

온라인 게임 내 유저 간 상호작용 분석을 통한 유저 이탈 예측

Churn Prediction in Online Games through User Interaction Analysis

저자 (Authors)	문주연, 김휘강, 우지영 Joo Yeon Moon, Huy Kang Kim, Jiyoung Woo
출처 (Source)	정보과학회 컴퓨팅의 실제 논문지 24(9), 2018.9, 433-441(9 pages) KIISE Transactions on Computing Practices 24(9) , 2018.9, 433-441(9 pages)
발행처 (Publisher)	한국정보과학회 The Korean Institute of Information Scientists and Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07531011
APA Style	문주연, 김휘강, 우지영 (2018). 온라인 게임 내 유저 간 상호작용 분석을 통한 유저 이탈 예측. 정보과학회 컴퓨팅의 실제 논문지, 24(9), 433-441
이용정보 (Accessed)	한신대학교 218.155.***.152 2019/07/19 17:02 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

온라인 게임 내 유저 간 상호작용 분석을 통한 유저 이탈 예측

(Churn Prediction in Online Games through User Interaction Analysis)

문 주 연 [†] 김 휘 강 ^{††} 우 지 영 ^{†††}
(Joo Yeon Moon) (Huy Kang Kim) (Jiyoung Woo)

요 약 본 연구에서는 대규모 다중 사용자 온라인 롤플레이 게임(MMORPG) 내에서 일어나는 소셜 활동이 유저 이탈에 미치는 영향에 대해 분석하였다. MMORPG 내에서는 유저 간 상호작용이 중요한 역할을 하는 경우가 많고, 따라서 이를 제공하는 커뮤니티 시스템이 발달하였다. 대표적으로 “퀘스트” 등의 미션 협업을 위한 파티와 그보다 지속적인 소셜 활동을 제공하는 길드 등이 그에 속한다. 본 연구는 MMORPG의 실제 데이터를 이용하여 일정 기간 내에 이탈한 유저와 지속적으로 게임을 이용한 유저를 추출하고, 두 유저의 이탈 직전 활동 로그를 소셜 행위를 중심으로 분석하여 이탈한 유저에게서 나타나는 특성을 추출하였다. 이후 추출한 특성을 기반으로 하여 활동 로그를 정제해 피처를 정의하고, 유저 이탈 예측 모델을 제시하였다. 본 논문에서는 관찰할 활동 로그 수와 분류 모델 학습에 사용되는 피처의 개수를 줄여 빠르게 유저 이탈을 예측할 수 있는 방법을 제시하였다.

키워드: 온라인 게임, 유저 상호작용, 유저 이탈 예측

Abstract This study analyzes the effect of social activities in massive multi-user online role-playing games (MMORPG) on user churn. In MMORPG, interactions between users often play an important role, and community systems that provide users' social behaviors have developed. Typically, there are parties for mission collaboration, such as "Quest," and guilds that provide more and longer social activities. In this study, we obtained the actual data from a major MMORPG. Using this data, we extract users who stop playing the game within a certain period of time, namely, churn users, and analyze their characteristics. Then we extract features from the extracted characteristics of churn users to present a user churn prediction model. In this paper, we propose a lightweight and quick prediction model by reducing the number of action logs to be observed and the number of features used in the classification model.

Keywords: online game, user interaction, user churn prediction

· 본 연구는 순천향대학교 학술연구비 지원으로 수행하였습니다.
(과제번호: 20170903)

[†] 학생회원 : 고려대학교 정보보호대학원
jymoon94@korea.ac.kr

^{††} 종신회원 : 고려대학교 정보보호대학원 교수
cenda@korea.ac.kr

^{†††} 정 회 원 : 순천향대학교 빅데이터공학과 교수
(Soonchunhyang Univ.)
jywoo@sch.ac.kr
(Corresponding author)

논문접수 : 2017년 10월 20일

(Received 20 October 2017)

논문수정 : 2018년 6월 26일
(Revised 26 June 2018)

심사완료 : 2018년 7월 9일
(Accepted 9 July 2018)

Copyright©2018 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회 컴퓨팅의 실제 논문지 제24권 제9호(2018. 9)

1. 서론

대규모 다중 사용자 온라인 롤플레이 게임(MMORPG)이란 온라인 게임의 한 장르로 각 유저는 게임 내의 가상 세계에서 한 직업을 맡아 게임을 진행한다. 유저가 몬스터를 잡거나 시나리오 상의 퀘스트를 진행하는 등 어떠한 목적을 달성하면 경험치, 아이템 등의 보상이 주어지고 일정 수준의 경험치가 모이면 레벨(level)이 오르는 방식으로 게임의 스토리가 진행된다. 이 때 직업별 능력치의 차이가 존재하기 때문에 다양한 직군의 협업이 중요한 요소로 작용하기도 한다. 또한 각 게임은 하나의 가상세계를 이루고 있어 그 안에서 유저 간 상호작용이 빈번하게 나타난다. MMORPG에는 일반적으로 길드, 클랜 등 친목 도모 목적의 커뮤니티 시스템이 존재하며 유저끼리의 아이템 매매, 교환 등이 이루어지기도 하는 등 다양한 형태의 상호작용이 이루어진다. 이러한 유저 간 상호 작용을 게임 내 가상 세계에서의 소셜 행위라고 본다면, 유저 간의 협업과 친목 도모가 게임 스토리 진행에 큰 비중을 차지하는 MMORPG의 경우 게임 이용 중 다양한 유저들의 소셜 행위가 동반되는 장르라고 볼 수 있다. 특히 파티(party)와 같은 짧은 협업 관계가 아닌 비교적 지속 기간이 긴 길드의 경우 실제 레벨을 올리는 데에 도움이 되지 않다 하더라도[1] 유저들로 하여금 게임을 지속하게 한다[2].

가상 세계 내 사회적 공간에서의 상호작용이 사회적 실재감이 긍정적인 피드백을 한다는 연구 결과도 있을 정도로 사람들이 가상 세계 내에서 사회적 실재감을 찾지만[3], MMORPG 내에는 혼자서 이용할 수 있는 많은 콘텐츠가 있으며 혼자 게임을 이용하더라도 최고 레벨까지 도달할 수 있기 때문에[1] 혼자서 게임을 즐기는 유저들도 당연히 존재한다. 이들 중에는 성격적, 시간적 이유 등으로 인해 처음부터 소셜 행위를 거의 동반하지 않고 게임을 이용하는 유저들도 존재하고, 모종의 이유로 게임 이용 중 소셜 행위에 대한 부정적인 이미지를 가지게 된 유저들도 존재할 것이다. 주로 “파티 플레이”라고 부르는 일반적인 단기간 협업의 경우 미션이 종료되면 경험치와 아이템 등을 나누어 가지는데 이러한 과정에서 소셜 활동에서의 이탈이 발생할 가능성은 충분하다. 또한 길드 내에서 레벨의 차이가 나타나 길드의 응집력을 떨어뜨리기도 하는데[2] 이렇게 소셜 행위에서 나타나는 다양한 유형의 이탈 현상은 게임 자체의 이탈 현상에 영향을 미칠 수 있다. 따라서 본 연구에서는 이탈한 유저와 이탈하지 않은 유저가 각각 일정 기간 동안 겪은 소셜 행위를 분석하여 이탈한 유저들이 가지는 특성을 추출하고, 추출한 특성을 바탕으로 한 유저 이탈 예측 모델을 제시하였다.

본 연구는 정액제 형식으로 서비스되는 NCSOFT사의 AION의 실제 로그 데이터를 이용해 진행되었다.

2. 관련 연구

2.1 유저 간 상호작용 및 소셜 행위에 대한 연구

Helena Cole 등은 이전의 연구에서 게이머들은 사회적으로 활동적이지 않다고 추정했으나 실제 설문문을 통해 많은 게임 이용자들이 온라인 게임 내에서 좋은 친구(good friend)를 사귀어 본 적이 있다고 답하는 등 MMORPG 유저들은 매우 활발하게 사회적으로 상호작용한다는 조사 결과를 통계적으로 제시하였다[4].

유저들은 기본적으로 레벨을 올리는 것에서 성취감을 얻으며 게임을 이용하지만, 협업 활동과 소셜 네트워킹 등 또한 게임 이용에서 중요한 역할을 차지하고 있다. Nicolas Ducheneaut 등의 연구에 따르면 퀘스트 등을 수행하기 위한 짧은 기간의 협업(파티 플레이 등)보다 길드 등 긴 기간의 소셜 커뮤니티가 게임 이용에 중요한 영향을 미치며, 길드는 게임을 더 자주, 더 정기적으로 이용하는 것을 가능케 한다[2].

이러한 길드 활동과 유저들의 레벨에 대해 분석한 Nathaniel Poor의 연구에서는 길드에 소속되어 있어도 플레이어의 레벨을 올리는 데에는 도움이 되지 않으며, 길드에 소속되어 있는 것은 우리가 사회적 존재(social being)이기 때문에 게임 내에서의 채팅 등으로 사회적인 상호작용을 이어가기 위해서라는 분석 결과를 제시하였다[1].

Micheal P. McCreedy 등의 연구에서는 가상 세계 안에서 사회적 공간 상호작용(socio-spatial interactivity)과 사회적 실재감(social presence) 사이에는 긍정적인 피드백 작용이 존재함을 보이는 결과를 제시하였다[3].

2.2 유저 이탈에 관한 연구

Kunwoo Park 등은 온라인 게임 내 유저가 최고 레벨을 달성하면 행동 패턴이 어떻게 변하는지를 분석하여 최고 레벨 유저의 행동 특징에 따른 이탈 여부 예측을 진행하였다. 사회적 상호작용 동기를 가지고 게임을 이용하는 유저들이 최고 레벨 달성 이후에도 이탈하지 않고 게임을 이용하는 경향을 보였다[5].

하지만 이러한 사회적 상호작용이 항상 유저들이 게임을 지속적으로 이용하는 데에 긍정적인 영향을 주는 것은 아니다. Nicolas Ducheneaut의 연구에 따르면 길드 내의 레벨의 차이가 길드의 응집력을 줄이고 유저들의 이탈 비율에 영향을 주기도 한다[2].

Hsiu-Yu Liao 등은 폭스 게임 플랫폼 내의 장기적인 이탈 관리와 유저의 지속적인 이용이 최고의 핵심 마케팅 전략이라고 하며 게임 Roomi 내의 가상 세계 데이터를 기반으로 화폐 소비, 활동 에너지, 소셜 분석을 결합하여 이탈 예측 기법을 제안하고, 평가하였다. 또한 제안한 예

측 기법이 금융 산업 등에 기존에 적용되었던 예측 기법에 비해 가상 세계에 적합하다는 평가 결과를 제시하였다[6].

Y. Richter 등은 유저 간의 상호 작용을 분석하여 이탈 예측의 정확성을 향상시키기 위해 소셜 지표를 이용하여 유저의 그룹을 식별하고 구성원의 이탈을 예측하는 새로운 프레임워크를 제시하였다. 또한 제시한 프레임워크를 평가하기 위해 실제 통신 사업자의 데이터를 기반으로 한 실험 결과를 제시하였다. 실험 결과에 따르면 제시한 프레임워크가 대규모 집단에서 유저의 이탈을 빠르고 정확하게 예측할 수 있음을 알 수 있다[7].

Seungwook Kim 등은 관측 기간과 이탈 예측 기간에 따라 이탈 예측 성능의 차이가 다양하게 나타난다는 연구 결과를 세 가지 게임의 실제 데이터를 통해 제시하였다. 이들의 연구 결과에 따르면 게임의 특성에 따라 관측 기간, 이탈 예측 기간을 설정하는 것은 이탈 예측의 중요한 요소이다[8].

JiHoon Jeon 등은 모바일 소셜 게임에서 유저의 인지 심리학적 피처를 추출하여 이탈을 예측하였다. 실험 결과, 인지 심리학적 피처를 이용한 이탈 예측 모델은 유저의 구매 정보 없이도 좋은 성능을 보였다[9].

Miloš Milošević 등은 모바일 소셜 게임 내의 이탈을 예측하고, 이탈할 것으로 예상되는 유저에게 적절한 푸시 알림을 보내도록 하여 유저의 이탈을 28% 줄이는 것이 가능하다는 연구 결과를 제시하였다[10].

3. 데이터셋

3.1 AION

분석 대상이 되는 AION은 NCSOFT사에서 제공하는 서비스로 2008년 11월 25일부터 정식 서비스가 시작되어 현재까지 한국, 북미, 일본, 대만, 유럽, 중국, 러시아 등에서 서비스 되고 있는 MMORPG 장르의 게임이다. AION은 출시 이후 현재까지 19800원의 서비스 이용료를 지불하면 30일 동안 300시간을 이용할 수 있는 정액제 형식의 수익 모델을 취하고 있다[11].

3.2 데이터셋

유저가 게임 내에서 취하는 행동들은 시간 정보와 함께 로그로 저장된다. 본 연구에서는 NCSOFT사로부터 제공받은 AION의 로그 데이터를 분석하였으며, 분석 기간은 2015년 12월 30일부터 2016년 6월 15일까지의 24주이다. 해당 기간 동안의 총 유저 수는 64564명, 캐릭터 수는 119306개이며, AION의 제 1서버인 시엘 서버의 데이터를 사용하였다.

4. 방법론

4.1 이탈 유저 및 지속 유저 선정

유저 이탈을 분석하기 위해 먼저 이탈을 정의하였다.

N주 간 접속을 하지 않은 유저는 이탈한 유저로 볼 때, N값을 정하기 위해 실험을 진행하였다. 이 때 i주차에는 접속 로그가 있지만 i+1주차부터 접속 로그가 없는 경우 i+1주차에 이탈한 유저로 판단하였다. 유저 행동 분석 기간 설정은 데이터셋의 제한 때문에 유저의 이탈 직전 8주간의 데이터를 분석할 데이터로 정했다. AION은 30일 단위로 결제하는 형식의 게임으로 이는 이용권 결제 단위의 약 2배에 해당하므로 이탈 징후를 찾고 대응하기에 충분한 기간이다.

분석 기간 중 9주차에 이탈한 유저부터 13주차에 이탈한 유저를 대상으로 N값을 1부터 12까지 변화시켜 가며 이탈 유저 수의 합을 기록하였다. 또한 해당 기간에 이탈한 유저들 중 분석 기간 내에 되돌아온 유저 수를 기록하였다. $C_c(N)$ 을 N주 동안 접속기록이 없는 유저 수로 정의하고, $C_r(N)$ 을 N주 이후에 다시 접속 기록이 발생한 유저 수로 정의하였다. $C(N)$ 을 N주차에 이탈한 유저 중 분석 기간 내에 되돌아온 유저로 정의하여 Fig. 1의 그래프를 도출하였다.

$$C(N) = \frac{C_r(N)}{C_c(N)}$$

Fig. 1의 그래프에서 뚜렷하게 선형할 만한 N값이 존재하지 않으나 N이 9에서 10으로 증가할 때 $C(N)$ 의 감소폭이 큰 편이며 처음으로 0.2 이하로 감소하였다. 따라서 10주간 접속 로그가 없는 경우 이후의 게임 이용 여부와 관계없이 이탈한 유저로 판단하였다. '이탈 유저'는 총 6894개의 계정이 추출되었으며, 해당 계정들의 이탈 직전 8주간의 데이터를 분석하였다.

본 연구에서 이탈 유저의 행동 양상 분석 시 이탈 직전 8주간의 데이터를 분석하였기 때문에 i+1주차부터 i+8주차까지 8주간 이탈하지 않은 유저는 i주차에서 이탈의 징후를 보이지 않는다고 판단하였다. 따라서 i주차부터 적어도 8주간 이탈하지 않은 유저를 i주차에서 이탈의 징후를 보이지 않고 게임을 지속하고 있는 '지속 유저'로 판단하였다. i주차에서의 지속 유저는 i주차에서의 이탈 유저가 어떤 행동 양상을 보이는지 분석할 때에 비교 대상으로 활용하기 위해 추출하였다. 지속 유저는 이탈 유저와 함께 분석할 때 업데이트 및 이벤트 등 특정 주에 일어난 사건으로부터 받는 영향을 줄이기 위해 9주차부터 13주차까지 해당 주에 이탈한 유저와 같은 수만큼 중복되지 않도록 무작위로 추출하였다. 따라서 이탈 유저와 마찬가지로 총 6894개의 계정이 추출되었으며, 계정들이 지속 유저로 판단된 주를 기준으로 직전 8주간의 데이터를 분석하였다. 본 논문에서는 각 계정이 이탈한 주를 '이탈 시점', 지속 유저로 판단된 주를 '지속 시점'으로 표현하였다.

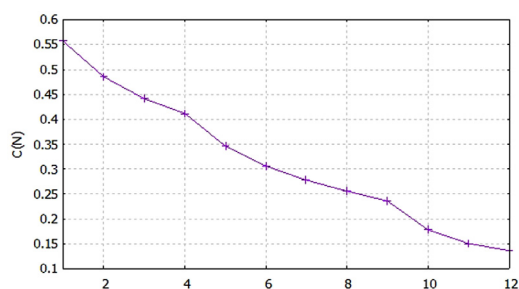


그림 1 N주 이후 되돌아온 유저 수/N주 동안 접속 기록이 없는 유저 수

Fig. 1 Number of churners who return after N weeks later/number of churners defined in terms of N weeks

4.2 이탈 유저 및 지속 유저 행위 분석

분석할 데이터를 추출한 후 유저의 소셜 행위를 파티 활동, 거래, 길드 활동, 친구 등록 네 분야로 나누어 이탈 유저와 지속 유저가 각 분야에서 어떤 특성을 보이는지 분석하였다.

먼저 파티 활동의 경우 퀘스트와 파티 활동에서의 차이를 분석하기 위해 이탈 유저와 지속 유저 각각에 대하여 파티에 참가하는 활동 로그, 파티에서 탈퇴하는 활동 로그, 퀘스트를 시작하는 활동 로그, 퀘스트를 완료하는 활동 로그를 추출하였다. Fig. 2는 이탈 시점으로부터 직전 8주 동안 이탈 유저의 각 활동 로그의 개수를 주 단위로 합하여 나타낸 그래프이고, Fig. 3은 지속 시점으로부터 직전 8주 동안 지속 유저의 각 활동 로그의 발생 수를 마찬가지로 주 단위로 합하여 나타낸 그래프이다. x축은 이탈 시점 또는 지속 시점을 기준으로 하여 표현하였으며, 가장 왼쪽의 -8이 이탈 또는 지속 시점으로부터 8주 전을 나타낸다.

Fig. 2를 보면, 이탈 유저의 퀘스트 활동은 이탈 시점에 가까워질수록 증가하는 반면, 파티 활동은 감소하였다. 이탈 직전 8주 전반에 걸쳐 파티 활동을 하지 않고 퀘스트 활동 위주로 게임을 이용하며, 이탈 시점에 가까워질수록 파티 활동보다 퀘스트 활동을 선호하는 정도가 증가하였다. 반면 Fig. 3을 보면, 지속 유저의 경우 파티 활동은 약간 감소했다가 증가하며 퀘스트 활동은 조금씩 증가하였다. 지속 유저의 파티 활동량과 퀘스트 활동량의 관계는 시간이 지나더라도 큰 변화를 보이지 않았다. 또한 이탈 유저의 경우 8주 전반에서 퀘스트 활동량이 파티 활동량의 약 4배로 나타나지만 지속 유저의 경우 반대로 파티 활동량이 퀘스트 활동량의 약 3.5배로 나타나 큰 차이를 보였다.

다음은 길드 활동에서 고립되어 길드로부터 이탈하는 경우를 분석하기 위해 길드 탈퇴 활동 로그의 개수를

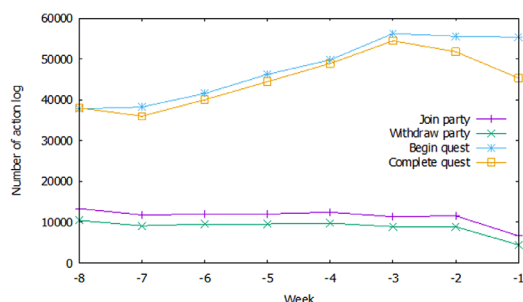


그림 2 이탈 유저의 파티, 퀘스트 로그 수

Fig. 2 Number of party and quest-action log of churning

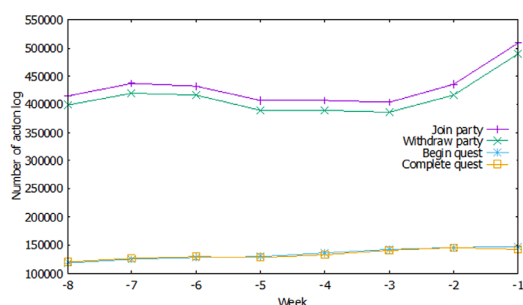


그림 3 지속 유저의 파티, 퀘스트 로그 수

Fig. 3 Number of party and quest-action log of non-churning

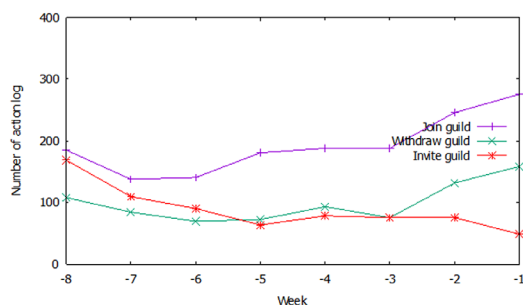


그림 4 이탈 유저의 길드 로그 수

Fig. 4 Number of guild-action log of churning

지속 유저와 이탈 유저로 나누어 측정하였다. Fig. 4는 이탈 시점으로부터 직전 8주 동안 이탈 유저의 길드 가입, 탈퇴, 초대 활동 로그의 개수를, Fig. 5는 지속 시점으로부터 직전 8주 동안 지속 유저의 길드 가입, 탈퇴, 초대 활동 로그의 개수를 각각 주 단위로 합하여 나타낸 그래프이다.

Fig. 4와 Fig. 5 모두에서 길드 가입과 탈퇴 활동이 조금씩 감소하였다가 증가하여 두 그래프의 추세 변화가 큰 차이를 보이지 않기 때문에 길드 가입과 탈퇴가 유저 이탈에 큰 영향을 미치지 않는다는 사실을 알 수

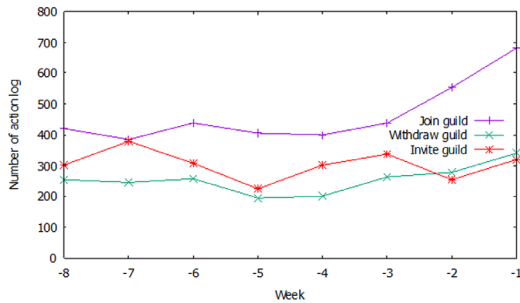


그림 5 지속 유저의 길드 로그 수
Fig. 5 Number of guild-action log of non-churner

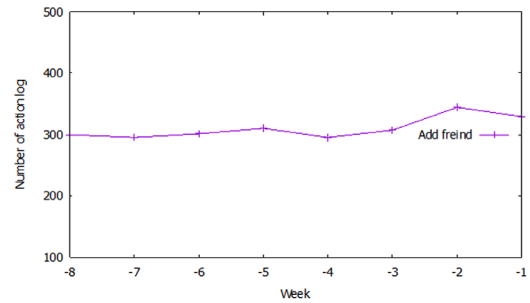


그림 8 이탈 유저의 친구 추가 로그 수
Fig. 8 Number of friend-addition action log of churner

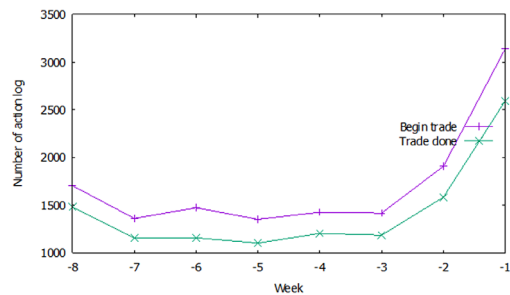


그림 6 이탈 유저의 거래 로그 수
Fig. 6 Number of trade-action log of churner

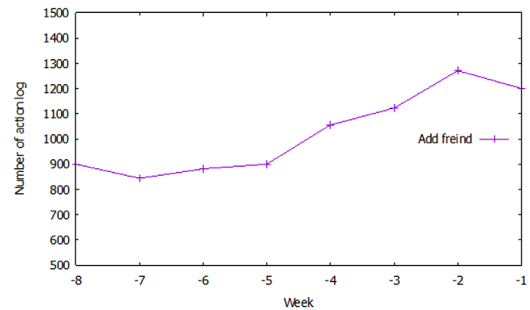


그림 9 지속 유저의 친구 추가 로그 수
Fig. 9 Number of friend addition-action log of non-churner

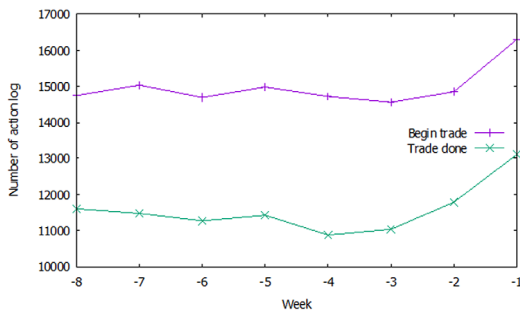


그림 7 지속 유저의 거래 로그 수
Fig. 7 Number of trade-action log of non-churner

있다. 지속 유저의 길드 탈퇴 기록이 이탈 유저의 길드 탈퇴 기록의 약 2배로 나타나는데, 이는 지속 유저가 일반 유저에 비해 게임 전반에 걸쳐 활발한 활동을 보이기 때문인 것으로 사료된다. 또한 Fig. 4의 길드 가입 활동 로그 수를 보면 이탈 직전에도 새로운 길드에 가입하는 유저가 많다는 사실을 알 수 있다. 반면 길드 초대 기록은 지속 유저와 이탈 유저에서 각각 다른 양상을 보이며, 이탈 시점에 가까워질수록 다른 유저를 길드로 초대하는 유저가 감소함을 볼 수 있다. 다만 길드 활동의 경우 파티 활동에 비해 장기간에 걸쳐 이루어지는

소셜 행위이기 때문에 변화가 적어 8주 동안 데이터가 충분히 모이지 못한다는 한계점이 있다.

다음은 이탈 유저와 지속 유저의 거래 활동 로그를 분석하였다. Fig. 6은 이탈 시점으로부터 직전 8주 동안 이탈 유저의 거래 시작과 완료 활동 로그의 개수를, Fig. 7은 지속 시점으로부터 직전 8주 동안 지속 유저의 거래 시작과 완료 활동 로그의 개수를 각각 주 단위로 합하여 나타낸 그래프이다.

이탈 유저의 경우 전체 기간 동안 대체로 거래 활동 로그가 적지만 이탈 시점으로부터 2주 전부터 급격하게 증가했고, 지속 유저의 경우 전체 기간에 걸쳐 고른 거래량을 보였다.

마지막으로 이탈 유저와 지속 유저의 8주간 친구 등록 횟수를 분석하였다. Fig. 8은 이탈 시점으로부터 직전 8주 동안 이탈 유저의 친구 등록 활동 로그의 개수를, Fig. 9는 지속 시점으로부터 직전 8주 동안 지속 유저의 친구 등록 활동 로그의 개수를 각각 주 단위로 합하여 나타낸 그래프이다.

지속 유저의 친구 등록 활동 로그가 이탈 유저의 친구 등록 활동 로그의 약 3배인 것을 알 수 있지만, 이탈 유저의 경우 지속 유저에 비해 전체 활동 로그가 적기 때문에 지속 유저가 이탈 유저에 비해 사회적이라고 판

단하기는 어렵다. 또한 이탈 직전에도 친구 등록을 하는 유저가 꾸준히 나타나는 것을 알 수 있다.

4.3 피쳐 선정

위에서 분석한 내용을 바탕으로 유의미한 로그를 선정하고, 선정한 로그를 정제해 유저 이탈 예측에 사용될 피쳐를 추출한다. 많은 종류의 활동 로그를 사용하면 정확도 및 정밀도가 증가할 가능성이 높지만, 필요한 활동 로그를 추출해 데이터를 정제하는 비용이 증가한다. 분류 시 많은 피쳐를 사용하는 방법도 정확도와 정밀도를 높이는 데에 유리하지만 분류 모델 설계 및 분류의 비용이 증가한다. 본 논문에서는 분류에 쓰이는 활동 로그와 피쳐 개수를 줄여 적은 비용으로 준수한 정확도와 정밀도를 보이는 분류 모델을 제시하였다.

첫 번째 피쳐로 고립 정도를 측정하였다. 이탈 유저는 파티 활동에 비해 퀘스트 활동을 많이 진행하고, 지속 유저는 퀘스트 활동에 비해 파티 활동을 많이 진행하는 경향이 있기 때문에 파티 활동과 퀘스트 활동 횟수 비율은 분류 시 유의미하게 작용할 것으로 판단하였다. 따라서 파티 가입 로그 수에 대한 퀘스트 시작 로그 수의 비율을 고립 정도로 정의하였다.

두 번째 피쳐로는 퀘스트 활동량 변화 정도를 측정하였다. 지속 유저는 지속 시점에 가까워질 때 퀘스트 활동량이 서서히 증가하는 데에 비해 이탈 유저는 이탈 시점에 가까워질수록 퀘스트 활동량이 눈에 띄게 증가하였다. 따라서 이탈 또는 지속 시점을 기준으로 8주 전부터 5주 전까지의 퀘스트 시작 로그 수에 대한 4주 전부터 1주 전까지의 퀘스트 시작 로그 수의 비율을 퀘스트 활동량 변화 정도로 정의하여 피쳐로 활용하였다.

세 번째 피쳐로는 파티 활동량 변화 정도를 측정하였다. 지속 유저의 파티 활동량은 퀘스트 활동량과 마찬가지로 서서히 증가하지만, 이탈 유저의 경우 이탈 시점에 가까워질수록 파티 활동량이 감소하였다. 따라서 파티 활동량 변화 정도를 이탈 또는 지속 시점을 기준으로 8주 전부터 5주 전까지의 파티 가입 로그 수에 대한 4주 전부터 1주 전까지의 파티 가입 로그 수의 비율로 정의하여 피쳐로 활용하였다.

네 번째 피쳐로는 파티, 거래, 퀘스트 활동량 감소 정도를 측정하였다. 파티, 거래, 퀘스트 활동 정도의 경우 이탈 유저와 지속 유저 모두에게서 전반적으로 증가하는 추세를 보여 단독으로는 활용도가 떨어질 수 있지만 다른 피쳐들과 함께 활용한다면 이탈 또는 지속 유저 유형을 파악하는 데에 도움이 될 것이라고 판단하였다. 파티, 거래, 퀘스트 활동량 감소 정도는 이탈 또는 지속 시점을 기준으로 1주 전부터 4주 전까지의 파티 가입, 거래 시작, 퀘스트 시작 로그 수에 대한 5주 전부터 8주 전까지의 파티 가입, 거래 시작, 퀘스트 시작 로그 수의

비율로 정의하였다.

다섯 번째 피쳐로는 파티, 거래, 퀘스트 활동량을 측정하였다. 전체 활동량을 측정하는 것은 비용적으로 비효율적이라고 판단하여 본 논문에서 주요 활동으로 간주한 세 가지 활동의 활동량으로 전체 활동량을 대신하였다. 파티 가입, 거래 시작, 퀘스트 시작 로그 수의 합으로 정의하였다.

5. 결과 분석 및 평가

5.1 분류 대상 추출

이탈 유저의 경우 활동 로그 개수가 지속 유저에 비해 현저히 낮은 경우가 많은데 이들은 활동을 거의 하지 않고 있는 유저이기 때문에 해당 유저들의 데이터는 결과 분석에 활용하지 않았다. 따라서 이탈 유저, 지속 유저 모두에 대해 활동 로그 개수가 일정 수 이상인 유저들을 선정하였다. 또한 본 연구는 퀘스트 활동과 파티 활동을 중점적으로 다루기 때문에 퀘스트 활동과 파티 활동이 있는 유저들의 데이터를 결과 분석에 활용하였다.

실제 이탈을 예측할 때에는 이탈 징후에 대응하는 시간이 필요하므로 이탈 시점보다 3주 정도 일찍 이탈을 예측할 수 있어야 한다. 따라서 i-8주차부터 i-1주차까지의 로그를 분석하여 해당 기간 직후(1주 후)인 i주차 뿐만 아니라 2, 3, 4주 후인 i+1, i+2, i+3주차에 이탈할 유저를 예측해보았다. 분석 기간으로부터 1주, 2주, 3주, 4주 후의 이탈 예측 평가를 위해 추출한 이탈 유저 및 지속 유저의 수는 각각 Table 1과 같다.

표 1 분류에 사용된 데이터 셋

Table 1 Dataset for classification

	Churner	Non churner	Sum
1 week later	315	1825	2140
2 weeks later	318	1427	1745
3 weeks later	324	1554	1878
4 weeks later	296	1535	1831

5.2 분석 및 평가

오픈 소스 기계학습 툴인 weka를 사용하여 분류 모델을 학습시키고 평가하였다. 분석할 유저 수가 적기 때문에 학습과 평가에 서로 다른 데이터 셋을 사용하지 않고 k-몹을 교차 검증법(k-fold cross-validation)을 평가 방법으로 선택하였다. 데이터 셋을 임의로 10개의 균등한 부분 집합으로 나눈 후 n번째 집합을 제외한 9개의 집합으로 분류 모델을 학습시키고 n번째 데이터로 모델을 평가하는 과정을 n이 1에서 10이 될 때까지 반복한 후 10개의 평가 결과의 평균을 총 평가 결과로 선택하는 방법이다[12].

표 2 각 분류 알고리즘의 TP/FP 비율(1주 후)
Table 2 TP/FP rate of each classifier(1 week later)

		Churn	Non churn
Random Forest	TP	0.679	0.984
	FP	0.016	0.321
BayesNet	TP	0.546	0.836
	FP	0.164	0.454
Random Committee	TP	0.727	0.967
	FP	0.033	0.273
JRip	TP	0.698	0.974
	FP	0.026	0.302
via Regression	TP	0.622	0.975
	FP	0.025	0.378

표 3 각 분류 알고리즘의 정확도와 정밀도(1주 후)
Table 3 Accuracy and precision of each classifier(1 week later)

	Accuracy	Precision
Random Forest	0.939	0.936
BayesNet	0.793	0.833
Random Committee	0.932	0.930
JRip	0.933	0.930
via Regression	0.923	0.919

먼저 i-8부터 i-1까지 8주차의 로그에서 추출한 피처를 사용하여 1주 후인 i주차의 이탈 예측을 진행하였다. 분류 알고리즘은 Random Forest, BayesNet, Random Committee, JRip, Classification via Regression 등 총 5개를 사용하였으며, 평가 결과는 Table 2, Table 3과 같다. Table 2는 각 알고리즘의 이탈 유저, 지속 유저 각각에 대한 TP(true positive), FP(false positive)를 나타낸다. 이탈 유저를 예측하는 데에 있어서 게임 이용을 지속한 유저를 지속 유저로 판단하는 것보다 실제 이탈한 유저를 이탈 유저로 판단하는 것이 더 중요하므로 Random Committee 알고리즘을 사용하는 것이 좋은 분류 모델을 생성한다고 볼 수 있다.

Table 3은 각 알고리즘의 정확도와 정밀도를 나타내며, Random Forest 알고리즘이 가장 높은 정확도와 정밀도를 보였다.

이후 같은 방법으로 i-8부터 i-1까지 8주차의 로그에서 추출한 피처를 사용하여 각각 2주 후, 3주 후, 4주 후에 유저가 이탈할 것인지 예측하는 실험을 진행하였다. 분류 알고리즘은 위 실험과 같이 총 5개를 사용하였으며, 각각 이탈 유저 및 지속 유저에 대한 TP/FP와 정확도, 정밀도를 구하였다.

2주 후의 이탈 예측 평가 결과는 Table 4, Table 5와 같다. 마찬가지로 Random Committee 알고리즘을 사용하는 것이 가장 좋은 분류 모델을 생성한다.

표 4 각 분류 알고리즘의 TP/FP 비율(2주 후)
Table 4 TP/FP rate of each classifier(2 weeks later)

		Churn	Non churn
Random Forest	TP	0.730	0.975
	FP	0.025	0.270
BayesNet	TP	0.626	0.792
	FP	0.208	0.374
Random Committee	TP	0.701	0.951
	FP	0.049	0.299
JRip	TP	0.692	0.971
	FP	0.029	0.257
Classification via Regression	TP	0.657	0.969
	FP	0.031	0.343

표 5 각 분류 알고리즘의 정확도와 정밀도(2주 후)
Table 5 Accuracy and precision of each classifier(2 weeks later)

	Accuracy	Precision
Random Forest	0.931	0.929
BayesNet	0.762	0.813
Random Committee	0.905	0.903
JRip	0.920	0.917
Classification via Regression	0.912	0.909

표 6 각 분류 알고리즘의 TP/FP 비율(3주 후)
Table 6 TP/FP rate of each classifier(3 weeks later)

		Churn	Non churn
Random Forest	TP	0.698	0.973
	FP	0.027	0.302
BayesNet	TP	0.642	0.802
	FP	0.198	0.358
Random Committee	TP	0.747	0.951
	FP	0.049	0.253
JRip	TP	0.747	0.924
	FP	0.039	0.253
Classification via Regression	TP	0.614	0.974
	FP	0.026	0.386

표 7 각 분류 알고리즘의 정확도와 정밀도(3주 후)
Table 7 Accuracy and precision of each classifier(3 weeks later)

	Accuracy	Precision
Random Forest	0.925	0.923
BayesNet	0.774	0.827
Random Committee	0.916	0.915
JRip	0.924	0.922
Classification via Regression	0.912	0.907

3주 후의 이탈 예측 평가는 Table 6, Table 7과 같다. 3주 후의 경우 1주 후보다 이탈 유저의 예측 확률이

표 8 각 분류 알고리즘의 TP/FP 비율(4주 후)
Table 8 TP/FP rate of each classifier(4 weeks later)

		Churn	Non churn
Random Forest	TP	0.730	0.969
	FP	0.031	0.270
BayesNet	TP	0.676	0.814
	FP	0.186	0.324
Random Committee	TP	0.764	0.950
	FP	0.050	0.236
JRip	TP	0.743	0.958
	FP	0.042	0.257
via Regression	TP	0.628	0.978
	FP	0.022	0.372

표 9 각 분류 알고리즘의 정확도와 정밀도(4주 후)
Table 9 Accuracy and precision of each classifier(4 weeks later)

	Accuracy	Precision
Random Forest	0.931	0.928
BayesNet	0.791	0.845
Random Committee	0.920	0.921
JRip	0.924	0.922
via Regression	0.921	0.918

표 10 기존 연구와의 피쳐 수 및 정밀도 비교
Table 10 Comparison with previous studies

	Number of Features	Precision
Early Churn Prediction	12	0.76
Cognitive Psychological	7	0.923
Proposed	5	0.928

올라가는 것을 볼 수 있다. 4주 후의 이탈 예측 평가는 Table 8, Table 9와 같으며, 1주차에 비해 3, 4주차에서 지속 유저의 예측 확률이 줄어든 반면 이탈 유저는 더 잘 선별해내는 것을 알 수 있다. 이는 유저가 이탈할 것처럼 행동하더라도 실제 이탈까지는 어느 정도 시간이 걸린다고 해석할 수 있다. 또한 실험 전반적으로 다른 세 분류 알고리즘에 비해 BayesNet, Classification via Regression 두 알고리즘의 성능이 낮은 것을 볼 수 있다.

Table 10은 제안한 이탈 예측 모델과 기존 연구의 이탈 예측 모델의 성능 및 비용을 비교한 표이다. 비교할 기존 연구는 빠른 이탈 예측을 목표로 하는 연구[10]와, 적은 수의 피쳐로 이탈을 예측하는 연구[9]로 선정하였다. Miloš Milošević 등이 제안한 방법은 early churn prediction으로[10], JiHoon Jeon 등이 제안한 방법은 cognitive psychological로 표시하였다[9]. Miloš Milošević 등은 총 12개의 피쳐를 사용하였으며, JiHoon Jeon 등은 96종의 로그를 사용하여 생성한 7개의 피쳐를 사용하였다. 전자에서는 여러 분류 알고리즘을 사용하였으나

정밀도가 가장 높은 분류 알고리즘은 Random Forest이며, 후자에서는 Random Forest만을 사용하였으므로 Random Forest를 사용한 성능을 비교하였다. 두 논문에서 공통적으로 precision을 기재하였으므로 성능은 정밀도로 나타내었으며, 제안한 이탈 예측 모델의 4주 후의 이탈 예측 평가와 비교하였다. 제안한 이탈 예측 모델이 적은 수의 피쳐로 높은 정밀도를 보이는 것을 확인할 수 있다.

6. 결론

본 연구에서는 실제 게임 데이터 중 한 서버의 24주 동안의 데이터를 분석에 활용하였다. 유저들은 이탈 전에 게임을 많이 이용하지 않는 경향을 보이기 때문에 전체 활동 로그 수가 매우 적어서 이탈 유저의 특성을 추출하는 데에 어려움이 있었다. 분석 데이터의 범위나 기간을 늘려서 연구에 활용할 수 있다면 보다 많은 특성을 추출할 수 있을 것이다. 또한 Table. 2를 보면 본 연구에서 제시한 유저 이탈 예측 모델은 지속 유저를 예측할 확률은 매우 높지만 이탈 유저의 예측 확률은 다소 떨어진다는 한계점이 존재한다. 따라서 사용하는 로그와 피쳐 수를 조금 증가시키더라도 이탈 유저 예측 확률을 높이도록 해야 한다.

MMORPG 게임 내에서 유저 간 상호작용 분석을 통해 유저들이 이탈 전에 보이는 소셜 행위를 추출하고, 지속적으로 게임을 이용하는 유저에 비해 이탈 유저가 어떤 차이를 보이는지를 분석하였다. 이탈하기 직전의 유저들에게서 소셜 행위 활동 감소를 볼 수 있을 것으로 예상하였는데, 파티 가입 및 탈퇴 로그는 감소하였지만 길드 가입, 탈퇴와 친구 등록 활동을 보이는 유저의 수가 이탈 직전에도 꾸준히 나타난 것을 확인할 수 있었다. 분석한 내용을 기반으로 하여 이탈 유저와 지속 유저에게서 차이를 보일만한 로그를 추출하고, 피쳐를 생성해 유저 이탈 예측 모델을 제시하였다. 본 연구에서 제시한 유저 이탈 예측 모델은 활동 로그 중 3가지의 로그만 사용하고, 분류 시 사용되는 피쳐 또한 5개로 적은 비용으로 빠르게 이탈 유저를 판단하는 것을 가능케 한다. 따라서 실제 게임에 적용한다면 단독으로 유저 이탈 예측 모델로써 활용되는 것이 아니라 다른 알고리즘에 피쳐를 통합하여 사용하거나 빠른 판단을 통한 빠른 피드백이 필요할 때 사용할 수 있다. 세 가지의 타입의 로그만을 관찰하여 4주 후까지 유저의 이탈 여부를 90%이상의 정확도 및 정밀도로 판별하기 때문에 다른 알고리즘에 비해 빠르고 간편하게 사용할 수 있다는 장점이 있다.

유저의 이탈을 빠르게 예측한 후에는 이탈을 막기 위해 이벤트 시행, 아이템 지급 등 적절한 조치를 취하는

것이 중요하다. 향후 업데이트, 이벤트 전후 발생하는 유저 이탈을 분석하여 유저의 이탈에서 발생하는 손실과 이탈을 막기 위해 게임사가 감수하는 비용에 대한 연구를 진행할 계획이다.

References

- [1] Nathaniel Poor, "What MMO Communities Don't Do: A Longitudinal Study of Guilds and Character Leveling, Or Not," *Ninth International AAAI Conference on Web and Social Media*, pp. 678-681, 2015.
- [2] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, Robert J. Moore, "Alone together?: exploring the social dynamics of massively multiplayer online games," *CHI '06 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 407-416, Apr. 2006.
- [3] Michael P. McCreery, David B. Vallett, Cynthia Clark, "Social interaction in a virtual environment: Examining socio-spatial interactivity and social presence using behavioral analytics," *Computers in Human Behavior*, Vol. 51, Part A, pp. 203-206, May 2015.
- [4] Helena Cole and Mark D. Griffiths, "Social Interactions in Massively Multiplayer Online Role-Playing Gamers," *Cyberpsychology & Behavior*, Vol. 10, No. 4, pp. 575-583, 2007.
- [5] Kunwoo Park and Meeyoung cha, "Churn Analysis of Maximum Level Users in Online Games," *Journal of KIISE*, Vol. 44, No. 3, pp. 314-322, Mar. 2017.
- [6] Hsiu-Yu Liao, Kuan-Yu Chen, Duen-Ren Liu, Yi-Ling Chiu, "Customer Churn Prediction in Virtual Worlds," *2015 IIAI 4th International Congress on Advanced Applied Informatics*, pp. 116-120, 2015.
- [7] Y. Richter, E. Yom-Tov, and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups," *SIAM International Conference on Data Mining*, pp. 732-741, 2010.
- [8] Seungwook Kim, Daeyoung Choi, Eunjung Lee, Wonjong Rhee, "Churn prediction of mobile and online casual games using play log data," *PLoS ONE*, Vol. 12, No. 7, 2017.
- [9] JiHoon Jeon, DuMim Yoon, SeongIl Yang and KyungJoong Kim, "Extracting Gamers' Cognitive Psychological Features and Improving Performance of Churn Prediction from Mobile Games," *IEEE Conference on Computational Intelligence and Games 2017*, pp. 150-153, Aug. 2017.
- [10] Miloš Milošević, Nenad Živić and Igor Andjelković, "Early churn prediction with personalized targeting in mobile social games," *Expert Systems with Applications*, Vol. 83, pp. 326-332, Oct. 2017.
- [11] ncsoft AION homepage, <http://aion.plaync.com/>
- [12] P. Refaeilzadeh, L. Tang, H. Liu, "Cross-Validation," [Online]. Available: <http://www.public.asu.edu/~ltang9/papers/ency-cross-validation.pdf>



문 주 연

2017년 고려대학교 컴퓨터학과 졸업 (학사). 2017년~현재 고려대학교 정보보호대학원 석사과정. 관심분야는 온라인 게임 보안, 유저 행위 분석, 이상 탐지 등

김 휘 강

정보과학회 컴퓨팅의 실제 논문지 제 24 권 제 5 호 참조

우 지 영

정보과학회 컴퓨팅의 실제 논문지 제 24 권 제 5 호 참조