Extracting Gamers' Cognitive Psychological Features and Improving Performance of Churn Prediction from Mobile Games

JiHoon Jeon
Sejong University
Seoul, South Korea
rhd00513@gmail.com

DuMim Yoon
Sejong University
Seoul, South Korea
krad@hanmair.com

SeongIl Yang
ETRI
Daejeon, South Korea
siyang@etri.re.kr

KyungJoong Kim*
Sejong University
Seoul, South Korea
kimkj@sejong.ac.kr

Abstract—**With the continued growth of the mobile game market, many game companies aim to make money through mobile games. In this situation, knowing the tendency of gamers and predicting the churn in advance can maximize profit through effective game services. For this reason, much study has been conducted for the purpose of gamer analysis and churn prediction. However, the study was mainly conducted using surveys, bio-signals, and PC online game logs, which are likely to make detailed information. In this study, we extracted seven cognitive psychological features from the game logs of Crazy Dragon, a commercial mobile RPG game, and used these to predict the churn. In addition, we analyzed the effect of purchasing feature by comparing the churn prediction performance according to presence or absence of purchase feature. In the conclusion, we obtained higher performance when predicting the churn using cognitive psychological features than using the basic raw logs. Also, we obtained high churn prediction performance using only cognitive psychological features without purchase feature.**

Keywords—cognitive psychological feature, churn prediction, purchase feature, mobile game log

I. Introduction

Currently, the game market is steadily moving toward mobile games since the advent of smartphones. Numerous companies are entering mobile games, and thousands of mobile games are released every year. In this context, many game companies want to know how gamers feel about their services. In other words, they want to know how much fun gamers have for their services and how much engagement they will have. This is because the more gamers are satisfied with the service and the more engaged in the game, the more profitable the company will be. Also, many game companies want to predict churn. This allows us to actively respond to the churn.

Many researchers have been studying the topic of gamer analytic and churn prediction. Game Analytics Company analyzed the Chinese gamers using analytical tools and services for gamers' behavior on various games including mobile games [1]. Another game analyst firm Quantic Foundry Company had presented 12 empirical models based on a survey of more than 220,000 gamers [2][3]. According to these results, gamers mainly have three tendencies, which are Openness to experience, Conscientiousness and Extraversion [4]. In addition, Borbora and Srivastava proposed a method to classify game players for churn prediction using log data of massively multiplayer online role-playing game(MMORPG) [5]. Nozhnin conducted a study to predict the churn by character level using MMORPG log data [6].

But previously studies have some limitation. researchers used surveys to obtain opinions from gamers [7][8][9] and analyzed the gamer's behavior using log data [10][11][12]. Also, they analyzed gamers by measuring various bio-signals using wearable equipment [13]. However, the analysis using survey and bio-signals is relatively more accurate than the analysis using log data, but it is difficult to obtain a large amount of data, and it is not realistic to actually apply it to a commercial game. Therefore, most of the studies for practical applications extract the behavior of the gamer in log data and use it to analyze the gamer or predict the behavior. In addition, most studies have mainly focused on PC online games. PC online games are relatively freer in-game activities than mobile games, so it is possible to log the various behaviors including the interaction and social activities among the gamers. But, the behavior of gamers is relatively limited in mobile games, making it difficult to leave detailed log data due to limitations of mobile device or network.

In this study, we extracted the cognitive psychological features of gamers by using log data of the mobile game, which is relatively limited information. Also, we predicted the churn of gamers using cognitive psychological features and compared the importance of features. In addition, we analyzed the effect of purchasing feature by comparing the churn prediction performance according to presence or absence of purchase feature.

II. Gamer Modeling Based on Cognitive Psychology

Cognitive psychological features are hidden in the various behaviors of gamer. Behaviors of gamer are represented by factors such as play style, strategy and behavior pattern, and actual behaviors are recorded as log data. The relationship between them is shown in "Fig. 1". As such, we must deduce the cognitive psychological features from observable behaviors of gamer or game log data. Therefore, we need to be able to record the behaviors of gamer as closely as possible. However, it is possible to lose information for various reasons in the process of recording behaviors of gamer. From cognitive psychological features to gamer behavior and game logs, the data are changed step by step. Even if it starts with a simple cognitive psychological feature, it can be manifested in various forms as it is embodied. In the opposite direction, specific information is made up of more abstract information. In the abstraction process, specific information disappears, but if the hidden cognitive psychological features underlying the gamer's behavior is discovered, it is likely to help predict the future behavior of the gamer. From this point of view, using cognitive psychological features to analyze gamers has three major advantages.

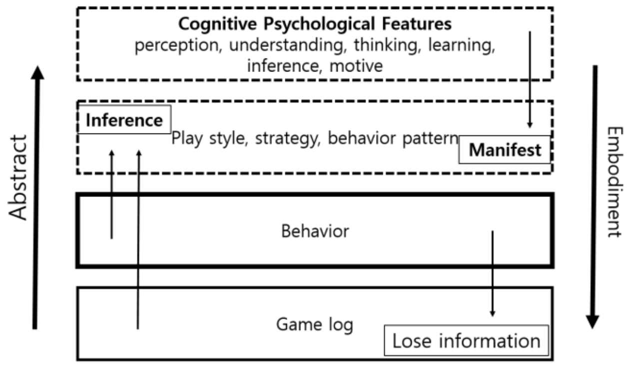

Fig. 1. Relationship between cognitive psychological features and game log

First, there is a high possibility of designing a simple and high-performance model. Since most of the actual observable behaviors are expressed from cognitive psychological features, cognitive psychological features can be seen as latent variables. Latent variables represent actual but unobservable factors, but they also contain abstract concepts that do not actually exist, such as motives or mental states. One of the biggest advantages of using latent variables is the reduced dimensionality of data. It is also possible to combine observable sub-symbolic data and change it into easy-to-understand symbol data.

Second, it is likely to find game-independent features. In most of the previous studies, much of the data used game-dependent features. There are game-independent features such as login times and access times, but most of the behavioral data is about specific interaction with the game environment, so it's hard to apply it directly to data from other games. However, because cognitive psychological features are revealed by abstract features such as individual player's personality, tendency, intention, and information processing process, it is more likely to find common factors to many games Since collecting large amounts of game log data in one game is a very difficult task, making a generic model that can describe data from a variety of games is not only academic but also very useful in industry.

Finally, using cognitive psychological features, it is possible to model not only predicting churn/purchase but also higher order decision making process. The main reason for the mainstream of the current study is predicting the churn/purchase because both of these studies are practical. However, the ultimate goal of the study is not to predict churn/purchase, but to maximize the interest and fun of gamers. Simple prediction of the churn can be too late to prevent gamers from leaving, but if the model can be made of what the gamers are experiencing in the game and how they feel, and if the response can be predicted, gamer behavior can be responded to more quickly and actively.

III. EXTRACT FEATURES

In this study, we conducted experiments based on the log of Crazy Dragon, a commercial mobile RPG game. we used the log data collected from March 21, 2016, to July 29, 2016, for a total of 130 days, when the Crazy Dragon was served and recorded the gamer's behavior using 96 kinds of raw logs. A total of 295,107,602 logs were collected, and 192,361 gamers played Crazy Dragon.

A. Extract raw features

We extracted 45 features that are assumed to be related to the churn from 96 kinds of raw logs as "TABLE I"

TABLE I. 45 FEATURES THAT ARE ASSUMED TO BE RELATED TO THE CHURN

45 kinds of features	
Login log	9 kinds of buy/sell logs
5 kinds of character logs	2 kinds of mail logs
8 kinds of reinforcement/compose logs	4 kinds of payment logs
3 kinds of mission logs	2 kinds of achievement logs
4 kinds of battle logs	5 other logs
2 kinds of PVP logs	

B. Extract Cognitive psychological features

We used only 96 kinds of raw log data to extract cognitive psychological features of gamers. We extracted three characteristics that are 'engagement', 'activity', 'personality' from the raw log that considered to be related to the psychology of the gamer and the activity information of the gamers difficult to measure in the mobile game. The 'engagement' is a factor that indicates how much the gamer is engaged in the game. We extracted a feature that is the number of actions per minute (APM) of the gamer as engagement's factor. The 'activity' is a factor that indicates the active level of gamers, we extracted two features that are the daily average play time (MPD) and the log-in rate (LR) during the entire game play period as activity 's factors. The 'personality' is a factor that indicates the personality of the gamer, we extracted four features that are daily average completing sub-mission (MCPD), the single game ratio (INT), the competitive game ratio (COM), and main character investment ratio (AF) as personality's factors. In addition, we included the gamer's unique number (ID) for user identification and the number of purchases (PAY) related to in-app payment. Two of the nine features used for gamer analysis (MCPD, AF) are dependent features and the other seven features are independent features applicable to other games (TABLE II).

TABLE II. EXTRACT GAMER'S COGNITIVE PSYCHOLOGICAL

Factor	Feature	Description	Method
unique number	ID	gamer's unique number	-
engage ment	APM	the number of actions per minute	Total action count / total gameplay time(minute)
activity	MPD	daily average play time (minute)	Total play time(minute) / total game play day
	LR	log-in rate	total gameplay day / (last gameplay day – first gameplay day)
personal ity	MCPD	daily average completing sub-mission (Conscientiousness)	Total complete sub-mission count / total gameplay day
	INT	single game ratio (Extraversion)	Total single play count / total fight count (PK, guild war, single, etc..)
	COM	competitive game ratio (Extraversion)	Total multi-play count / total fight count (PK, guild war, single, etc..)
	AF	main character investment ratio (Openness)	Total main character log / total character log
payment	PAY	the number of purchases	Total purchase in-app payment count

IV. EXPERIMENTS AND RESULTS

A. Data set

To define the churn in this experiment, the gamer was assumed to be a churn if there was no recorded log between the 7th day and the 14th day after the first login. The data model for gamers is shown in "Fig. 2".

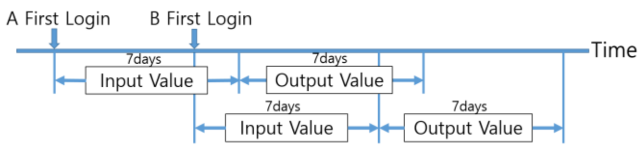

Fig. 2. Data model

We extracted 45 raw features recorded for a week from the player's first login as an input value. The output value was determined by whether the log for the gamer has appeared from 7 days to 13 days after the first login. We excluded new members who had joined during the last two weeks (July 14 ~ July 29) because we could not decide whether or not they were leaving. Eventually, 176,544 of the total 192,361 gamers were actually used to predict churn. Also, 153,465 out of 176,544 gamers went out, and only 23,079 gamers did not leave. With 86.93% of total gamers leaving, it is possible to get a high accuracy of 86.93% even if it is assumed that all gamers will leave. In order to reduce this error and evaluate the actual prediction performance more precisely, we solved the imbalance of data by using under-sampling and various evaluation indexes such as accuracy, precision, recall, f1-score, and area under the curve (AUC).

B. Churn prediction base on raw log

We used decision tree and random forest (an option of the number of trees is 15.) model with 45 extracted features to predict the churn. The results are shown in "TABLE III". The results are the mean and standard deviation of 10-fold cross-validation for each model.

TABLE III. 10-FOLD CROSS-VALIDATION RESULT FOR LEARNING MODEL

	Accuracy	Precision	Recall	F1-score	AUC
Decision tree	90.4 % ±0.04	65.4 % ±0.12	72.1 % ±0.10	67.5 % ±0.10	82.7 % ±0.06
Decision tree + under-sampling	84.7 % ±0.09	85.8 % ±0.43	83.0 % ±0.22	84.4 % ±0.10	84.9 % ±0.21
Random forest	93.7 % ±0.02	84.7 % ±0.17	67.8 % ±0.16	74.5 % ±0.09	95.9 % ±0.02
Random forest + under-sampling	89.0 % ±0.11	89.3 % ±0.13	88.7 % ±0.14	89.0 % ±0.11	95.9 % ±0.04

For each model, comparing the results of imbalanced and under-sampled data shows that the accuracy is somewhat lower, but the results of precision, recall, F1-score, and AUC are improved. It can be said to be a performance improvement from the viewpoint of reliable churn prediction for under-sampling models. As a result, if we have sufficient imbalance data, it can be expected that learning from the balance data extracted through sampling rather than learning the whole data can improve the prediction performance.

C. Churn prediction base cognitive psychological features

We predicted the churn using seven previously-extracted cognitive psychological features. Furthermore, in order to analyze the effect of the PAY feature, which is assumed to be highly related to the churn, we compared the performance of churn prediction and the importance of each feature according to the presence or absence of the PAY feature. The combination of these features is shown in "TABLE IV". Data was balanced through under-sampling and random forest model was used. For reliability of the experimental results, we performed 10-fold cross-validation 10 times and then calculated the mean and standard deviation.

TABLE IV. THREE FEATURE SETS

Feature Set	Count	Description
RAW	45	45 raw logs
With PAY	8	APM, MPD, LR, MCPD, INT, COM, AP, PAY
Without PAY	7	APM, MPD, LR, MCPD, INT, COM, AP

The experimental results are shown in "TABLE V", and the churn prediction using cognitive psychological features has improved about 5% more than raw data alone. Also, there is almost no difference in churn prediction performance depending on the presence or absence of the PAY feature. This shows that using only cognitive psychological features can predict the churn with high performance.

TABLE V. CHURN PREDICTION PERFORMANCE OF THREE FEATURE

	RAW	With PAY	Without PAY
Accuracy	89.01% ±0.11	94.08% ±0.10	94.02% ±0.11
Precision	89.25% ±0.13	92.24% ±0.15	92.30% ±0.14
Recall	88.70% ±0.14	96.26% ±0.05	96.05% ±0.09
F1-score	89.00% ±0.11	94.20% ±0.09	94.14% ±0.10
AUC	95.88% ±0.04	98.34% ±0.05	98.32% ±0.04

In addition, we measured the importance of the features used in the churn prediction and ranked it as "TABLE VI". According to this ranking, the complete mission count was the most important feature in the raw features and the login ratio (LR) was the most important feature in the churn prediction for features including cognitive psychological features. Also, as a result of measuring the importance of feature set including PAY feature, it is considered that the top three cognitive psychological features are more helpful for churn prediction than PAY feature because PAY feature does not exist in 3rd place.

TABLE VI. TOP 3 FEATURES OF EACH FEATURES SET

Feature Set	1st	2st	3st
RAW	Complete mission count	Obtain mercenary count(mission)	login count
With PAY	LR	APM	AF
Without PAY	LR	APM	AF

V. CONCLUSION & FURTHER WORK

We extracted seven cognitive psychological features of the gamer from the game log and experimented to predict the churn. The results of this experiment show that using the cognitive psychological features of the gamer is more helpful in predicting the churn than the basic raw. In addition, comparing the performance of churn prediction with the presence of absence of the PAY feature, the performance of the churn prediction was high even if only cognitive psychological features were used without the PAY feature. From this study, game companies will be able to extract cognitive psychological features using limited mobile log data and predict the churn more accurately. Furthermore, by analyzing the cognitive psychology of the churn, it is possible to cope with the churn more effectively.

However, there are still some issues that need additional study. The first issue is a prediction of gamers' various behaviors. Churn prediction is one of the basically considered ways of predicting gamer's behavior, but it is not the ultimate goal. We will be able to predict various behaviors such as purchase prediction, event participation prediction. Second, in this study, we extracted the cognitive psychological features of the gamers on the gamer's total game play period, However, since the gamer's propensity may be changed during the game play period, it is possible to extract it by considering the period. Finally, we use the cognitive psychological features of the gamer to estimate the psychology of the gamer and demonstrate the experiment through the gamer's survey. To do this, it is necessary to make a comparative analysis through the psychological analysis of the gamer and the questionnaire of the gamer.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2017. *: corresponding author

REFERENCES

[1] Gameanalytics, "http://www.gameanalytics.com/," gameanalytics, 20 1 2016. http://www.gameanalytics.com/blog/chinese-gamers-in-game-behaviour-of-players-from-china.html.

[2] THIS IS GAME, "www.thisisgame.com," THIS IS GAME, 29 3 2016. http://m.post.naver.com/viewer/postView.nhn?volumeNo=3904593&memberNo=24985926.

[3] N. Yee, "http://quanticfoundry.com/," 7 4 2016. http://quanticfoundry.com/2016/04/07/gdc-talk/.

[4] O. P. john, L. P. Naumann and C. J. Soto, "Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues.," in Handbook of personality: Theory and research, New York, NY: Guilford Press, 2008, pp. 114-158.

[5] Z. H. Borbora and a. J. Srivastava, "User Behavior Modelling Approach for Churn Prediction in Online Games," in 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, 2012.

[6] D. Nozhnin, "Predicting Churn: Data-Mining Your Game," 17 5 2012. http://www.gamasutra.com/view/feature/170472/predicting_churn_data mining_your_.php.

[7] R. M. Ryan, C. S. Rigby and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach," Motivation and Emotion, vol. 30, no. 4, pp. 344-360, 12 2006.

[8] A. K. Przybylski, C. S. Rigby and R. M. Ryan, "A motivational model of video game engagement," Review of General Psychology, vol. 14, no. 2, pp. 154-166, 2010.

[9] M. V. Birk, C. Atkins, J. T. Bowey and R. L. Mandryk, "Fostering Intrinsic Motivation Through Avatar Identification in Digital Games," the 2016 CHI Conference on Human Factors in Computing Systems, pp. 2982-2995, 2016.

[10] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting and C. Bauckhage, "Predicting player churn in the wild," 2014 IEEE Conference on Computational Intelligence and Games, pp. 1-8, 2014.

[11] A. Drachen, C. Thurau, R. Sifa and C. Bauckhage, "A comparison of methods for player clustering via behavioral telemetry," in arXiv, 2014.

[12] A. Drachen, N. Ross, J. Runge and a. R. Sifa, "Stylized Facts for Mobile Game Analytics.," in 2016 IEEE Conference on Computational Intelligence and Games (CIG), santorini, 2016.

[13] A. Clerico, C. Chamberland, M. Parent, P.-E. Michon, S. Trembla, T. H. Falk, J.-C. Gagnon and P. Jackson, "Biometrics and Classifier Fusion to Predict the Fun-Factor in Video Gaming," in 2016 IEEE Conference on Computational Intelligence and Games(CIG), Santorini, 2016.