International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2018, 3-5 September 2018, Belgrade, Serbia

Hardcore Gamer Profiling: Results from an unsupervised learning approach to playing behavior on the Steam platform

Florian Baumann*, Dominik Emmert, Hermann Baumgartl, Ricardo Buettner

Aalen University, Germany

Abstract

Based on a very large dataset of over 100 million Steam platform users we present the first comprehensive analysis of hardcore gamer profiles by over 700,000 hardcore players (users playing more than 20 hours per week) covering more than 3,300 games. Using an unsupervised machine learning approach we reveal the specific behavioral categories of hardcore players, i.e. First Person Shooter, Team Fortress 2 player, Action game player, Dota 2 player, Strategy and action combiner, Genre-switching player. Subsequently we derive individual patterns of hardcore gamers in the categories found, such as strategy-action games combiner or game switching players. Our results are useful for computer science and information systems scholars interested in individual differences in user behavior as well as practitioners interested in game-designing.

© 2018 The Authors. Published by Elsevier Ltd.

Keywords: computer gaming behavior; hardcore gamer; player profiling; k-means cluster analysis

1. Introduction

Computer science and information systems scholars argue for the need to take individual differences into account in order to increase our understanding of information systems user behavior [1-3]. One of the most interesting areas for retrieving unbiased and large amounts of data is the computer gaming sector [4].

Online distribution networks like Valve’s Steam, one of the largest gaming networks for computer games in the world, provide an effective opportunity to analyze game behavior in an unbiased way, because the Steam platform community records actual gaming behavior without distortion across a broad spectrum of different players and games.

While a substantial amount of studies analyzing game playing behavior exist these generally focus on overall playing behavior [9-16]. However, in-depth analyses of specific types of players have been largely neglected. Since the so-called “hardcore gamer” spends most of his/her lifetime and money in gaming [8, 9], it is valuable for computer science and information systems scholars to understand a player’s intentions and behavior.

* Corresponding author. E-mail address: florian.baumann@studmail.htw-aalen.de

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Selection and peer-review under responsibility of KES International.

10.1016/j.procs.2018.08.078
That is why our study focuses on the behavior of “hardcore gamers” in a lot more detail. This type of player is more dedicated to gaming in almost every way, for example due to their high level of involvement in games, quantified by time spent playing and the scale of their respective in-game achievements [5-7]. Hardcore gamers can be described as people who play as a lifestyle preference and invest substantial amounts of time and money on games [8, 9]. They constitute the pioneers of a particular game, despite being the smallest group of players among the total player-base, and they help to define the experience for their fellow players through their own actions and behavior [7]. By identifying and analyzing their playing patterns it is possible to see how games are perceived by these influential players [7]. Such information can help support improvements in game design and game development [4, 7]. It is also important for the game industry to know and better understand their most influential players in terms of marketing and sales-promotional activities.

In this paper, we provide interesting results from an unsupervised learning approach to the analysis of the playing behavior of over 700,000 hardcore players, covering more than 3,300 games. Our analysis is based on a very large dataset, collected by O’Neil et al. [10]. The dataset was originally used to analyze the gaming behavior of over 100 million Steam users in general, and offers numerous possibilities for follow-up research [10]. For this reason, in our study we focus on a subset of players with the highest playtime in the dataset. In order to fit into the subset, Steam players classified as “hardcore gamers” where selected. According to Poels et al. [17] “hardcore gamers” play 19 hours per week on average. O’Neil et al. [10] also demonstrated that the 95th percentile of gamers has a total playtime of 1,233.9 hours, while the 99th percentile has 2,660.1 hours of total playtime. Based on these results, players with a minimum total playtime of 2,000 hours and a minimum two-week playtime of 40 hours across all games were selected as “hardcore gamers”. This subset represents the active Steam community and is characterized by the large amount of time the players spend playing computer games.

Our results identify six behavioral subtypes of hardcore gamers as well as the games and genres they play.

In the next section we examine related work in the field and outline our methodology, before presenting results in-depth and a discussion of them. We then conclude by highlighting limitations and future work.

2. Research Background

While scholars are becoming more and more interested in hedonic information systems, the focus on computer games as a serious field of research was established within the last two decades. As a result there are comprehensive general overviews of gaming behavior [4, 16, 18, 19] and a lot of studies on individual games and their related communities [11, 13]. The authors of these studies limit their findings since data were regularly biased due to the single game focus. Authors of studies analyzing gaming behavior using questionnaires limit their findings due to self-selection and social desirability biases.

For these reasons scholars acknowledge the desirability of capturing unbiased actual behavior data of a broad range of players using gaming platform data. Since the Steam platform (store.steampowered.com) is the largest digital computer gaming platform with an open application programming interface, analyzing Steam data offers potential insights into realistic gaming behavior. Scholars [10, 21] have demonstrated that Steam data is unbiased as they have also found well-known real-world phenomena [22] in gaming behavior data from Steam. For instance, Becker et al. [21] demonstrated that the network shows small world graph characteristics by studying the structure of Steam’s gaming network. They used Steam network data of 9 million users, 82.3 million friendships, 1.98 million groups and 1,824 games using a web crawling program [21]. O’Neill et al. [10] also demonstrated the power law distribution in Steam data.

However, while gaming behavior was analyzed either based on biased data from questionnaires, or unbiased data from gaming platforms but with a single-game focus or small data, no gamer type specific analysis using massive unbiased data exists – with one exception: Sifa et al. [12] analyzed cross-games behavior based on a Steam sub-sample of 6 million players and 3,000 games. As a result, Sifa et al. [12] demonstrated that gaming behavioral data
can be clustered according to the users’ playtime. Cross-games behavior is a term used by Sifa et al. [12] to refer to cross-sectional analyses of relationships between players / game ownership, games and playing time.

We extend the approach adopted by Sifa et al. [12] using cluster analysis but based on massive data from all 108.7 million Steam user accounts and their 384.4 million owned games. In addition, to offer in-depth results we cluster them based not only on playtime but also on game type/genre.

3. Methodology

3.1. Dataset

The dataset we use is from a comprehensive analysis of the Steam gaming network, comprising all 108.7 million user accounts and 384.4 million owned games, a scale that makes this dataset unique in terms of both magnitude and focus [10]. The dataset [10, 22] covers gaming behavior across several dimensions, for example, social connectivity, playtime, game ownership, genre affinity and monetary expenditure. Compared to a lot of other studies collecting (small and biased) samples of Steam user data via web crawling, the dataset we use here captured the data directly from Steam’s application programming interface.

3.2. Data cleansing

The data set has been adjusted according to the definition of “hardcore gamers”. In order to classify “hardcore gamers”, all Steam-IDs were filtered by their playtime over the previous two weeks and their added total playtime for all games. It was also necessary to consider games only, and no other software such as developer tools. All lines were removed that did not correspond to the game type or download content (dlc). Some entries were not games but developer or streaming software, even though they had the entry of a specific game type. To cater for this, we removed all entries with genres that were not playable, examples being “Video Production” and “Utilities”. The final dataset consists of 707,477 players and 3,366 different games.

3.3. Clustering method

To analyze a gamers’ behavior we use K-means clustering. The Player Characteristic (SteamID) and Game Characteristic (GameID) correlates with the age of an account, where lower IDs indicate older accounts. We use Euclidean distance for measuring the similarity. Related work shows that K-means clustering is an appropriate method for analyzing play behavior and creating player profiles [12, 18, 26]. It is necessary to find the best number of clusters to run the K-means algorithm with the cleaned-up dataset. There are two methods to search for the best number of clusters; these are the silhouette and the elbow method [24]. It is also necessary to encode the categorical data in the final table using Label Encoding since the features Player Characteristic and Game Characteristic are categorical and the K-means algorithm can only be used with numerical data [25]. To preserve information about the account age we sorted the Player Characteristic and Game Characteristic in ascending order. This process makes possible the correct usage of the K-means algorithm, and the algorithm clustered the Player Characteristic by games and their respective playtime. Most clustering performance metrics such as Adjusted Rand Index, Homogeneity, V-measure or Fowlkes-Mallows score require knowledge of the ground truth labels. Since these labels are unknown for the given hardcore gaming dataset, we report the average silhouette score for each cluster in order to evaluate cluster quality. With further analysis it is possible to explain these clusters in more detail by creating player profiles for each cluster [18]. These profiles characterize the playing behavior of hardcore gamers on the Steam platform [26].

For cluster naming and explanation we use all the characteristics the dataset provides.
4. Results

4.1. Most played games and genres

Since 10 games drew more playtime than all the other 3,537 games together (Fig. 1), we found a power law-like distribution of playtime in Steam data.

A genre-related analysis revealed that in general action games dominate gaming behavior (Figure 2).

Figure 1. Game-related playtime distribution [y-axis shows the total playtime from zero to 1 billion hours].

Figure 2. Genre-related playtime distribution without Dota 2 since Dota 2 is related to both genres action and strategy [y-axis shows the total playtime from zero to 1 billion hours].
4.2. Gamer profiles

K-means is an unsupervised algorithm that clusters data entries based on the given number of clusters. The most appropriate number of clusters for a given dataset can be calculated using the silhouette method or the elbow method. Both methods show that gaming behavior can be portioned in six clusters (separation index: 0.4). For this reason we built six clusters to meaningfully group the behavior of hardcore gamers (Table 1).

<table>
<thead>
<tr>
<th>No.</th>
<th>Cluster name</th>
<th>Number of players</th>
<th>Share of playtime</th>
<th>Playtime in hours</th>
<th>Silhouette score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First-Person Shooter</td>
<td>221,507</td>
<td>13.5 %</td>
<td>3,293.0</td>
<td>2,089.0</td>
</tr>
<tr>
<td>2</td>
<td>Team Fortress 2 player</td>
<td>96,362</td>
<td>9.2 %</td>
<td>2,956.6</td>
<td>1,567.1</td>
</tr>
<tr>
<td>3</td>
<td>Action game player</td>
<td>295,588</td>
<td>11.2 %</td>
<td>2,938.5</td>
<td>1,855.0</td>
</tr>
<tr>
<td>4</td>
<td>Dota 2 player</td>
<td>14,438</td>
<td>39.5 %</td>
<td>2,638.3</td>
<td>1,542.3</td>
</tr>
<tr>
<td>5</td>
<td>Strategy and action combiner</td>
<td>1,008</td>
<td>15.5 %</td>
<td>5,409.9</td>
<td>3,368.4</td>
</tr>
<tr>
<td>6</td>
<td>Genre-switching player</td>
<td>78,574</td>
<td>11.1 %</td>
<td>3,080.7</td>
<td>2,357.7</td>
</tr>
</tbody>
</table>

We used the K-means algorithm with the features Player Characteristic (N=707,477), Game Characteristic (N=3,366), and (summarized) Playtime per game in hours (N=2,222,468,173, M=45.96, STD=292.05).

**First Person Shooter (Cluster 1)**: Hardcore gamers in this cluster use most of their time to play shooters such as Counter-Strike (Global Offensive, and Source), Call of Duty, Left for Dead (2) or DayZ. While users of this category also play Dota 2 and Garry’s Mod, the playtime of these two games is small compared to the time used for shooters. With the exception of Dota 2, all games played by hardcore gamers in this category are not free to play and are sometimes very cost-expensive.

**Team Fortress 2 player (Cluster 2)**: Hardcore gamers in this category use most of their time to play Team Fortress 2 followed by Skyrim, Civilization V and Terraria. Players in this cluster tend to use free-to-play games such as Team Fortress 2 or War Thunder.

**Action game player (Cluster 3)**: Users in this category are attracted to action games such as Dungeon Defenders or Grand Theft Auto (IV). None of the Top 50 games in this cluster are free to play with the exception of Dota 2.

**Dota 2 player (Cluster 4)**: This category is dominated by players who spend heavy playing time on Dota 2 (420 million playing hours in total). Dota 2 is a free-to-play multiplayer online battle arena computer game. With a big gap in terms of playtime, a series of ego-shooters is also played by users in this category (Fig. 3).
Strategy and action combiner (Cluster 5): Users in this category play games which simultaneously combine strategy and action gaming elements. Users here often switch between Civilization V, Counter-Strike, Dota 2, Garry’s Mod and Skyrim (alphabetical order).

Genre-switching player (Cluster 6): Hardcore gamers in this category move from genre to genre (role-playing, adventure, indie, strategy, action, etc.) when playing games (Fig. 4). It seems that these players are very open to new gaming types and elements.

5. Discussion
Our results demonstrate that hardcore gamers can be grouped by separate clusters to reflect their individual playing behavior. These clusters seem to reflect individual differences in players, which is an interesting finding for computer science and information systems scholars [1-3]. For example, Jansz and Tanis [37] found that first-person shooters (cluster 1 in our study) scored highest on motives with respect to competition, and challenge. In addition, it can be speculated that strategy and action combiners (cluster 5) score higher on consciousness, and that genre switching players (cluster) score higher on openness to experience [38, 39].
In addition we found that the free-to-play property of games has a substantial impact on the playing of these games. Many top games, and even games with less playtime, are free to play games. These make up a very large proportion of the content on Steam. Examples include the big names Dota 2 and Team Fortress 2, as well as increasingly popular free-to-play games like Path of Exile [27]. One of the reasons for the popularity of these free to play games is that you can basically play them for free without any restrictions on gameplay. Spending money on these titles does not provide any competitive advantage over fellow players. There are also titles where this is not the case, but this type of free-to-play game is not noticeable in the analysis. Free-to-play games usually use microtransactions to finance them. The player is thus less inhibited from first testing the game and only spends money if he/she wants to [29].

It is noticeable that many of Steam's top titles are played in the Electronic Sports League (ESL). These include games like Dota 2, the Counter-Strike series and titles from the Call of Duty series [30-32]. Games and teams found here will be famous in the context of this eSport which provides a good marketing opportunity. Free-to-play games and games that can be purchased through the acquisition of software licenses are equally popular on Steam and in the ESL [31, 32]. Another aspect that makes these games become popular is expenditure on them from social media platforms. The most games played in the ESL are played from streamers on YouTube, Twitch and other streaming-platforms. This has a large potential to, first, increase the distribution of games and, second, to analyze game popularity [33].

The clusters also show which types of games are played most frequently via the gaming platform Steam. Genres such as car racing games or general sports games are relatively less well represented compared to the others. Players tend to be attracted more to real-time strategy games like Dota 2, Shooters or more generally Action, Strategy, RPG, Indie or Adventure titles. Certain titles might be more interesting to users of personal computers than others. One reason for this could be that consoles are easy to use via a controller that is well suited for sports games. Keyboard and mouse work well for shooters and real-time strategy games, as they provide a greater variety of key combinations and engagement with the monitor is very precise [34]. Steam tries to close this gap through the use of its proprietary controller [35]. It is possible to connect popular controllers for the Xbox series and PlayStation series to the PC and use it as a gaming interface. Another reason could be that very popular and famous sport titles like the Fifa series are Electronic Arts (EA) games. EA has its own game distribution platform for personal computers called Origin.

Another notable feature of heavily used games is that these games have multiplayer capability (e.g., almost all top shooters, strategy games like Civilization V, etc.).

5.1. Limitations

A limitation is related to the fact that we did not observe the behavior when a game was being playing, but we retrieved behavioral data summarizing facts such as the playtime of each gaming session. However, the approach used here is in line with several studies analyzing gaming behavior [40].

Since numerical and categorical data is present in the dataset, a mixture of similarity measurements such as Hamming and Euclidean distance could yield better clustering results [41].

Despite analyzing playing behavior using massive data (over 700,000 hardcore players covering more than 3,300 games) our results are internally valid for Steam platform users [10, 23, 35] but eventually slightly biased (single data source selection bias) [42].
5.2. Future Work

Based on the clustering results derived from this work we intend to analyze cluster-personality relationships in one future study and personality-mining chances in another study. Mining information about a user’s personality based on gaming behavior is very interesting for recruiting [43-47].

Furthermore, while we used K-means algorithm due to its good performances in previous studies that also used clustering gaming data [12, 16, 18], and due to its low time complexity as well as good applicability on a large scale [48], future work could use different clustering algorithms like DBSCAN or k-prototypes and genre-specific clustering. In addition, this study could be expanded by examining how gaming behavior changes over time.

6. Conclusion

Based on a very large dataset of over 100 million Steam users we analyzed the gaming behavior of over 700,000 hardcore players covering more than 3,300 games. Using K-means clustering as an unsupervised learning approach we identified specific behavioral categories (First Person Shooter, Team Fortress 2 player, Action game player, Dota 2 player, Strategy and action combiner, Genre-switching player), and subsequent individual patterns of gaming behavior. Our results are useful for computer science and information systems scholars interested in individual differences in user behavior [1-3] as well as practitioners interested in game-designing.

References

Based on gaming behavior is very interesting for recruiting and personality-mining chances in another study. Mining information about a user's personality can identify specific behavioral categories (First Person Shooter, Team Fortress 2 player, Action game player, Dota 2 player) and differences in user behavior as well as practitioners interested in game-designing. In addition, this study could be expanded by examining how gaming behavior changes over time.

Future Work

In the future, we intend to analyze cluster-personality relationships based on the clustering results derived from this work. We will focus on analyzing cluster-personality relationships based on a very large dataset of over 100 million Steam users. We will analyze the gaming behavior of over 700,000 casual and hardcore gamers. Based on the clustering results, we will attempt to analyze cluster-personality relationships.

References