

게이머 행동 예측을 위한 다양한 모바일 게임 로그 기반의 일반화된 특징 추출 방법

노재윤^o 전지훈 배청목 김경중*

세종대학교 컴퓨터공학과

jaeyoon937@naver.com, rhd00513@gmail.com, cjdahrl@gmail.com,

kimkj@sejong.ac.kr

Generalized Feature Extraction Method Based on Various Mobile Game Logs for Predicting Gamer Behavior

JaeYoon Noh^o JiHoon Jeon CheongMok Bae Kyung-Joong Kim*

Department of Computer Science and Engineering, Sejong University

요 약

현재 게임 시장은 모바일 게임 중심으로 나아가고 있다. 모바일 게임은 휴대가 가능하여 시간적, 공간적 제약을 받지 않아 게이머가 모바일 게임으로의 진입과 이탈이 쉬운 특징을 가지고 있다. 이러한 상황에서 게임 회사들이 게이머의 이탈을 사전에 예측할 수 있다면 효율적인 게임 서비스를 게이머들에게 제공함으로써 수익을 극대화할 수 있다. 본 연구에서는 모바일 RPG 게임인 크레이지 드래곤과 소울 게이지의 게임 로그를 이용하여 6 가지의 일반적인 속성을 추출하였고 이를 사용하여 게이머의 이탈을 예측하였다.

1. 서 론

많은 게임들이 온라인으로 개발되고 있는 현재의 게임 시장에서 게이머의 행동 예측 및 모델링은 높은 중요도를 가진다. 게이머의 행동 예측은 게임의 생명력과 매출로 직결되는 중요한 요소이며, 게이머의 행동을 모델링하고 예측할 수 있다면, 실용적으로도 매우 의미가 있다. 본 연구에서는 게이머들이 이탈하기까지 여러 단계의 과도기를 가질 것으로 예상하였다. 게임에 대한 흥미가 올라가는 구간, 또 게임에 대한 흥미가 떨어지는 구간 직후의 게이머 행동 패턴을 분석하였다. 이러한 과도기들의 분석을 통해 게이머의 이탈을 예측하는 것이 게이머의 행동 예측에 있어 높은 성능을 나타낼 것으로 기대하였다. 또한, 게이머가 게임에서 이탈에 이르기까지의 기간 동안 나타나는 이탈의 징조는 게이머의 이탈 직전의 짧은 시간 동안의 행동 패턴에서 나타날 확률이 높다고 추정하였다. 본 연구에서는 이러한 유저들의 이탈을 예측하기 위해 사용한 게임 로그, 공통 속성, 이탈 정의, 데이터 전처리, 알고리즘 Random Forest, Decision Tree, Extra Tree에 대한 자세한 설명은 2, 3장에서 기술한다.

2. 배경

2.1 게임 로그 분석

본 연구에서는 모바일 RPG 게임 ‘소울 게이지’와 ‘크레이지 드래곤’의 로그를 분석하였다. 소울 게이지는 약 3개월(107일)동안 수집된 로그 데이터를 가지고 분석하였다. 부가적인 로그를 모두 포함하여 총 117 종류의 게임 로그를 이용하여 698,552,485개의 로그 데이터가 기록 되었다. 이를 보아 소울게이지는 비교적 짧은 기간임에도 불구하고 굉장히 많은 수의 로그 데이터가 기록 되었으며, 이는 게이머의 행동을 비교적 자세하게 기록한 것으로 추정된다. 전체 유저 수 또한 189,042명으로 매우 많은 게이머를 가지고 있는 매우 유용한 데이터이다. 크레이지 드래곤은 약 4개월(130일) 동안 수집된 로그데이터를 가지고 분석하였다. 부가적인 로그를 모두 포함하여 총 96 종류의 로그를 이용하여 295,107,602개의 로그가 기록 되었고 192,361명의 게이머가 해당 게임을 플레이 하였다. 하지만 소울 게이지에 비해 기간과 유저 수 대비 기록된 로그가 절반 이하이기 때문에 자세한 로그를 기록하지는 못했을 것으로 추정된다.

표 1. 기본적인 게임 로그 분석

	소울 게이지	크레이지 드래곤
기간	2017.05.25 ~ 2017.09.08	2016.03.21 ~ 2016.07.29
전체 게이머 수	189,042 명	192,361명
로그 타입 종류	117종	96종
로그 개수	698,552,485개	295,107,602개

2.2 게이머 이탈 정의

게이머의 이탈을 예측하기 위해, 게이머 이탈에 대한 정의가 필요하다. 본 연구에서 사용한 이탈의 정의는 게이머 각각의 개인별 시간을 기준으로 학습되는 데이터를 동일하게 가져가는 방법이다. 예를 들어, 각각의 게이머들의 첫 로그인으로부터 일주일간의 행동 데이터를 학습 데이터로 사용하고, 그 이후의 이탈 여부를 판별할 수 있다. 이러한 방법은 행동패턴을 분석하기 용이하다는 장점이 있다.

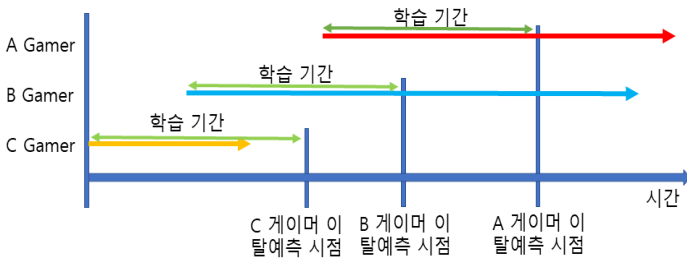


그림 1. 개인별 시간 기준 게이머 이탈 예측 모델 예시

게이머의 특정구간만을 예측하는 것이 아닌, 게이머가 이탈에 이르기까지의 전체 구간을 학습할 수 있도록 슬라이딩 윈도우 방식으로 게이머의 학습 데이터를 이동하면서 게이머의 이탈을 예측한다.

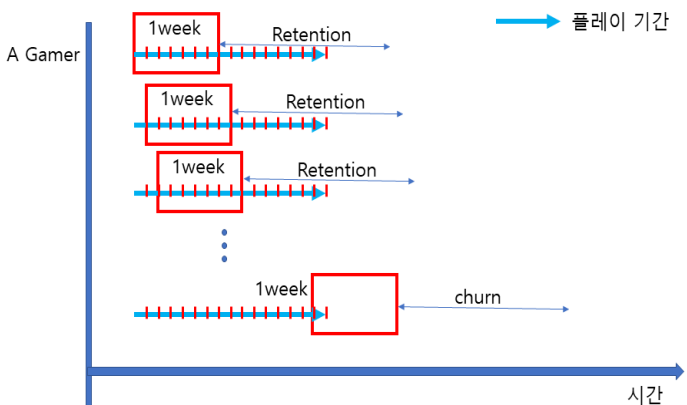


그림 2. 학습 윈도우의 이동 예시

이와 같은 방법을 게이머의 행동이 존재하지 않는 구간까지 진행하여 한 명의 게이머로부터 여러 개의 데이터를 얻는 부스팅 효과를 얻을 수 있으며 버려지는

데이터를 극적으로 줄일 수 있다.

2.3 일반적 속성

게이머의 행동을 예측하기 위해서는 게이머의 행동 데이터가 필요하다. 이러한 행동 데이터는 게임 속 환경과 상호작용을 통하여 게임 로그로 기록되기 때문에, 그 형태와 종류가 게임마다 모두 다를 수밖에 없다[2]. 본 연구에서 사용할 일반적인 속성은 표 2와 같다.

표 2. 게임에 독립적인 속성 6가지

독립적 속성	설명
Play Time (PT)	순수 게임 플레이 시간의 총합
Active Days (AD)	게임 접속 일수
Login Cnt (LC)	로그인 횟수
Pay	결제 금액의 합
Quest Clear (QC)	퀘스트(임무) 클리어 횟수
Has Clan (HC)	클랜이나 길드 가입 여부

3. 데이터 전처리

3.1 데이터 부스팅

앞서 2.2에서 서술하였듯이, 본 연구에서는 슬라이딩 윈도우 방식을 사용하여 게이머의 이탈 직전의 행동 패턴을 분석한다. 이 기법은 특정 기간의 슬라이딩 윈도우가 존재하며, 해당 윈도우가 시간의 흐름에 따라 게이머 속성의 변화를 기록하게 된다. 슬라이딩 윈도우의 크기는 일주일로 지정하였으며, 일주일 동안의 게이머 행동을 기록하고 있다. 슬라이딩 윈도우는 하루만큼 이동하며, 게이머의 행동이 존재하지 않을 때까지 이동하며 데이터를 생성하게 된다.

$$\text{Count(Data)} = \sum_{i=0}^n \text{PlayDay}_i, n = \text{게이머수}$$

수식 1

따라서 원래라면 하나의 게이머 당 한 개의 데이터가 존재했겠지만, 이 기법을 통하여 게이머가 플레이한 일수만큼 곱해진 수의 데이터를 생성하게 된다.

표3. 슬라이딩 윈도우를 통한 데이터 부스팅 효과

데이터 셋	데이터 수	이탈 데이터	잔류 데이터
소울게이지	189,041	182,122	6,919
소울게이지(부스팅)	394,648	261,371	133,277
크레이지 드래곤	192,361	160,021	32,340
크레이지 드래곤(부스팅)	818,066	340,674	477,392

4. 결과 및 결론

본 연구에서 사용하는 데이터 학습 모델은 Decision Tree, Random Forest, Extra Tree를 사용하고 각 모델의 성능을 정확하게 평가 하기 위해 Accuracy, Precision, Recall, F-1 Score, AUC를 평가 지표로 사용한다. 추출한 데이터는 소울 게이지 게이머의 일주일 데이터이며, 그 이후 2주 동안의 로그인 여부로 이탈 및 잔류를 예측한다. 추출된 속성은 게임에 독립적인 속성 Play Time(PT), Active Days(AD), Login Cnt(LC), Quest Clear(QC), Has Clan(HC) 6가지를 사용한다. 이렇게 추출된 데이터를 이용하여 Decision Tree(DT), Random Forest(RF), Extra Tree(ET)의 3가지 모델을 사용하여 게이머의 이탈을 예측한다.

표4. 표준화 데이터 셋을 이용한 소울 게이지 이탈 예측 성능

normalization Dataset	Accuracy	precision	recall	f1 score	AUC
DT	0.85925	0.78933	0.72295	0.75468	0.83546
RF	0.87925	0.82602	0.75602	0.78947	0.93337
ET	0.87302	0.82654	0.72895	0.77468	0.92506

소울 게이지의 경우 표 4과 같은 결과를 얻었으며

정규화를 통한 성능 향상은 없었지만, 성능 하락 또한 없었다. 이는 크로스 플랫폼에 있어 큰 이점이 되며, 향후 연구에 유의미한 결과를 기대할 수 있는 가능성을 얻었다.

크레이지 드래곤 데이터 또한 소울게이지와 마찬가지로 게이머의 일주일 데이터를 추출했고, 그 이후 2주 동안의 로그인 여부로 이탈 및 잔류를 예측한다. 마찬가지로 같은 속성 6 가지와 DT, RF, ET의 3가지 모델을 사용하여 게이머의 이탈을 예측한다.

표5. 표준화 데이터 셋을 이용한 크레이지드래곤 이탈 예측 성능

normalization Dataset	Accuracy	precision	recall	f1 score	AUC
DT	0.85593	0.85719	0.74601	0.79774	0.81935
RF	0.87706	0.88323	0.78035	0.82861	0.91825
ET	0.87429	0.88188	0.77352	0.82415	0.91491

크레이지 드래곤의 결과 또한 표 5와 같은 결과를 얻었으며 소울 게이지 데이터와 마찬가지로 두드러진 성능 향상은 없었지만 성능 하락 또한 일어나지 않았다.

5. 감사의 글

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초 연구 사업임 (2017R1A2B4002164)

참고 문헌

[1] Crosby, L.A. and N. Stephens, "Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry," Journal of Marketing Research, Vol.24, No.4, pp.404-411, 1987

[2] Africa Peri ´ a´nez, Alain Saas, Anna Guitart, Colin Magne "Churn Prediction in Mobile Social Games: Toward a Complete Assessment Using Survival Ensembles", 2017