compare user search behavior across different device types, namely tablets and mobiles. Next, we develop models for interaction prediction in location search.

A.2 Location Search Dataset

People use location search on mobile, tablet, or GPS devices to locate places such as airports, markets, gas stations, etc., and then navigate to these places. After submitting a query, the user is presented with a list of location results and can click on them to see the map centered on the result, the result's phone number and web site address. Using the interface, the user can connect to the chosen location by phone, query for more information, share information, or plan a route. What sets location search services apart of other search services are three elements: (1) the goal of user, namely navigation, (2) the type of information resources, namely maps, and (3) the type of interactions, namely routing.

For this study, we sampled the log of location search of a popular navigation application during the period from February to June 2014. We considered search sessions from the USA and UK and filtered out non-English queries. Sessions were logged on the following devices: iPhone ("mobile"), iPad ("tablet"), and GPS devices.

Each session consists of one or multiple queries. Sessions are separated by a period of inactivity of more than 30 minutes or based on closing the application. Overall, we collected 445,446 search sessions consisting of 670,417 queries: 21,936 sessions and 38,129 queries for tablet, 423,509 sessions and 632,288 queries for mobile and 170,017 sessions and 170,017 queries for GPS devices. The uneven distribution of the number of sessions and queries between tablet, mobile, and GPS devices are due to the difference in device usage frequency in the sampled part of our log.

The dataset includes approximately 4.1M query-result pairs, including approximately 3.6M pairs that did not receive clicks. The query-result pairs which did receive clicks include 35K pairs that are clicked, but for which the user did not ask a route, and 360K pairs for which the users did ask a route. Our dataset contains approximately 3.2M query-result pairs for mobile, 800K query-result pairs for GPS-devices, and 180K for tablet devices. The number of USA query-result pairs is close to the number of UK ones.

Table A.2.1 shows the number of sessions and queries in the location search service in tablets and mobiles. For each device, we represent number of interactions. When users submit a query, they then either click or don't click (no-click) on a result. If a user clicks on a result,

Table A.2.1: User search behavior statistics for the location search in a GPS-navigation system on tablet and mobile devices compared with web search on desktop, tablet, and mobile [84]. All statistics for the tablet are significantly different from those for the mobile ($p < 0.01$).

	#sessions (%)	#queries (%)
Web search		
Desktop [84]	N/A	13,928,038
Tablet [84]	N/A	8,423,111
Mobile [84]	N/A	9,732,938
Tablet location search		
All	21,936	38,129
Click	15,770 (72%)	21,208 (56%)
No click	6,166 (28%)	16,921 (44%)
Route	15,277 (70%)	19,580 (51%)
Mobile location search		
All	423,509	632,288
Click	305,104 (72%)	360,343 (57%)
No click	118,405 (28%)	271,945 (43%)
Route	296,568 (70%)	340,953 (54%)

then route interaction is possible. Table A.2.1 also compares the same statistics for the dataset that Song et al. [84] used for web search.

In our dataset, for each session, there are a unique session ID, device type, and query ID. Each query sample includes a query ID, query term, date and time of query submission in UTC time-zone, country, latitude and longitude of the user when submitting a query, and session ID. For every result item, we have a unique result ID, Point Of Interest (POI) ID, query ID, and rank of this item on the result page. Since users can do different interactions on each result item, there is a log of interactions for each entry: this includes a unique action ID, interaction type, POI ID, name of POI, query ID, the rank of the related result in the result page, and session ID. For each POI, besides ID and name, latitude and longitude are available.

We removed samples without known values for country name, device name, and date and time related information. Then we localized date and time for samples from the US, to bring their time zones in line with the UK.

A.3 User Behavior in Tablet Versus Mobile

In this section, we analyze user interactions with a location search service. First, we compare user interactions with location search to those with general web search. Then we compare user interactions with location search on tablets vs. user interactions with location search on mobile devices.

Table A.3.1 shows user search statistics: the number of sessions, number of queries, the average number of queries per session, the average session length in minutes, and the average query length in words. The first block of table A.3.1 shows the statistics for general web search on desktop, tablet and mobile devices, as reported by Song et al. [84].

The second block reports the statistics of user search sessions in tablet and mobile location search. The first row for each device type shows the overall user search statistics. The second row presents the statistics for sessions and queries in which a user clicked on one or more results. The third row shows the statistics for sessions and queries in which a user did not click on any result. Since the goal of location search is to help users plan a route to a desired POI, the last row shows the statistics for sessions and queries that contain the "route to" action. Absence of the route action does not mean that a user is not satisfied with the search results – in many cases users are interested in checking the results without navigating to them (e.g., pre-trip planning). The differences between the corresponding tablet and mobile location search statistics in table A.3.1 are statistically significant according to the

Table A.3.1: User search behavior statistics for location search in a GPS-navigation system on tablet and mobile devices, compared to that in standard web search on desktop, tablet and mobile [84]. All statistics for the tablet location search are significantly different from those for the mobile location search ($p < 0.01$).

	avg. queries per session	avg. session length (mins)	avg. query length (words)
Desktop [84]	1.89	8.61	2.73
Tablet [84]	1.94	9.32	2.88
Mobile [84]	1.48	7.62	3.05
Tablet location search			
All	1.74	2.69	1.93
Click	1.82	3.22	1.84
No click	1.53	1.34	2.05
Route	1.79	3.16	1.83
Mobile location search			
All	1.49	1.86	1.87
Click	1.49	2.22	1.78
No click	1.49	0.94	1.99
Route	1.47	2.18	1.78

Mann-Whitney U-test at the 0.01 level.

Note that the number of sessions in the mobile location search is much larger than the number of sessions on tablets. This is due to the fact that location search is mostly used on the go and, therefore, users tend to prefer mobile to tablet. Also, the form factor of mobile phones makes them much more popular for in-car navigation, which is further stimulated by the availability of phone docking stations.

In the following, we first compare tablet/mobile location search with general web search, and then compare tablet location search with mobile location search.

Location Search vs. Web Search According to table A.3.1, users submit more queries per session while performing web search on tablets compared to location search on tablets. The opposite is true when users interact with mobile devices but the difference is much smaller.

Users spend less time interacting with location search than performing web search: three times less on tablets and four times less on mobile, even though the average number of queries per session is roughly the same. This observation can be interpreted as users of a location search mostly being on the move and having less time for searching compared to the web search scenario. Also, users can easily understand if a location is relevant or not, while in web search users spend more time on examining results.

In general, queries in location search are shorter than in web search. This can be explained by the fact that queries in location search are limited to places as opposed to web search queries, which can be about anything.

Tablet vs. Mobile in Location Search The number of sessions and queries indicate that mobile location search is used much more often than tablet location search. However, the average number of queries per session, average session length, and average query length for tablet location search are all larger than those for mobile location search. These observations can be explained as follows. Tablets are more often used for pre-trip planning, while mobile phones are used on the go. In trip planning, people spend more time and use more queries because they want to explore all possible results (e.g., finding appropriate hotels, restaurants, etc.). In contrast, people on the move execute more targeted searches and are mainly looking for the nearest available POI that solves their direct needs (e.g., petrol station, parking, fast-food, etc.).

We note that the above behavior is similar to what is found in web search (see the first block of table A.3.1). This means that the different form factor between tablet and mobile devices has a similar effect on how people use them for location search and for web search.

When we consider the percentage of sessions that have at least one click, location search on tablets and mobiles appear to be similar. In user interactions with location search, the routing action is a strong signal of user satisfaction. The percentage of routing in tablet and mobile location search reaches 70% of sessions (97% of sessions with clicks). Therefore, if a user clicks on a result, it is almost certain that their intent is to plan a route somewhere. In the remainder of the sessions, the user was either unable to locate relevant POIs or did not want to plan a route. This can mean that a click is a reliable indicator of user intent while interacting with a location search.

In sessions with both click and route actions (which we assume to be successful), the average number of queries per session and the average session length are usually larger than the average for all sessions. This can be explained as follows: users who do not click anywhere give up fast and submit a few queries; users who are more persistent in finding relevant POIs click on returned results and submit more queries.

We compared users' behavior in tablets and mobile location search along the temporal dimension. The query frequency distribution during the day is shown in figure A.3.1. The graph shows that users who do location search have a slight preference to use mobiles during working hours (from 11 AM till 7 PM) and a slight preference to use tablets while at home (from 9 PM to 10 AM). This observation is unsurprising, because users usually carry their mobiles with them, but may keep their tablets at home. Moreover, tablets are used more for pre-trip planning, usually done during non-working hours, while mobiles are used for actual navigation. We also analyzed the query frequencies for different days of the week and found that the relative number of queries in mobile location search is lower than on tablets during weekdays, but larger during weekends. A potential explanation is that people use

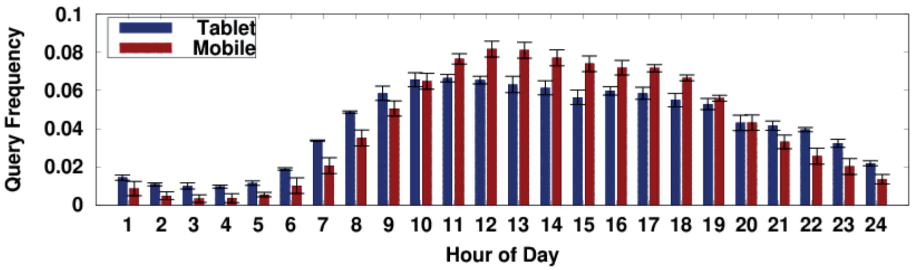

Figure A.3.1: Query frequency distribution in a GPS-navigation system on tablet and mobile devices.

weekends to go on trips and need mobile navigation for that.

A.4 Interaction Prediction

Interaction prediction includes both click prediction and route prediction. Click prediction has been studied in mobile local search [51]. Route prediction, as far as we know, has not been studied at all. In our study, we develop models for predicting click and route interactions in location search. In this section we explain the features that we used in this study. Then, we explain the models that we developed for interaction prediction in location search.

Different studies have reported the influence of the context on the users' behavior in mobile local search [9, 48, 51, 84, 104, 113]. Researchers have found that distance information and popularity of POIs have more effect than the other contextual features on the users' clicks [2, 34, 48, 51, 92].

Next to the baseline features, i.e., the features used by other researchers, we propose new features that we use in our prediction models. The features that we employ in our research are listed in table A.4.1. The first part of this table shows the baseline features that are used for click prediction in mobile local search [51]. The second part of the table includes the features proposed by us for interaction prediction in location search.

Baseline features are:

- *rank* of the result in the results list, whereby the best results have the lowest numbered ranks;
- *distance* between location of user and location of POI in kilometers;
- *popularity (clicks)* of each POI as the sum of all clicks that this POI received, as derived from our logs;
- *time-frame*, i.e., time slice of the day as a six-hour period, namely from 12 AM to 6 AM, from 6 AM to 12 PM, from 12 PM to 6 PM, or from 6 PM to 12 AM;
- whether it concerns a *weekday* or *weekend day*, stored as a binary feature;
- *device*: mobile, tablet, or GPS;
- *country*, either US or UK.

The second part of the table A.4.1 presents the features proposed by us. These are:

- for *scroll*, we counted the number of result pages that users scrolled to see the result items;
- *popularity (interactions)*, which represents the popularity of a POI as the number of all interactions (both click and route);
- *popularity (routes)*, which represents the popularity of a POI as the number of route interactions only – this means that popularity (routes) plus popularity (clicks) equals popularity (interactions);

- *avg-distance-country* is the difference between the current distance and the average distance for all searches in the log files, where the average distance is calculated separately for US and UK;

- *avg-distance-click* is similar to *avg-distance-country*, whereby the average distance is calculated not for all searches in the log files, but only for those that are clicked;

- *avg-distance-route* is similar to *avg-distance-country*, whereby the average distance is calculated not for all searches in the log files, but only for those that are routed.

In our study, we compare the success of prediction models which use only the baseline features, and which use both baseline features and proposed features.

Location search provides a variety of interactions. The prediction models (classifiers) we develop attempt to predict which interaction a user will engage in. The possible interactions that we study, schematically shown in figure A.4.1, are the following:

1. Click/no-click prediction: For each query-result pair, we predict whether a user will click on a result or not. The class label for this pair is no-click if the user does not click on any result, otherwise, it is labeled as a click.

2. Route/no-route prediction: For each query-result pair, we predict whether a user will engage in the routing interaction, i.e., whether the user will click on a result and then route. The class label is no-route if the user does not route, and otherwise, it is labeled as route.

3. Click/route prediction For each query-result pair in which the user clicked on a result, we predict whether the user engages in the route interaction. The class label is click if the user only clicked but did not route, and otherwise it is labeled as the route.

We built the prediction models using the L1-regularized Logistic Regression (LR) model [10] and a Gradient Boosting Regression Trees (GBRT) [33], using 5-fold cross-validation. We used the Scikit-learn package in Python [67]. We discuss each of the models below.

Table A.4.1: Features used in the prediction models in location search.

Feature	**Description**
Baseline Features	
rank	Rank of the result.
distance	Distance between the user location and POI.
popularity (clicks)	Popularity of POI as number of clicks.
time-frame	6 hours time-frame.
weekend-or-not	Weekday or weekend.
device	Device type (mobile, tablet, or GPS-navigation).
country	Country (US or UK).
Proposed Features	
scroll	Number of result pages that are scrolled.
popularity (interactions)	Popularity of POI as number of interaction types.
popularity (routes)	Popularity of POI as number of route interactions.
avg-distance-country	Difference between distance of user to POI and average of distances per countries.
avg-distance-click	Difference between distance of user to POI and average of distances of clicked POIs per countries.
avg-distance-route	Difference between distance of user to POI and average of distances of routed POIs per countries.

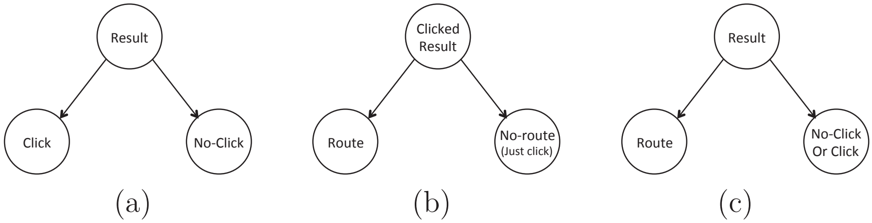

Figure A.4.1: Prediction tasks. (a) Click/no-click prediction (b) Click/route prediction (c) Route/no-route prediction.

1) Click/no-click prediction We formulate the click prediction problem as a binary classifier to predict a user's response (1 for click and 0 for no-click).

Table A.4.2 shows the achieved Precision at different Recalls of the click/no-click models trained using both baseline-only and baseline plus proposed features. The average precision does not improve much when the proposed features are added.

To evaluate our model we plotted the Receiver Operating Characteristic (ROC) curve. Figure A.4.2 shows the ROC curve for GBRT and LR models for both baseline-only and baseline plus proposed features. Again, the ROC curves demonstrate that there is almost no difference between the effectiveness of the two compared models.

The performance of click prediction models reported by Lymberopou-

Table A.4.2: Precision across different recall values for click/no-click classifiers with and without proposed features. (B): Baseline features (P): Proposed features.

	Avg. Precision			
Recall Value	GBRT(B)	GBRT(B+P)	LR(B)	LR(B+P)
50%	0.96	**0.97**	0.93	0.92
60%	**0.96**	**0.96**	0.92	0.92
70%	**0.94**	**0.94**	0.90	0.89
80%	**0.92**	**0.92**	0.84	0.85

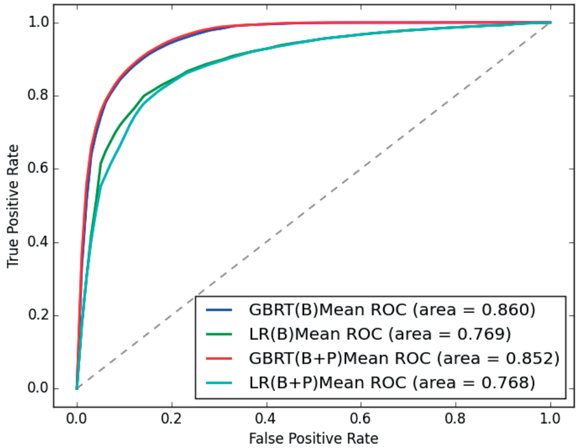

Figure A.4.2: Click/no-click ROC curve. (B): Baseline features (P): Proposed features.

los et al. [51] in terms of Precision at different Recall values is between 0.20 and 0.52. This is much lower than the performance of the click prediction in location search found in our research, where Precision is higher than 0.90, even when trained only on baseline features. A possible explanation for the highly improved results is that location search is limited to the queries about locations, while the queries in mobile local search are more general. For instance, if the query is 'gasoline,' location search just shows the locations of gasoline stations near the user's location, while in mobile local search the system provides more general results such as the definition of 'gasoline,' gasoline prices, and news related to gasoline.

2) Route/no-route prediction We formulate the route prediction problem as a binary classifier to predict whether a user will engage in the route interaction after seeing the search results (1 for clicking a result and then routing and 0 for all other cases).

The samples that received route interaction, are labeled as route (1) and the samples that are just clicked and did not receive the route

Table A.4.3: Precision across different recall values for route/no-route classifiers with and without proposed features. (B): Baseline features (P): Proposed features.

	Avg. Precision			
Recall Value	GBRT(B)	GBRT(B+P)	LR(B)	LR(B+P)
50%	0.96	**0.97**	0.92	0.92
60%	**0.96**	**0.96**	0.92	0.92
70%	**0.95**	**0.95**	0.90	0.90
80%	**0.93**	**0.93**	0.85	0.86

interaction are labeled as no-route (0).

Table A.4.3 shows the achieved Precision at different Recalls of the route/no-route model trained using both baseline-only and baseline plus proposed features. The average precision does not improve much when the proposed features are added.

Figure A.4.3 shows the ROC curve for GBRT and LR models for both baseline-only and baseline plus proposed features. Again, the ROC curves demonstrate that there is almost no difference between

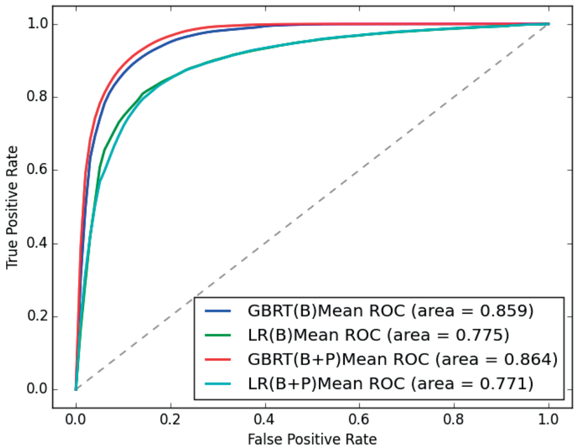

Figure A.4.3: Route/no-route ROC curve. (B): baseline features (P): Proposed features.

Table A.4.4: Precision across different recall values for click/route classifiers with and without proposed features. (B): Baseline features (P): Proposed features.

	Avg. Precision			
Recall Value	GBRT(B)	GBRT(B+P)	LR(B)	LR(B+P)
50%	0.74	**0.96**	0.72	0.88
60%	0.67	**0.95**	0.70	0.87
70%	0.63	**0.93**	0.65	0.82
80%	0.60	**0.89**	0.61	0.72

the effectiveness of the two compared models.

3) Click-route prediction We formulate the click-route prediction problem as a binary classifier to predict a user's interaction after having clicked a result (1 for click and 0 for route). Because a route interaction always follows a click, we removed from our dataset the query-result pairs that did not receive a click.

The performance of prediction models for route interactions are shown in table A.4.4. The GBRT and LR models that are trained by baseline plus proposed features show considerably higher Precision at Recall values compared to the models that employed baseline features only. Moreover, there is a clear increase in the reliability of prediction when we used the proposed features in GBRT and LR models (figure A.4.4).

A possible explanation for the improved results using the proposed features is the following. The baseline features were originally chosen for local search [51]. One of the main differences between location search and local search is the possibility to choose the route interaction. The proposed features specifically address the route interaction by including distance measures and popularity indications of the results. Thus, we may conclude that user modeling in location search services is clearly different from user modeling in local search or web search, and therefore

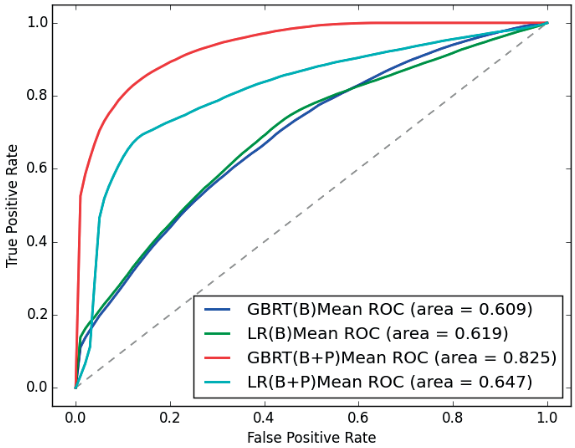

Figure A.4.4: Click/route ROC curve. (B): baseline features (P): Proposed features.

it makes sense to customize prediction models based on the available interaction types.

A.5 Discussion

In this work, we analyzed location search logs of a popular GPS-navigation system and compared user interaction with location search to that of web search and local search.

In section A.3, we showed that user interaction with location search has certain similarities to web search, but also some differences. The similarities include the number of queries per session and the relative session length on tablets compared to mobile. On the other hand, due to specific usage scenarios of location search, sessions, and queries are shorter in location search compared to the web search. Our observations on location search vs. web search have implications for the interaction design and underlying technology for location search.

Our statistical observations also showed similarities and differences between tablet and mobile location search. People use the mobile location search more, especially during working hours. In addition, mobile location search sessions and queries are shorter than comparable sessions on tablets. A possible explanation is that tablets are used more often for pre-trip planning, while mobile phones are used "on the go." The design of the different search services should take these differences into account to improve usability.

In section A.4, we compared location search to mobile local search, attempting to predict user interactions. We proposed several new features to be used for such predictions, and found that the proposed features considerably improved predictions of click/route interactions. A possible explanation for the differences in prediction accuracy can be found in the users' intent. In location search the intent of the user is to route to a POI, while in mobile local search the intent of the user may be more varied, e.g., the user often only wants to find information on the POI. Due to the more focused intent of the user, in location search predictions about interactions can become more accurate, at least, if the proposed features are included.

A major use of these results is that they can potentially be employed to improve the ranking of search results. Using the classifiers developed for click/route prediction, for each search result it can be predicted with high accuracy whether the user will route once this search result is clicked. Therefore, rankings can be improved by giving a higher rank to those results for which it is likely that the user will route after clicking.

A.6 Summary

In this appendix, we discussed modeling user interactions for click and route prediction in location search, finding that search platforms influence a user's interactions, and that accurate prediction of click/route interactions can be made for location search, using new features that we proposed.

The core text of this thesis investigates user models in the virtual environments of video games, called player models. While video game environments generally provide more features and more possible interactions than location search services, some of the lessons learned in the studies discussed in this appendix apply to the studies in the core text. In particular, we found that in most studies, the platform of choice and/or the user's intent have a considerable influence on user behavior, and thus on predictions of that behavior.

B
Studied Video Games

In this appendix, we explain the game play for the video games used in this research, namely *StarCraft*, *Destiny*, *Battlefield 4*, *Counter-Strike*, and *Dota 2*. These games have been popular in recent years and they have millions of players across the world. Moreover, they provide different character types with various functionalities. As a result, there are different ways to play a match. The variations in playing styles provide the opportunity to study players' behavior across the world.

We mainly concentrate on the games *StarCraft* and *Destiny*, as they are used in multiple chapters in the thesis. The other three games only get a brief treatment, as they are used in one chapter only.

Figure B.1.1: StarCraft Zerg versus Protoss.

B.1 StarCraft

Among AI researchers, Real-Time Strategy (RTS) games have been a popular research domain in the past decade. In particular, the complex, partially observable, and dynamic environments of RTS games motivate AI researchers to study different approaches and techniques to create strong AI, analyse the games, and model players. In particular, winner prediction is a highly relevant topic of AI research into RTS games. In *StarCraft*, winner prediction is challenging because players have many action choices, in a discrete environment where players manage their units concurrently. Moreover, the strategy of players depends on the match type. These factors increases the complexity of winner prediction.

StarCraft has been a popular RTS game since 1998. Figure B.1.1 shows a battle in StarCraft. In *StarCraft*, players gather resources to strengthen their economy. To provide military power, they must spend resources to construct buildings, research new technologies, and training units. The goal of the game is to destroy all of the opponent's bases and armies. *StarCraft* has various maps that differ in dimensions, arrangement of resources, and the areas that are build-able and walk-able.

StarCraft includes three different playable races: Terran, Zerg, and Protoss. The player chooses one of the races to play at the start of a match. While the races are well-balanced, they each need a different playing style. Very generally speaking, Terran are defensively strong and employ slow units, Zerg are offensively strong and employ fast units, and Protoss form a middle ground. Some of the commands (such as move and build) are available for all races, while some other commands are available only for one race (e.g., the siege command is limited to some units of the Terran race). Moreover, considering the race type, different units are available. For instance, Protoss has Carrier as an air unit for transporting ground units, while Terran has Medivac Dropship as an air unit for transporting ground units and healing them.

Variations of the units and commands for each race provides the opportunity to choose different ways of playing. Since each race needs a different playing style, using a playing style that is not fitting the chosen race may very well lead to defeat. It is possible to recognize a playing style already early in a game, meaning that sometimes it is possible to accurately predict the ultimate winner of a match already in the first few minutes of the match – if one of the players is employing a strategy unsuitable for the chosen race.

Winner prediction in matches where a player faces an AI bot tends to be relatively easy: AI bots are fairly weak and tend to lose against players who have some experience with the game. Consequently, considerable research has been invested in this area to make AI bots stronger. In human vs. human matches, winner prediction is not as straightforward.

Figure B.2.1: Destiny 1 PvP.

B.2 Destiny

Destiny is a FPS game with strong influences of Massively Multi-player Online Role-Playing Games (MMORPG). *Destiny* has a science fiction story that merges the characteristics of different game genres. It provides a wide range of human Player versus Environment (PvE) and PvP game modes. Figure B.2.1 shows a PvP game mode in *Destiny*.

In *Destiny*, players can participate in missions, events and raids. They engage in combat and other activities to gain new abilities, more powerful guns, and to level up their character. Players can run, jump, crouch, shoot, and use melee weapons. *Destiny* includes three main character classes: Hunter, Titan, and Warlock. Each has different strengths and weaknesses, with access to various abilities. The player chooses one of these classes at the start of a match. Each class includes subclasses that determine the specific upgrades and improvements of the main class.

In *Destiny*, PvP content is accessed via the Crucible, which is a hub for PvP content in the game. In the Crucible, players can choose different modes of play, with varying rules and objectives. In most of the game modes levels and gear are disabled, meaning that bonuses conferred by these are equalized among the players. Weapon stats and abilities are generally enabled. Some of the game modes are only available during specific events.

Points are generally scored by killing (with bonuses for particular kinds of kills), and assisting and supporting team mates. In particular game modes points can also be scored for capturing or neutralizing zones, reviving team mates, and deploying or neutralizing probes. Points for particular actions may vary between game modes, and there is also variation in how many points go to the team, and how many points go to individual players. Most game modes represent win-loss matches, i.e., players win as a team. A few game modes are free-for-all matches, in which each player receives a rank at the end to compare his or her performance with the performances of the other players.

There are 13 PvP modes in *Destiny* at the time of writing: Skirmish, Control, Salvage, Clash, Trials of Osiris, Doubles, Iron Banner, Elimination, Rift, Mayhem Clash, Zone Control, Rumble and Supremacy. PvP modes could be available as permanent, weekly, or event based. For reasons of space not all of these are described here, but they include the following:

- Skirmish: Skirmish is a 3v3 PvP mode whereby the first team which earns 5,000 points wins the match. The objective is to keep the teammates alive and fight the enemy.

- Control: Control is a 6v6 PvP mode. In this mode, three flags are scattered around the map, and teams must capture flags and defend them.

- Salvage: Salvage is a 3v3 mode. The goal is to capture a target point and collect secrets. The team that did not capture the target point, must interrupt the first team. The team that collects more secrets in a limited time wins the match.

- Clash: Clash is a 6v6 PvP mode in which players team up in 2 teams. Teams fight to earn 10,000 points by getting kills and assists.

- Trials of Osiris: Trials of Osiris is an 3v3 Elimination game type. This mode requires a purchased pass that is valid for three losses. Players need to make their team. In this mode, player's Light and gear advantages enabled.

- Doubles: Double is a 2v2 game mode with teammate revival.

- Iron Banner: Iron Banner is a 6v6 with limited-time game mode that is available once every month and lasts for one week. Level advantages are enabled in this mode

- Elimination: Elimination is a 3v3 game mode. Each player has one life, but can be revived. The first team to kill the entire other team wins the round, and the first team to win five rounds wins the match.

- Rift: Rift is a 6v6 capture the flag game mode. Players must capture the flag in the center of the map and deliver it to the enemy rift.

- Mayhem Clash: Mayhem Clash is 6v6 game mode. in this mode, super, grenade, and melee energy recharge frequently.

- Zone Control: Zone control is 6v6 mode. Only captures of zones count towards victory.

- Rumble: Rumble is a free-for-all game mode with 6 players game mode. First one to reach the score cap wins.

Figure B.3.1: Battlefield 4.

- Supremacy: Supremacy is 6v6 game mode with limited time. Players must pick up enemy ”crests” to score points. The team with more points wins.

B.3 Battlefield 4

Battlefield 4 is a FPS game, which is considered to be a representative game for multi-player FPS games, a very popular subgenre of video games [93]. Figure B.3.1 shows the game environment.

The most import goal of a player of *Battlefield 4* is to win a match. The game enables many options for different strategies and tactics. Points can be earned by reaching goals (e.g., winning a match) and sub-goals (e.g., supporting the team). Players can choose to forgo per-

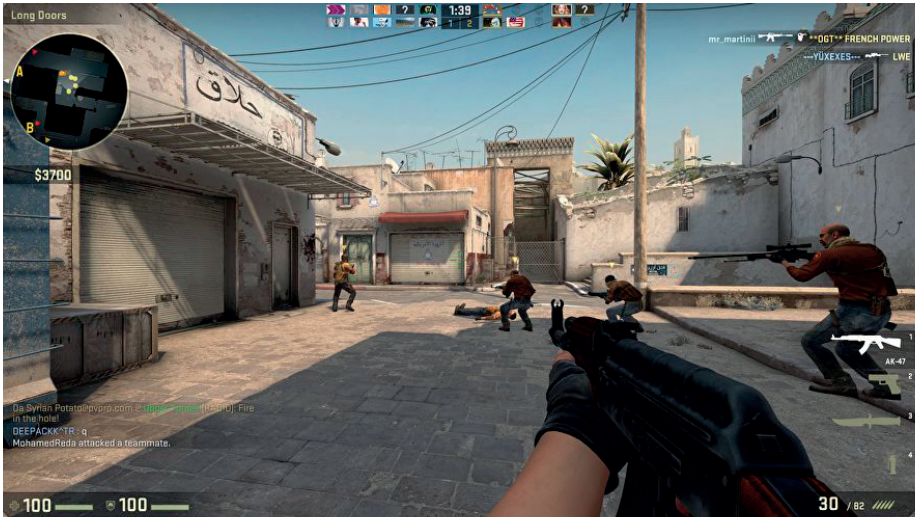

Figure B.4.1: Counter-Strike.

sonal achievements in favor of the team by supporting teammates and defending objectives, or they can focus on achieving individual goals. The game also stimulates collaboration within a team by, for instance, awarding points for kills based on teamwork.

B.4 Counter-Strike

Like *Battlefield 4*, *Counter-Strike* is a FPS game, mostly played online on PC. *Counter-Strike* is one of the most popular games on the digital distribution platform Steam. The latest edition, *Counter-Strike: Global Offensive*, is used in this study. Figure B.4.1 shows the game environment. In *Counter-Strike*, two teams, namely terrorists and counter-terrorists, compete against each other on different maps and try to complete objectives, which depend on the game scenario/game mode. For example, in the bomb scenario the terrorists must plant a bomb at a specific bomb site, while the counter-terrorists need to prevent this. In the hostage scenario, the counter-terrorists must rescue the hostages

Figure B.5.1: Dota 2.

and the terrorists must defend them. At the end of each round, players earn in-game money based on their performance. The money can be spent on more powerful weapons. By completing some specific objectives such as killing enemies, bonus money can be earned.

B.5 Dota 2

Dota 2 (Defense of the Ancients) is a Multiplayer Online Battle Arena (MOBA) game, a genre in which the player controls a single character in one of two teams with the main goal of destroying the units of the competing team and destroying the opponents' "Ancient," which is a heavily defended building in the opponents' base. *Dota 2* is only available on PC, and just like *Counter-Strike* one of the most popular games on Steam. Figure B.5.1 shows a battle in *Dota 2*.

Two teams consisting of five players play against each other. At the start of the game each player must choose one of the (at the time of this research) 115 playable heroes, although in some game modes the selection is limited. The heroes are broadly divided into three categories: strength, agility, and intelligence. Strength heroes have more health and health generation than other heroes, agility heroes have greater attack speed and armor, and intelligence heroes have more mana regeneration and a greater number of abilities. Besides these categories, heroes can have nine in-game roles: carry, nuker, initiator, disabler, durable, escape, support, pusher, and jungler. Each hero can have one or more of these roles. It is important to get balance in a team by choosing heroes with different roles. For example, carry heroes are weak in the beginning of the game and often need help of support heroes, but later in the game they are usually the strongest and most useful heroes.

Summary

In this research, we study players' interactions in video games to understand player behavior. The first part of the research concerns predicting the winner of a game, which we apply to *StarCraft* and *Destiny*. The second part of the research concerns distinguishing playing styles of players of *StarCraft* and *Destiny*. In the final part of the research, we investigate to what extent playing styles are related to different demographics, in particular to national cultures. We investigate this for four popular MMOGs, namely *Battlefield 4*, *Counter-Strike*, *Dota* 2, and *Destiny*. Below we briefly discuss the results of the research.

In chapter 2, we investigate to what extent it is possible to predict the winner of a *StarCraft* match, regardless of the races that are involved. We developed models for individual match types, and also general models for predicting the winner of non-symmetric matches, symmetric matches, and general matches. The research results in (1) a generic and relatively accurate model for winner prediction in *StarCraft*, and (2) a detailed analysis of which features are the principal components in accurately predicting the winner in this complex game. Specifically, our results show that we can predict the winner of a match with an accuracy of more than 63% on average over all time slices, regardless of the time slice and the combination of the match types. We also show that the economic aspects of *StarCraft* matches are the strongest predictors for winning, followed by the use of micro commands.

In chapter 3, we investigate the question whether it is possible to predict match outcomes in *Destiny*, a FPS game which mixes multiple video game genres. Two groups of models are presented for predicting match results: one group predicts match results for each individual game mode and the other group predicts match results in general, without considering specific game modes. Models achieve a performance between 63% and 99% in terms of average precision, with a higher performance recorded for the models trained on specific multi-player game modes, of which *Destiny* has several. We also analyzed performance metrics and their influence for each of the models. The results show that many key shooter performance metrics such as Kill/Death ratio are relevant across game modes, but also that some performance metrics are mainly important for specific competitive game modes. The results indicate that reliable match prediction is possible in FPS-type esports games.

In chapter 4, we investigate how different human players approach the popular *StarCraft* game in terms of preferences and strategies that may be inferred from game observations. In particular, we investigate how distinct match-types relate to the different playable races in the game. To this end, we propose features that reflect playing style, and uncover unique variations in playing style by means of Principal Component Analysis (PCA). Findings of experiments with clustering player styles of *StarCraft* players reveal that playing styles can indeed be distinguished in different match types. While one may expect playing styles to affect the chance of winning, results reveal that win probability is not significantly affected by playing style. The length of a match, however, is of significant influence.

In chapter 5, we aim to categorize players of *Destiny* into groups which are characterized by the way that they learn. Using archetype analysis, we discovered three player archetypes in the area of learning behavior, which seem to distinguish fast-learners, supportive-learners, and slow-learners.

In chapter 6, we examine playing styles in four popular Massively Multiplayer Online Games (MMOGs), namely *Battlefield 4*, *Counter-Strike*, *Dota 2*, and *Destiny*. We investigate to what extent national culture influences playing styles, and whether players from countries with similar cultures exhibit similar playing styles as well. We gathered playing style information from hundreds of thousands of players of these games, and applied correlation and clustering algorithms to relate playing styles to nationalities and to Hofstede cultural dimensions. We found that playing styles are influenced by nationality and cultural dimensions, and that there are clear similarities between the playing styles of similar cultures. In particular, the Hofstede dimension "Individualism" explained most of the variance in playing styles between national cultures for the games that we examined.

From our research and our findings in this thesis, we conclude that player models can be constructed from interaction data for video games. While the accuracy and effectiveness of the models depends on parameters such as the provided information, variations of interactions, and the complexity of the environment, we found that the models have predictive capabilities with regards to a variety of player features in video games, such as playing style, ability to be victorious, learning behavior, and cultural background.

Chovgan, Chowgan or Chogan, is a sporting team game with horses that originated in ancient Iran (Persia). It was later adopted in the Western World, known today as polo. The game has over 2000 years of history and it was played by both men and women. The game was accompanied by music and storytelling. Fragments of the game were periodically portrayed in ancient miniatures. This design is inspired by ancient miniatures and reflects the diversity in players and their equipment.