Analyzing User Behavior Data in a Mobile Tennis Game

Maxim Mozgovoy
School of Computer Science and Engineering
The University of Aizu
Aizu-Wakamatsu, Japan
mozgovoy@u-aizu.ac.jp

***Abstract*—The users of online multiplayer games generate vast amount of data that can be later used to fine-tune the game and optimize user experience. The present paper is dedicated to the analysis of user behavior in a mobile tennis game World of Tennis: Roaring '20s. It implements a pseudo-multiplayer game mode by introducing AI-controlled opponents trained on real user data and mimicking human behavior. We analyze user recordings to ascertain they exhibit a variety of consistent play styles that can be learned by the AI system to ensure challenging and lasting entertainment for the players.**

Keywords—data mining; behavior capture; game AI.

I. Introduction

Free-to-play games constitute the overwhelming majority of mobile games currently on the market [1, 2]. This business model requires the game designers to devote more efforts to the creation of a sustainable *metagame* that would justify giving away for free most core game loops [3]. As a result, free-to-play games are typically designed *"for a (very) long duration of play"* to increase in-app spending [4].

One of the elements of a good metagame is the system of upgrades and virtual items that supports player day-to-day progress. Particular genres and peculiarities of specific games, however, may introduce additional issues into this process. This is the case of *World of Tennis: Roaring '20s*, a mobile lawn tennis game developed with active involvement of the author of this paper.

Many free-to-play mobile games implement some variation of online multiplayer play. Player-versus-player mode is one of natural cornerstones of a good metagame, able to deliver encounters with diverse opponents and certain social elements, such "play with or against your friends" modes, group chats, and player club/clan activities.

Tennis is by nature a player-versus-player game, so multiplayer matches look like must-have functionality. However, in practice virtually all existing multiplayer tennis games suffer from *connection lags* and *complicated matchmaking*. Tennis is a fast-paced game, so even minor connection drops might cause pauses and slowdowns during the matches. Furthermore, mobile devices are often used in transport or in public spaces, where internet connection might be slow or unstable. Since the players should compete against the opponents who are currently online, have acceptable data roundtrip time, and possess comparable ranks, matchmaking can be difficult as well.

To remedy the situation, we decided to implement a machine learning-based AI system that would control all the opponents in the game [5]. By playing tennis matches, people *train* their virtual characters (avatars) that can later substitute them in the game. In other words, people in *World of Tennis: Roaring '20s* compete with AI agents rather than with real online opponents.

Our preliminary experiments showed that the resulting AI system is skillful, reliable, and mimics human behavior reasonably well. However, it was difficult for us to test AI performance in the long run, i.e., as a part of a metagame. In particular, we were interested to find out whether AI-controlled players are able to provide diverse experiences and long-term fun, and to play on par with really good players, whose skills exceed the skills of ourselves and of our beta testers.

The goal of the present paper is to discuss some of the preliminary findings of user behavior analysis. We focus on human-controlled rather than AI-controlled characters to prove that the game itself is rich enough to provide lasting player engagement that we strive to support with the AI system.

As of today, it is clear that people are comfortable playing against an AI system if it is implemented properly. We have users who joined the game 1.5 years ago (at the time of the first public beta releases) and completed over than 4000 tiebreak matches[1]. We believe that our experience can be useful for the researchers and game developers, interested in metagame analysis and game AI.

II. The Structure of User Data

World of Tennis: Roaring '20s allows the users to play three different types of matches:

1. Quick matches. A game against a random opponent of a similar rank.
2. League matches. A player is always placed into a *league* with nine other AI-controlled characters. A

[1] By default, all matches in the game are tiebreaks, and last for 2-3 minutes on average. The user can buy the "custom match duration" item to play full matches. This is one of the most popular purchasable items in the game for now.

league match is the game against the next opponent in the league. After nine matches the player will be placed into another league with nine random opponents of comparable ranks.

3. Practice matches. Users can play a training game with a coach character. The coach can mimic the behavior of any league opponent, so the players can use this capability to prepare for their next league game.

An overwhelming majority of matches in the game are played according to the rules of tiebreak. The game ends when two conditions are satisfied: 1) one of the opponents scores at least seven points, and 2) the difference between the scores is at least two points. Until recently, all played matches were stored on the game backend in the "full format" that includes all actions made by both opponents (we used this information to analyze players' behavior and fix bugs). This backend data represents user matches "in the wild": we have a mixture of quick, league, and practice matches against diverse opponents of different skill levels. Our matchmaking algorithm forms player league so that it includes 2/3 opponents of lower rank, thus giving the player more chances to finish in the top three and be awarded a trophy. Our beta tests show that people feel frustrated if they cannot win any trophies for a long time, so we are trying to keep challenge on acceptable/enjoyable level.

These decisions, however, complicate certain types of analysis. For example, it is difficult to check whether AI-controlled characters play on par with human players, since on average people play against deliberately chosen weaker opponents, and thus score more victories.

III. Data Analysis

A. Dataset Description

In the present study, we will consider a dataset that includes the recordings of 213 randomly chosen users, who played 20 or more matches within one month (September, 2017). Unsurprisingly, the recordings are distributed among the users very unevenly. While the most dedicated user played over 1000 matches, the median number of recordings per user is only 56 (see Fig. 1).

Since this data shows a one-month snapshot of user activity, it includes the users of different skills and experience. The most crude skill/experience indicator is "player level" that is increased automatically as player progresses through the game and earns "skill points" (needed to improve individual abilities, such as shot accuracy, player speed, and so on).

Player level is roughly proportional to the logarithm of the number of victories. Player level in the data set ranges from 1 to 58 with a median value of 10.

B. Behavior Comparison Procedure

One of our primary interest in dataset analysis is to compare player behavior to reveal whether human players exhibit enough diversity of playstyles, and whether our AI system is able to reproduce them adequately to keep the users motivated to continue playing. To compare game recordings (and thus player behavior), we rely on the procedure described in [5]. In brief, it works as follows. The basic game process consists of two primary actions: move the player to a certain court location, and shot the ball to a certain location on the opponent's side. We can use these locations to build a heatmap of player actions, reflecting the popularity of particular locations for the user's playstyle (see Fig. 2). Heatmaps are converted into vectors and compared with other vectors using a dot product, yielding a similarity ratio in a [0, 1] range.

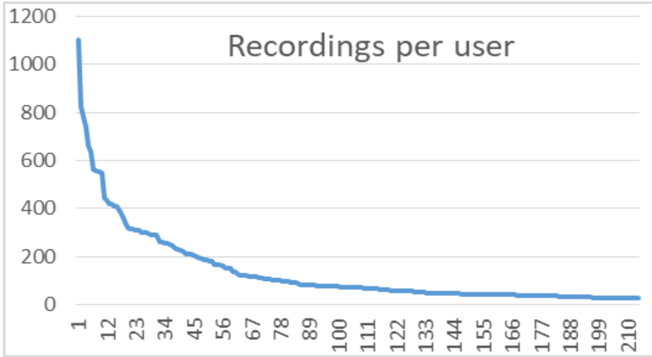

Fig. 1. Distribution of recordings among the users

C. Comparing Matches of the Same User

First, we wanted to find out whether an individual user exhibits similar play styles when playing against different opponents, while staying on the same player level. If players significantly alter their behavior to particular opponents, it means that our AI system has to distinguish opponents, too. To find out the answer, we selected 13 most dedicated users, having at least 400 completed matches in our dataset, and compared their recordings, corresponding to the same player level. All recordings were randomly separated into roughly equal groups, then all recordings belonging to the same group were concatenated, and finally two resulting recordings were compared using the heatmap method.

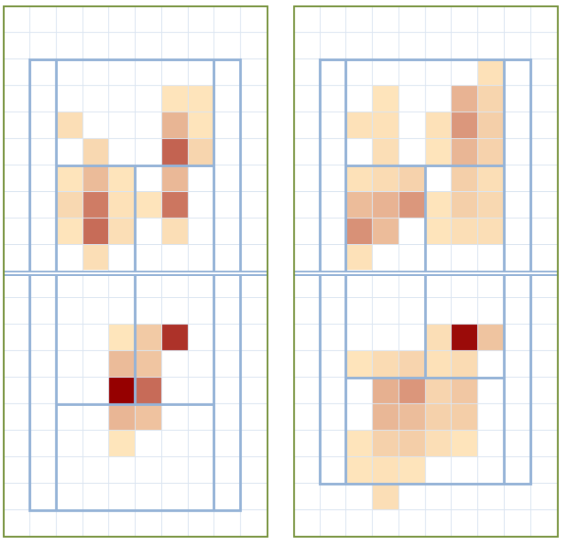

Fig. 2. Heatmaps of two distinct players. Darker locations correspond to more frequent targets of move and shot actions.

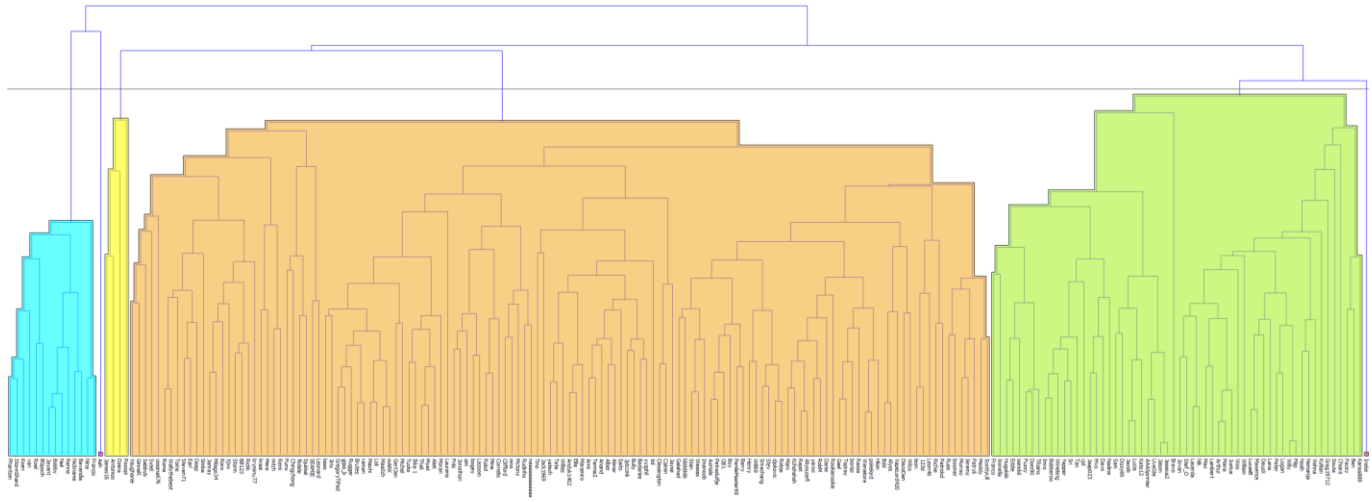

Fig. 3. Identified clusters of users. To measure the distance between two clusters, the algorithm uses the average distance between cluster elements ("average linkage" setting in Orange).

The resulting similarity values were 96.8% on average, with standard deviation not exceeding 0.13. Therefore, we decided that a single behavioral profile would be sufficient to play against any opponent.

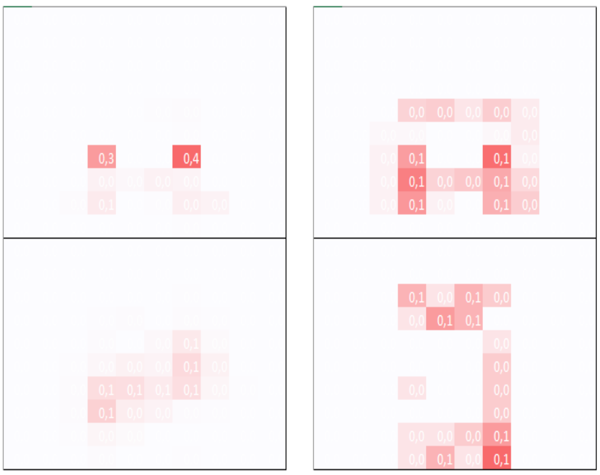

Fig. 4. Heatmaps of the users from clusters A and B.

Next, we decided to check how behavior changes due to user progress by comparing recordings corresponding to different player levels. As in the previous experiment, we concatenated all the recordings belonging to the same group (i.e., player level in this case) and compared the resulting recordings against each other.

Somewhat surprisingly, it turned out that the users adhere to the same playstyle even in the long run: character upgrades do not change the basic features of their behavior. The largest changes were detected on very low player levels (up to level 10). Since level 10 can be achieved relatively quickly (within several hours of play), we believe that these changes represent growing mastery of the game: beginners just learn how to play, and experiment with different actions. By level 10 or so they already achieve maturity and stick to their chosen style.

Only 1.9% of recordings of a player of level 5 or above got similarity score less than 50%, while 72.1% of recordings were at least 80% similar. This trend continues as players reach higher level. Just 0.1% of recordings of a player of level 10 or above have similarity lower than 50%, while 81.8% of recordings got the score of 80% or higher.

We should also note that the current game design does not enforce the users to experiment with their behavior. One might assume that upgraded player abilities introduce new elements of strategy. In reality, however, they often merely compensate higher abilities of the opponents. For example, one's shots might lose efficiency when the opponents upgrade their *speed* ability to run after the ball quicker, so it becomes necessary to upgrade own *power* ability to restore the balance. In turn, higher shot power reduces accuracy, so the player is forced to upgrade *accuracy*, and so on.

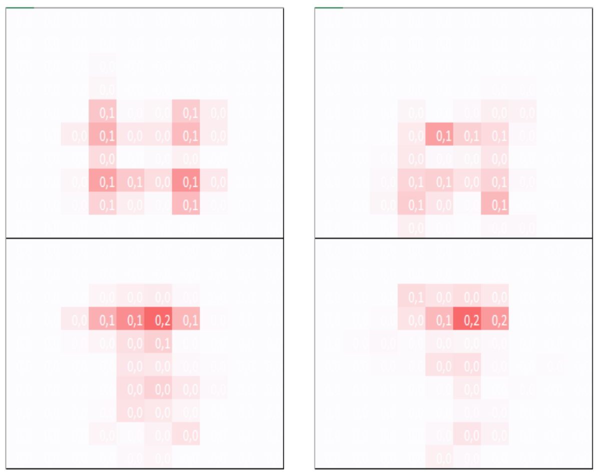

Fig. 5. Heatmaps of the users from clusters C and D.

D. *Comparing Matches of Different Users*

Our next goal was to ensure that the game process is rich enough to let the users exhibit enough diverse behaviors types, and thus enjoy meeting new opponents. We randomly selected 20 recordings of each user, concatenated them, and compared with concatenated recordings of the rest of the users, thus obtaining a 213×213 symmetrical similarity matrix.

Fig. 6. Heatmaps of the two outlier users (level 7 and level 19).

The average similarity in the matrix is 53.80% (standard deviation = 0.16), which is significantly lower than typical similarity between two different recordings of the same user. It was also interesting for us to check whether users can be clustered into groups, sharing similar behavioral patterns. We have to note that the algorithm of similarity assessment was initially designed as a simple method of checking whether certain users are “similar” or “distinct”, and it cannot be considered a reliable similarity function for clustering. In particular, it does not take into account sequences of actions, forming actual player tactics. Still, we used “Hierarchical clustering” capability of Orange [6] to identify clusters of players according to our similarity matrix (see Fig. 3).

The system identified four isolated clusters of users (we will refer to them as A, B, C, and D). In addition, two users were classified as outliers, not belonging to any cluster. We visualized the heatmaps of some players belonging to different clusters to understand the difference in playstyles. Two leftmost clusters A and B on the Fig. 3 are mostly comprised of novice users with median player level 5 and 6 respectively. The players of cluster A tend to stick to few move/hit locations, while the players of cluster B explore a wider range of movements. However, they seem to ignore certain key areas of the court, such as the middle zone of the own side of the court, widely used by more experienced players. The clusters C and D consist of players of median level 9 and 12 respectively. These users tend to adhere to fewer key locations of the court, but still use the whole court area (see Fig. 4 and Fig. 5). The remaining outlier users exhibit highly idiosyncratic behavior. They focus on few locations on the court, unusual for most users belonging to large clusters (see Fig. 6).

IV. Discussion and Conclusion

Our experiments show the first attempt to analyze a relatively large sample of user recordings in *World of Tennis: Roaring ’20s.* In our previous work [5] it was demonstrated that the proposed AI system mimics a small group of players reasonably well, which gave us confidence to use the experimental AI engine in the final game release. However, we also wanted to ensure that the specific design of mobile tennis would provide fun and diverse gameplay for the users. The results of our study show that the users are indeed diverse in their approach to tennis tactics, and they form clusters of somewhat similar behavior patterns. However, the currently used similarity function is too crude to reflect the nuances of user behavior, so there is a room for further investigation.

Our data suggest that the users are willing to play hundreds and even thousands of matches against the AI system if it provides them enough diversity and fun, so relying on a learning by observation-based AI can be one of the ways to overcome the difficulties of true online multiplayer gameplay development. Current user feedback make us believe that the most sensitive issue in the present game is matchmaking. The users are frustrated if they face unexpectedly strong opponents or feel that their gradual progress comes to a sudden stop due to increasing difficulty (and they often attribute it to a deliberately erected “paywall”, even if it is not the case).

One of the possible directions of the future research is also the analysis of the best players’ behavior profiles to reveal the most successful tennis strategies. We think that this kind of analysis can be helpful for the game developers in particular to reveal loopholes allowing the players to find unexpected winning strategies.

References

[1] Clairfield International, *Gaming Industry — Facts, Figures, and Trends (2018).* Available: https://bit.ly/2sv6QSg.

[2] K. Alomari, T. Soomro, and K. Shaalan, “Mobile Gaming Trends and Revenue Models,” in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2016, pp. 671–683.

[3] M. Sicart, “Loops and Metagames: Understanding Game Design Structures,” in *FDG*, 2015.

[4] P. Luban, *The Design of Free-To-Play Games.* Available: http://www.gamasutra.com/view/feature/6552/the_design_of_freetoplay_games_.php.

[5] M. Mozgovoy, M. Purgina, and I. Umarov, “Believable Self-Learning AI for World of Tennis,” in *2016 IEEE Conference on Computational Intelligence and Games*, 2016, pp. 1–7.

[6] J. Demšar *et al,* “Orange: data mining toolbox in Python,” *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 2349–2353, 2013.