

# 트리 기반 분류기의 모델 전이를 이용한 모바일 게임에서의 사용자 이탈 예측

배청목<sup>○</sup> 김경중<sup>†</sup>  
광주과학기술원 융합기술학제학부  
cmbae0307@gist.ac.kr<sup>○</sup> kjkim@gist.ac.kr<sup>†</sup>

## Churn Prediction for Mobile Game Players using Model Transfer of Tree-based Classifiers

Cheong-mok Bae<sup>○</sup> Kyung-Joong Kim<sup>†</sup>  
School of Integrated Technology, GIST

### 요 약

게임에서의 사용자 이탈 예측은 게임에서 데이터 마이닝이 시작된 이래로 다양한 방법으로 시도된 문제이다. 최근의 연구에서는 사용자 이탈 예측에 중요한 게임 내적 요소의 패턴을 찾아 내거나, 게임 사용자의 패턴을 프로파일링 하는 등의 시도를 통해 다양한 게임에 대해 일반적으로 모델을 학습할 수 있는 방법을 찾기 위한 노력이 계속되고 있지만, 수집한 데이터의 소수만을 사용한다는 점과 결국 하나의 게임 데이터를 활용해 학습 모델을 다른 게임에 적용할 수 없다는 모델의 재활용성 문제를 해결하지 못하고 있다. 본 연구에서는 하나의 게임 로그 데이터로부터 학습한 트리 기반 분류기 모델을 다른 게임의 사용자 이탈 예측에 활용하여, 후속될 다양한 게임에 대한 사용자 이탈 예측 모델의 가능성을 탐구한다.

### 1. 서 론

게임 분석(Game Analytics)은 게임 개발과 연구를 위한 분석 방법으로, 원격측정(Telemetry)을 이용 해 수집한 데이터를 목적에 따라 분석하여 게임과 관련된 전반적인 분야에 걸쳐 의사 결정을 돕는 것을 목적으로 한다[1].

사용자 이탈 예측은 게임 분석의 사용자 메트릭(User Metric) 분석의 하나로, 게임 플레이어-게임 간의 상호작용에 대한 분석을 통해 게임 플레이어가 게임을 떠날 것인지 계속 플레이 할 것인지를 예측하는 것이다[1].

게임에서의 사용자 이탈 예측을 위한 연구는 2011년 Z. Borbora의 MMORPG(Massively Multiplayer Online Role-Playing Game) Everquest에서의 사용자 이탈 예측 연구를 시작으로, 2014년 F. Hadji가 F2P(Free-to-Play 혹은 Freemium) 게임에서 사용자 이탈 예측 연구, 2016년 A. Perianez가 모바일 소셜 게임에서의 사용자 이탈 예측을 시도하는 등 다양한 장르의 게임에 시도되었다[2]-[4]. 특히, 2016년 A. Perianez는 모바일 소셜 게임을 대상으로 진행한 연구에서는 게임의 장르와 구조에 관계없이 어느 게임에나 적용할 수 있는 게임에 종속되지 않는 속성(Game-independent features)와 게임에 장르나 형태에 따라 적용의 여부가 결정되는 게임에 종속되는 속성(Game-dependent features)가 정의되었다[4].

게임에서의 사용자 이탈 예측 문제에서는 성능의 문제와 별개로 모델을 구축 및 학습하는데 필요한 각각 게임에 대한 데이터 수집의 문제와 이를 활용하여 구축한 모델이 개별 게임에만 적용할 수 있다는 모델의 재활용성 문제를 안고 있다. 연구가 진행됨에 따라 다양한 게임에서 일반적으로 활용 가능한 모델을 구축하기 위한 노력이 계속되고 있고,

이를 위해 게임에 종속되지 않는 속성을 활용하여 더 유의미한 데이터를 추출하거나 사용자 패턴을 발굴하는 등의 방향으로 연구가 진행되었다[5]-[7]. 또한 전통적인 트리 기반의 기계학습 기법을 벗어나 심층 신경망이나 앙상블 모델을 활용하는 등의 시도도 계속 되고 있다[7].

본 연구에서는 게임에서의 사용자 이탈 예측 발생하는 모델의 재활용성 문제의 해결과 다양한 게임에 대한 일반적인 모델 설계의 가능성 탐구를 위해 트리 기반 분류기의 모델 전이를 이용해 모바일 게임 사용자 이탈 예측 연구를 진행하였다.

### 2. 실험 설정

#### 2.1 게임 로그 데이터

본 연구에서는 모바일 크레이지 드래곤 과 소울 게이저 로그 데이터를 활용해 실험을 진행했다. 두 개의 게임 모두 RPG(Role Playing Game) 장르에 속하여 비교적 비슷한 게임이라고 생각할 수 있으며, 각 데이터의 수집 기간 및 기타 정보는 표 1과 같다.

표 1. 게임 정보와 로그 데이터 정보

|       | 크레이지드래곤                             | 소울게이저                              |
|-------|-------------------------------------|------------------------------------|
| 제작사   | MGame                               | DigitalFrog                        |
| 수집 기간 | 2016.03.21. ~<br>2016.07.29. (130일) | 2017.05.25. ~<br>2017.09.08 (109일) |
| 유저 수  | 192,361 명                           | 189,042 명                          |
| 로그종류  | 96 가지                               | 117 가지                             |
| 로그 수  | 295,107,602 개                       | 698,552,485 개                      |

게임 별로 기록된 데이터의 로그 종류가 상이하기 때문에 게임에 종속되지 않은 속성을 정제한 5가지의 속성(플레이 시간, (관측기간 내) 활성일 수, 로그인 횟수, 결제 금액)을 사용했으며, 게임에 종속되는 속성이지만 두 개의 게임 모두에 나타나는 길드(혹은 클랜)를 가졌는지 여부에 대한 로그를 포함해 총 6가지 속성을 모델 학습에 사용했다. 또한 서로 다른 게임에서 나타날 로그 양상의 차이를 감안해 플레이 시간, 로그인 횟수, 결제 금액의 데이터는 각 게임에 대해 정규화하여 사용했다.

2.2 이탈 모델 정의 및 데이터 증대

이탈자 예측에 앞서 예측에 사용될 이탈자에 대한 정의가 필요하다. 본 연구에서는  $k$ 일로부터  $k+14$ 일간의 데이터를 관측하고  $k+15$ 일과  $k+21$ 일 사이에 해당 유저가 게임 내의 활동을 했는지 아닌지를 이탈 여부 판단에 사용했다. 해당 기준에 따라 기록일의 마지막 7일 시점에서 이탈 여부를 레이블하면 각 게임의 이탈/비이탈자의 극적인 불균형을 확인할 수 있는데, 이를 각 유저 별 데이터를 그림 1과 같이 슬라이딩 윈도우를 통해 데이터를 증대하여 데이터 불균형 문제를 완화함과 동시에 예측 시점으로부터 가까운 시일의 데이터가 이탈 여부와 관련이 있다는 것을 모델에 반영할 수 있도록 했다. 아래 표 2에서 데이터 증대 전후의 이탈 비이탈 데이터의 수와 비율을 확인할 수 있다.

그림 1. 슬라이딩 윈도우를 통한 유저 데이터 수집 및 증대

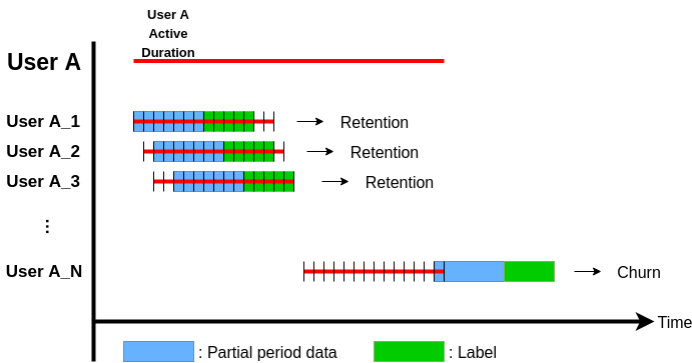


표 2. 데이터 증대 전후의 이탈 비이탈 비율

|          | 원본            |              | 데이터 증대        |               |
|----------|---------------|--------------|---------------|---------------|
|          | 이탈            | 비이탈          | 이탈            | 비이탈           |
| 크레이지 드래곤 | 160,021 (83%) | 32,340 (17%) | 989,759 (62%) | 606,479 (38%) |
| 소울 게이지   | 182,122 (96%) | 6,919 (4%)   | 382,030 (70%) | 164,065 (30%) |

2.3 모델 전이

본 연구에서는 사용자 이탈 예측에 대한 모델 전이의 가능성과 성능을 테스트하기 위해 아래와 같이 하나의 게임에서 자체적인 데이터를 활용해 일자 별 모델을 구성하는 Standalone 모델과 하나의 게임의 전체 데이터를 활용해

모델을 구성한 후, 이를 이용해 다른 게임 데이터의 이탈 예측을 하는 Transferring 모델에 대한 실험 설정을 했다.

✓ Standalone

- 1) 주어진 게임A의 각 유저에 대해  $k \sim k+14$ 일의 데이터를 취합하고, 이에 대한 이탈 여부 레이블은  $k+15 \sim k+21$ 일 사이에 활동 여부로 정한다.
- 2) 이탈 여부의 레이블이 가능한 로그 기록 마지막 21일 전까지 모델을 구성하여, 일자 별 모델의 성능을 확인한다.

✓ Transferring

- 1) Standalone에서 마지막 일자에 학습된 모델(게임 A의 전체 데이터를 활용해 학습 한 예측 모델)로 새로운 게임B 유저의  $k \sim k+14$ 일의 데이터에 대해 일자별로 예측하고 이에 대한 성능을 비교한다.

학습에 사용된 모델은 결정 트리, 랜덤 포레스트, 엑스트림 랜덤 트리의 세 가지 트리 기반 분류기를 사용했고, Transferring 과정에서는 각 분류기에 대해 Standalone과정에서 주어진 데이터를 10-fold 교차 검증을 통해 가장 좋은 성능을 보인 모델을 활용했다.

3. 결과 및 분석

그림 2은 제안한 방법을 이용해 소울게이지 Standalone 모델의 일자 별 예측 성능과 크레이지드래곤의 마지막 일자 데이터를 활용해 학습한 모델에 소울게이지 일자 별 데이터를 예측한 성능을 비교한 것이다. 결과를 보면, 정확도는 전체 구간에 걸쳐 점차 낮아지는 것을 확인할 수 있는데, 이는 모델 구성 초기에 대비해 시간이 지날수록 학습 및 예측해야하는 데이터가 많아지기 때문이다. 또한 정밀도와 재현률, F1-Score 부분에서는 학습의 초기에는 Standalone이 Transferring 모델보다 정확도가 낮지만, 시간이 지남에 따라 Standalone 모델의 예측 성능이 Transferring 모델을 넘어서는 것을 확인할 수 있다. 이는 Transferring에 활용한 모델은 비록 다른 게임의 Standalone 모델 이지만, 하나의 게임 데이터 전체를 모델을 학습하는 데 활용했기 때문에, 모델을 구성하는 데 사용된 데이터 총량이 새로운 게임의 초기 Standalone 모델을 학습하는 데 사용된 데이터보다 많기에 더욱 더 견고한 모델이 구성되기 때문으로 보여진다. 시간이 흐를수록 자체 게임의 데이터가 Standalone 모델의 학습에 투입되어 스스로에 더 적합해지면서 Transferring 모델을 넘어서는 성능을 갖게 되는 것으로 해석할 수 있다.

최종적으로는 자체적인 데이터를 활용해 모델을 학습하는 경우가 더 좋은 성능을 보이기는 했지만, 충분한 양의 데이터를 수집하지 못했거나, 기존에 비슷한 다른 게임에 대한 데이터가 있을 경우에 이를 활용하여 사용자 이탈 예측에 대해 활용할만한 성능을 얻을 수 있다는 것을 알 수 있다.

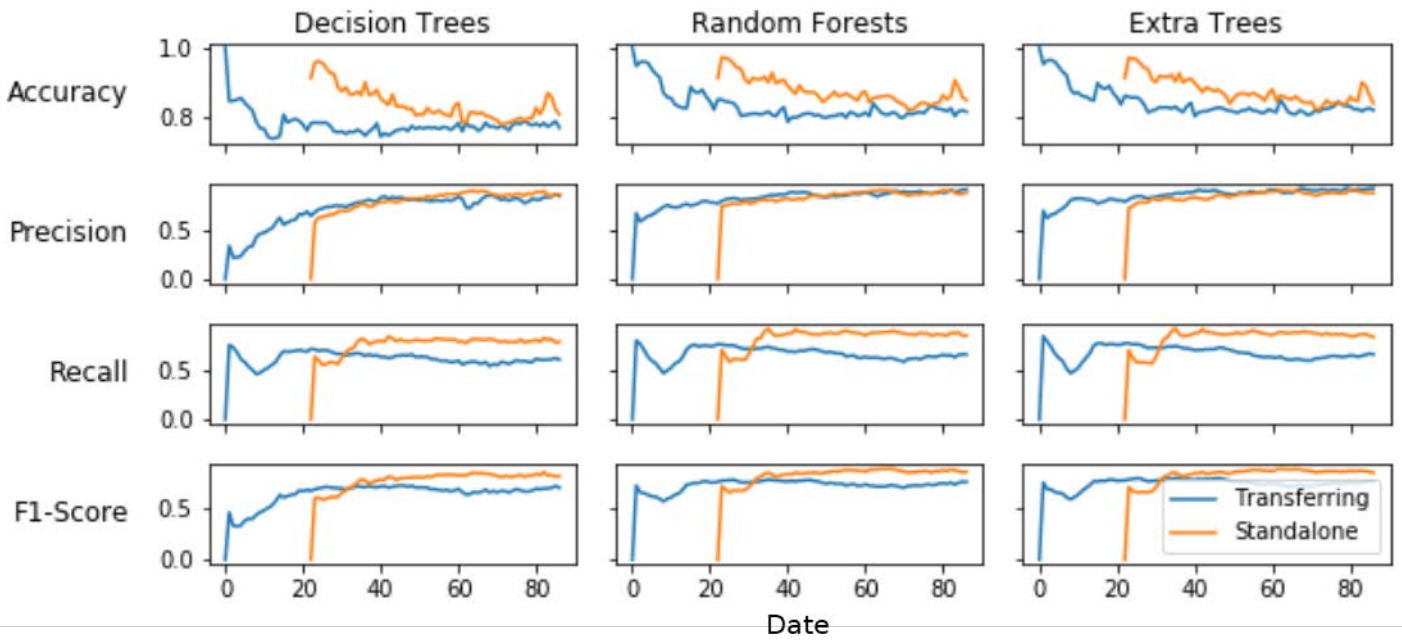


그림 2. 소울게이지 Standalone 모델의 일자 별 예측 결과(주황색)와 크레이지드래곤 마지막 일자의 Standalone 모델로 소울게이지 데이터를 일자 별로 예측한 결과(파란색) 비교.

4. 향후 연구

설정된 실험에 따라 서로 비슷한 게임에서의 사용자 이탈 예측 모델이 서로 전이가 가능할 것이라는 가능성을 점칠 수 있지만 몇 가지 문제점을 발견했고 이를 통해 추후 연구에 대한 기회를 찾을 수 있었다.

첫째, 트리 기반 분류기는 데이터의 증가에 따른 점진적인 학습이 불가능하기 때문에 일자 별 모델을 따로 학습을 해야했으며, 이 과정에서 많은 시간이 필요하게 되어 시간적인 효율성 측면에서는 오히려 손실이 있다. 이를 해결하기 위해 증분 데이터에 대해서만 학습이 가능한 분류기를 활용하거나 인공지능망을 통해 학습을 시도해볼 필요가 있다.

둘째, 모델 학습에 사용된 속성으로 수집한 로그에 대비해 실제 모델 구성에 활용한 로그의 종류는 소수에 불과하며, 게임에 비종속적인 속성만을 활용하여 모델을 학습하는 것은 각 게임의 특징을 충분히 반영하지 못할 수 있다는 점이다.

셋째, 수행한 실험의 모델 전이는 같은 장르의 게임에 대해 수행한 것으로, 완전히 다른 종류의 게임에 대해서는 성능을 알 수 없다는 점이다. 또한 두 게임이 비슷하다는 것의 정의가 게임의 장르에만 국한되며, 객관적으로 두 개의 게임이 비슷하다는 것을 정의할 수 없다는 점이 있다.

5. 감사의 글

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 중견연구자지원사업임(NRF-2017R1A2B4002164).<sup>†</sup>교신저자

[1] London, 2013.

[2] Z. Borbora, J. Srivastava, K.-W. Hsu, and D. Williams, "Churn Prediction in MMORPGs Using Player Motivation Theories and an Ensemble Approach," in *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 2011, pp. 157-164.

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

[4] A. Perianez, A. Saas, A. Guitart, and C. Magne, "Churn prediction in mobile social games: Towards a complete assessment using survival ensembles," in *Proceedings - 3rd IEEE International Conference on Data Science and Advanced Analytics, DSAA 2016*, 2016, pp. 564-573.

[5] A. Guitart, P. Chen, and Á. Periañez, "The Winning Solution to the IEEE CIG 2017 Game Data Mining Competition," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, pp. 252-264, Dec. 2018.

[6] A. Saas, A. Guitart, and A. Perianez, "Discovering playing patterns: Time series clustering of free-to-play game data," in *IEEE Conference on Computational Intelligence and Games, CIG*, 2017, pp. 1-8.

[7] W. Yang *et al.*, "Mining Player In-game Time Spending Regularity for Churn Prediction in Free Online Games," in *2019 IEEE Conference on Games (CoG)*, 2019, pp. 1-8.

참고 문헌

[1] M. S. El-nasr, *Game Analytics*. London: Springer