

AI-based early detection to prevent user churn in MMORPG

Minhyuk Lee^a, Sunwoo Park^a, Sunghwan Lee^a, Suin Kim^a, Yoonyoung Cho^b,
Daesub Song^b, Moonyoung Lee^b, Yoonsuh Jung^{1,a}

^aDepartment of Statistics, Korea University; ^bData Analytics Lab, Kakao Games

Abstract

Massive multiplayer online role playing game (MMORPG) is a common type of game these days. Predicting user churn in MMORPG is a crucial task. The retention rate of users is deeply associated with the lifespan and revenue of the service. If the churn of a specific user can be predicted in advance, targeted promotions can be used to encourage their stay. Therefore, not only the accuracy of churn prediction but also the speed at which signs of churn can be detected is important. In this paper, we propose methods to identify early signs of churn by utilizing the daily predicted user retention probabilities. We train various deep learning and machine learning models using log data and estimate user retention probabilities. By analyzing the change patterns in these probabilities, we provide empirical rules for early identification of users at high risk of churn. Performance evaluations confirm that our methodology is more effective at detecting high risk users than existing methods based on login days. Finally, we suggest novel methods for customized marketing strategies. For this purpose, we provide guidelines of the percentage of accessed users who are at risk of churn.

Keywords: churn prediction, deep learning, game analytics, MMORPG

1. 서론

Massive multiplayer online role playing game (MMORPG) 시장에서 이탈 예측은 중요한 과제이다. 유저가 게임 자체를 구매하여 수익이 발생하는 패키지 게임과는 달리, Free-to-play 게임에서는 유저가 게임을 플레이 하는 과정에서 현금으로 구매하는 인게임 재화를 통해 수익을 얻는다. 따라서, 유저가 게임을 지속적으로 플레이 할 수록 기대 수익이 더 커지는 구조를 가지고 있다. 이런 이유로 게임사 입장에서 이탈 위험이 있는 유저를 조기에 파악하고 선제적으로 대응하는 것이 중요하다. Zhao 등 (2023)은 실제 게임사 고객을 대상으로 A/B test를 수행하여 이탈 위험 유저에 대해 적절한 팝업 알림과 선물을 제공하면 이탈률을 낮추는 것에 효과적이라는 것을 보였다. MMORPG 게임에서는 유저가 이탈했는지 여부를 명시적으로 확인하는 것이 불가능하다. 접속을 하지 않는 유저가 아예 게임을 떠난 것인지, 잠시 일이 생겨 쉬는 것인지 구분할 방법이 없기 때문이다. 이에 따라, 확인 가능한 지표를 통해 이탈을 재정의하고 이를 목적 변수로 삼는 방식을 취하는 것이 적절하다. 가장 많이 쓰이는 방법은 미리 정의된 기간 동안 연속적으로 미접속하는 행동을 이탈로 정의하는 것이다. 이렇게 사전 정의된 기준으로 이탈 여부를 이항 변수화한 후, 기계학습 모델을 이용해 예측하는 연구들이 활발히 진행되고 있다. 예를 들면, 로지스틱 회귀, support vector machine (SVM), 랜덤 포레스트 (Lee 등,

Jung's work has been partially supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (MIST) (No. 2022R1F1A1071126 and No. 2022M3J6A1063595).

¹Corresponding author: Department of Statistics, Korea University, 145 Anam-ro Seongbuk-Gu, Seoul 02841, Korea. E-mail: yoons77@korea.ac.kr

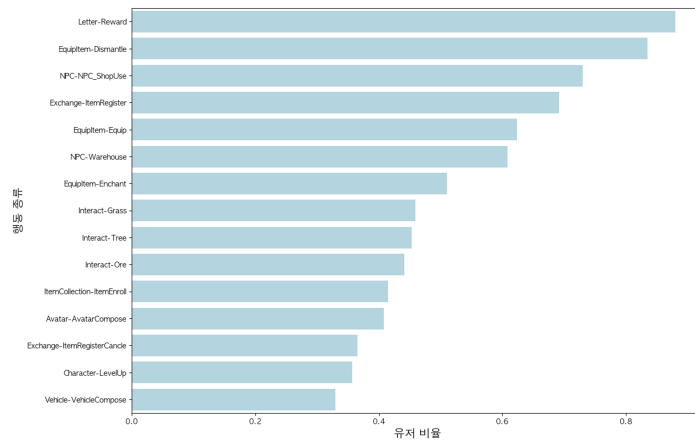


Figure 1: The proportion of users who show at least one action for each behavior.

2016), 나이브 베이즈, gradient boosting (Milovsevic 등, 2017) 등이 있다. 또한, Mena 등 (2019)과 Kristensen과 Burelli (2019)는 시간이 흐름에 따른 유저의 행동 변화를 감안하기 위하여 유저 로그 기록을 시퀀스 변수화하고, recurrent neural network (RNN)이나 long short term memory (LSTM) (Hochreiter와 Schmidhuber, 1997)을 기반으로 한 모델을 구성하여 분석을 진행하였다. Viljanen 등 (2017)은 생존분석 모델을 적용하여 유저가 얼마나 더 게임을 플레이할지 예측하는 접근법을 제시하였고, Khodadadi 등 (2020)은 유저의 이탈 여부가 아닌 재접속까지 며칠이 걸리는지를 목적 함수로 사용하였다.

이탈 유저의 예측 문제에서는 예측의 정확도도 중요하지만, 이탈의 징후를 얼마나 빠르게 파악할 수 있는지도 또한 중요하다. 이미 유저가 떠난 후에는 다시 게임으로 잘 돌아오지 않기 때문이다. 본 연구에서는 조기에 이탈 위험 유저군을 탐지하고, 이 유저군에 대한 그룹별 마케팅 시기를 제안하고자 한다. 게임사에서 제공이 가능한 인센티브를 고려하여 게임 유저들을 유지할 수 있는 여러가지 방법을 제안한다. 이를 위하여 유저별 잔존 확률을 기반으로 확률의 변화 패턴을 살펴보고, 이탈 가능성이 높은 유저를 조기에 감지할 수 있는 경험적 규칙을 제시한다. 유저 재접속 여부를 이항 변수로 정의하여 이에 대한 분류 모형을 학습하는 점에서 기존 방법과 동일하다. 기존 연구들과의 차별점은 학습된 분류 모형을 통해 예측된 이탈 확률을 직접적으로 사용하여 결론을 내는 기존의 방법과는 달리, 일일 확률의 변동을 통해 이탈 위험 유저를 특정하는 점이다. 본 논문은 국내 게임사의 MMORPG 로그 데이터를 모델링하여 이를 기반으로 한 실제적인 마케팅 전략을 제안하는 것에 초점이 맞춰져 있다. 이 논문이 게임 분야의 연구에 기여하는 바를 요약하면 다음과 같다.

첫째, 이 논문에서 제안하는 방법을 이용하면 기존의 접속일 기반으로 한 방법들보다 이탈 위험 유저를 더 빠른 시기에 식별할 수 있다. 둘째, 기존의 이탈 정의 모델은 게임을 잘하다가 갑자기 그만두는 유저를 잘 찾아내지 못하는데 반해, 여기서 제시하는 방법은 이러한 유저들도 효과적으로 탐지할 수 있다. 셋째, 유저마다의 잔존가치가 다르다는 사실에 기반하여 유저의 과금량을 기반으로 잔존가치를 측정하고, 이를 통해 유저 그룹별로 맞춤형 마케팅 시기를 제안한다. 마지막으로 유저가 접속했을 때 해당 유저가 이탈할 가능성을 실시간으로 예측하는 것이 대응의 차원에서는 매우 중요하다. 접속해 있는 상태에서 이탈 위험이 있는 유저에게 즉시 프로모션을 제공하는 것은 유저를 유지하기 위한 전략의 핵심이기 때문이다. 이러한 게임 현업의 현실적인 요구사항을 반영하여 본 연구는 접속 유저 대상으로 이탈 위험 유저군에 대한 식별을 위한 가이드라인을 제공한다.

본 논문의 구성은 다음과 같다. 2절에서는 분석에 사용된 데이터에 대한 설명과 모델에서 사용한 변수들을 소개한다. 3절에서는 본 논문에서 제안하는 유저 이탈을 조기에 탐지하는 방법을 설명한다. 4절에서는 3

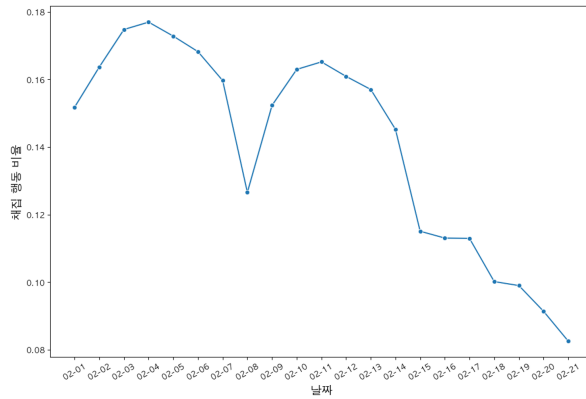


Figure 2: Mean of the proportion of collection behavior from the users who left on February 22nd.

절에서 소개한 방법을 실제 MMORPG 게임 데이터에 적용한 결과와 활용방안에 대하여 서술한다.

2. 데이터 소개 및 EDA

이 논문은 국내 MMORPG의 게임에서 2023년 2월과 3월 동안 수집된 로그 데이터를 이용한다. 데이터의 구조는 일반적인 테이블 형식이다. 각 행은 게임 유저를, 각 열은 설명 변수를 구성한다. 원본 데이터는 접속 기록 로그, 플레이 행동 로그, 재화 변동 로그 등 다양한 종류의 로그로 나뉘어져 있다. 탐색적 자료 분석 과정을 거쳐 로그 중에서도 유저 이탈과 관련이 있을 것으로 생각이 되는 로그들을 선택한 후, 해당 로그들을 유저 단위로 다시 정리하여 일일 변수로 추출한다. 유저 행동 로그는 유저가 행동을 취할 때마다 이에 대한 정보가 기록이 되는 로그이다. 이 로그를 기준으로 특정 유저가 어느 날짜에 접속했는지 여부를 변수화 한다. 재화 변동 로그에서는 유저의 재화가 변동이 생길 때마다 변화된 재화의 종류, 원인, 변화량 그리고 변화 이후 잔액이 기록되어 있다. 재화 중에서도 인게임 재화와 현금 거래 재화로 구분하여 이에 대한 변수화를 진행하며 두 재화에 대하여 일일 재화 획득량과 소모량을 변수화한다. 변수들은 유저 행동을 분석한 변수들과 재화의 변동을 기록한 로그로 이해할 수 있으며 그 자세한 사항은 다음과 같다.

유저 행동 로그는 유저가 행동을 취할 때마다 이에 대한 정보가 기록이 되는 로그이다. Figure 1에서 유저들이 많이 하는 행동들을 확인할 수 있다. 주요 행동들이 나타나는 빈도를 통해 유저 행동 패턴을 설명할 수 있다. 예를 들어, 가장 비율인 높은 행동인 Letter-Reward는 편지를 열어 보상을 받는 행동이다. 접속 이벤트 등의 보상을 받는 행동이므로 게임을 계속 할 마음이 있는 유저들은 대부분 하는 행동이다. 그 밑으로 Dismantle, Shopuse, ItemRegister 등의 행동들은 일반적으로 게임 플레이를 하는 과정에서 필수적으로 하게 되는 행동들이다. 해당 행동들이 나타나는 빈도를 통해 유저의 플레이 패턴을 파악할 수 있다. Interact 변수는 채집 행동을 나타낸다. 채집은 RPG게임에 메인 콘텐츠인 전투와 직접적으로 관련된 행동은 아니지만, 인게임 재화를 획득할 수 있기에 꾸준히 하는 유저들이 존재한다. 따라서, 모든 유저들이 하는 행동은 아니지만 게임에 대한 충실도를 확인할 수 있는 행동이라고 볼 수 있으므로 변수화 한다. Figure 2는 (임의로 선택한) 22일에 이탈한 유저들이 날짜별로 채집 행동을 한 비율 나타낸다. 전체적인 트렌드를 봤을 때, 이탈 날짜에 가까워질 수록 채집의 비중이 점점 감소함을 확인할 수 있다. Figure 3은 이탈 유저와 비이탈 유저의 행동들 중 비중의 차이가 큰 행동들을 보여주고 있다. 채집 행동과 아이템 거래 관련 행동들이 상위권을 차지하고 있다. 그 외에도 길드 탈퇴, 아이템 해제 등의 행동들이 나타나 있다. 따라서, 해당 행동들의 변동을 파악하면 유저가 행동하기 전 징후를 파악할 수 있을 것으로 기대된다.

재화 변동 로그에서는 유저의 재화가 변동이 생길 때마다 변화된 재화의 종류, 원인, 변화량 그리고 변

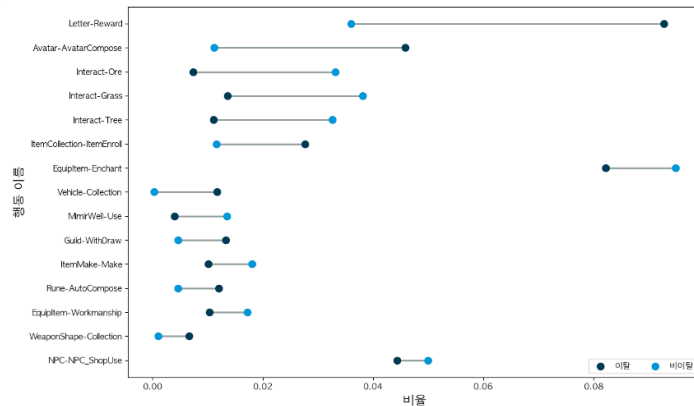


Figure 3: Difference of the rate for each behavior between the churn users and non-churn users. The rates are calculated before the churned users' leave.

화 이후 잔액이 기록되어 있다. 재화 중에서도 가장 많이 사용되는 인게임으로 획득 가능한 재화(“골드”)와 현금 거래 재화(“캐쉬”)로 로그들을 구분하여 일일 재화 획득량과 소모량을 변수화 하여 사용한다. 단순 재화 획득과 소모량만 보서는 유저의 행동을 모두 반영할 수는 없다. 따라서, 어떤 행동에 의해서 재화 변동이 발생했는지 세분화하여 이를 변수로 추가한다.

특정 로그가 하루에 나타난 숫자를 그대로 변수로 사용 한다면 유저들의 일일 행동 패턴 변화를 반영할 수 있다는 장점이 있다. 그러나, 유저 별로 플레이 타임이 달라서 행동 로그가 기록되는 횟수가 다르기에 유저 별 스케일이 크게 다를 수 있다는 문제가 발생한다. 이를 고려하여 행동 로그가 나타난 횟수를 바로 사용하기 보단 유저의 하루 전체 행동 중 해당 행동의 비율을 변수화한다. Table 1은 변수의 탐색과 가공 과정을 거쳐서 나온 최종 시계열 변수를 나타낸다.

현재 해당 게임사에서는 7일간 미접속 시 이탈로 판단하고 있으나 7일 이상 미접속한 유저들에게 이탈 대응의 효과는 미비하다. 따라서 이탈 유저를 7일보다 조기에 정확하게 파악할 수 있는 방법을 제공하고자 한다. 또한, Free-to-play (F2P) 모바일 게임의 재화, 활동 등 다양한 로그 데이터를 바탕으로 이탈 위험 유저군을 식별하고, 해당 유저군을 대상으로 한 그룹별 마케팅 시기를 제안한다. 추가적으로, 접속 유저 대상으로 이탈 위험 유저군에 대한 가이드라인을 제공하고자 한다.

분석을 시작하기 전에 붓일 가능성이 있는 신규 유저는 분석대상에서 제외하며, 2월 내에 접속 기록이 있는 유저 214,069명을 분석 대상으로 하되, 현재 접속한 유저를 대상으로 그 이후의 이탈을 예측하는 경우에는 접속한 유저만 사용한다. 예를 들어 3월 1일에 접속한 유저는 164,812명이며 3월 8일에 접속한 유저는 164,984명이다. 신규 유저를 판단하는 기준은 해당 유저의 가입일이다. 과금을 하는 유저의 이탈은 게임사의 수익에 큰 영향을 미치기 때문에, 과금 유저에 대한 모니터링이 필수적이다. 유저의 잔존가치란 유저가 회사에 가져다주는 예상 전체 수익을 의미한다. 이는 유저가 처음 서비스를 이용하기 시작한 순간부터 그들이 서비스를 더 이상 이용하지 않을 때까지의 기간 동안 회사에 제공하는 총 가치를 나타낸다. 본 연구는 유저의 잔존가치를 그들의 과금 내역을 기반으로 분석하고자 한다. 유저들을 무과금, 저과금, 중과금, 고과금, 초고과금 유저로 5개의 종류로 분류한다. 이 분류는 분위수를 기준으로 하여 이루어지며, 2Q 값은 67,000원, 90%Q는 997,000원, 그리고 99%Q는 4,713,480원으로 확인된다. Figure 4는 유저들의 2월 과금량 분포를 나타내며, Table 2는 과금 수준을 4가지의 카테고리로 정의한다.

‘구매 주기(purchase lag)’는 유저별로 발생하는 모든 구매 주기의 평균 시간을 의미하며, Figure 5와 같이 과금량이 많은 유저일수록 구매 주기가 상대적으로 짧은 경향을 보인다.

Table 1: Variables obtained from the daily log

Variable name	Details
Connection status	Whether user is connected on a certain day (0:no, 1:yes)
Play time	Play time in seconds
Dismantling item	Dismantling unused item and make materials
Visit a shop	Visit NPC shop in a game
Gathering	Acquire materials by interacting with dedicated objects
Synthesis Avatar	Synthesis Avatar to make a new avatar
Item registration	Register items in the auction house for user-to-user trading
Item enhancement	Enhance items by consuming materials and resources
Accept rewards	Accept rewards from events and/or challenges
Gold	Total gold obtained
Cash usage	Total amount of consumed cash
Challenge	Successful in challenge
Elite dungeon	Successful in elite dungeon mission
Party dungeon	Successful in party dungeon mission
Quest	Successful in quest
Use gold	Consumption of gold in NPC shop for buying item
Use gold2	Consumption of gold for rapid movement
Use cash	Consumption of cash for buying item
Cash acquisition	Acquired cash in an auction house by selling items
Use cash2	Consumption of cash in an auction house by buying items

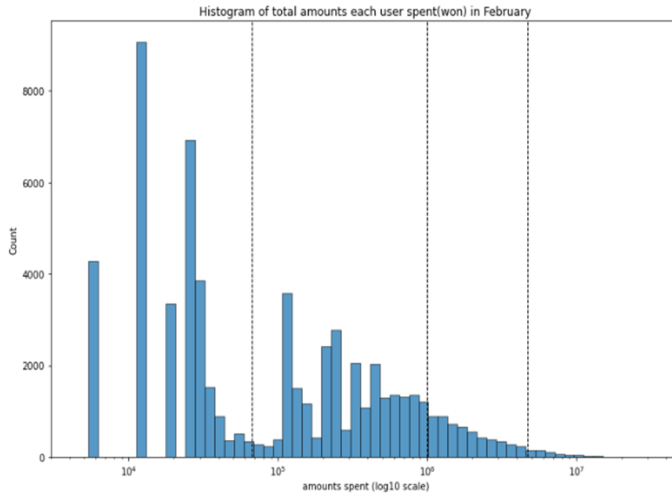


Figure 4: Distribution of expenditure in February. Vertical lines are the quantiles of interest: 2Q (67,000won), 90% (997,000won), 99% (4,713,480won)).

고과금 및 초고과금 유저를 대상으로 (상위 10%의 과금) 유저 주별 과금일 수와 유저 주별 과금량을 확인한다. 2023년 2월의 마지막 주 (2월 22일 ~ 2월 28일) 동안 접속하지 않는 유저를 이탈 유저로 정의하며, 이에 따라 고과금 이탈 유저는 60명, 초고과금 이탈 유저는 4명이다. 고과금 및 초고과금 유저군에서는 이탈 직전의 주에 주별 과금일 수와 주별 과금량이 급감하는 특징을 보인다. 또한 이탈하지 않는 유저군에 비해

Table 2: Number of users in each category of expenditure

Payment category	Payment in February (won)	Number of users
None	0	157,877
Small	1 – 67,000	28,217
Medium	67,000 – 997,000	22,022
Large	997,000 – 4,713,480	5,344
Super large	More than 4,713,480	609

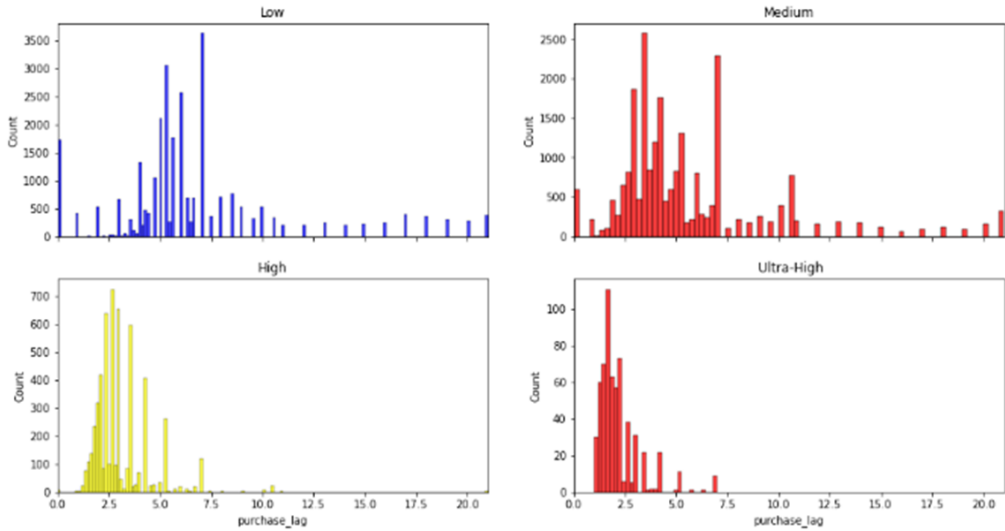


Figure 5: Interval of expenditure among the users except for those who do not pay.

이탈을 하는 유저군은 이탈 직전 주에 과금 행동이 확연히 감소한다.

3. 유저 이탈 방지 및 프로모션을 위한 방법론들

3.1. 유저의 군집화에 기반한 프로모션

기초적인 분석 방법으로 K-평균 군집 분석을 통해 고과금 유저들의 기본적 특성을 확인한다. 최적의 군집 개수를 3으로 선정하였으며 군집 0, 군집 1, 군집 2로 명명한다. 군집 변수로는 앞에서 제시한 시계열 변수와 비시계열 변수(예: 최대 연속 접속 일수, 최대 연속 미접속 일수, 최초 접속 이후 비접속 횟수)가 사용된다. 각 군집의 특징을 파악하여 해당 군집 내의 유저들이 이탈 위험군으로 분류될 경우, 마케팅 전략이나 프로모션 제안에 유용한 정보를 제공할 수 있을 것으로 기대된다. 시계열 데이터는 절대적인 값에 중점을 두는 것이 아닌, 시점 변화에 따른 추세를 파악하는 것이 중요하다. 우선 2월 1일부터 2월 21일까지의 시계열 변수들을 식 (3.1)의 방법으로 표준화한다.

$$X_{ij}^* = \frac{X_{ij} - \hat{\mu}_i}{\hat{\sigma}_i}, \quad (3.1)$$

X_{ij} 는 유저 i 의 t 시점의 j 번째 시계열 변수를 의미하며, $\hat{\mu}_i$ 는 유저 i 의 1일부터 21일까지의 시계열 변수의 평균을 뜻한다. $\hat{\sigma}_i$ 는 유저 i 의 1일부터 21일까지의 시계열 변수의 표준편차를 의미한다. 또한, 모든 변수들의

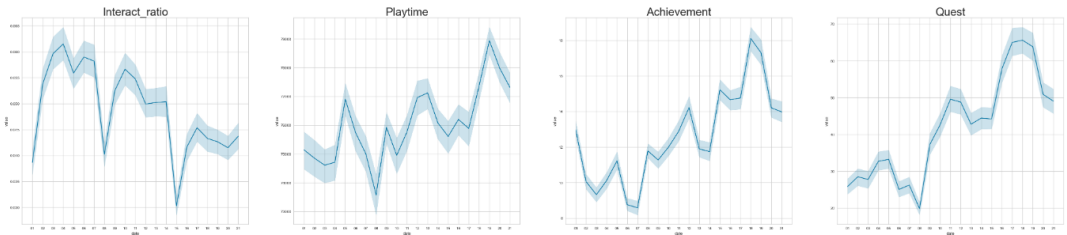


Figure 6: Variables showing clear trend over time in Cluster 2.

분포와 스케일 차이를 조정하기 위해 변수별 표준화를 진행한다.

Figure 6은 여러 시계열 변수들 중에서 2월 1일부터 2월 21일까지의 기간 동안 특히 두드러진 패턴을 보이는 ‘interact_ratio’, ‘playtime’, ‘achievement’, ‘quest’ 변수들의 시계열 추세를 나타낸다. 군집 분석과 EDA를 통해 파악한 군집들의 특징은 다음과 같다.

군집 0: 특정 활동에 몰입하는 유저군 (2977명) 군집 0의 유저들은 ‘interact_ratio’가 증가하는 반면, ‘playtime’, ‘achievement’, ‘quest’는 감소하는 경향을 보인다. 채집 작업이 게임 내에서 많은 시간을 소모하는 것을 고려하면, 이 군집의 유저들은 게임의 효율성보다는 특정 활동에 몰입하는 경향이 있을 것이라고 판단된다.

군집 1: 활동 빈도가 낮은 유저군 (151명)

통해 군집 1에 속하는 유저들은 다른 군집에 비해 최대 연속 접속 일수가 적고, 최대 연속 미접속 일수와 최초 접속 이후 비접속 횟수(inactive_count)가 높은 것이 확인된다. 이러한 특징은 일정 기간 동안 활동하지 않는 경향이 있는 유저군을 나타낸다. 더욱이, 이 군집에 속하는 유저들 중 이탈 고과금 유저 41명 중 30명이 포함되어 이탈 가능성이 다른 군집에 비해 높다. 이 군집의 유저들은 게임 내에서의 성취를 추구하기보다는 과금을 통해 게임의 성공을 추구하는 경향이 있는 것으로 보인다.

군집 2: 다양한 게임 활동을 선호하는 유저군 (2827명)

군집 2에 속하는 유저들은 ‘interact_ratio’가 감소하는 반면, ‘playtime’, ‘achievement’, ‘quest’는 증가하는 경향을 보인다. 이들은 채집보다는 다양한 게임 활동에 더 많은 시간을 투자하며, 게임의 다양한 측면을 즐기는 경향이 있다. 따라서, 이들에게는 다양한 퀘스트와 성취를 통한 보상을 제공하는 전략이 효과적일 것으로 예상된다.

3.2. 이탈 예측에 기반한 프로모션

이 절에서는 유저의 이탈 예측 및 이에 기반한 마케팅 전략 수립을 위한 분석 프로세스를 제시한다. 분석 방법은 다음과 같은 순서로 진행된다.

1. **유저의 다음날 접속 여부 예측:** 학습한 모델을 바탕으로 분석 시점 이후의 날짜에 대해 각 유저별로 다음날 접속하지 않을 확률을 예측한다.
2. **확률 패턴 분석 및 시각화:** 예측된 확률 값들의 패턴을 시각화하여 분석한다.
3. **이탈 위험 유저군 식별할 규칙 생성:** 분석된 확률 패턴을 바탕으로 규칙을 생성하여 이탈 위험 유저군을 식별한다.
4. **과금 수준 기반 마케팅 시기 제안:** 유저의 과금 수준을 기반으로 잔존가치를 고려한 적절한 마케팅 시기를 제안한다.

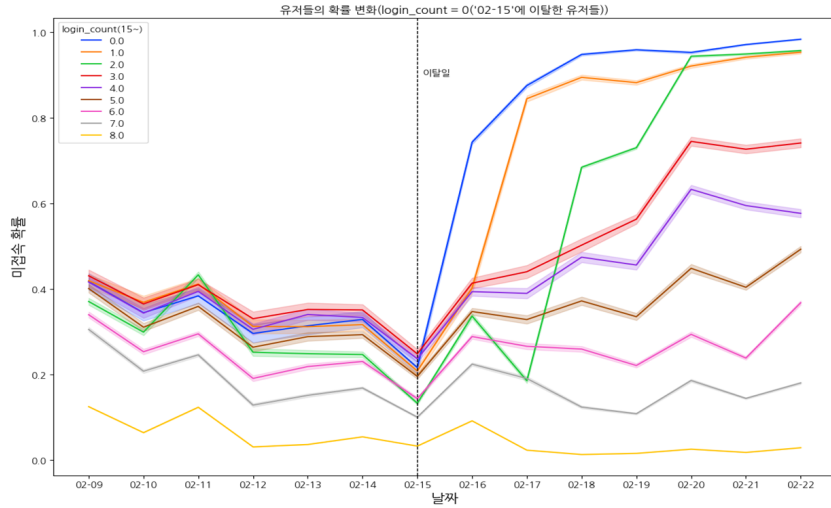


Figure 7: Trend in the probability of churning for those who visited game on February 14th. $\text{login_count} = 0$ stands for the probability of churn who left on Feb 15th, and $\text{login_count}=0$ for the probability of churn who left on Feb 16th, and so on.

위 방법의 목적은 일련의 시계열 변수들을 바탕으로 유저의 다음날 접속 여부를 예측하는 것이다. 모델의 독립 변수는 Table 1에서 제시한 시계열 변수들이며, 분석일까지의 모든 시계열 변수들을 사용한다. 내일 유저가 접속할 지에 대한 라벨이 존재하지 않기 때문에, 오늘(분석시점)의 접속 여부를 라벨로 사용하여 이전 일까지의 데이터(분석 시점 이전)로 모델을 훈련한다. 이렇게 훈련된 모델은 다음날의 접속 여부를 예측하는 데 사용된다.

모형의 아키텍처 측면에서는 시계열 데이터에 적합한 시퀀스 기반의 LSTM 네트워크에 기반한 딥러닝 모델을 사용한다. 우선 테이블 데이터 형식을 딥러닝 모델에 맞게 3D 텐서 형식으로 변환한다. 이 모델은 3개의 LSTM 레이어를 적층적으로 구성하며, 각 레이어의 유닛 개수는 [32, 64, 128] 중 가장 높은 검증 F1 score를 가지는 값을 최적으로 선정한다. 그 결과 각 레이어는 64개의 유닛을 포함한다. 이렇게 설계된 구조는 복잡한 시간적 패턴을 효과적으로 학습하도록 도와준다. 각각의 LSTM 레이어에는 과적합을 방지하기 위해 0.3의 비율로 Srivastava 등 (2014)이 제안한 dropout과 recurrent dropout을 설정한다. LSTM의 출력을 분류 작업에 적합하게 변환하기 위해 64개의 유닛과 ReLu 활성화 함수를 가진 Dense 레이어를 사용하며, 이 레이어에서도 0.3의 비율로 Dropout을 적용한다. 모델의 출력 레이어는 1개의 유닛과 시그모이드 활성화 함수를 사용하여, 다음날 사용자의 미접속 확률을 예측한다. 이 출력 레이어에는 계수가 과도하게 커지지 않도록 L1 정규화가 적용되어 있다. ($L1 = 0.001$)

이 모델을 이용하여 유저가 다음날 미접속할 확률을 매일 예측할 수 있으며, 이렇게 예측된 확률의 변화를 확인하여 이탈 징후를 파악할 수 있다. 시각화를 통해 이탈 전 추세를 파악하고 이를 바탕으로 이탈 위험군으로 분류할 규칙을 생성하고자 한다. Figure 7은 2월 14일에 접속한 유저들을 대상으로, 그 이후 15일부터 22일까지의 접속 일수에 따른 미접속 확률 변화를 매일 예측하여 그 예측된 값을 그래프로 보여준다.

Figure 7에서 파란색 선은 14일이 마지막 접속일이며, 15일날 이탈한 유저들을 의미한다. 이 그룹의 유저들에서는 미접속 확률에 급격한 상승이 발생하며, 이 확률은 이후에도 상승하거나 유지되는 것으로 보인다. 반면에, 접속 일수가 많아지는 유저들은 밀어 갈수록 미접속 확률이 상대적으로 낮게 나타난다. 이는 접속 일수가 많은 유저들은 이탈 위험이 낮다고 해석할 수 있다. 이탈 추세에 대한 변화를 명확하게 파악하기 위해,

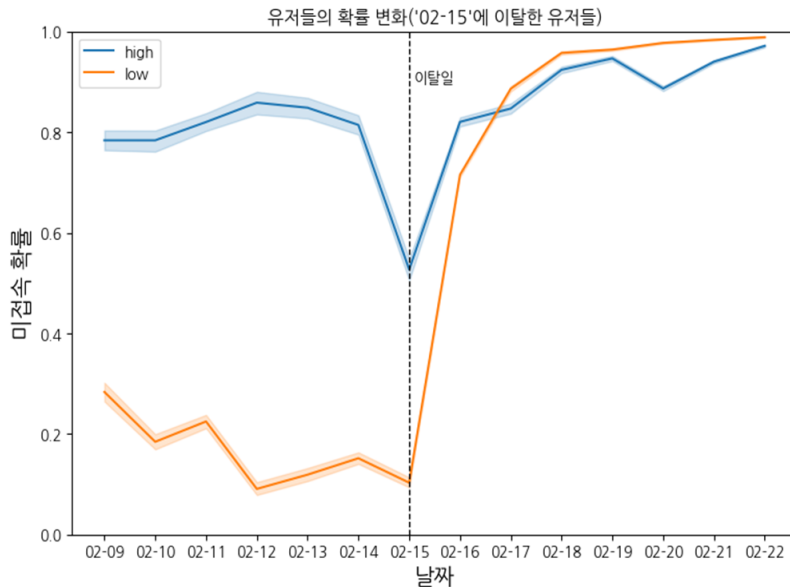


Figure 8: Two patterns in the predicted probability of not connecting to the game. This predicted probability is obtained from those who left on Feb 15th. The predicted probability is calculated using the data until the prior day of prediction.

유저들을 두 가지 경우로 분류하여 미접속 확률의 변화를 확인하고자 한다. ‘high user’는 이탈 전날까지의 평균 미접속 확률이 0.6 이상인 유저로, 원래부터 접속 빈도가 높지 않았던 유저군이다. 그리고 ‘low user’는 이탈 전날까지의 평균 미접속 확률이 0.6 미만인 유저로, 원래부터 접속 빈도가 높았던 유저군을 의미하며 이 기준값이 되는 0.6은 경험적으로 찾아낸 수치이다. Figure 8은 두 이탈 유저 그룹의 패턴이 다를 것을 보여준다. ‘high user’의 경우 V자 형태의 패턴을 보이고, ‘low user’는 급격한 확률 상승의 패턴이 보인다. 이러한 패턴을 바탕으로 유저의 이탈 징후를 파악하고, 이탈 위험 유저군을 식별할 규칙을 생성한다. 해당 규칙은 아래와 같으며 0.25, 0.4와 같은 수치는 경험적으로 판단한 값이다. 판단을 위해 탐색된 값들은 [0.05, 0.10, ..., 0.50]이며, 이 중에서 검증 데이터셋에 대하여 위음성(false negative)을 최소화 하는 값을 선택한다. 위음성을 기준으로 정한 이유는 게임사의 입장에서는 이탈하지 않는 유저를 이탈한다고 잘못 예측하는 것보다는 이탈하는 유저를 이탈하지 않는다고 잘못 예측하는 것이 더 문제가 되기 때문이다.

이탈 위험 유저군의 식별 규칙

- ‘High user’ (기준일 전 미접속 확률의 평균이 0.6 이상)인 경우 다음을 모두 만족: 1. 미접속 확률 하락폭이 0.25 이상, 2. 다음 날 미접속 확률 상승폭이 0.25 이상, 3. 증가한 확률이 계속 증가 or 하락폭이 0.1 이하
- ‘Low user’ (기준일 전 미접속 확률의 평균이 0.6 미만)인 경우 다음을 모두 만족: 1. 미접속 확률 상승폭이 0.4 이상, 2. 이후 증가한 확률이 계속 증가 or 하락폭이 0.1 이하

총 214,069명의 유저 중 19,585명이 이탈 유저를 대상으로 이탈 위험 유저의 확률 변화를 관찰한 후, 이를 바탕으로 이탈 여부를 판단한다. Table 3는 위의 식별 규칙을 1-5일간 관찰하여 적용한 결과이다.

단 하루의 관찰을 통해 이탈을 평가하면, 총 214,069명의 사용자 중 48,199명을 이탈 위험 유저군으로 식별한다. 이들 중 잘못 예측된 유저의 비율이 많다. 반면, 관찰일 수를 3-4일로 늘린다면, 이탈 위험 유저군으로 예측되는 수가 약 20,000명으로 줄어들고, 잘못 판단되는 비율이 크게 감소한다. 짧은 관찰 기간은 이탈 가능

Table 3: The summary of the churn prediction depending on the days of observation

Observed days	1	2	3	4	5
Number of users predicted to churn	48,199	28,698	23,811	21,686	20,531
Correctly predicted users	15,725	15,535	15,486	15,448	15,421
Wrongly predicted users	32,474	13,163	8,325	6,238	5,110

The longer we observe, the more accurate results can be obtained. Because the early prediction is preferred, there is a tradeoff between the early detection and accurate prediction.

Table 4: Prediction accuracy of the user churn depending on their expenditure category

Payment category	Total user	Churn user	Observed days							
			1		2		3		4	
			TP	FP	TP	FP	TP	FP	TP	FP
None	157,877	19,004	15,293	28,522	15,112	12,107	15,065	7,759	15,028	5,866
Small	28,217	335	263	2,433	259	659	259	368	259	252
Medium	22,022	221	151	1,360	148	367	146	185	145	114
Large	5,344	23	16	150	16	25	16	12	16	6

성이 높은 사용자를 조기에 발견하는 것을 가능하게 하므로, 적시에 프로모션을 진행하면 그 효과는 커진다. 그러나 이는 높은 확률로 잘못 판단될 가능성이 있어 게임사의 마케팅 비용이 비효율적으로 사용될 위험이 존재한다. 반대로, 더 긴 관찰 기간은 더 정확한 예측을 제공하지만, 개입이 늦어지는 단점이 있다. 비용과 자원의 잘못된 할당 위험을 줄여주지만, 사용자를 유지하기 위한 조치를 취할 수 있는 시기를 늦추게 된다. 그러므로 프로모션 개입의 시점성과 이탈 예측의 정확성 사이에는 trade-off가 존재한다. 따라서 관찰일 수를 전략적으로 선택함으로써, 프로모션 자원을 효율적으로 사용하는 것이 중요하다. 이러한 방법 중의 하나는 유저의 과금 수준에 따라 그룹을 나누어 수익면에서 중요도가 높은 유저에 대해서는 손해를 감수하고 최대한 빠르게 대처를 하는 동시에 과금 수준이 낮은 유저는 조금 더 보수적으로 판단을 내리는 것이다.

구체적으로, 과금이 없는 유저는 '무과금', 과금량 하위 50% 이하는 소과금, 상위 50%부터 상위 10% 유저는 중과금, 나머지 상위 10% 이상 유저는 고과금으로 정의한다. 이렇게 나눈 과금 수준에 따른 유저 수, 이탈 유저 수 및 예측 결과는 Table 4에 나타나 있다. 고과금 유저의 경우 하루만 관찰하고 이탈 유저를 판단 내리게 되면 166명 중 150명이 실제로는 이탈하지 않는 유저로 나타나게 된다. 비율 상으로 이는 매우 좋지 않은 성과로 보일 수 있으나, 해당 유저층은 과금이 상위 10% 수준인 고과금 유저이기에 16명의 유저의 이탈을 방지할 수 있다면 150명에게 헛되게 프로모션을 제공하는 것도 손해가 아닐 수 있다. 반면, 무과금 유저는 숫자가 많아 절대적인 예측 수가 많을 수 밖에 없다. 이들은 잔존 가치가 높은 유저도 아니기에 더 보수적으로 이탈을 판단하는 것이 수익에 도움이 될 것임을 예상할 수 있다.

현재 게임사에서는 7일 연속 미접속을 이탈로 정의하고 있다. 일률적으로 적용하는 7일 이상의 미접속 기준은 사실상 게임 유저를 유지하기에 거의 불가능하다.

3.3. 접속중인 유저를 대상으로 프로모션

게임에 접속하지 않은 이탈 위험 유저에게 개인 메시지로 선물이나 메시지를 보내는 방법은 유저가 무시할 가능성이 매우 높다. 이에 비해 접속 중인 유저에게 인게임 알림을 통해 메시지를 보내면 유저에게 메시지가 전달될 확률을 높일 수 있다. 이를 위해서는 이탈하기 전 마지막 접속 시점에 이탈 유저를 파악할 수 있는 새로운 방법이 필요하다. 만일 유저가 게임에 접속 중일 때 이탈 위험을 판단하여 즉각적인 프로모션이 가능하다면, 유저의 유지 및 재활성화의 효과를 극대화할 수 있다. 이 절에서는 이러한 상황을 고려하여 게임에

Table 5: Hyperparameters attempted to tune for LightGBM and XGBoost

Hyperparameter	Range	Intervals of tuning
n_estimators	[0,1000]	Tuning by 100 units then by 5 units
max_depth	[1, 2, 3, ..., 13]	By one unit
scale_pos_weight	[1, 3, 5, ..., 41]	Every two units
num_leaves	[2, 4, 6, ..., 40]	Every two units
reg_alpha, gamma	[0,20]	Tuning by 1 unit then by 0.1 units
min_split_gain, subsample, colsample_bytree	[0, 0.1, 0.2, ..., 1]	By 0.1 unit

접속 중인 유저를 대상으로 유저의 다음날 이후의 이탈을 실시간으로 예측하는 방법론을 제안한다.

이를 위하여 다양한 이탈 기간의 연속 미접속 기준을 고려하였는데, 구체적으로 이탈을 하루 미접속부터 7일 연속 미접속까지 다양하게 정의한다. 이렇게 정의한 이유는, 미접속 날짜가 길어질수록 예측의 성능이 저하될 것으로 예상되어 랜덤한 예측에 비교하여 연속 며칠까지 유의미한 예측을 하는지 확인하고자 함이다. 예측 모델에 사용되는 설명 변수는 분석 시점 날짜의 데이터를 포함하지 않고, 분석 시점 전날까지의 행동 데이터만을 포함한다. 설명 변수의 종류는 3.2 절에서 사용한 변수들과 동일하다. 이러한 설계는 사용자가 게임에 접속한 직후 그들의 이탈 확률을 예측하기 위한 것으로, 접속한 날의 데이터는 사용하지 않고 접속했다는 사실만 이용한다.

이탈의 예측을 위해 고려하는 모형들은 Stacked-GRU (Dey와 Salem, 2017), Stacked-LSTM (Hochreiter와 Schmidhuber, 1997), LSTM-CNN (Xia 등, 2020), LightGBM (Ke 등, 2017), XGBoost (Chen과 Guestrin, 2016) 이다. 데이터는 날짜별로 7 : 3의 비율로 학습 데이터와 검증 데이터로 나눈다. 현재 접속일을 3월1일부터 3월 20일 중의 하루로 정의하고, 2월 1일부터 현재 접속일 이전까지의 데이터를 모두 이용하여 모델을 훈련한다. 예를 들어, 3월 2일 접속한 유저에 대한 데이터는 2월 1일부터 3월 1일까지의 데이터를 사용한다. 분석이 가능한 데이터가 제한적이어서 데이터를 최대한 이용하기 위하여 3월 20일 접속한 유저의 경우에는 2월 1일부터 3월 19일까지의 데이터를 모두 이용한다. 이 과정을 5번을 반복하여 오차의 평균과 표준오차를 계산한다. 딥러닝 모델의 아키텍처 최적화를 위해 고려한 하이퍼파라미터는 다음과 같다. Node의 수는 [16,32,64,128], dropout은 [0,0.1,0.3] 중 선택되며, L1 regularizer은 0.001로 고정한다. 마지막으로 learning rate과 epoch 수는 학습 과정에서 자동적으로 조절된다. Table 5는 튜닝에 사용된 초모수들의 범위를 나타낸다.

이탈 예측 모델의 성능을 비교하기 위하여 정확도(accuracy), 정밀도(precision), 재현율(recall), F1 score 를 이용한다. Table 6 는 이탈 정의 기준 별 결과를 보여준다. 전체적으로, 이탈의 기준이 되는 연속 미접속일이 길어질수록 예측 성능이 낮아짐을 확인할 수 있다. 예측해야 하는 연속적인 미접속일이 늘어남에 따라, 이탈하는 유저 비율이 더 낮아진다. 이에 따라 발생하는 이탈 유무 클래스의 불균형이 심화되어 점점 예측이 어려워지는 것으로 이해할 수 있다. Baseline 모델은 설명변수들을 이용하지 않고 학습 데이터에서 유저의 접속 비율만 이용하여 베르누이 분포에서 랜덤하게 데이터를 발생한 후 이를 이용하여 접속유무의 예측을 진행한다. 연속미접속 일수가 늘어남에 따라 전반적인 모델의 예측은 낮아짐을 알 수 있다. 그럼에도 불구하고 7 일까지의 연속 미접속을 이탈로 정의한 경우에도 Baseline 모델에 비하여 논문에서 고려한 방법들의 예측력이 더 높은 것을 확인할 수 있다. 특히, 클래스 불균형으로 인해 F1 score를 기준으로 보면 baseline 모형에 비해 제안하는 방법론을 사용할 경우 큰 폭의 상승을 기대할 수 있다.

게임사에서도 몇 일의 미접속을 이탈로 볼 것인가에 대한 명확한 정의가 없다. 따라서 1일에서 7일까지의 미접속을 이탈로 간주하여 이탈 확률(즉, 접속 다음날 1일간부터 7일간 미접속할 확률)을 예측하였다. Table 6를 기반으로 이탈 위험 유저군에 대한 가이드라인을 각 미접속에 대한 정의별로 제시하면 다음과 같다. 접속한 유저들을 대상으로 이탈할 확률을 계산한 후 이탈 확률이 1일 미접속을 이탈로 간주한 경우에는 상위 5.8%, 2일은 상위 3.4%, 3일은 상위 2.4%, 4일 이상은 상위 1.0%에 해당하는 유저들을 이탈 위험 유저군으로

판단한다. 이탈 예측 확률을 현재 게임에 접속한 유저들을 대상으로 계산하였기 때문에 이탈 위험 유저군으로 판단이 되면 게임에 접속한 직후에 여러가지 인센티브나 마케팅 프로모션을 제공할 수 있는 기회가 있다.

4. 결론 및 향후 과제

본 연구의 방법론의 기여는 다음과 같다. 첫째, 기존의 접속일 기반의 방법보다 조기에 이탈 위험 유저를 식별할 수 있다. 둘째, 게임을 잘하다가 갑자기 그만두는 유저도 효과적으로 탐지할 수 있다. 셋째, 유저의 과금량을 기반으로 잔존가치를 측정하고 이를 통해 그룹별 마케팅 시기를 제시하여 비용 대비 게임 수익을 증대시킬 수 있다. 마지막으로, 접속 유저를 대상으로 이탈 위험 유저군 가이드라인을 제공해 유저가 접속한 시점에서 프로모션의 효과를 극대화시킬 수 있다.

모바일 게임 산업에서 유저 이탈은 서비스의 지속 가능성과 수익에 결정적인 영향을 미친다. 이탈 예측 문제에서는 예측의 정확도뿐만 아니라, 이탈의 조기 징후를 얼마나 빠르게 파악할 수 있는지도 중요하다. 이를 위해 본 연구에서는 LSTM모델을 활용하여 모바일 게임의 로그 데이터를 분석하고, 유저별 잔존 확률을 일별로 예측하였다. 분석된 확률 패턴을 바탕으로 규칙을 생성하여 이탈 위험 유저군을 식별하였다.

본 연구의 성능 평가 결과, 다양한 과금 수준의 유저를 대상으로 관찰 일수에 따른 이탈 판정 결과를 분석하였다. 이를 통해 관찰일 수와 예측 정확도의 trade-off가 존재함을 확인하였다. 특히, 무과금 유저에 대한 조기 대응은 비효율적일 수 있음을 확인하였으며, 반면에 고과금 유저에 대해서는 즉각적인 대응이 중요함을 인식하였다. 이에 따라 본 연구는 유저의 잔존가치를 과금량을 기반으로 측정하고, 잔존가치에 따라 관찰일수를 다르게 설정하여 그룹별 마케팅 시기를 제안한다. 결과적으로 무과금 유저는 4일, 저과금, 중과금 유저의 경우 2일, 고과금 유저는 1일을 관찰일 수로 마케팅 시기를 제안한다. 추가적으로 게임에 접속한 유저 중 상위 몇 퍼센트가 이탈 위험 유저군에 속하는지 가이드라인을 제공한다.

이 연구는 온라인 게임 산업에 중요한 기여를 할 것으로 보이나, 데이터의 한계로 인해 추가적인 연구가 필요하다. 특히, 분석에 사용된 데이터의 기간이 제한적이어서, 더 장기간의 데이터를 분석할 필요가 있다.

Table 6: Mean of prediction results and its standard error (in the parenthesis) for the users who are currently connecting at game

Day	Model	Accuracy	Recall	Precision	F1 score
1	Baseline	0.892 (0.00152)	0.056 (0.00148)	0.056 (0.00152)	0.056 (0.00147)
	Stacked-GRU	0.920 (0.00131)	0.321 (0.00473)	0.312 (0.00519)	0.316 (0.00481)
	Stacked-LSTM	0.920 (0.00135)	0.312 (0.00491)	0.306 (0.00529)	0.309 (0.00495)
	LSTM-CNN	0.923 (0.00132)	0.330 (0.00533)	0.331 (0.00547)	0.330 (0.00528)
	LightGBM	0.928 (0.00109)	0.371 (0.00435)	0.372 (0.00479)	0.371 (0.00439)
	XGBoost	0.933 (0.00087)	0.412 (0.00432)	0.412 (0.00455)	0.411 (0.00421)
2	Baseline	0.942 (0.00130)	0.030 (0.00149)	0.030 (0.00155)	0.030 (0.00150)
	Stacked-GRU	0.954 (0.00117)	0.268 (0.00620)	0.253 (0.00620)	0.259 (0.00598)
	Stacked-LSTM	0.954 (0.00118)	0.259 (0.00671)	0.247 (0.00674)	0.252 (0.00657)
	LSTM-CNN	0.956 (0.00107)	0.266 (0.00721)	0.264 (0.00709)	0.265 (0.00702)
	LightGBM	0.961 (0.00083)	0.353 (0.00776)	0.350 (0.00749)	0.351 (0.00743)
	XGBoost	0.962 (0.00081)	0.354 (0.00775)	0.351 (0.00740)	0.352 (0.00737)
3	Baseline	0.958 (0.00117)	0.022 (0.00161)	0.022 (0.00157)	0.022 (0.00158)
	Stacked-GRU	0.966 (0.00098)	0.245 (0.00632)	0.227 (0.00690)	0.234 (0.00645)
	Stacked-LSTM	0.965 (0.00103)	0.238 (0.00707)	0.215 (0.00750)	0.224 (0.00705)
	LSTM-CNN	0.967 (0.00094)	0.239 (0.00763)	0.235 (0.00780)	0.237 (0.00759)
	LightGBM	0.971 (0.00079)	0.317 (0.00887)	0.312 (0.00901)	0.314 (0.00878)
	XGBoost	0.971 (0.00077)	0.327 (0.00791)	0.321 (0.00796)	0.323 (0.00775)
4	Baseline	0.966 (0.00125)	0.016 (0.00152)	0.015 (0.00148)	0.015 (0.00149)
	Stacked-GRU	0.970 (0.00103)	0.215 (0.00705)	0.190 (0.00700)	0.199 (0.00664)
	Stacked-LSTM	0.969 (0.00115)	0.215 (0.00652)	0.186 (0.00693)	0.196 (0.00641)
	LSTM-CNN	0.972 (0.00099)	0.203 (0.00677)	0.199 (0.00718)	0.200 (0.00681)
	LightGBM	0.975 (0.00081)	0.267 (0.00973)	0.262 (0.01000)	0.263 (0.00972)
	XGBoost	0.975 (0.00078)	0.283 (0.00968)	0.278 (0.00987)	0.280 (0.00960)
5	Baseline	0.970 (0.00113)	0.017 (0.00152)	0.017 (0.00150)	0.017 (0.00149)
	Stacked-GRU	0.972 (0.00106)	0.198 (0.00676)	0.166 (0.00662)	0.178 (0.00624)
	Stacked-LSTM	0.972 (0.00112)	0.194 (0.00677)	0.162 (0.00673)	0.173 (0.00632)
	LSTM-CNN	0.975 (0.00094)	0.187 (0.00654)	0.184 (0.00707)	0.185 (0.00663)
	LightGBM	0.977 (0.00076)	0.243 (0.01063)	0.237 (0.01067)	0.239 (0.01048)
	XGBoost	0.978 (0.00075)	0.269 (0.00948)	0.264 (0.00970)	0.265 (0.00941)
6	Baseline	0.972 (0.00114)	0.016 (0.00158)	0.016 (0.00162)	0.016 (0.00158)
	Stacked-GRU	0.974 (0.00103)	0.186 (0.00686)	0.151 (0.00690)	0.163 (0.00644)
	Stacked-LSTM	0.973 (0.00105)	0.182 (0.00650)	0.144 (0.00638)	0.157 (0.00599)
	LSTM-CNN	0.977 (0.00095)	0.165 (0.00707)	0.163 (0.00738)	0.163 (0.00709)
	LightGBM	0.979 (0.00075)	0.221 (0.01105)	0.215 (0.01137)	0.217 (0.01110)
	XGBoost	0.979 (0.00075)	0.245 (0.01054)	0.240 (0.01053)	0.241 (0.01037)
7	Baseline	0.975 (0.00102)	0.014 (0.00161)	0.014 (0.00176)	0.014 (0.00164)
	Stacked-GRU	0.976 (0.00107)	0.171 (0.00700)	0.140 (0.00647)	0.150 (0.00619)
	Stacked-LSTM	0.974 (0.00105)	0.168 (0.00668)	0.129 (0.00591)	0.142 (0.00573)
	LSTM-CNN	0.978 (0.00090)	0.149 (0.00647)	0.147 (0.00702)	0.147 (0.00660)
	LightGBM	0.979 (0.00070)	0.209 (0.01010)	0.199 (0.01084)	0.202 (0.01038)
	XGBoost	0.981 (0.00073)	0.236 (0.00987)	0.232 (0.01027)	0.233 (0.00990)

We define the churn based on one day to seven days of not connecting to the game. 'Day' stands for consecutive days to define churn.

References

- Chen T and Guestrin C (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Dey R and Salem F (2017). Gate-variants of gated recurrent unit (GRU) neural networks, *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 1597–1600.
- Hochreiter S and Schmidhuber J (1997). Long short-term memory, *Neural Computation*, **9**, 1735–1780.
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, and Liu T (2017). Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, **30**, 3146–3154.
- Khodadadi A, Hosseini SA, Pajouheshgar E, Mansouri F, and Rabiee HR (2020). Choracle: A unified statistical framework for churn prediction, *IEEE Transactions on Knowledge and Data Engineering*, **34**, 1656–1666.
- Kristensen JT and Burelli P (2019). Combining sequential and aggregated data for churn prediction in casual freemium games. In *Proceedings of 2019 IEEE Conference on Games (CoG)*, London, 1–8.
- Lee S, Hong S, Yang S, and Lee H (2016). *Predicting churn in mobile free-to-play games* 2016 international conference on information and communication technology convergence (ictc), 21046–1048.
- Mena CG, De Caigny A, Coussement K, De Bock KW, and Lessmann S (2019). *Churn prediction with sequential data and deep neural networks. a comparative analysis* arXiv preprint arXiv:1909.11114.
- Milošević, M., Živić, N., and Andjelković, I. (2017). Early churn prediction with personalized targeting in mobile social games, *Expert Systems with Applications*, **83**, 326–332.
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, and Salakhutdinov R (2014). Dropout: A simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, **15**, 1929–1958.
- Viljanen M, Airola A, Heikkonen J, and Pahikkala T (2017). Playtime measurement with survival analysis, *IEEE Transactions on Games*, **10**, 128–138.
- Xia K, Huang J, and Wang H (2020). LSTM-CNN architecture for human activity recognition, *IEEE Access*, **8**, 56855–56866.
- Zhao S, Wu R, Tao J, Qu M, Zhao M, Fan C, and Zhao H (2023). perCLTV: A general system for personalized customer lifetime value prediction in online games, *ACM Transactions on Information Systems(T-G)*, **41**, 1–29.

Received April 08, 2024; Revised May 28, 2024; Accepted June 17, 2024

MMORPG 게임의 이탈 유저에 대한 인공지능 기반 조기 탐지

이민혁^a, 박선우^a, 이성환^a, 김수인^a, 조운영^b, 송대섭^b, 이문영^b, 정윤서^{1,a}

^a고려대학교 통계학과; ^b카카오게임즈 데이터분석랩

요약

Massive multiplayer online role playing game (MMORPG)은 국내 게임에서 큰 비중을 차지하는 게임 장르이다. MMORPG에서 유저 이탈 예측은 중요한 과제 중 하나이다. 인게임 결제가 수익 비중이 높기에 유저 잔존율이 서비스 수명 및 수익과 깊이 연관되기 때문이다. 만약, 특정 유저의 이탈을 사전에 예측할 수 있다면 프로모션을 통해 해당 유저의 잔존을 유도할 수 있을 것이다. 따라서, 이탈 예측 문제에서는 예측의 정확도도 중요하지만 이탈의 징후를 얼마나 빠르게 파악할 수 있는지 또한 중요하다. 본 논문에서는 이탈 징후를 조기에 탐지하기 위하여, 유저별 잔존 확률을 일별로 예측하고 이 예측된 확률 값들을 활용하여 유저 이탈 징후를 조기에 파악하는 방법을 제안한다. 이를 위해, 국내 게임사의 유저 로그 데이터로 여러 모형을 학습하고 유저별 잔존 확률을 구하여 잔존 확률의 변화 패턴에 대한 분석을 통해 이탈 가능성이 높은 유저를 조기에 감지할 수 있는 경험적 규칙을 보인다. 최종적으로, 성능 평가 결과를 통해 기존에 접속일을 기반으로 한 규칙보다 제시한 규칙을 이용할 시 이탈 유저를 조기에 감지할 수 있음을 확인한다. 추가적으로, 유저가 이탈하기 전 시점에 게임 접속 중 유저 이탈을 예측하는 방법과 유저의 게임 스타일에 따른 프로모션 방안을 제시한다.

주요용어: MMORPG, 딥러닝, 유저 이탈 방지, 조기 이탈 예측

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1F1A1071126, No. 2022M3J6A1063595).

¹교신저자: (02841) 서울특별시 성북구 안암로 145, 고려대학교 통계학과. E-mail: yoons77@korea.ac.kr