Exploiting General-Purpose In-Game Behaviours to Predict Players Churn in Gameful Systems

[1]Enrica Loria
[1] Francesco Paissan
[1] Annapaola Marconi
[1]Fondazione Bruno Kessler (FBK)
Via Sommarive, 18, 38123 Povo (TN), Italy
eloria@fbk.eu

Abstract

The value of a game is assessed by measuring the intensity of the level of activity of its players. No matter how thoroughly though the design is, the litmus test is whether players keep using it or not. To reduce the number of abandoning players, it is important to detect in time the subjects at risk. In the literature, many works are targeting this issue. However, the main focus has been on entertainment games, from which articulated indicators of in-game behaviors can be extracted. Those features tend to be context-specific and, even when they are not, they are proper of full-featured games, and thus, impossible to adapt to other systems such as games with a purpose and gamified apps. In this preliminary work, we fed to an Artificial Neural Network general-purpose in-game behaviors, such as participation data, to predict when a player will definitively leave the game. Moreover, we study the appropriate amount of information, in terms of players' history, that should be considered when predicting players' churn. Our use case study is an on-the-field long-lasting persuasive gameful system.

Introduction

The richness of information deriving from real-time logging systems is often shadowed by the elevated amount of data itself, which can be noisy and difficult to treat. The very same phenomenon occurs in games and gameful systems, where every in-game action can be instantaneously stored. This gold mine of knowledge has to be cautiously handled to avoid drawing biased and misleading conclusions (Hooshyar, Yousefi, and Lim 2018). Nevertheless, gameplay data should be exploited to monitor, evaluate and improve players' experience. Studying this data enables, among the other analyses, is the prediction and prevention of churn, in particular in the industry - e.g., (Datta et al. 2000; Hadden et al. 2006; Coussement and Van den Poel 2008; Kawale, Pal, and Srivastava 2009; Periáñez et al. 2016). It is, in fact, of crucial importance to control players abandonment so that contingency strategies can be applied. Players' retainment is also important in other application domains, such as serious games and gameful systems. This is particularly true especially because of the escalation of interest that gamification has suffered in the last few years (Warmelink et al. 2018). Using elements generally proper of games to pursue an ulterior motive is an appealing concept (Huotari and Hamari 2012; 2017) and is drawing plenty of attention. In gameful systems, since the fun component must go hand in hand with the message that the designer is communicating through her system, keeping users engaged is even more challenging than in traditional games. Even though several studies have shown that gameful systems have generally a positive effect (Xi and Hamari 2019), players, as human beings, are unique, and hence, answer to the same mechanics differently (Orji, Mandryk, and Vassileva 2017). To comply with the many facets of the personality of the users involved in the system, many theoretical models, which aim at classifying them, have been proposed (Marczewski 2015; Yee 2016; Caillois 2001), together with data-driven approaches (Hooshyar, Yousefi, and Lim 2018) to study in-game behaviors. The latter represents a more objective source of information, which could be used in the design and development phases. Nevertheless, this type of data lends itself to interpretation. An example is the difficulty of evaluating the goodness of players' experiences (Yannakakis and Togelius 2011; 2018), which is highly related to the prevention of players churn: if the player is engaged enough, she will keep playing.

In this work we contribute to the Game User Research community by investigating whether gameplay data can be exploited to predict if and when a player will abandon the game, by structuring it in general-purpose behaviors monitoring the participation level, using an Artificial Neural Network. Since such behaviors can be monitored in real-time, this approach can result in a powerful online tool. Moreover, we test the boundaries of this methodology by identifying the amount of information needed to make the most accurate prediction, in terms of the players' history.

Related Works

The application of analytics to game-like data and research, known as Game Analytics (El-Nasr, Drachen, and Canossa 2016), is assuming a fundamental role in the support of the design and the development of games. Analytic procedures allow the evaluation of games, both online and offline, resulting in a strategic means to gather knowledge about players by observing their in-game behaviors. Observing the way players interact with the game, if done properly, could help in measuring the goodness of players' experience (Loh, Sheng, and Ifenthaler 2015). To this end, more and more developers and researchers made use of a plethora of sources to gain knowledge about their players (Isbister and Schaffer 2008). The objective nature of this approach led to a massive use of gameplay metrics to assist and improve the design and development of games (Kim et al. 2008). The sources from which game metrics can be gathered are several, an example being telemetry. The most common form of telemetry data is gameplay data, which is very knowledgeable when measuring the player experience is the purpose (El-Nasr, Drachen, and Canossa 2016). Moreover, it can be very easily retrieved through an automatic logging system (Kim et al. 2008). The possible employments of this data are several, from the extraction of basic learning indicators (Kiili, Moeller, and Ninaus 2018) to the instruction of difficulty adjustment mechanisms, and the evaluation of player performance (Mellon 2009). Modeling players' experiences through telemetry data (Gameplay-based PEM) (Yannakakis and Togelius 2011) is driven by the assumption that player experience can be extrapolated from real-time actions. Being able to monitor players' in-game behaviors gives immediate feedback on how they interact with the game, without the biases that are intrinsic in more subjective methods (Sangin 2018). Besides, it enables the designers to monitor and fine-tune the game online. Gameplay-based PEM can either be theory-driven or data-driven. Although theory-driven approaches can supply a rationale to the choices of the model, they lack in adaptability, which counts as a key feature in contexts such as games, that are versatile by definition. As a consequence, data-driven approaches are preferred. However, even these methods have their downsides, such as the generalization issues, algorithmic efficiency constraints and the ability to consider player variation - e.g., skills development - in time (Hooshyar, Yousefi, and Lim 2018).

Understanding whether players are living a pleasant experience while playing a game can be of use in multiple contexts. In player churn prediction analysis, gameplay data is exploited to assess whether (and when) a player will leave the game, as long as to identify the behaviors that can be symptoms of an abandonment (El-Nasr, Drachen, and Canossa 2016). Many studies have been conducted to inspect whether churn is indeed predictable and which are the features that should be studied to help to prevent it (e.g., (Periáñez et al. 2016; Hadiji et al. 2014; Melhart et al. 2019; Kawale, Pal, and Srivastava 2009; Mahlmann et al. 2010)). In terms of results, the latest work presents very well-performing models (Melhart et al. 2019). However, in most cases, the features used by researchers are context-specific game metrics (Mahlmann et al. 2010) and/or generalized metrics applicable in full-featured games (Hadiji et al. 2014). As a result, those solutions do not apply to contexts with a lower-level domain complexity, such as gameful apps. Besides, the vast majority of studies on churn prediction use aggregate data (Mahlmann et al. 2010; Hadiji et al. 2014), instead of analyzing temporal data. This is an aspect that should be carefully considered since it has been shown that players, rather than having a static in-game behavior, tend to change it throughout their gameplay (El-Nasr, Drachen, and Canossa 2016). Exception are the study of (Hadiji et al. 2014) and if (Demediuk et al. 2018). (Hadiji et al. 2014) defined a generic model that considers temporal features, while (Demediuk et al. 2018) used survival analysis to predict churn by considering information related to matches, such as match length and time between matches. However, the focus is on F2P and MOBA games which, together with entertainment games, have been the sole application domains of the studies on churn prediction. It derives that concepts like matches are often not found in gamified apps and, thus, other employable features should be investigated, which is one of the main motivations that have guided our study.

The Use Case Scenario

Play&Go is a mobile gameful system whose aim is to incentivize citizens to assume sustainable mobility behavior. The game is active in the city of *Trento* and *Rovereto* (*Italy*) and, with the sponsorship of the local municipalities, a 6-month edition is conducted every year. In this work, we considered gameplay data retrieved from the 3rd and 4th (from September 2017 to March 2018, and from October 2018 to April 2019) editions.

The game allows players to track their sustainable movement within the region of *Trentino* by specifying the transportation mean employed (walk, bike, bus or train). Moreover, an automatic system validates the trip by inferring the veracity of what has been stated by analyzing the history of physical quantities in the duration of the journey - e.g. the speed - and by checking whether the journey is feasible - e.g. checking the bus timetable. For each validated trip, Green Leaves points are awarded based on the length of the journey and the level of sustainability of the mean used. Players are also given weekly challenges - tuned according to both their preferences and their skills -through a recommendation system. The goal is to keep players in a state of flow (Nakamura and Csikszentmihalyi 2014), by, on the one hand, motivating them to improve (or maintain) their performance, and, on the other hand, promoting (if needed) different and more sustainable transportation means. For instance, even though a player has already a high performance, if she tends to mostly use the bus, the system would try to motivate her to walk or use the bike. Each won challenge awards players additional Green Leaves points. Those points are used to compute players' leaderboards.

Goals and Methodology

In this work we primarily aim at answering the following research question:

> Can we predict player churn by inspecting players' participation level, measured through in-game behaviors?

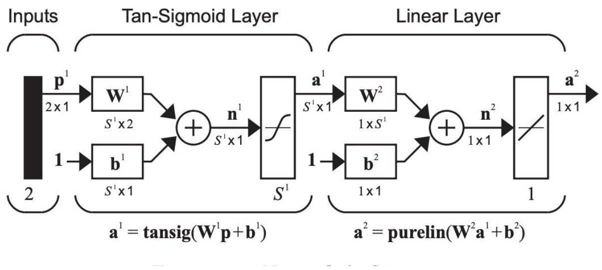

Figure 1: **Network structure of the universal approximator (Hagan et al. 2016).**

Then, we researched the amount of gameplay history needed to consider when predicting churn. We analyzed gameplay data gathered from a persuasive gameful system. We divided the whole gameplay in time-frames since we needed to discretize the gameplay period to observe players in well-defined and equidistant points of time. The rationale is that the analysis of in-game interactions is meaningful if conducted iteratively. Thus, the period considered should be long enough to let players exhibit their behaviors, without, on the other hand, risking flattening data by aggregating information gathered in a too wide period. This consideration derives from the idea that behaviors assumed by players, instead of being static over time, tend to dynamically change (El-Nasr, Drachen, and Canossa 2016). The most natural choice, in our application domain, was to set the length of the time-frames to 7 days since the concept of the week is highly tied to the game itself. The whole gameplay is already structured in weeks: we have weekly challenges and weekly leaderboards.

Our approach is based upon Loria and Marconi's (Loria and Marconi 2018) abstract behaviors, which are abstract concepts meant to be instantiated by associating to each one of them the appropriate gameplay items to monitor. More specifically, we instantiated the behaviors related to the evaluation of players' participation: `Committed` and `Active` behaviors. The `Committed` behavior is defined to give a measure of the regularity of the gameplay over time - e.g., does the player play every day? The *Active* behavior, on the other hand, inspects the intensity of the participation by, for example, measuring the amount of points obtained.
We instantiated the behaviors as follows. To monitor the `Committed` behavior we calculated the percentage of days in which the player was active over the week. A day is considered as active if at least a game action -i.e., tracking of a trip - has been performed. To monitor the `Active` behavior, which measures the intensity of the participation, we calculated the amount of Green Leaves Points gathered and the number of Actions performed during the week.

Both items measure the level of activity, one (Actions) regardless of the ulterior motive the app is fostering, while the other (Green Leaves points) is proportional to the level

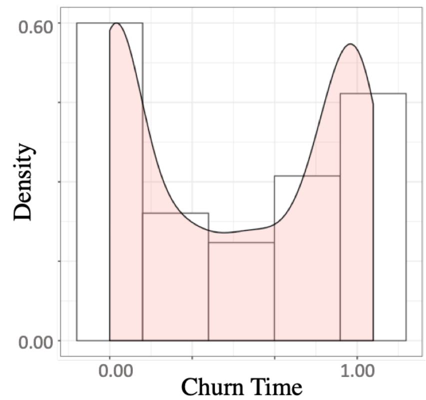

Figure 2: **Churn Time distribution over players.**

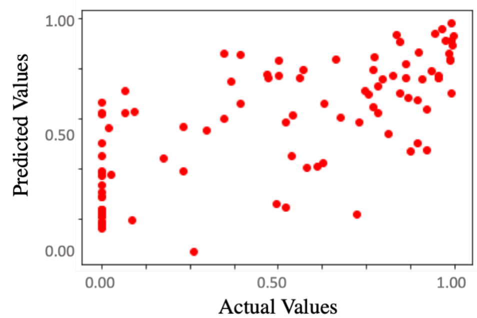

Figure 3: **Real values against predicted values, considering 3 weeks memory.**

of the sustainability of the means employed to track the trip and its length.

We analyzed 3 datasets, considering 2, 3, and 4 weeks of memory (including the current). To avoid introducing a bias by considering multiple observations per user, each player is represented by only one entry in the dataset. To this end, we picked a random point (i.e., time-frame) in the gameplay (between the logged weeks) to consider the current time-frame. From this point, we computed the behavioral features for every week in the history considered (2, 3 or 4 weeks). The features selected were the commitment, the activity indicators and the percentage of gameplay time elapsed. Each entry has been labelled with a real value between 0 and 1, indicating how soon the player will leave (`Churn Time`), calculated as the number of time-frames the player will remain in the game over the total number of time-frames (weeks) left before the gamification campaign, hence the game, ended. A value of `Churn Time` of 0 meant that the player never abandoned the game, rather it stayed for the whole gamification campaign. A value of 1 indicated that the player will abandon it in the following time-frame. Both input vectors and target vectors have been normalized in the range (-1,1), which is one of the standard methods used (Hagan et al. 2016). To predict the `Churn Time` value, we built an Artificial Neural Network (ANN). The design of the network has been guided by the structure

	Metric	Min	Mean	Max
Regression	R^2	0.10	0.37	0.53
	MSE	0.26	0.40	0.56
	MAE	0.40	0.51	0.60
Binary Classification	*Acc*	0.63	0.70	0.85
	Prec	0.58	0.72	0.87
	Rec	0.62	0.77	0.94

Table 1: **Outcomes of the repeated tests, considering 3 weeks memory.**

	Metric	Min	Mean	Max
2 weeks	R^2	0.08	0.27	0.31
	MSE	0.20	0.42	0.55
	MAE	0.45	0.55	0.59
4 weeks	R^2	0.14	0.31	0.56
	MSE	0.26	0.40	0.56
	MAE	0.40	0.60	0.60

Table 2: **Outcomes of the repeated tests, considering 2 weeks memory.**

depicted in Figure 1, which has been proven to be a universal approximator of continuous functions (Hagan et al. 2016). We adapted the ANN to our problem by defining a custom loss function, which is a linear combination of the mean absolute error (MAE) and the mean squared error (MSE). The combination of the two quantities is designed to consider them as equally significant. We opted for this loss function since MAE is more robust (less sensitive to outliers), while MSE gives higher importance to unexpected values, which in our case are the under-represented Churn Time values in the middle of the interval (see the distribution in Figure 2). This loss function represented a good compromise between the two.

Results and Discussion

To understand whether in-game behaviors are informative in the prediction of when and if a player churn occurred, we conducted a regression analysis using as features 3 items evaluating 2 in-game behaviors (Committed and Active), together with the current length of the player's gameplay up to that moment. We analyzed the same indicators, considering three different memory spans: 2,3 and 4 time-frames (i.e., weeks).

In the two editions of the game, we counted 595 players that kept playing for more than 2 weeks, 484 for more than 3 weeks and 418 for more than 4 weeks, which are respectively the size of the datasets employed for the predictions with 2, 3 and 4 weeks of memory.

We found that indeed inspecting in-game behaviors provides insights on the players' churn. More into detail, we obtained the best results when we considered 3 weeks memory (current week included), with R^2 = .37, MSE = .40, and MAE = .51 (Figure 1). Either when we increased and decreased

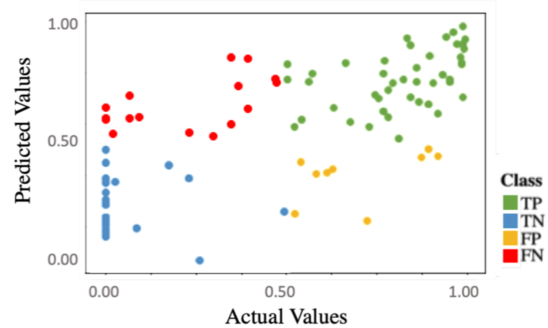

Figure 4: **Binary Classification of Churn, considering 3 weeks memory.**

memory, the outcomes worsened (R^2 = .27, MSE = .42, and MAE = 0.55 considering 2 weeks memory; and R^2 = .31, MSE = .40, and MAE = 0.60 considering 4 weeks memory), as shown in Figure 2. Figure 3 shows as the prediction is more accurate at the extremes, which is justified by the distribution of the value of the Churn Time (Figure 2) in the dataset. In our game either players immediately abandoned the game, or they remained for the whole campaign, resulting in very few in-between values. Consequently, learning to predict the churn time of players that abandoned the game halfway through the gameplay was harder. At confirmation of that, we brought the perdition to a more coarse-grained level, by moving towards binary classification. Churn is labeled as TRUE when the Churn Time ≥ 0.5. The classifier built worked generally well (accuracy = .76, and precision = 0.77, recall = .75). Figure 4 shows the predicted values against the real values of the test set (as in Figure 3) colored by the class of belonging. It is even more evident that the successfully predicted values are the ones at the extremes. Despite having multiple observations of every player, we selected only one entry for the user, to prevent introducing dependencies among the data. The point of time considered for every player was selected randomly and, to reduce the possibility to have good (or bad) results due to chance, we repeated the process 100 times. Every time we picked a point in time randomly. Table1 shows the best, the worse and the average values obtained for each metric, both for the regression and the classification analysis, considering the dataset with 3 weeks memory. Table 2 shows the statistics of the results for 2 and 4 weeks of memory in the regression analysis. The main limitation of this study lied in the size of the dataset and in the uneven distribution of the Churn Time values. In fact, despite having more than 450 players active in a real gamification campaign is already a challenging goal to achieve, these numbers are just enough to conduct a prediction analysis. Therefore, the outcomes should be seen as a motivation to further explore this approach, since such easy and generalizable in-game behaviors can be very insightful. Thus, in spite of having a limited variety of data, in comparison to traditional games, churn can also be predicted in gameful systems. These preliminary results are a small step towards novel gameful design, more user-centered and aimed at fostering long-term engagement.

Conclusion

This preliminary study leads the way in addressing the problem of detecting churn in gameful systems where complex indicators are unavailable, either because not retrievable or not implemented. We suggested using general-purpose in-game behaviors, which the designers can easily extract and analyze from the gameplay data. Although in this study we conducted our analysis offline, this approach is immediately applicable to an online setting.

Despite having obtained a first piece of evidence that monitoring participation behaviors hold insightful information, the paths that can be walked from this point onward are many. Shortly, we plan to improve our prediction models by increasing the size of our datasets with gameplay data from the next edition of the game and by comparing the new model with others - e.g., (Demediuk et al. 2018). Moreover, we will migrate from an offline to an online approach, by integrating churn prediction as a metric to continuously monitor players' experience to identify players at risk of abandonment and actuate a contingency strategy to keep them in the game. The goal is to reduce the players' drop rate.

References

Caillois, R. 2001. *Man, play, and games*. University of Illinois Press.

Coussement, K., and Van den Poel, D. 2008. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert systems with applications* 34(1):313–327.

Datta, P.; Masand, B.; Mani, D.; and Li, B. 2000. Automated cellular modeling and prediction on a large scale. *Artificial Intelligence Review* 14(6):485–502.

Demediuk, S.; Murrin, A.; Bulger, D.; Hitchens, M.; Drachen, A.; Raffe, W. L.; and Tamassia, M. 2018. Player retention in league of legends: a study using survival analysis. In *Proceedings of the Australasian Computer Science Week Multiconference*.

El-Nasr, M. S.; Drachen, A.; and Canossa, A. 2016. *Game Analytics*. Springer.

Hadden, J.; Tiwari, A.; Roy, R.; and Ruta, D. 2006. Churn prediction using complaints data. In *World academy of science, engineering and technology*.

Hadiji, F.; Sifa, R.; Drachen, A.; Thurau, C.; Kersting, K.; and Bauckhage, C. 2014. Predicting player churn in the wild. In *CIG'14*, 1–8. IEEE.

Hagan, M.; Demuth, H.; Beale, M.; and De Jesus, O. 2016. Neural network design 2nd. *Ed., Lexington, KY*.

Hooshyar, D.; Yousefi, M.; and Lim, H. 2018. Data-driven approaches to game player modeling: a systematic literature review. *CSUR* 50(6):90.

Huotari, K., and Hamari, J. 2012. Defining gamification: a service marketing perspective. In *MindTrek 16*, 17–22. ACM.

Huotari, K., and Hamari, J. 2017. A definition for gamification: anchoring gamification in the service marketing literature. *Electronic Markets* 27(1):21–31.

Isbister, K., and Schaffer, N. 2008. *Game usability: Advancing the player experience*.

Kawale, J.; Pal, A.; and Srivastava, J. 2009. Churn prediction in mmorpgs: A social influence based approach. In *CSE'09*, volume 4, 423–428. IEEE.

Kiili, K.; Moeller, K.; and Ninaus, M. 2018. Evaluating the effectiveness of a game-based rational number training-in-game metrics as learning indicators. *Computers & Education* 120:13–28.

Kim, J. H.; Gunn, D. V.; Schuh, E.; Phillips, B.; Pagulayan, R. J.; and Wixon, D. 2008. Tracking real-time user experience (true): a comprehensive instrumentation solution for complex systems. In *CHI'08*, 443–452. ACM.

Loh, C. S.; Sheng, Y.; and Ifenthaler, D. 2015. Serious games analytics: Theoretical framework. In *Serious games analytics*. Springer. 3–29.

Loria, E., and Marconi, A. 2018. Player types and player behaviors: Analyzing correlations in an on-the-field gamified system. In *CHIPlay'18 Extended Abstracts*, 531–538. ACM.

Mahlmann, T.; Drachen, A.; Togelius, J.; Canossa, A.; and Yannakakis, G. N. 2010. Predicting player behavior in tomb raider: Underworld. In *CIG'10*, 178–185. IEEE.

Marczewski, A. 2015. *Even Ninja Monkeys Like to Play*. Blurb.

Melhart, D.; Azadvar, A.; Canossa, A.; Liapis, A.; and Yannakakis, G. N. 2019. Your gameplay says it all: Modelling motivation in tom clancy's the division. *arXiv preprint arXiv:1902.00040*.

Mellon, L. 2009. Applying metrics driven development to mmo costs and risks. *Versant Corporation*.

Nakamura, J., and Csikszentmihalyi, M. 2014. The concept of flow. In *Flow and the foundations of positive psychology*. Springer. 239–263.

Orji, R.; Mandryk, R. L.; and Vassileva, J. 2017. Improving the efficacy of games for change using personalization models. *TOCHI* 24(5):32.

Periáñez, Á.; Saas, A.; Guitart, A.; and Magne, C. 2016. Churn prediction in mobile social games: Towards a complete assessment using survival ensembles. In *DSAA'16*, 564–573. IEEE.

Sangin, M. 2018. *Games user research*. Oxford University Press.

Warmelink, H.; Koivisto, J.; Mayer, I.; Vesa, M.; and Hamari, J. 2018. Gamification of the work floor: A literature review of gamifying production and logistics operations.

Xi, N., and Hamari, J. 2019. Does gamification satisfy needs? a study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management* 46:210–221.

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

Yannakakis, G. N., and Togelius, J. 2018. *Artificial intelligence and games*, volume 2. Springer.

Yee, N. 2016. The gamer motivation profile: What we learned from 250,000 gamers. In *CHIPlay'16*, 2–2. ACM.