Profit Optimizing Churn Prediction for Long-Term Loyal Customers in Online Games

Eunjo Lee, Boram Kim, Sungwook Kang, Byungsoo Kang, Yoonjae Jang, and Huy Kang Kim, *Member, IEEE*

***Abstract*—To successfully operate online games, gaming companies are introducing the systematic customer relationship management model. Particularly, churn analysis is one of the most important issues, because preventing a customer from churning is often more cost-efficient than acquiring a new customer. Churn prediction models should, thus, consider maximizing not only accuracy but also the expected profit derived from the churn prevention. We, thus, propose a churn prediction method for optimizing profit consisting of two main steps: first, selecting prediction target, second, tuning threshold of the model. In online games, the distribution of a user's customer lifetime value is very biased that a few users contribute to most of the sales, and most of the churners are no-paying users. Consequently, it is cost-effective to focus on churn prediction to loyal customers who have sufficient benefits. Furthermore, it is more profitable to adjust the threshold of the prediction model so that the expected profit is maximized rather than maximizing the accuracy. We applied the proposed method to real-world online game service, Aion, one of the most popular online games in South Korea, and then show that our method has more cost-effectiveness than the prediction model for total users when the campaign cost and the conversion rate are considered.**

***Index Terms*—Churn prediction, cost-benefit analysis, customer lifetime value, data mining, game analytics.**

I. Introduction

ONLINE games have become one of the most successful online services with cumulative user base reaching around 40% of the global online population [1] and a global online gaming market valued at around $20 billion [2]. To successfully operate an online game, it is crucial not only to create a fun game but also to provide proper user management. Therefore, gaming companies are introducing various customer relationship management techniques to monitor and analyze the behavior of game users [3].

In general, preventing a customer from churning is more cost-efficient than acquiring a new customer [4]. For high retention rates, companies implement various strategies, such as launching promotional campaign incentives to remain loyal. Considering the cost associated with such incentivizing policy, correctly identifying and targeting potential churners has become essential as the incorrect selection will lead to unnecessary expenses (Type 1 error) as well as insufficient outcomes (Type 2 error). For a successful churn prevention campaign, the accurate identification of potential churners has become increasingly crucial, and with continuously advancing data mining techniques, many researchers have studied various classification methods [5].

Churn prediction's main goal is pursuing high accuracy to find possible churners, but a churn prediction model should consider maximizing the profit derived from churn prevention as well. There are various studies for a churn prediction technique and tools that account for the associated costs and derived earnings [4], [6]–[8]. While these studies propose general frameworks or methods, several issues should be considered upon applying them to an online game.

First, churn should be appropriately defined unlike traditional subscription based services, such as telecommunications or banking services, online game services do not require formal withdrawal procedures, such as deleting accounts. In fact, many online game users stop playing the game while maintaining their accounts, as there is no cost associated with maintaining accounts. In the case of Aion, extensively analyzed throughout this paper, only 0.8% of users who have not resumed playing over a year deleted their accounts. Many users also resume playing the game after a long-term hiatus—about 50% of the Aion users resume playing after a one-month break. Consequently, to minimize the cost from false-positive and undetected cases, correctly defining churn is crucial.

Second, a churn prediction target should be appropriately selected. Because the ultimate goal of churn prediction is to prevent loss derived from churning, users who add profits when retained should be the primary target of churn prediction models. Considering additional marketing and operating costs to switch churners, it is not cost effective to target all users in predicting churn. Also, increasing the survival rate of users who access a game temporarily due to curiosity may be a valuable strategy to pursue a newly released game title, but for a mature game title, it is more useful to focus on preventing the churn of loyal users. Furthermore, some users affect a game service negatively, that is, gold farming group (GFG) using game bots [9] or cyber bullies who disturb others online [10]. If these users are not filtered out correctly, a churn prevention strategy can

Manuscript received August 29, 2017; revised January 18, 2018 and June 23, 2018; accepted September 13, 2018. Date of publication October 8, 2018; date of current version March 17, 2020. This research is partially supported by NCSOFT. *(Corresponding author: Eunjo Lee.)*

E. Lee, B. Kim, S. Kang, B. Kang, and Y. Jang are with the Data Analysis and Modeling Team, NCSOFT, Seongnam 13494, South Korea (e-mail: gimmesilver@ncsoft.com; sweetieally@gmail.com; swkang@ncsoft.com; bsoo414@gmail.com; yoonjaej@gmail.com).

H. K. Kim is with the Graduate School of Information Security, Korea University, Seoul 02841, South Korea (e-mail: cenda@korea.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TG.2018.2871215

have an adverse effect to the game service. Consequently, it is necessary to select an appropriate target before churn prediction modeling.

Third, it is necessary to quantitatively estimate the profit generated when a potential user is retained. Accurate expected benefit estimates are essential information for determining churn prevention strategies since the cost of them may exceed the expected profit.

Therefore, we propose a churn prediction process and a detailed methodology considering the factors mentioned above. The characteristics of the proposed method are as follows.

1) We extract "long-term loyal customers" and analyze them for churn prediction. To do this, users are assigned to loyalty grades by using their in-game activities and payment patterns, and the sequence of their loyalty levels are analyzed for approximately six months.
2) We estimate the expected profit of preventing user churn using cost-benefit analysis. Then, the churn prediction model is optimized for maximizing the profit.

We have applied the proposed method to the live data of Aion, one of the most popular massively multiplayer online role-playing games (MMORPGs) in South Korea that has been serviced for nine years from 2009.

The key contributions of this paper are as follows.

1) We propose a churn prediction framework considering the expected profit in online games. The framework includes predictive modeling as well as the related prerequisite works, such as churn definition and selection of prediction targets and performs a cost-benefit analysis of the prediction model. To the best of our knowledge, this is the first study that has considered overall procedures concerning expected profit in the online game field.
2) The cost-benefit analysis techniques in other fields have been modified for the application of the online game domain. The modified method was used to optimize the prediction model via simulations, using real data and comparing it with the results by only considering prediction accuracy. We expect our result to serve as an excellent benchmarking reference for other churn prediction models regarding online games.
3) We analyze the game activities for one and a half years of approximately half a million users and find that the social events in the game are highly related to user churn. Additionally, trends and volatility of in-game activities can also be an essential factor for churn prediction. Our results can constitute a reference for churn analysis in other online games.

The remainder of this paper is organized as follows. In Section II, we review related studies. Section III presents the dataset and criteria for churn analysis. Section IV outlines the proposed methodology for churn prediction. Section V describes the characteristics of churners via exploratory data analysis. In Section VI, we evaluate the performance of the prediction model and show the result of profit optimization. Section VII discusses limitations of our study and issues that need to be improved via future works. Finally, we summarize our findings and conclude with Section VIII.

II. Related Literature

Churn analysis using data mining is a subject actively studied in various fields [11].

In the mobile communications industry, [12] and [13] are pioneering papers regarding churn analysis. Particularly, Mozer *et al.* analyzed churn from the viewpoint of service satisfaction, credit risk, and service provider profitability of subscribers. On the other hand, Baumann *et al.* [14] proposed a prediction method, focusing on profit growth through churn prevention from an actual marketing viewpoint, and estimates the per user profit that can be obtained through experiments using five public communication data. Moreover, Dasgupta *et al.* [15] proposed a technique for predicting churners through the call network analysis of mobile phone users. This network analysis technique was also used in several online games [16]–[18].

Glady *et al.* [19] and Mavri and Ioannou [20] targeted churn analysis in the financial industry. Mavri and Ioannou investigated customer churn using Greek bank data and analyzes the determinants of the churn rate increase by using the hazard proportional model and a survival analysis methods. Alternatively, Glady *et al.* proposed a method for churn prediction by defining customer loyalty via customer lifetime value and analyzing customer behavior, which is expected to decrease future profits. It is similar to [13] and [14], in terms of studying customers' churn from the viewpoint of revenue obtained from them.

Churn analysis has also been actively conducted in the field of online games. Runge *et al.* [21] studied the technique of churn prediction of users who have a high profit in casual games, and Hadiji *et al.* [22] proposed prediction framework on free-to-play game. Milošević *et al.* [23] showed an impressive result in predicting early churners and preventing their departure using A/B testing. Tamassia *et al.* [24] proposed a hidden Markov model to utilize the features of time series data. Periáñez *et al.* [25] and Viljanen *et al.* [26] proposed applying survival analysis methods for churn prediction. Borbora *et al.* [27] proposed a method that clusters game users based on motivation theory and applies a model suitable for each group. Although clustering is similar to our study in that it divides the type of users, we analyze the churn only for the groups with the highest expected profit, whereas Borbora *et al.* proposed the model for each type. We think that their study will be a good reference when proceeding the churn prediction for more diverse types in the future.

Furthermore, there are various studies on the effect of social relations on user churn. Kawale *et al.* [28] studied the social relationship between users who perform a quest together, and it analyzes the effect of one user's churn on another. Park *et al.* [29] revealed that in-game achievements are primary factors for users with characters in a growth period to continue playing the game whereas communications and social interactions among users are important for users who already completed growth period. Moreover, Shores *et al.* [10] analyzed the effect of malicious users bullying other users by using user reputation data for League of Legends. Our study also shows that the factors related to social activities have a significant influence on the user churn, which supports the results of the above-mentioned studies.

Feng *et al.* [30] and Debeauvais *et al.* [31] showed that users have a change in play time or frequency just before leaving via study of popular MMORPGs. According to our study, Aion also shows that trend and volatility of gaming activity are highly correlated with churn.

Finally, there are studies of cost-benefit analysis for the churn prediction model [4], [6]–[8]. Theses describe how to measure the performance and compensate the limit of commonly used receiver operating characteristic (ROC) curves when measuring the performance of classification model—such as churn prediction—from the viewpoint of expected profit and cost. Specifically, Hand proposed H-measure in terms of minimizing costs, and Verbraken *et al.* proposed an indicator designed from the perspective of maximizing profits by expressing an expected profit that takes the operating cost and conversion rate of a campaign execution into account. Bahnsen *et al.* [7] and Verbraken *et al.* [4] refer to [6], and proposed a method for predictive modeling aiming to maximize profit rather than accuracy. Our formula used for profit evaluation is created by referring to [6] and [7]. Additionally, while the previous studies mainly focus on the prediction modeling stage, we propose an overall procedure consisting of churn definition, predicting target selection, prediction modeling, and cost-benefit analysis.

III. Dataset

We use the dataset of the MMORPG Aion, which launched in November 2008. Specifically, we used the game activity logs of one million users from November 2014 to July 2017.

A. Time Unit for Data Aggregation

Many Aion users have a weekly cycle of play pattern. While there are users who access the game daily, most users access it only on weekends. Moreover, Aion has a few hours of downtime every Wednesday morning for game updates or server maintenance. Consequently, all statistics are based on data weekly aggregated from a Wednesday morning to the subsequent Wednesday before downtime.

B. Churn Definition

For online games, unlike telecommunications or financial services, it is inappropriate to use withdrawal of membership as a definition of churn. In Aion, only 0.8% of users who have not accessed the game for more than a year have explicitly unsubscribed. If we define membership withdrawal as churn, most users who have effectively churned will be classified as nonchurners. Therefore, we should identify a user as a churner using inactivity period. If we set the period to define a churner to be too long, the expected benefits of churn prevention will be reduced because a user who is actually churning will not be identified until it is too late. On the other hand, if the period is too short, most users who return after a short break will be misclassified as churners, and unnecessary costs will be incurred.

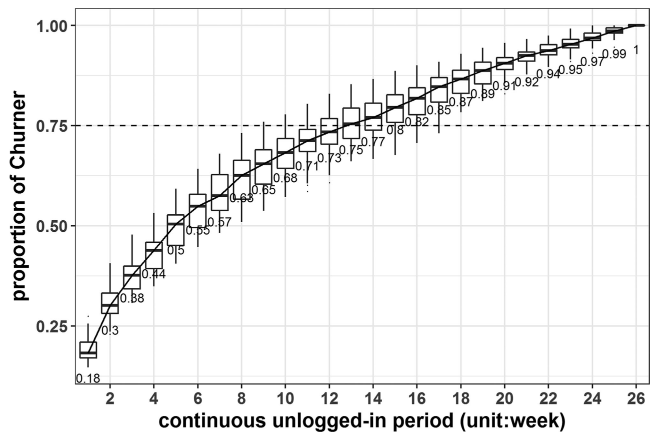

Fig. 1. Variations of P_n (churn) according to n.

Consequently, for optimizing profit via churn prediction, it is necessary to set the appropriate period to define a churner.

Aion usually goes through a significant contents update every six months. Therefore, the developing team of Aion defines 26 weeks as the maximum possible dormant period for a user because a user who does not respond to the major update should be regarded as having lost interest in the game entirely. However, it is necessary to define a shorter period because a half-year is too long when considering the expected profit related to user churn.

To find an appropriate period, we calculated $P_n(\text{churn})$, the fraction of users who have not accessed the game for 26 weeks among users who have not accessed the game for at least n weeks

$$P_n(\text{churn}) = \frac{F(26)}{F(n)}, 1 \leq n \leq 26 \tag{1}$$

where $F(n)$ represents the number of users who have not logged in for at least n weeks.

To avoid bias associated with specific dates, we used the median of $P_n(\text{churn})$ measured using multiple 26-week-sequences from November 5, 2014, to December 23, 2015 (see Fig. 1). Although the "elbow method" is commonly used to determine an appropriate parameter, the outcome as shown in Fig. 1 did not present any outstanding point. Consequently, we use $P_n(\text{churn}) = 0.75$ as an "elbow" and define a churner as a user who does not access the game for more than 13 weeks. In other words, we compromise between certainty and profit; the compromise is that among the players we define as churners, 25% are not going to churn according to Aion definition.

IV. Methodology

This section describes the detailed methodology of the proposed churn prediction process. The overall process consists of the following points:

1) selection of prediction targets;
2) feature engineering;
3) expected profit evaluation.

TABLE I
FEATURES FOR IN-GAME ACTIVITIES

Name	Feature
DICE	The count of rolling dice
DT	The number of days a user plays
EXP	The amount of experience a user obtains
FORTRESS	The number of combats in a fortress a user participates
AP	The achievement points a user obtains
GP	The glory points a user obtains
GLIDE	The number of glide activities a user play
DUNGEON	The number of dungeons a user plays
PLAYTIME	Play time (unit: hour)
PVP	The number of combats against another player
PVE	The number of combats against NPC
HARVEST	The number of materials a user obtains in the field
QUEST	The number of quests a user completes
TELEPORT	The number of teleports a user casts
SCROLL	The count of scrolls which a user uses
GETMONEY	The amount of game money a user obtains from other users (unit: millions)
GIVEMONEY	The amount of game money a user gives to other users (unit: millions)
PARTY	The number of parties a user joins
PARTYMEMBER	The number of users who are in the same party

A. Prediction Target Selection

To identify long-term loyal customers who are expected to generate profit if retained, we first assign loyalty grades based on the user's weekly gaming activities and payment amounts, then track the grade change for thirty weeks to select a user group that maintains a high grade for a sufficiently long period. The details are as follows.

1) Loyalty Grade Assignment: First, user's in-game activities described in Table I are aggregated weekly and clustered into nine clusters using the k-means clustering algorithm. We then analyzed the characteristics of each clusters using the average value of the in-game activity features and extracted the following additional indicators to understand the characteristics of the clusters better.

ARPU: Average revenue per user. ARPU is expressed as a relative value so that the average ARPU of all users is 1 instead of the actual one to mask the actual ARPU.

BOT: The ratio of users detected as game bots via the detection system [9] among users in each clusters.

IP: The number of other users accessed to IP address that a user accesses the most frequently.

Table II summarizes the characteristics of each cluster type. Clusters 1–5 show positive correlations between in-game activities and ARPU. On the other hand, clusters 6 and 7 have very low ARPU despite the extremely long PLAYTIME. The activities related to the production of goods, such as PVE, HARVEST, and GETMONEY are very high, while activities related to the consumption of goods or interaction between users, such as PVP, DICE, GLIDE, SCROLL, and GIVEMONEY have relatively low values. Besides, these types have a very high ratio of game bots compared to other types. Consequently, we defined these types as game bots.

Cluster 9 has a low level of overall activities, but it has a large number of users accessing the game using the same IP address. Although this type has a low ratio of detected game bots, we identified these users to be malicious accounts of "GFGs" that operate numerous game bots or cheap laborers to obtain virtual assets and monetize. According to our previous study [32], some GFGs manage multiple accounts with little playtime to avoid bot detection. Finally, cluster 8 is an outlier type with a deficient proportion of actual users and extreme values for several game activities.

After that, nine clusters are reassigned into six grades in descending order based on their profitabilities to simplify user types—grade 1 represents those with the highest profitability. Clusters 1–5 are assigned to grades 1–4 ordered by the number of in-game activities and payment amount. Game bots (Clusters 6, 7, and 9) are integrated as grade 5 because they have negative impacts on other game users. Cluster 8 is assigned as grade 6—the lowest grade—to be excluded from further analysis. Table III shows the result of reassigning for loyalty grades.

2) Long-Term Loyal Customer Extraction: In addition to the six grades, a user who did not access the game for a week or just recently subscribed was given a grade 9 (dormant state) and grade 0 (unsubscribed state) for that respective period. With the additional two grades as well as the six grades mentioned above, all users have a 30-weeks-long data sequence consisting of eight possible weekly grades.

Sequence clustering is then performed to extract "long-term loyal customers." We use the optimal matching algorithm [33] to measure the distance between the sequences of users and categorize them using a hierarchical clustering algorithm.

The optimal matching algorithm is a technique to obtain dissimilarity between two sequences. It calculates the minimum number of insertion, deletion, or replacement required to generate another sequence from one sequence. The minimum value is then defined as the dissimilarity between two sequences.

The resulting dissimilarities of all sequence pairs were then clustered into five groups using hierarchical clustering using the following algorithm.

1) At first, each element is constructed a separate set.
2) For all pairs, the pairs with the highest similarity are combined and form a bottom layer.
3) The dissimilarity is measured for the pairs of sets in the bottom layer generated in the previous step. The dissimilarity between the sets A and B—$d(A, B)$—is measured as follows:

$$d(A, B) = \max\{d(x, y) : x \in A, y \in B\}. \quad (2)$$

4) Repeat steps 2) and 3) until all elements are agglomerated into a whole set.

The time complexity of the optimal matching and hierarchical clustering are $O(n^2)$ and $O(n^3)$, respectively. Because both algorithms require a large amount of computation, instead of using data of all users, we extract the data of 285 136 users who have been classified as grades 1 and 2 at least once for thirty weeks, and then sampled 7540 of them. In the clustering result, a group corresponding to the "long-term loyal customer" (red box in Fig. 2) is extracted, and then the sequence pattern is analyzed to obtain a simple filtering rule. After that, prediction targets are extracted from all users using the filtering rule. We

TABLE II
SUMMARY OF THE USER CLUSTERING RESULT

Cluster No.	1	2	3	4	5	6	7	8	9
Rate (%)	0.97	4.08	8.54	6.33	20.93	15.51	8.09	0.04	35.52
PLAYTIME	79.82	66.90	37.33	15.57	2.94	152.06	138.24	56.42	1.73
PVP	442.98	261.27	88.70	8.70	3.54	0.88	0.71	4.32	0.02
PVE	2,515.14	2,278.43	1,204.67	187.28	94.57	31,215.15	7,598.14	141.02	14.75
HARVEST	8.95	16.87	17.68	34.50	2.84	0	557.12	10.999.47	0.76
SCROLL	1,942.62	1,412.52	623.37	77.59	38.87	47.80	14.81	41.40	0.54
GETMONEY	113.48	93.69	42.51	5.71	4.29	171.50	320.09	60.88	1.70
GIVEMONEY	392.31	284.38	110.71	46.59	19.93	59.38	74.91	154.41	7.73
DICE	667.53	464.80	180.04	18.85	10.06	0.80	2.04	13.43	0.09
GLIDE	538.50	415.40	195.98	27.13	15.78	2.12	5.33	14.65	0.19
SCROLL	1,942.62	1,412.52	623.37	77.59	38.87	47.80	14.81	41.40	0.54
ARPU	7.93	6.00	4.31	2.50	0.44	0.09	0.31	0.94	0.06
BOT (%)	0.52	1.48	1.20	0.10	0.00	15.74	60.12	0.00	0.05
IP	2.36	2.60	2.35	1.94	2.66	17.97	18.56	10.35	641.57

TABLE III
CUSTOMER LOYALTY GRADE

Grade	Cluster No.	Description
1	1	Users who had many in-game activities and spent significant amounts
2	2 and 3	Users who had some in-game activities and spent moderate amounts
3	4	Users who had some in-game activities and spent small amounts
4	5	Users who had a few in-game activities and spent almost no money
5	6, 7 and 9	Users who are suspected game bots
6	8	Outliers

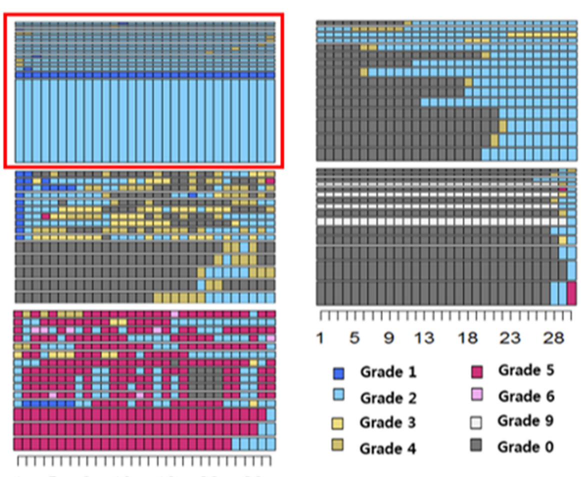

Fig. 2. Categorizing customers via sequence clustering. Users of red box (left top) are identified as long-term loyal customers because they are given a grade 1 or 2 for most periods.

note that this is just a trick we use to circumvent the constraints on the system resource, but it is not a required step in the process we propose.

Algorithm 1 is the filtering rule. $C(n, g)$ is the function of counting how often a user obtains grade g for n weeks. mode_grade(n) is the function that returns the grade a user obtains most frequently for n weeks, and max_grade(n) is the function that returns the highest grade a user obtains for n weeks.

TABLE IV
FEATURES OF THE SOCIAL RELATION INDICATOR

No.	Feature
1	# of friends who played the game over the last 4 weeks
2	# of friends who were in the same party over the last 4 weeks
3	# of legion members who played the game over the last 4 weeks
4	# of legion members who were in the same party over the last 4 weeks
5	# of times a user moved legion over the last 12 weeks

Algorithm 1: Prediction Target Selection – Only Choose Players that Have Never been Grade 5 or 6, and have been Most Often One of Grade 1 to 3, and Have Played for at Least 10 Weeks.

if $C(13,1)+C(13,2)+C(13,3)+C(13,4) \geq 10$ AND $mode\_grade(13) \leq 3$ AND $max\_grade(13) \leq 4$ **then**
  $extract\ the\ user$
else
  $filter\ out\ the\ user$
end if

B. *Feature Engineering*

Features used in the churn prediction model are as follows.

1) Social Relationship—Friend and Legion: Aion provides various social activities in the virtual world. Game users can make friends and join a guild called "legion," which consists of dozens to hundreds of members that cooperate and compete against other legions. A negative social experience, such as a friend quitting a game, reduction or disbandment of the legion due to legion members retiring affects a user's churn rate [28]. Table IV shows the features accounting of social ties among friends and legion members.

2) Social Relationship—Party: Users can also temporarily organize a party of two to six members to complete a quest or explore a dungeon. Given most activities in Aion require collaboration among party members, party activities also influence user churn [28]. We use social network indicators to extract

TABLE V
PARTY NETWORK INDICATORS

Feature	Description
Clustering coefficient	The likelihood that three nodes will become fully connected forming a clique
Graph density	$\frac{2E}{N(N-1)}$, (E, N): the number of edges and nodes
Community size	The number of nodes in a community

party-related features. The feature engineering process for the party network is as follows.

1) Party network is constructed. Nodes represent users and edges represent party relationship. An edge is generated only when the party is performed more than or equal to twice between the same users over a week to prevent the excessive network construction
2) Party network is divided into multiple communities so that the edges between the nodes in the same community is denser than the edges between nodes in other communities. The more detailed process is explained in the next paragraph.
3) The indicators described in Table V are calculated for each community network.

The algorithm proposed in [34] is used to divide communities on the party network. To do this, an indicator called "modularity" is calculated as follows:

$$Q = \sum_i (e_{ii} - a_i^2) \tag{3}$$

where e_{ii} is the fraction of edges between nodes in ith community to all edges in the network and a_i is the fraction of edges connecting to a node in ith community to all edges in the network.

The modularity Q is larger as each community includes the more edges and the edges between communities is the fewer. The communities are extracted as follows.

1) At first, each node is constructed a community.
2) The modularity Q is calculated as per (3).
3) For each community pair ΔQ, the difference in Q when two communities are combined with the Q of the previous step is calculated.
4) Combine the pairs where ΔQ is the maximum.
5) Repeat steps 2)–4) until ΔQ is no more positive.

3) Trend and Deviation for In-Game Activities: Changes in play time or frequency are highly correlated with a user's churn [31]. Notably, long-term loyal customers tend to access and play the game steadily and consistently. Therefore, we assume that these users gradually lose interest in the game rather than suddenly leaving the game, so, the number of their activities or the number of connections in the game is likely to decrease as well gradually. Various statistics related to the amount of in-game activity for each user are collected on a weekly basis for the 12 weeks to measure trends and volatilities. The specific in-game actions are described in Table I.

The methods for measuring trend and deviation for the statistics are described as follows.

1) Moving average convergence divergence (MACD), which is calculated by the difference between the moving average for 4 weeks and 12 weeks. If this value is a positive number, it indicates an increasing trend and a negative number indicates a decreasing trend.
2) Coefficient of variation, which is the variance divided by the average statistics for 12 weeks. A high-value means that a user's in-game activity is irregular.

All features are scaled via the standardization as follows:

$$x_i' = \frac{x_i - \mu}{\sigma} \tag{4}$$

where μ and σ is a mean and a standard deviation for each feature, respectively.

C. Profit Evaluation

When the retention campaign is applied, the expected profit per user can be described as Fig. 3.

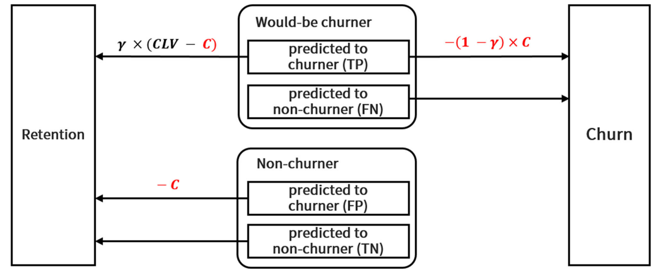

Fig. 3. Expected profit and cost upon application of the churn prediction model.

In Fig. 3, CLV is the expected customer lifetime value that the company gains if a user continues to play the game, and C is the campaign cost. TP and TN denote the proportion of users who are accurately classified as churners and nonchurners among all users, and FP and FN represent the proportion of users misclassified as churners or nonchurners among all users.

Campaign cost (C) represents the price of discount coupons or reward items that are provided to change the mind of potential churners and retain them. If a game company gives reward items to users who are classified as churners via the prediction model, the total cost of such campaign is proportional to the number of targeted users. Consequently, a campaign cost is expressed as TP and FP multiplied by C. γ is the rate of users who are retained via the retention campaign among potential churners. If a churner is correctly identified, the expected profit per detected churner would be γ of CLV (i.e., $\gamma \times$ CLV).

In the classification model, the rate of true positives or false positives depends on the threshold used to determine whether a user has churned. Consequently, TP and FP can all be expressed as a function of threshold t, and then the expected profit can be shown as follows:

$$\text{Profit}(t) = \text{CLV}\{\gamma \text{TP}(t)\} - C\{\text{TP}(t) + \text{FP}(t)\}. \tag{5}$$

The CLV can be calculated as the product of the user's service life expectancy and the average revenue generated by the user per time unit. Since it is difficult to estimate the service life expectancy of a user accurately, it is generally calculated using

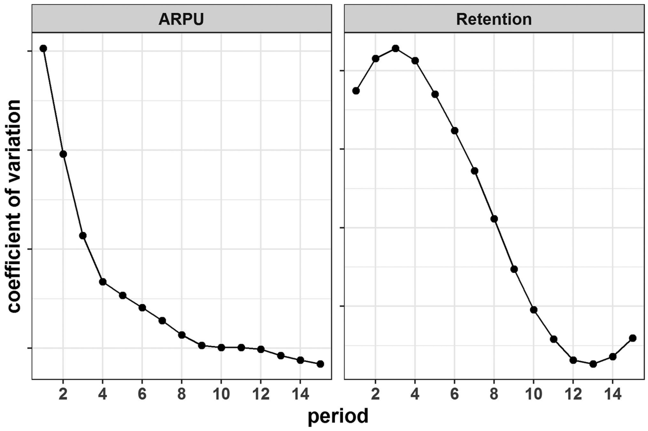

Fig. 4. Change of variance for ARPU_n and r_n by n. The variance of ARPU decreases as n increases. The variance of the retention rate, on the other hand, is the least where n is 13.

the infinite series of a retention rate as follows [35]:

$$\mathrm{CLV} = \sum_{i=0}^{\infty} (\mathrm{ARPU} \times r^i) = \frac{\mathrm{ARPU}}{1-r}. \tag{6}$$

In (6), ARPU stands for the average revenue per user, and r for the retention rate. The original equation includes the interest rate and operating cost per time unit, but these two parameters are omitted for simplicity in this research because for online games, the interest rate and operating cost are far less than revenue generated from a user.

The above-mentioned equation assumes that the retention rate and ARPU per time unit are constant. However, weekly ARPU and retention rate vary depending on the time of aggregation. Therefore, we used the following method to determine the ARPU and the retention rate with change over time.

1) The ARPU and the retention rate for n weeks are calculated. n ranges from 1 to 15 with one increments. These values are called ARPU_n and r_n, respectively.
2) The variances of ARPU_n and r_n for the various time points are calculated. These are called $\mathrm{var}(\mathrm{ARPU}_n)$ and $\mathrm{var}(r_n)$, respectively.
3) Determine ARPU_n and r_n at the point where $\mathrm{var}(\mathrm{ARPU}_n)$ and $\mathrm{var}(r_n)$ are smallest.

Fig. 4 shows the variance of ARPU_n and r_n for long-term loyal customers. The $\mathrm{var}(\mathrm{ARPU}_n)$ continues to decline as the aggregate period increases, but after the point where n is 9, there is an insignificant difference. The variance of the retention rate, on the other hand, is in the form of a cubic line with the least variance point where n is 13. Based on these data, we choose 13 weeks as a unit period for ARPU and r to calculate CLV.

Furthermore, we calculate the individual CLV of each user instead of just calculating the average CLV for the entire user base to estimate the expected profit precisely when the predicted model is applied. For this, we use the payment amount for the last 13 weeks for each user instead of ARPU. After dividing the whole user base into long-term customer group and the rest, the retention rate for each group is used because the retention rate is not available for individual measurement.

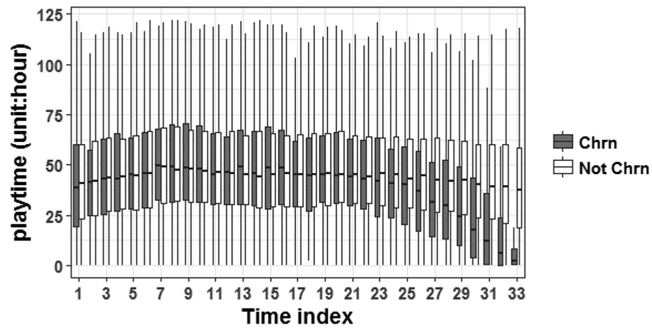

Fig. 5. Comparison of play trend between churners and nonchurners. The playtime of churners gradually decreases from about ten weeks before churning.

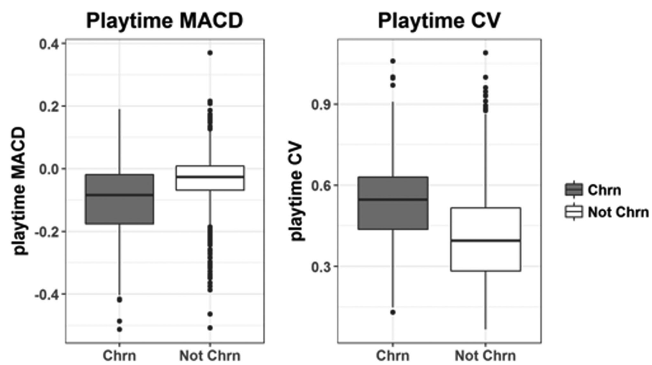

Fig. 6. Comparison of playtime trend and variation between churners and nonchurners. (left) Playtime of churners decreases while nonchurners' playtime remains stable. (right) The variance of the playtime of a churner is relatively larger than a nonchurner's one.

We use the function in (5) to measure the expected profit of a churn prediction model and to find the threshold that maximizes the expected profit. The detailed process is described in Section VI.

V. Exploratory Data Analysis

This section introduces some parts of the exploratory analysis to compare the differences between churners and nonchurners. The features described in Section IV-B were selected through the exploratory analysis described in this section.

A. Trends and Variations in Gaming Activities

According to our analysis of weekly playtime of users, churners showed gradually decrease in playtime from about ten weeks before churning (see Fig. 5), and playtime for churners became irregular as churning point approached while it remained constant for nonchurners; hence, higher coefficient of variation of playtime for churners. The MACD for playtime also shows a clear difference between the churners and nonchurners, and such quantitative difference in playtime between churners and nonchurners are depicted in Fig. 6. Consequently, the trends and variations of users' in-game activities are measured by the various methods described in Section IV-B3 and are used as features of the prediction model.

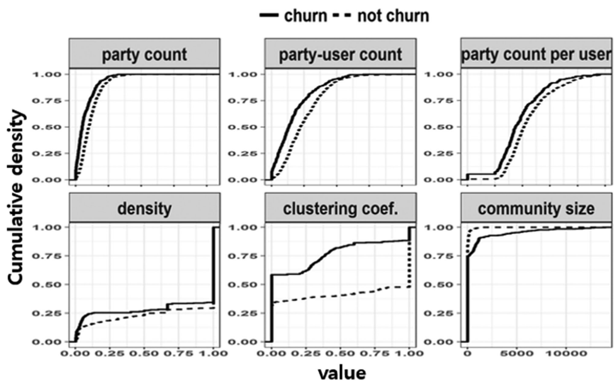

Fig. 7. Comparison of party network indicators between churners and nonchurners.

B. Party Activity

As mentioned in Section IV-B2, party activities are critical social activities that users need to play the game. Many of the contents provided by the game require collaboration with other users through party activities.

We constructed a social network for cases in which users formed a party with the same members for at least two times and extracted party communities using the method described in Section IV-B2. We then calculated network indicators, such as community density, clustering coefficient, and community size (see Fig. 7).

The clustering coefficient of communities for churners appears to be much lower than that of nonchurners, while the size of the party community for churners is larger than nonchurners. This characteristic is related to the sociality of churners. The clustering coefficient is the rate at which any three nodes within the community form a clique and are completely connected to each other. Therefore, the low clustering coefficient of communities for churners means that party members who churners formed a party with more than once are not mutual party members. On the other hand, party members of nonchurners tend to be shared party members since nonchurners repeatedly perform party activities with members from previous parties. Consequently, we can assume that the nonchurners mainly perform party activities with friends while churners tend to perform party activities with random users.

For example, in the case of Aion, a party usually consists of six players. Therefore, the size of the closed community that only performs parties with the same members will not exceed 6, and the clustering coefficient is 1. On the other hand, in the case of the community that consists of members who tend to form party no more than once with the same member, the community size is a maximum of 6^n, and the clustering coefficient is closed to 0 (see Fig. 8).

Previous studies have shown that social networks, such as Twitter and Facebook, have higher clustering coefficients than random networks [36]. This study supports our assumptions.

C. Legion

Legions also affect the users' experience. Members of good legions will be more attached to the game due to the additional benefits and sense of belonging they attain from their legions. The exact opposite is the case for members of small or inactive legions as they feel isolated instead of feeling attached, hence more likely to churn.

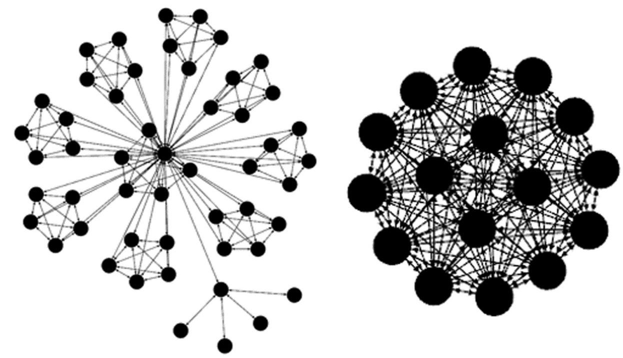

Fig. 8. Party community samples for churner (left) and nonchurner (right).

TABLE VI
RATES OF LEGION JOINING FOR CHURNERS AND NONCHURNERS

	Legion	No legion
Churner	33%	67%
Non-churner	42%	58%

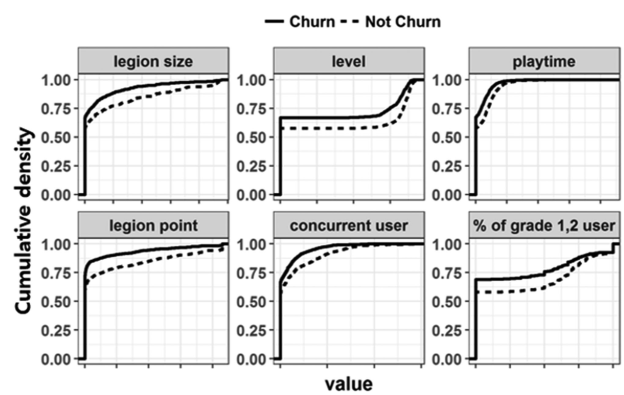

Fig. 9. Comparison of legion activity between churners and nonchurners.

Table VI shows that churners tend not to join a legion compared to nonchurners. Moreover, churners' legions have lower activity values—members' play time, loyalty to the legion, and legion points compared to nonchurners' legions (see Fig. 9).

Another interesting finding is that churners tend to drift from a legion to another more frequently than nonchurners (see Fig. 10). Such frequent transfer of legions indirectly indicates a low sense of belonging, and previous studies have shown that game users tend to leave legions if they have little benefit or minimal communication with legion members [18]. The desire for such personal accomplishments and interaction with other users is a crucial factor of player retention [29].

VI. Experiments

A. Total Churner Prediction Versus Loyal Churner Prediction

We aggregated users who played the game at least once from March to June 2017 for the experiments. We then tagged a long-term loyal customer and a churner who left a game within four

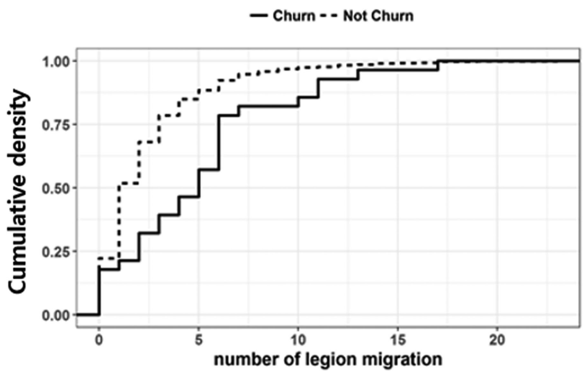

Fig. 10. Comparison of legion migration counts between churners and nonchurners. Churners tend to change legion more often than nonchurners.

TABLE VII
COMPARISON OF THE NUMBER OF USERS, THE PERCENTAGE OF ESCAPE USERS, AND THE CUSTOMER LIFETIME VALUE FOR LOYAL CUSTOMERS AND THE REST OF CUSTOMERS

	Long-term loyal	Others	Total
# of users	15,797	637,527	653,324
# of churners	1,801	480,664	482,465
ratio of churners	11.40%	75.40%	73.85%
CLV (total)	4,750,029	637,527	5,387,556
CLV (per user)	300.69	1	8.25

weeks. CLV was calculated separately for long-term loyal customers and the rest. All features, calculations, and tags were generated using the process described in Section IV. For confidentiality, CLV has been normalized so that the value for the rest set to one.

Table VII shows that the long-term loyal customers are 2.4% of the total users, and only 11% of them leave a game within four weeks. However, CLV per user of long-term loyal customers is about 300 times higher than that of nonloyal customers. This means that preventing one loyal customer from churning will yield an expected profit comparable to that of preventing 300 nonloyal customers from churning. Consequently, the total CLV for long-term loyal churners is larger than the value for the rest, while the number of them is only 0.37% of total churners.

We constructed two training sets. One is the training set extracted from total users, and another is obtained from only long-term loyal customers. Both training sets were sampled as the ratio of churners to nonchurners becomes one-to-one. We then constructed test sets for both the entire users and the long-term loyal customers to compare the expected profit of the prediction models. In the test set I, the ratio of the long-term loyal customers and the ratio of the churners were adjusted to be the same as those in the Table VII to accurately estimate the expected profit. All datasets were randomly sampled from the dataset described in Table VII. The test set II consisted of only long-term loyal customers who extracted from the test set I. We then created three prediction models for the training set I and the training set II, respectively, using random forest, XG boost, and generalized boosting regression as training algorithms. Table VIII describes the size of training sets and test sets constructed for our experiments. Maximum expected profit is the value that can

TABLE VIII
TRAINING SET AND TEST SET

	Training		Test	
	Set I	Set II	Set I	Set II
# of non-churners	10,000	1,000	26,150	2,126
# of churners	10,000	1,000	73,850	274
# of total users	20,000	2,000	100,000	2,400
maximum expected profit			105,621	67,382

Set I was sampled from entire users and set II was sampled from only long-term loyal customers.

be obtained when all churners are switched from churners to nonchurners.

To calculate the expected profit, we should do a campaign that provides reward items to the predicted churners and then measure the conversion rate. Unfortunately, we were not able to apply the churn prevention service to the live game service. Consequently, we compared the differences between models by carefully selecting γ and C.

Table IX shows the comparison of the results of six prediction models. Columns of test set I are the result of prediction models using the test set I and columns of test set II are the result of prediction models using the test set II. RF, XGB, and GBM denote random forest, XG boost, and generalized boosting regression, respectively. The expected profits were calculated assuming γ and C are set to the values written in Table IX. Each value for γ and C was assigned as realistic as possible with the help of domain experts. C and the expected profit value were normalized in the same manner as CLV in Table VII.

Table IX shows that the prediction accuracy results of the test set II are lower than the accuracies of the test set I, whereas the expected benefits of test set II are significantly higher than the test set I, even though maximum expected profit of the test set I is higher than the test set II. The reason is that CLV for true positive users of the test set II intend to be much higher than those of the test set I (see Fig. 11). If a retention campaign is progressed using the results of the test set I, the game company will take a monetary loss except for the case in which the campaign cost is close to zero. When a churn prediction model was constructed for all users, any long-term loyal churners were not detected while only churners with very low CLV intended to be detected. The reason is that the long-term loyal churners are only about 0.3% of the total users, so a training algorithm treated them as an outlier in the training process. As a result, while the expected profit obtained from true positive users of the test set I is meager, the cost to be spent is enormous because the total number of churners is large. Consequently, the total expected profit is very small or even harmful. On the other hand, in the test set II, even though the number of true positive users is tiny, the CLV per target user of retention campaign is very high, and the total cost for them is low. Consequently, The total expected profit is very high compared to the test set I.

B. Threshold Optimization

As described in the previous section, it is crucial to detect as many as possible of the users who have high CLV. Therefore, it is advantageous to reduce the number of false negative

TABLE IX
COMPARISON OF THE PREDICTION PERFORMANCE AND EXPECTED PROFIT BETWEEN TEST SET I (FOR TOTAL CUSTOMERS) AND TEST SET II (FOR ONLY LONG-TERM LOYAL CUSTOMERS)

			Test set I			Test set II		
			RF	XGB	GBM	RF	XGB	GBM
# of true positive users			63,778	63,249	61,932	212	213	201
# of false positive users			3,917	4,307	4,713	568	615	586
# of true negative users			22,233	21,843	21,437	1,558	1,511	1,540
# of false negative users			10,072	10,601	11,918	62	61	73
accuracy			0.8601	0.8509	0.8337	0.7375	0.7183	0.7254
precision			0.9421	0.9362	0.9293	0.2718	0.2572	0.2554
recall			0.8636	0.8565	0.8386	0.7737	0.7774	0.7336
F1 score			0.9012	0.8946	0.8816	0.4023	0.3866	0.3789
AUC			0.9358	0.9264	0.9067	0.8296	0.8129	0.8159
expected profit	γ:0.1	C:0	704	503	488	5,192	5,169	4,908
		C:0.01	27	-173	-178	5,184	5,161	4,900
		C:0.1	-6,065	-6,253	-6,177	5,114	5,086	4,830
		C:1	-66,991	-67053	-66,157	4,412	4,341	4,121
	γ:0.05	C:0	352	251	244	2,596	2,585	2,454
		C:0.01	-325	-424	-422	2,588	2,576	2,446
		C:0.1	-6,417	-6,504	-6,421	2,518	2,502	2,375
		C:1	-67,343	-67,305	-66,401	1,816	1,757	1,667
	γ:0.01	C:0	70	50	49	519	517	491
		C:0.01	-607	-625	-618	511	509	483
		C:0.1	-6,699	-6,705	-6,616	441	434	412
		C:1	-67,625	-67,506	-66,596	-261	-311	-296

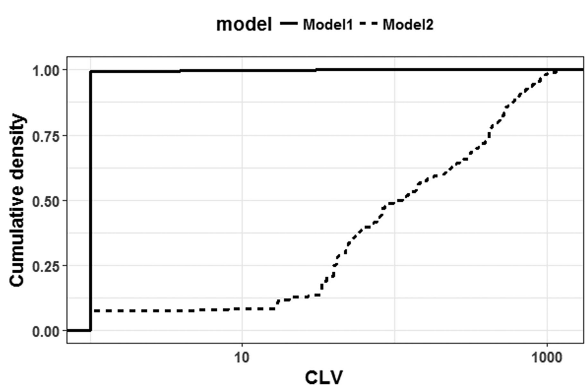

Fig. 11. Comparison of CLV per user between true positive users of the test set I and the test set II. CLV s for most true positive users of the test set II have then 10, while most CLV s for the test set I are close to 0.

users as many as possible even if the proportion of false positives is increased in the churn predictive model for long-term loyal customers with high expected profit. Consequently, it is a better decision to keep the threshold for classifying churners lower. Table X shows how much the predicted profit increases when the threshold is adjusted to maximize the expected profit using the three models of the test set II used in the previous experiment. In Table X, the base profit is the predicted profit when the threshold set to 0.5 in the model. This value is equal to the expected profit in the test set II in Table IX. The adjusted threshold is the threshold adjusted to maximize expected profit. When the campaign cost is zero, the adjusted threshold is zero because there is no cost to the user, so it is the best strategy to campaign for all users. In most cases, the adjusted threshold is lower than 0.5. The optimized profit is the predicted profit when the threshold is adjusted. The increasing rate is a variable by the gap of the optimized profit and the base profit. When we optimized the threshold, we could check that the profit increased by approximately 10%–30%.

VII. DISCUSSIONS

In this section, we introduce a limitation of our method and an idea that we can further improve the profit optimization model in the future works.

A. *Limitation of a Binary Classifier*

The prediction model for long-term loyal customers had low precision in our experiments. To analyze the reason, we compared the characteristics of false positive users and true negative users.

In Fig. 12, left box plots show the changes of the playtime before and after predicted churn point. False positive users tend to decrease their play time later than before the predicted time. On the other hand, there is no significant difference in play time before and after the expected churn time in the group of nonchurners.

In Fig. 12 (right), around 26% of false positive users churned within the subsequent four weeks, while only 8% of true negative users churned. Additionally, approximately 50% of false positive users decreased in their loyalty grade after the time of the expected churning, while loyalty grades of only 20% of true negative users decreased. These findings indicate that misclassified users are likely to leave the game soon, or at least to lose

TABLE X
COMPARISON OF EXPECTED PROFIT BETWEEN THE THRESHOLD OPTIMIZING MODEL AND THE BASE MODEL (THRESHOLD SET TO 0.5) IN TEST SET II

γ		0.1				0.05				0.01			
C		0	0.01	0.1	1	0	0.01	0.1	1	0	0.01	0.1	1
RF	base profit	5,192	5,184	5,114	4,412	2,596	2,588	2,518	1,816	519	511	441	-261
	adjusted threshold	0	0.11	0.11	0.25	0	0.11	0.16	0.5	0	0.11	0.25	0.89
	optimized profit	6,738	6,717	6,523	4,989	3,369	3,348	3,163	1,816	674	652	499	9
	increasing rate	0.2978	0.2956	0.2754	0.1309	0.2978	0.2935	0.2563	0	0.2983	0.2764	0.1314	-
XGB	base profit	5,169	5,161	5,086	4,341	2,585	2,576	2,502	1,757	517	509	434	-311
	adjusted threshold	0	0.01	0.03	0.29	0	0.01	0.03	0.41	0	0.03	0.29	0.94
	optimized profit	6,710	6,689	6,504	4,903	3,355	3,334	3,151	1,858	671	650	490	40
	increasing rate	0.2981	0.2961	0.2787	0.1295	0.2978	0.2943	0.2594	0.0575	0.2978	0.2777	0.1298	-
GBM	base profit	4,908	4,900	4,830	4,121	2,454	2,446	2,375	1,667	491	483	412	-296
	adjusted threshold	0	0.26	0.26	0.32	0	0.26	0.29	0.39	0	0.26	0.32	0.73
	optimized profit	6,738	6,716	6,517	4,957	3,369	3,347	3,155	1,860	674	652	496	0
	increasing rate	0.3729	0.3706	0.3494	0.2029	0.3729	0.3684	0.3285	0.1160	0.3723	0.3494	0.2032	-

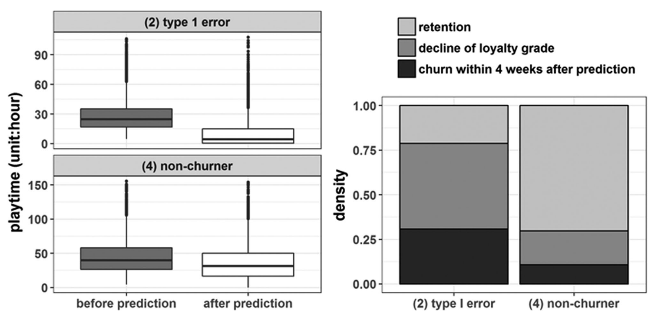

Fig. 12. (left) Comparison of change in playtime before and after the expected churn time. False positive users show a significant decrease in playtime after the time. (right) Comparison of change in user status after prediction point.

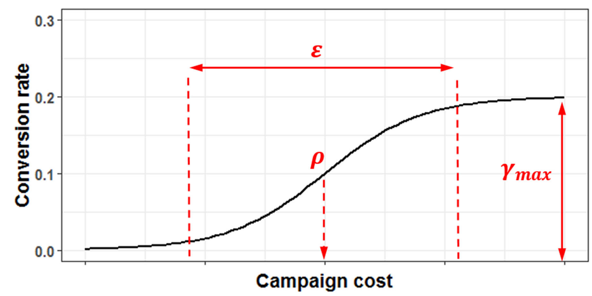

Fig. 13. Estimation of conversion rates by cost.

their game loyalty, even if they do not leave precisely within four weeks after prediction point.

This result shows that a binary classifier is not a smart approach for churn prediction. To overcome the limitation, a multiclass classifier, which can predict several categories according to churning point, could be better. Considering flexibility, it will be a good choice to apply a regression model that predicts user's service life expectancy instead of a classifier to predict user's churn. In this case, since there is a right censoring issue that cannot be checked when nonchurners leave after the observation point, life expectancy cannot be labeled accurately for a training set. Survival analysis should be used to address this problem. Good references of applying survival analysis to churn prediction for online game are [25] and [26]. We also have the plan to apply survival analysis to our method in future works.

B. Cost Optimization

In Section VI, we treated campaign costs and conversion rates as mutual independent variables. However, the conversion rate γ is positively correlated with the incentive (i.e., campaign cost) for the user who would be a churner. In other words, the bigger the incentive is given to a user planning to leave, the more likely the user is to be retained. Consequently, conversion rate can be expressed as a function of the cost C. The function can be a logistic function considering the following assumptions.

1) As the cost of campaign increases, the conversion rate of users will increase.
2) At a certain point, no matter how much the cost is, the conversion rate will follow the law of diminishing marginal utility.
3) Until the campaign cost exceeds a certain level, the conversion rate of users will be small; however, at a certain point, the efficiency will be maximized. However, this efficiency will decrease again and converge to zero following 2).

We can define $f(C)$, which is a function of conversion rate for a cost as per (7) considering these assumptions

$$f(C) = \frac{\gamma_{\max}}{1 + e^{-\epsilon\times(\frac{C}{C_{\max}}-\rho)}}, C_{\max} = \gamma_{\max} \times \text{CLV} \quad (7)$$

where $\gamma_{\max}$ is the maximum conversion rate that can be obtained through the campaign for churn prevention, and $C_{\max}$ is the maximum campaign cost. ρ and ϵ are hyperparameters for adjusting the position and width of the interval, respectively, in which the efficiency is maximized.

ρ is a value between 0 and 1. If ρ is less than 0.5, it is efficient at low cost; otherwise it is efficient at a high cost. A large value of ϵ means that the conversion rate rises sharply in a narrow interval. For example, Fig. 13 is a graph when $\gamma_{\max}$, ρ, and ϵ are 0.2, 0.5, and 10, respectively. If it is possible to track changes of the conversion rate of churners with various reward conditions in a live game via A/B test, these parameters can be estimated

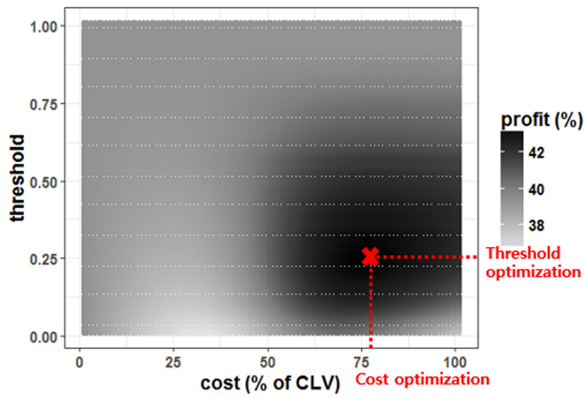

Fig. 14. Example of profit optimization applying $f(C)$.

approximately via least square method. We then can determine the optimal campaign cost by applying $f(C)$ to (5) (see Fig. 14).

VIII. Conclusion

The purpose of churn analysis is to prevent losses caused by user churn. Consequently, churn prediction is required to not only improve prediction accuracy but also maximize expected benefits. To the best of our knowledge, there is no study for churn prediction considering the expected profit in the online game, while there are numerous studies in other fields [4], [6], [7].

We propose a churn prediction process considering the expected profit of the online game by referring to the existing research methods and apply it to the live game that has been in service for over nine years to verify its effectiveness. There are three main features of our proposed method. First, we define churn via analyzing the access patterns of users. Second, long-term loyal customers with a high benefit are identified and used for churn prediction. Finally, we calculate the expected profit per user via cost-benefit analysis and optimize the prediction model.

According to our experiments, only the users with little benefit are most likely to be detected when the churn prediction model is applied to the entire user. Therefore, considering the campaign cost, it is possible to incur a revenue loss. On the other hand, if the prediction model is applied only to loyal customers, high profit can be expected in most scenarios. Furthermore, optimizing the threshold of the predictive model can obtain an additional benefit of approximately 10%–30% over the optimized model for accuracy.

The churn prediction process we propose partially applied to the game data mining competition at CIG 2017 [37]. The training data used in the competition were constructed using the process described in Sections IV-A and VI.

Furthermore, we discovered some useful characteristics of churners in an online game. According to our analysis, social factors have a significant influence on loyal customers' churn in online games. The result supports the claims made by previous studies [18], [28], [29]. Additionally, we found that churners may show signals of a descending trend or increasing volatility of in-game activities from up to ten weeks before leaving. We believe that this result will be a useful reference for predicting user churn in other online game services.

Finally, we consider that the profit estimation method we used will be necessary for other researchers to analyze user churn in online game services. However, there is still the limitation that sufficient verification has not been achieved in practice. We plan to verify and improve the proposed method in subsequent studies rigorously.

References

[1] Infiniti Research Limited, "Global online gaming market 2014," 2014. [Online]. Available: http://www.marketwatch.com/story/global-online-gaming-market-2014–2014-06-25

[2] SUPERDATA, "The MMO & MOBA games market report, 2016," 2016. [Online]. Available: https://www.superdataresearch.com/market-data/mmo-market/

[3] M. S. El-Nasr, A. Drachen, and A. Canossa, *Game Analytics*. Berlin, Germany: Springer, 2016.

[4] T. Verbraken, W. Verbeke, and B. Baesens, "Profit optimizing customer churn prediction with Bayesian network classifiers," *Intell. Data Anal.*, vol. 18, no. 1, pp. 3–24, 2014.

[5] A. Saran Kumar and D. Chandrakala, "A survey on customer churn prediction using machine learning techineque," *Int. J. Comput. Appl.*, vol. 154, no. 10, 2016.

[6] T. Verbraken, W. Verbeke, and B. Baesens, "A novel profit maximizing metric for measuring classification performance of customer churn prediction models," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 961–973, May 2013.

[7] A. C. Bahnsen, D. Aouada, and B. Ottersten, "A novel cost-sensitive framework for customer churn predictive modeling," *Decis. Anal.*, vol. 2, no. 1, pp. 1–15, 2015.

[8] D. J. Hand, "Measuring classifier performance: A coherent alternative to the area under the ROC curve," *Mach. Learn.*, vol. 77, no. 1, pp. 103–123, 2009.

[9] E. Lee, J. Woo, H. Kim, A. Mohaisen, and H. K. Kim, "You are a game bot!: Uncovering game bots in MMORPGS via self-similarity in the wild," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2016.

[10] K. B. Shores, Y. He, K. L. Swanenburg, R. Kraut, and J. Riedl, "The identification of deviance and its impact on retention in a multiplayer game," in *Proc. 17th ACM Conf. Comput. Supported Coop. Work Social Comput.*, 2014, pp. 1356–1365.

[11] E. W. Ngai, L. Xiu, and D. C. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2592–2602, 2009.

[12] C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to churn prediction: A data mining approach," *Expert Syst. Appl.*, vol. 23, no. 2, pp. 103–112, 2002.

[13] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 690–696, May 2000.

[14] A. Baumann, S. Lessmann, K. Coussement, and K. W. De Bock , "Maximize what matters: Predicting customer churn with decision-centric ensemble selection," in *Proc. Eur. Conf. Inf. Syst.*, 2015.

[15] K. Dasgupta *et al.*, "Social ties and their relevance to churn in mobile telecom networks," in *Proc. 11th Int. Conf. Extending Database Technol., Adv. Database Technol.*, 2008, pp. 668–677.

[16] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore, "Alone together?: Exploring the social dynamics of massively multiplayer online games," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2006, pp. 407–416.

[17] S. Pang and C. Chen, "Community analysis of social network in MMOG," *Int. J. Commun. Netw. Syst. Sci.*, vol. 3, no. 2, pp. 133–139, 2010.

[18] T. Chung, J. Han, D. Choi, T. T. Kwon, H. K. Kim, and Y. Choi, "Unveiling group characteristics in online social games: A socio-economic analysis," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 889–900.

[19] N. Glady, B. Baesens, and C. Croux, "Modeling churn using customer lifetime value," *Eur. J. Oper. Res.*, vol. 197, no. 1, pp. 402–411, 2009.

[20] M. Mavri and G. Ioannou, "Customer switching behaviour in Greek banking services using survival analysis," *Manage. Finance*, vol. 34, no. 3, pp. 186–197, 2008.

[21] J. Runge, P. Gao, F. Garcin, and B. Faltings, “Churn prediction for high-value players in casual social games,” in *Proc. IEEE Conf. Comput. Intell. Games*, 2014, pp. 1–8.
[22] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage, “Predicting player churn in the wild,” in *Proc. IEEE Conf. Comput. Intell. Games*, 2014, pp. 1–8.
[23] M. Milošević, N. Živić, and I. Andjelković, “Early churn prediction with personalized targeting in mobile social games,” *Expert Syst. Appl.*, vol. 83, pp. 326–332, 2017.
[24] M. Tamassia, W. Raffe, R. Sifa, A. Drachen, F. Zambetta, and M. Hitchens, “Predicting player churn in destiny: A hidden Markov models approach to predicting player departure in a major online game,” in *Proc. IEEE Conf. Comput. Intell. Games*, 2016, pp. 1–8.
[25] Á. Periáñez, A. Saas, A. Guitart, and C. Magne, “Churn prediction in mobile social games: Towards a complete assessment using survival ensembles,” in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal*, 2016, pp. 564–573.
[26] M. Viljanen, A. Airola, T. Pahikkala, and J. Heikkonen, “Modelling user retention in mobile games,” in *Proc. IEEE Conf. Comput. Intell. Games*, 2016, pp. 1–8.
[27] Z. Borbora, J. Srivastava, K.-W. Hsu, and D. Williams, “Churn prediction in MMORPGS using player motivation theories and an ensemble approach,” in *Proc. IEEE Third Int. Conf. Soc. Comput. Privacy, Secur. Risk Trust*, 2011, pp. 157–164.
[28] J. Kawale, A. Pal, and J. Srivastava, “Churn prediction in MMORPGS: A social influence based approach,” in *Proc. IEEE Int. Conf. Comput. Sci. Eng.*, 2009, vol. 4, pp. 423–428.
[29] K. Park, M. Cha, H. Kwak, and K.-T. Chen, “Achievement and friends: Key factors of player retention vary across player levels in online multiplayer games,” in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 445–453.
[30] W.-C. Feng, D. Brandt, and D. Saha, “A long-term study of a popular MMORPG,” in *Proc. 6th ACM SIGCOMM Workshop Netw. Syst. Support Games*, 2007, pp. 19–24.
[31] T. Debeauvais, B. Nardi, D. J. Schiano, N. Ducheneaut, and N. Yee, “If you build it they might stay: Retention mechanisms in world of warcraft,” in *Proc. 6th Int. Conf. Found. Digit. Games*, 2011, pp. 180–187.
[32] E. Lee, J. Woo, H. Kim, and H. K. Kim, “No silk road for online gamers!: Using social network analysis to unveil black markets in online games,” in *Proc. Web Conf. Proc.*, 2018, pp. 1825–1835.
[33] A. Abbott and A. Tsay, “Sequence analysis and optimal matching methods in sociology: Review and prospect,” *Sociol. Methods Res.*, vol. 29, no. 1, pp. 3–33, 2000.
[34] A. Clauset, M. E. Newman, and C. Moore, “Finding community structure in very large networks,” *Phys. Rev. E*, vol. 70, no. 6, 2004, Art. no. 066111.
[35] P. D. Berger and N. I. Nasr, “Customer lifetime value: Marketing models and applications,” *J. Interactive Marketing*, vol. 12, no. 1, pp. 17–30, 1998.
[36] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, 2007, pp. 29–42.
[37] “Game data mining competition 2017,” *IEEE Conf. Comput. Intell. Games*, 2017. [Online]. Available: https://cilab.sejong.ac.kr/gdmc2017/

Eunjo Lee received the B.S. and M.S. degrees in computer science from Soong-sil University, Seoul, South Korea, in 2003 and 2007, respectively. He is currently working toward the Ph.D. degree in information security at Korea University, Seoul, South Korea.

He is currently a Leader of Data Analysis and Modeling Team, NCSOFT, Seongnam, South Korea. His research interests include game fraud detection, user behavioral pattern modeling, and time-series analysis in on-line games.

Boram Kim received the B.S. degree in statistics from Hankuk University of Foreign Studies, Seoul, South Korea, in 2006, and the M.S. degree in statistics from Seoul National University, Seoul, South Korea, in 2009.

From 2009 to 2017, she was with the Data Analysis and Modeling Team, NCSOFT, Seongnam, South Korea. She is currently a member of the Data Analysis Team, NEXON, Seongnam, South Korea. Her research interests include data mining and artificial intelligence.

Sungwook Kang received the B.S. degree in software engineering from Kumoh National Institute of Technology, Gumi, South Korea, in 2014, and the M.S. degree in information security engineering from Korea University, Seoul, South Korea, in 2016.

From 2016 to 2018, he was with the Data Analysis and Modeling Team, NCSOFT, Seongnam, South Korea. He is currently a member of the Data Analysis Team, NEXON, Seongnam, South Korea. His research interests include data mining and machine learning.

Byungsoo Kang received the B.S. degree in computer science from Dankook University, Yongin, South Korea, in 2014.

He is currently a member of the Data Analysis and Modeling Team, NCSOFT, Seongnam, South Korea. His research interests include the game fraud detection using deep learning algorithm.

Yoonjae Jang received the B.S. degree in statistics from Carnegie Mellon University, Pittsburgh, PA, USA, in 2015.

He has been a member of the Data Analysis and Modeling team, NCSOFT, Seongnam, South Korea, since 2017. His research interests include predictive modeling of game user behavior based on behavioral and survival analysis of in-game activities.

Huy Kang Kim (M’13) received the B.S. degree in industrial management and the M.S. and Ph.D. degrees in industrial and systems engineering from KAIST, Daejeon, South Korea, in 1998, 2000, and 2009, respectively.

He founded A3 Security Consulting, South Korea, the first information security consulting company in 1999. He was the Technical Director and the Head of the Information Security Department, NCSOFT, Seongnam, South Korea, from 2004 to 2010. He is currently an Associate Professor in the Graduate School of Information Security, Korea University, Seoul, South Korea. His research interests include solving security problems in online games based on the user behavior analysis and data mining.