

모바일 게이머의 인지심리속성 추출 및 구매예측 성능개선

(Extraction of Cognitive Psychological Features of Mobile Gamers and Improvement of Purchases Prediction Performance)

전지훈[†] 양성일^{**} 김경중^{***}
(Jihoon Jeon) (Seongil Yang) (KyungJoong Kim)

요약 게이머의 게임 내 구매는 게임회사의 매출과 직결되는 중요한 요소 중 하나이다. 평균적으로 95%의 게이머는 게임 내 구매를 하지 않으며, 극소수의 게이머가 게임회사 매출의 대부분을 책임지고 있다. 이러한 까닭에 게임 기업들은 이러한 소수의 구매 게이머들을 유지시키고 증대시킬 필요가 있다. 본 논문에서는 모바일 알파지 게임 '소울게이지'의 로그데이터를 사용하여 게이머의 인지심리를 추정할 수 있는 7가지 인지심리속성(경쟁성, 도전성, 충성도, 사회성, 활동성, 효율성, 성실성)을 정의하고 추출하였다. 또한, 결제금액을 기준으로 분류된 게이머들을 7가지 인지심리속성을 이용하여 분석하였고, 그 결과 인지심리속성과 결제금액과의 연관성이 있음을 보였다. 추가적으로, 7가지 인지심리속성을 이용하여 게이머의 구매를 비교적 높은 정확도로 예측하였다. 이러한 결과는 게임 로그로부터 게이머를 인지심리 기반하에 분석할 수 있고 게임 내 구매를 비교적 높은 성능으로 예측할 수 있음을 의미한다.

키워드: 인지심리, 구매 예측, 모바일 게이머, 게이머 분석, 소울게이지, 로그데이터

Abstract In-game purchases are one of the important factors that directly affect a company's revenue. In total, 95% of gamers do not pay for in-game purchases, meaning that a small number of gamers are responsible for most of the revenue of the company behind their games. For this reason, game companies must maintain and augment these few purchasing gamers. In this paper, we extracted seven cognitive psychological features (competitive, challenge, loyal, social, activity, efficient, and sincerity) that can be used to estimate the cognitive psychology of a gamer by using log data of a mobile RPG game. We analyzed the gamers, classified by payment amount, based on seven cognitive psychological features. As a result, the cognitive psychological features and payment amount of the gamers could be correlated. In addition, using seven cognitive psychological features, we predicted the purchasing behavior of gamers with high accuracy. This implies that gamers can be analyzed based on their cognitive psychology and the gamer's purchases can be predicted with comparatively high performance.

Keywords: cognitive psychology, purchase prediction, mobile gamer, gamer analysis, soulgauge, log data

· 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2018년도 문화기술연구개발 지원사업으로 수행되었으며, 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2017R1A2B4002164).

† 비회원 : 세종대학교 컴퓨터공학과 외부연구원
rhd00513@gmail.com

** 비회원 : 한국전자통신연구원 차세대콘텐츠연구본부 책임연구원
siyang@etri.re.kr

*** 정회원 : 광주과학기술원 융합기술학제학부 교수(GIST)
kjkim@gist.ac.kr
(Corresponding author)

논문접수 : 2018년 9월 10일
(Received 10 September 2018)
논문수정 : 2019년 6월 15일
(Revised 15 June 2019)
심사완료 : 2019년 6월 17일
(Accepted 17 June 2019)

Copyright©2019 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.
정보과학회논문지 제46권 제9호(2019. 9)

1. 서론

모바일 게이머의 게임 내 구매는 게임회사의 매출과 직결되는 중요한 요소 중 하나이다. 대부분의 게이머는 인 게임 구매를 하지 않으며, 평균적으로 전체 게이머 중 약 5%의 게이머가 게임 내 구매를 하고 있다. 게다가 그 5%의 구매유저 중 약 20% 정도가 전체매출의 90% 이상을 책임지고 있다. 만약 이러한 구매유저를 예측하고 인지심리기반 하에 분석할 수 있다면, 게임회사들은 게이머들에게 더 효율적이고 만족스러운 게임 서비스를 제공할 수 있을 것이다. 이러한 현황 속에서 많은 데이터 마이닝 분야의 연구자들은 게이머들을 분석하거나[1-3] 게이머의 행동 로그데이터를 기반으로 게이머의 행동을 예측하는[4-6] 연구가 이루어지고 있다. 다만, 기존의 대부분의 연구는 설문조사를 이용하여 게이머를 분석하거나 로그데이터의 통계적 수치만을 이용하여 게이머의 행동을 예측하고 있다. 설문조사를 이용한 게이머의 분석은 비교적 게이머의 성향과 같은 심리적 특징을 분석하긴 쉽지만, 많은 양의 데이터를 얻기 어렵고, 게임 내 행동을 게이머 각각의 주관적인 기억에 따라 얻기 때문에 정확한 정보를 얻기 어렵다. 또한, 게이머의 심리나 성향을 추론할 수 있는 속성을 추출하려 하지 않고, 로그인 횟수(Login count)와 같은 단순한 게이머의 게임 내 행동로그를 통계적 수치로 정형화하여 게이머의 행동을 예측하고 있다. 이러한 방법은 게이머의 행동을 예측하는 성능이 좋을 수 있지만, 게이머의 성향 및 심리적 특성을 분석하기엔 어려움이 있다. 이처럼 이러한 어려움을 해결하기 위해서는 게이머들의 행동이 기록된 게임 로그데이터로부터 게이머의 인지심리학에 기반한 분석이 필요하다.

본 논문에선 모바일 알피지 게임 '소울게이지'의 로그데이터를 이용하여 연구를 진행하였다. 소울게이지는 Digital Frog에서 개발한 모바일 롤플레이팅 게임으로, 2017년 5월 25일 오픈을 한 후, 동년 6월 27일까지 약 1개월 동안 수집된 로그데이터를 가지고 분석을 진행하였다. 약 18만명의 게이머의 행동을 총 117종류의 게임 로그를 이용하여 기록하였으며 총 7억개 가량의 로그데이터가 기록되었다. 이러한 가공되지 않은 게임 로그를 이용하여 게이머와 모바일 게임간의 상관관계를 고려한 7가지 인지/심리적 속성(경쟁성, 도전성, 충성도, 사회성, 활동성, 효율성, 성실성)을 정의하였다. 이렇게 정의된 7가지 인지/심리적 속성은 게이머의 심리적 특성이나 성향, 행동성을 나타낼 수 있을 것으로 고려되는 지표로써, 서로의 특성이 겹치지 않도록 게임에 특화되지 않은 특징을 뽑아내는 것을 목적으로 만들었다. 이러한 속성들은 게이머의 심리적 성향이나 특징을 수치화 할 수

있으며, 다른 게임에서 추출한 게이머의 인지 심리적 속성 수치와 비교 및 분석 할 수 있음을 의미한다.

2. 게임과 인지심리

2.1 근래의 게임 현황

현재, 게임시장은 스마트폰의 등장 이래 지속적으로 모바일 게임 중심으로 나아가고 있다. 시장조사업체 슈퍼데이터리서치(SuperData Research)에 따르면, 2016년 전 세계 모바일 게임시장 매출이 410억 달러로 집계된다. 이는 2015년 대비 18%가량 증가했으며, 아시아가 249억 달러의 매출을 기록하며 전체 시장의 50% 이상을 차지하고 있다[7]. 또한, 2015년에 비해 모바일게임에서 연 200억원 이상 매출을 낸 게임이 26개로 증가하면서, 전년 대비 73%가 증가하였다[8]. 이는 더 많은 기업이 모바일게임시장에 진입할 동기가 되며, 모바일게임시장은 지속적인 상승세를 보일 것으로 예상된다.

2.2 인지심리기반 게이머 모델링

게임회사들과 연구자들은 게이머들이 게임에 대해 어떻게 느끼는지를 알아내기 위해 다양한 방법을 사용해왔다. 설문조사를 이용해 게이머들로부터 직접 의견을 받기도 하고[1-3], 로그로 남겨진 게이머의 행위 기록들로부터 게이머의 만족을 측정하기도 했다[4-6]. 또한 각종 생체신호를 센서로 측정하여 게이머들을 분석[9]하기도 하였다.

생체신호를 직접 측정하면, 실제 게이머들이 얼마나 집중하고 있고, 만족하고 있는지를 비교적 정확하게 알 수 있지만, 실제로 상업용 게임에 적용하는 것은 아직은 비현실적이다. 따라서 실용적인 응용을 목적으로 하는 대부분의 연구는 게임의 로그 데이터에서 게이머가 어떤 행동을 했는지를 추출하고, 이를 이용해 게이머들이 얼마나 게임에 흥미를 느끼고 있는지를 추측하려한다. 실제 게이머들이 얼마나 흥미 혹은 재미를 느끼는가는 정량적으로 측정이 불가능하기 때문에, 보통 게이머들의 이탈 혹은 구매를 예측하여 간접적으로 게이머들의 내적 상태를 추론한다[6]. 또한 이탈과 구매는 게임 회사의 수익과 직접적으로 연관되는 핵심적인 사건이기 때문에 이를 예측하는 것은 실용적이기도 하다.

게이머들의 이탈 혹은 구매를 예측하는 대부분의 연구는 게임 로그로부터 알아낼 수 있는 직접적인 사건을 기반으로 하여 예측모델을 만드는 것을 목적으로 한다. 주로 로그인 횟수, 접속시간, 구매 금액 등, 이탈과 구매 사건과 직접적으로 연관이 있어 보이는 기본 사건들의 통계적인 정보(평균, 분산 등)를 사용하여, 게이머들이 이탈/구매하기 전에 미리 사건을 예측하는 모델을 만든다. 많은 연구들이 이와 같은 기본 사건들의 통계 수치를 사용해서 게이머들의 행동을 예측하거나 분류하지만,

아직 게이머들의 인지심리적인 정보를 사용한 연구는 초기단계이다.

3. 데이터

본 연구에서는 롤플레이 게임인 소울게이지의 로그를 분석하였다. 소울게이지는 Digital Frog에서 개발한 모바일 롤플레이 게임으로, 2017.5.25.~6.27동안 약 1개월(33일)동안 179,431명의 게이머로부터 수집된 로그데이터를 가지고 분석하였다. 추가적인 로그를 모두 포함하여 총 117종류의 게임로그를 이용하여 698,552,485 개의 로그데이터가 기록되었다.

3.1 구매 예측 모델

모바일 게임에서 게이머의 게임 내 구매는 게임회사의 매출과 직결되는 중요한 요소 중 하나이다. 하지만, 대부분의 게이머는 캐시아이템을 구매하지 않으며, 소수의 게이머만 캐시아이템을 구매 하여 게임회사의 매출을 책임지고 있다. 따라서, 게임회사는 대부분에 해당하는 게이머들을 분석하여, 게임의 매출을 올려줄 수 있는 구매유저가 되게끔 만들고, 구매 유저들 또한 지속적으로 캐시아이템을 구매할 수 있도록 만들어야 한다. 이를 위한 방법으로 게임 로그 분석을 통해 비 구매 게이머와 구매 게이머의 행동패턴을 분석하여 향후 구매여부를 예측할 수 있다. 각각의 유형은 두드러진 특징을 보일 것이며, 이러한 특징을 모델링 한다면, 새로운 게이머가 유입되어도 게이머의 특징분석을 통해 구매여부를 예측할 수 있을 것이다. 또한, 새로운 콘텐츠와 이벤트 등의 효과도 미리 예측해 볼 수 있어, 게임서비스에 있어 좀 더 효율적인 서비스가 가능하다. 이는 게임회사의 매출을 증대시키는 방법 중 하나가 될 것이다.

게이머의 구매예측을 위한 기계 학습 모델로는 결정 트리(DT, Decision Tree)의 카트(CART, Classification And Regression Tree), 랜덤 포레스트(RF, Random Forest) 그리고 엑스트라 트리(ET, Extra Tree)를 사용하고, 각 모델에 대한 성능을 보다 정확하게 평가하기 위해 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), 에프-원 점수(F1-Score)를 평가지표로 사용한다.

3.2 데이터 전처리

구매 예측에 앞서 게이머의 구매에 대한 정의가 필요하다. 본 연구에서는 게임 기업의 입장에서 보았을 때, 좀 더 실용적이고 효율적으로 사용할 것으로 생각되는 특정 기간 동안의 게이머들의 데이터를 사용하여 그 이후의 구매를 예측하였다. 따라서 소울게이지의 게이머 로그가 수집된 33일의 기간인 2017.05.25.부터 2017.06.27.기간 중 앞선 11일(2017.05.25.~2017.06.05.)을 학습데이터로 학습시켰고, 학습된 기간의 게이머를 기준으로 22일인 2017.06.05.부터 2017.06.27.기간 동안의 구매여

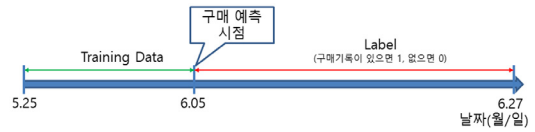


그림 1 구매 예측 모델

Fig. 1 Purchase prediction model

표 1 일정 시간이상 게임을 플레이한 게이머를 대상으로 한 구매게이머 비율

Table 1 Percentage of purchase gamers who have played for more than a certain amount of time

play time	total gamer	purchase gamer	ratio
1 hour	22540	625	2.77%
2 hours	15100	623	4.13%
3 hours	11735	620	5.28%
4 hours	9778	617	6.31%
5 hours	8376	614	7.33%
10 hours	5187	600	11.57%
24 hours	2730	553	20.26%
48 hours	1572	472	30.03%
72 hours	1044	381	36.49%

부에 따라 구매유저와 비 구매유저를 정의하였다. 구매 예측 모델은 그림 1과 같다.

모바일 게임에서 구매유저는 전체유저의 약 3~5% 내외로 매우 극소수의 유저만 게임 내 결제를 하고 있다. 이러한 까닭에 구매유저와 비구매유저의 비율은 매우 극단적인 비율을 가지고 있다. 또한, 대부분의 모바일 게임에서 많은 신규 유저는 게임에 매우 짧게 머무르고 나가는 형태를 보이고 있다. 이는 게임이 해당 게이머의 취향에 맞지 않기 때문에 그럴 수도 있지만, 광고성 보상으로 해당 게임을 설치하여 해당게임이 아닌 다른 무언가의 보상을 얻기 위해 게임에 진입하는 유저들도 많다. 이러한 유저들은 데이터의 신뢰성을 낮추는 노이즈 게이머로, 해당 게이머들을 제외할 필요가 있다. 따라서 구매 예측에 앞서 본 연구에서는 이러한 유저를 제외하기 위하여 일정 플레이타임 이하의 게이머를 제외한 후 구매/비구매 게이머의 비율을 비교해 보았다. 그 결과 표 1과 같은 결과를 얻었으며, 본 연구에서는 좋은 데이터 셋을 3~5시간의 게이머를 제외하였을 때가 노이즈 게이머가 적은 게이머 데이터로 추정하고 실험을 진행하였다.

하지만 3~5시간 이하의 플레이타임을 가지는 게이머를 제외하여도, 구매게이머와 비 구매게이머는 여전히 불균형한 비율을 띄고 있다. 본 연구에서 진행하는 기계 학습 기법들을 포함한 일반적인 기계학습 기법들은 학습데이터가 비슷한 비율로 구성되어 있다고 가정된 후 학습이 진행된다. 하지만, 데이터가 불균형한 데이터일 경우 적은비율의 데이터들을 잘못 분류될 가능성이 높

아진다[10]. 이를 해결하기 위해 본 연구에서는 언더샘 플링을 이용하여 데이터의 불균형을 해결한 후 실험을 진행하였다.

3.3 속성

소셜게이지의 전체 게이머 179,431명 중, 첫 11일인 2017.05.25.부터 2017.06.05.까지 플레이한 게이머를 대상으로 3~5시간이하 플레이타임 게이머를 제외한 결과 각각 11,735명, 9,778명, 8,376명의 게이머를 대상으로 구매 예측 실험을 진행하였다. 구매예측에 앞서 먼저 기본적으로 기록되는 raw 게임로그의 카운트를 속성으로 게이머의 구매를 예측하였으며, 117종의 raw 게임로그 중 실질적으로 사용될 수 있는 44종의 로그를 이용하였다.

3.4 결제 금액 별 게이머 분류

게이머를 분류하는 방법이나 기준은 그 목적이나 의도에 따라 천차만별로 달라질 수 있다. 그럼에도 불구하고 주로 사용되는 기준을 하나 뽑자면 게이머들이 게임에 사용하는 결제 금액의 양이다. 기본적으로 게임은 게이머들에게 즐거움을 주기 위한 서비스로써 운용되지만, 게임 회사에서 이런 서비스를 지속적으로 제공하기 위해서는 그에 필요한 인건비, 서버 운영비 등 자금이 필수적이다. 따라서 부분 유료화 모델이 일반적이게 된 요즘의 게임 회사에서는 게임 내 생태계 구성에 영향을 미치는 전체 또는 활성 게이머들의 수도 중요하지만 그 중에서도 매출에 직접적인 영향을 미치는 유료 결제 게이머를 더욱 잘 알아야할 필요가 있다. 본 연구에서는 '2015년 Google Play 게임 카테고리 총 결산 보고서'를 참고하여 로그를 수집한 마지막 날짜를 기준으로 유료 결제 게이머들을 고래(Whale) 게이머, 돌고래(Dolphin) 게이머, 피라미(Minnow) 게이머의 3 가지로 분류하였으며 각각 10만원 이상, 10만원 미만~1만원 이상, 1만원 미만 금액을 기준으로 나누었다[11]. 이를 기준으로 결제를 한 게이머들을 분류할 경우 그림 2, 3과 같이 각각 16%, 30%, 54%로 고래 유저들은 그 수가 적지만, 매출에서는 큰 비중을 차지하는 것을 볼 수 있었다.

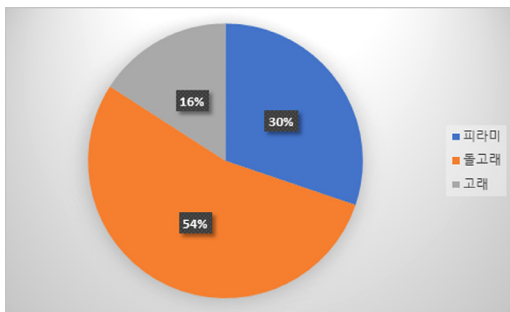


그림 2 고래/돌고래/피라미 게이머수의 비율
Fig. 2 Ratio of whale/dolphin/minnow gamer

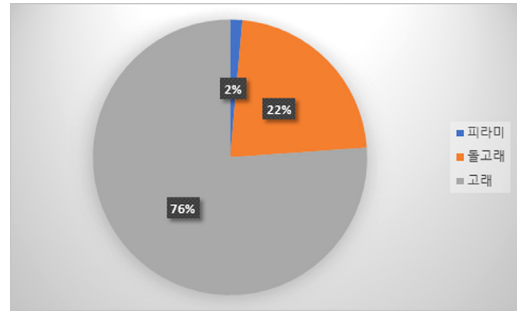


그림 3 고래/돌고래/피라미 게이머의 매출 비율
Fig. 3 Sales ratio of whale/dolphin/minnow gamer

4. 인지심리 속성 추출 및 분석

4.1 인지심리 기반의 분석을 위한 게이머 특징 추출

본 연구에서는 게이머의 인지심리 분석을 위해 게이머가 플레이 하면서 기록된 로그 데이터를 사용하였다. 게임 로그데이터는 게임의 특정 환경이나 게이머의 할 수 있는 행동의 제약으로 인해 기록되는 로그의 종류와 수는 게임마다 매우 다르고, 정형화 되어있지 않다. 또한, 모바일 게임의 로그는 종래의 PC용 온라인과 비교해 게임이 동작하는 환경이 다르기 때문에 로그의 수와 종류가 상대적으로 적어 모바일 게임의 특징에 맞는 정보를 인 게임 로그로부터 추출할 필요가 있었다. 그러한 이유로, 모바일 게임속의 게이머와 게임들의 관계에 대한 상관관계를 고려하여 게임에 특화되지 않은 게이머의 인지/심리적 특성을 나타낼 수 있는 7가지 인지/심리적 속성을 표 2와 같이 정의하였다.

앞서 제시된 7가지 인지/심리적 속성은 게이머의 심리적 특성이나 성향, 행동성을 나타낼 수 있을 것으로 고려되는 지표로써, 서로의 특징이 겹치지 않도록 게임에 특화되지 않은 특징을 뽑아내는 것을 목적으로 만들었다. 위 7가지 속성으로 게이머의 심리적 성향이나 특징을 수치화 할 수 있으며, 이는 다른 게임에 적용하여 추출한 게이머의 인지 심리적 속성 수치와 비교 및 분

표 2 일곱가지 인지/심리적 속성
Table 2 Seven cognitive psychological features

cognitive psychological feature	explain
Competitive	An indicator of how competitive a gamer is
Challenge	An indicator of how challenging a gamer is
Loyal	An indicator of how loyal gamer is in the game
Social	An indicator of the sociality of gamer
Activity	An indicator of the activity of gamer
Efficient	An indicator of how effective behavior gamer does
Sincerity	An indicator of how sincere a gamer is

표 3 인지심리분석을 위한 특징 추출

Table 3 Feature extraction for cognitive psychological analysis

cognitive psychological feature	33 kind of Primary features
Competitive	FildBattleLog/playday, Pay sum, Exp Increase, ArenaLog/playday, RankLog/playday, Battlezone Kill count
Challenge	Gacha/playday, Combine/playday, Enchant/playday, DungeonDead/dungeon Count, arenaLog/playday
Loyal	AttendanceRatio, Playtime/playday, DailyBonus/playday, PayLog/playday, Pay sum, VIP level, ChatLog/playday
Social	PartyLog/playday, ChatLog/playday, ClanLog count
Activity	QuestLog/playtime, DungeonLog/playtime, EnchantLog/playtime, ArenaLog/playtime, FarmLog/playtime, MoneyLog/playtime, CombineLog/playtime, CraftLog/playtime, PetItemLog/playtime
Efficient	FreeBitRatio, BonusLog/playday, EventBonusLog/playday
Sincerity	BonusLog/playday, PlayTime/playday, FreeGachaLog/playday, DailyQuestLog/playperiod

석할 수 있음을 의미한다. 더 나아가, 다른 장르의 인지심리 속성으로 정의된 게이머들을 추가한다면, 보다 일반적인 인지심리 기반의 속성으로 수치화된 게이머 데이터를 얻을 수 있을 것이다. 따라서, 본 연구에서 사용하는 소울게이지의 게임에 특화된 117종류의 raw 게임 로그로부터 7가지 인지/심리적 속성을 추출 및 가공할 필요가 있으며, 도메인지식에 의존하여 각 인지심리속성에 해당하는 것으로 추정되는 raw로그 기반의 가공된 1차 속성데이터들을 추출하였다. 그 속성들은 표 3과 같으며, 로그인 비율이나 평균 접속시간과 같은 통계적인 수치를 1차적 속성으로 추출한 후 이러한 1차 속성을 이용하여 인지/심리적 속성들로 통합하였다. 하지만, 각각의 1차 속성들의 값의 크기는 제각각이며, 이러한 수치들의 통합으로 인지심리속성이 만들어 지기 때문에 각각의 1차 속성들은 비교 가능한 규격화된 값으로 바뀔 필요가 있다. 그러한 이유로, 1차 속성들이 가우시안 정규분포를 따른다고 가정한 후, 각각의 1차 속성들을 100분위로 정규화를 진행하였다. 이는, 게이머 각각의 1차 속성들의 수치는 전체 게이머의 각각의 1차 속성에 대한 100분위로의 포함정도를 나타내는 비교 가능한 수치로 바뀌게 된다. 이는 게임에 특화되지 않고 비교 가능한 수치 추출과도 부합되며, 각각의 1차 속성에 가중치가 할당된 후 통합되어 인지심리속성들을 만들 수 있게 된다. 그 방법은 다음과 같다. 1단계에서는 먼저 게이머의 행동이 기록되는 기본적인 raw 로그로부터 7가지 인지심리적 속성(경쟁성, 도전성, 성실성, 충성도, 사회성, 효율성, 활동성)에 포함될 것으로 추정되는 1차 속성들을 추출한다. 2단계에서는 추출한 1차 속성들을 이용하여 게이머의 구매(Label)와 관련하여 가중치(Importance or Correlation)를 트리 기반 기계학습 알고리즘을 이용하여 추출한다. 3단계에서는 각각의 1차 속성들은 정규분포를 따른다는 가정 하에, 정규분포를 따르도록 100분위 수치로 정규화를 해준다. 4단계에서는 2단계에

서 추출한 각각의 1차 속성들의 가중치와 정규화된 1차 속성들을 곱하여 구매에 대한 가중치를 할당해준다. 마지막으로, 7가지 인지심리속성들에 해당하는 가중치가 할당된 1차 속성들을 모두 더하여 통합해준다.

이러한 구성은 게이머의 구매에 대한 성향을 보다 직관적으로 알 수 있고, 분석 가능하다는 장점이 있다. 또한 다른 게임에 적용하여도, 일반화된 수치 추출을 목적으로 만들었기 때문에 정확하지 않을 수 있지만 어느정도 비교 가능하다는 장점이 있다.

4.2 인지심리기반 결제 금액별 게이머 분석

먼저 게이머의 분석에 앞서 게이머를 군집화 할 필요가 있다. 특정한 정보를 기준으로 게이머를 유형화 한 후 분석한다면, 해당 유형 특징을 알고, 유형에 맞는 서비스를 기대할 수 있을 것이다. 따라서 앞서 3.4에서 게이머의 게임 내 결제 금액별로 게이머를 분류하였는데, 결제 금액을 기준으로 게이머의 유형별 인지심리 분석을 진행하였다. 결제 금액 별 분류는 게임 내 구매에 대한 게이머 특성을 가장 잘 나타내리라 기대하는 분류이며, 각각의 유형의 특징은 구별되고 차이를 보일 것으로 추측되었다. 그 결과 그림 4와 표 4와 같은 특징이 나타났다. 그림 4의 경우 4가지 그래프의 형태가 유사한 형태로 보이고 있는데, 이는 제외된 노이즈 게이머의 영향이 최소화되었음을 보여준다. 즉 최소 3시간이하 게임을 플레이한 게이머들을 제외하였을 때, 유의미한 게이머만이 남는다고 말할 수 있다.

이와 같이 게임 내 결제 금액별 게이머들을 인지심리 속성을 기반으로 분석해본 결과, 각 게이머별로 몇 가지 특징이 표 5와 같이 도출되었다.

5. 인지/심리적 속성 기반의 게이머행동 예측

5.1 인지/심리적 속성을 이용한 게이머 구매 예측

게이머 구매 예측을 위해서 앞서 이미 설명한 것과 같이 소울게이지 로그 데이터 117종 중 44종의 로그 카

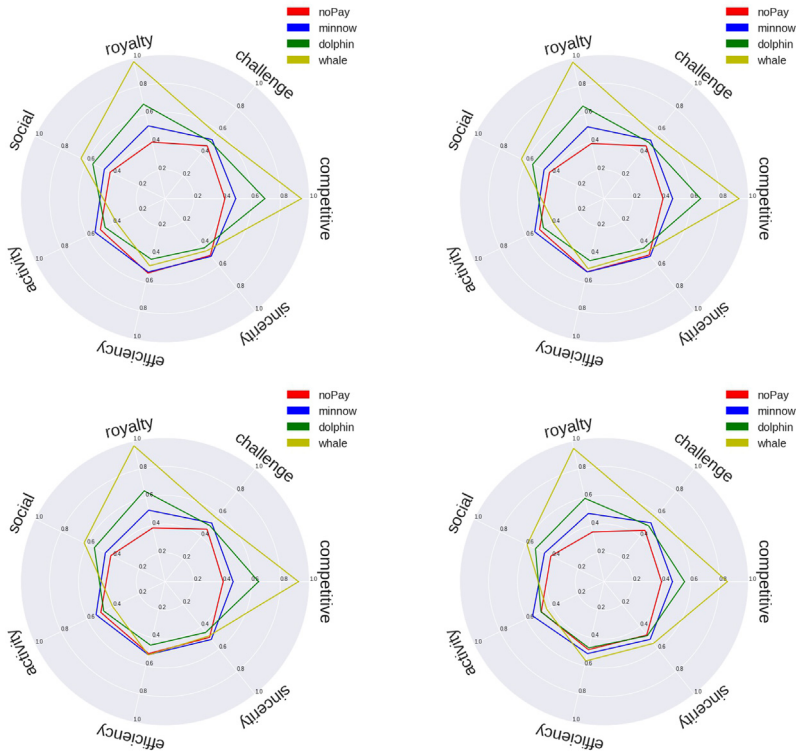


그림 4 무과금/피라미/돌고래/고래 게이머의 7가지 인지심리속성 수치(좌상: 3시간 이상 플레이 게이머, 우상: 4시간 이상 플레이 게이머, 좌하: 5시간 이상 플레이 게이머, 우하: 10시간 이상 플레이 게이머)

Fig. 4 Seven cognitive psychological features of no-pay/minnow/dolphin/whale (left up: over hours play gamer, right up: over hours play gamer, left down: over hours play gamer, right down: over 10 hours play gamer)

표 4 인지심리속성과 결제 금액과의 관계

Table 4 Relationship between cognitive psychological features and payment amount

cognitive psychological feature	Relationship with payment amount
Competitive	This feature is proportional to the payment amount. It is one of the features that show the biggest difference by payment amount.
Challenge	This feature is proportional to the payment amount. However, the difference between dolphin gamers are not large.
Loyalty	This feature is proportional to the payment amount, but the difference is not large.
Social	This feature is proportional to the payment amount. It is one of the features that show the biggest difference by payment amount.
Activity	No correlation.
Efficient	This feature is inverse proportional to the payment amount, but the difference is not large. However, minnow gamers are higher than the no-charge gamers. This is supposed to be because the no-charge gamer is not passionate about the game.
Sincerity	No correlation.

유틸리티 데이터들과 raw 로그를 기반으로 한 통계적 수치인 33개의 속성(TABLE 3)들을 사용하여 2017.06.05.~2017.06.27.동안 게임 내 결제 여부를 예측 하였다. 이 결과를 기준으로 7종의 인지심리속성들(경쟁성, 도전성, 성실성, 충성도, 사회적, 효율성, 활동성)을 사용하여

다시 예측을 수행하여 그 결과를 비교하였다. 구매예측에 사용된 데이터 셋은 표 6과 같고, 전체 게이머 수와 비교하여 구매 게이머들에 비해 비 구매 게이머들이 많아서 언더샘플링을 통해 데이터의 불균형을 맞췄으며 결정 트리, 랜덤 포레스트, 엑스트라 트리 기법으로 10-

표 5 결제금액 기준 게이머 유형별 인지심리 분석

Table 5 Cognitive psychological analysis by gamer type based on payment amount

type	Relationship with payment amount
noPay	This type has low cognitive psychological features overall, but 'sincerity' and 'efficient' features are relatively high. As a result, this type has little interest in the game and is enjoying the game lightly. However, in order to reduce the gap with other charging gamers, it seems that they are doing sincere and efficient actions in the game.
minnow	This type is a little higher cognitive psychological features overall than 'noPay' gamers but has lower cognitive psychological features than the other types. However, 'activity', 'efficient' and 'sincerity' features are relatively higher than the other types. This type is considered to be a gamer who enjoys the game with the minimum amount of payment in the game and play the game faithfully and efficiently in order to reduce the gap with other high charging gamers.
dolphin	This type has a highest 'social', 'loyal' and 'competitive' features except for whale gamer. However, the other features are relatively lower than the other types.
whale	This type has a highest 'social', 'loyal' and 'competitive' features, but the 'activity' and 'efficient' features are low. This is supposed to be an advance in competition with other gamers by replacing active benefits with in-game purchase.

표 6 구매 예측에 사용된 데이터 셋

Table 6 Data set used for purchase prediction

data	log	explain
Basic	41 kinds of logs	44 basic raw log counts
Primary	33 kinds of logs	33 statistical features
C&P	7 kinds of logs	7 cognitive psychological features

분할 교차 검증을 수행하였다.

먼저 3시간 이상 플레이시간 게이머를 기준으로 한 구매 예측 결과 7종의 인지심리 속성을 사용하는 것이 3개 알고리즘 모두 가장 높은 성능을 보였다. 특히 결정트리 알고리즘의 경우 매우 큰 성능 향상을 얻은 것을 볼 수 있다. 그러나 4시간 이상 플레이시간 게이머를 대상으로 구매 예측한 실험 결과에선 인지심리속성을 사용한 결정트리의 경우 raw로그를 기반으로 한 구매예측 보단 성능이 올랐지만, 33개의 1차 속성을 이용한 구매 예측보단 성능의 하락을 보였다. 다만, 결정트리가 아닌 그 외 알고리즘에선 모두 성능의 향상을 보이고 있다. 5시간 이상 플레이시간 게이머를 대상으로 한 구매예측 또한 인지심리속성을 사용한 구매예측에서도 전체적으로 약간의 성능향상을 이루었다. 비록 랜덤포레스트에서는 1차 속성대비 약간의 성능하락이 있었지만, 그 외 알고리즘은 모두 상승한 것을 볼 수 있다. 마지막으로 10시간 이상 플레이시간 게이머를 대상으로 한 구매예측 실험결과도 인지심리속성을 이용한 구매예측이 전체적으로 성능향상을 보이고 있는 것을 확인할 수 있다(표 7, 8, 9, 10). 결론적으로, 7개의 인지심리속성을 사용하여 구매예측을 하였을 때, 항상 성능이 상승하는 것은 아니지만, 대체적으로 우수한 성능을 보이고 있음을 확인할 수 있었다.

표 7 3시간 이상 플레이 게이머 기준 구매예측

Table 7 Purchase prediction of gamers with over three hours of play time

data set	algorism	accuracy	precision	recall	F1-score
Basic	DT	0.8129	0.8122	0.8161	0.8133
	RF	0.8774	0.8637	0.8984	0.8800
	ET	0.8653	0.8510	0.8871	0.8683
Primary	DT	0.8323	0.8390	0.8226	0.8304
	RF	0.8847	0.8715	0.9032	0.8866
	ET	0.8798	0.8608	0.9065	0.8827
C&P	DT	0.8452	0.8444	0.8468	0.8451
	RF	0.8871	0.8755	0.9032	0.8890
	ET	0.8831	0.8698	0.9016	0.8850

표 8 4시간 이상 플레이 게이머 기준 구매예측

Table 8 Purchase prediction of gamers with over four hours of play time

data set	algorism	accuracy	precision	recall	F1-score
Basic	DT	0.7966	0.7957	0.8006	0.7975
	RF	0.8533	0.8426	0.8704	0.8556
	ET	0.8363	0.8282	0.8509	0.8388
Primary	DT	0.8331	0.8369	0.8299	0.8322
	RF	0.8752	0.8676	0.8898	0.8773
	ET	0.8671	0.8558	0.8866	0.8698
C&P	DT	0.8145	0.8034	0.8348	0.8182
	RF	0.8801	0.8790	0.8833	0.8806
	ET	0.8777	0.8708	0.8898	0.8792

6. 결론 및 향후 연구

본 연구에서는 게임 내 로그로부터 게이머의 인지심리적인 분석을 위한 7가지 요소들을 추출하였고 이를 이용하여 게이머의 구매를 예측하는 실험을 진행하였다. 이 실험의 결과 게임 내에서 획득 가능한 기본 로그보다

표 9 5시간 이상 플레이 게이머 기준 구매예측
Table 9 Purchase prediction of gamers with over five hours of play time

data set	algorithm	accuracy	precision	recall	F1-score
Basic	DT	0.7882	0.7902	0.7866	0.7881
	RF	0.8461	0.8362	0.8632	0.8487
	ET	0.8265	0.8095	0.8567	0.8317
Primary	DT	0.8135	0.8149	0.8127	0.8136
	RF	0.8706	0.8677	0.8762	0.8715
	ET	0.8600	0.8577	0.8664	0.8610
C&P	DT	0.8219	0.8274	0.8145	0.8205
	RF	0.8682	0.8673	0.8713	0.8685
	ET	0.8641	0.8650	0.8664	0.8645

표 10 10시간 이상 플레이 게이머 기준 구매예측
Table 10 Purchase prediction of gamers with over 10 hours of play time

data set	algorithm	accuracy	precision	recall	F1-score
Basic	DT	0.7492	0.7495	0.7517	0.7498
	RF	0.8142	0.8002	0.8417	0.8197
	ET	0.7908	0.7847	0.8117	0.7961
Primary	DT	0.7617	0.7657	0.7567	0.7606
	RF	0.8308	0.8401	0.8217	0.8292
	ET	0.8150	0.8206	0.8067	0.8131
C&P	DT	0.7825	0.7728	0.8017	0.7863
	RF	0.8308	0.8459	0.8117	0.8274
	ET	0.8308	0.8426	0.8167	0.8282

게이머의 심리적 성향 특징이 고려된 추출정보를 활용하는 것이 게이머의 분석과 행동 예측에 더 도움이 된다는 것을 알 수 있었으며, 그 상세한 결과는 다음과 같다.

6.1 인지심리 분석

게이머의 기본적인 행동을 기록한 raw 게임 로그로부터 게이머의 인지심리속성들이 추출 가능하였다. 이를 통해, 직접적으로 드러나지 않는 게이머의 심리적인 성향이나 특징을 추정하여 분석하였고, 게이머의 분류에 따라 각각의 유형들은 각 유형만의 두드러진 특징을 보이고 있었다. 추가적으로, 추출한 인지심리 속성들은 게임에 특화되지 않은 일반적인 속성들로, 정규화(표준화)되어 있으며, 이를 통해 다른 게임에도 적용 가능하여 비교 가능할 것으로 생각된다.

6.2 게이머의 행동 예측의 측면

게이머의 구매 예측 실험을 진행한 결과, 7가지 인지심리 속성을 사용하여 구매예측 실험을 진행하였을 때 그 성능이 전체적으로 향상된 것을 보였다. 다만, 1차 속성을 이용한 구매 예측성능 대비 큰 성능향상을 보인 않지만, 속성들의 차원축소와 통합에 의의가 있다. 또한, 구매 예측에는 충성도와 활동성, 경쟁성이 가장 중요한 인지심리적인 요소로 나타났다.

6.3 추가적인 연구에서 고려될 수 있는 사항

본 연구를 통해 게임 내의 기본적인 로고데이터를 이용하여 게이머의 인지심리 요소들을 추출하여 분석이 가능하다는 것과 이를 이용하여 게이머의 행동 예측 성능을 높일 수 있음을 알 수 있었다. 그럼에도 불구하고 추가적으로 연구가 필요한 부분이 남아있다. 첫 번째로 고려될 수 있는 것은 추출한 게이머의 인지심리속성의 검증이다. 추출된 인지심리 속성이 해당 게이머의 인지심리적인 특성을 정확하게 반영하는지는 아직 검증되지 않았다. 따라서, 실제 추출한 게이머를 대상으로, 설문조사 등을 통하여 인지 심리적 특징을 추출하고, 해당 게이머의 기본적인 행동로그로부터 추출한 인지심리속성과의 비교분석을 통해 인지심리속성의 검증과정이 필요하다. 다음으로 고려할 수 있는 것은 다른 게임에 본 연구를 적용하여 인지심리속성을 추출하고 비교 분석하는 것이다. 본 연구에서 추출한 7가지 인지심리속성은 게임에 특화되지 않은 속성추출을 목적으로 하고 있으며, 다른 게임에도 적용 가능할 것이라는 가설을 가지고 있다. 또한, 게이머의 인지심리속성 가공에 있어서 다른 게임의 게이머 정보가 추가된다면, 보다 일반적인 게이머 인지심리속성이 만들어질 것이라는 기대가 있다. 다른 게임의 게이머들의 구매에 대해 학습이 이뤄지고, 이를 통하여 충분한 데이터가 없어 학습이 어려운 신규 게임이나 특별한 게임들의 게이머 구매예측을 시도해 볼 수 있다. 세 번째로, 단순한 트리기반 알고리즘이 아닌, 딥러닝과 같은 더 고차원적인 알고리즘을 통해 게이머의 이탈을 더 우수한 성능으로 예측해 볼 수 있다. 마지막으로, 7가지 인지심리속성의 추출 및 과정에 있어서 연구자의 도메인지식에 의존하는 추출을 할 수밖에 없었습니다. 이러한 방식은 연구자가 바뀌면 그 결과가 달라질 수 있기 때문에, 일관성 있는 7가지 인지심리속성의 추출 및 가공에 대한 연구가 필요합니다.

References

- [1] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motivation and Emotion*, Vol. 30. No. 4. pp.344-360, 2006.
- [2] A. K. Przybylski, C. S. Rigby, R. M. Ryan, "A motivational model of video game engagement," *Review of General Psychology*, Vol. 14. No.2. pp. 154-166, 2010.
- [3] M. V. Birk, C. Atkins, J. T. Bowey, R. L. Mandryk, "Fostering Intrinsic Motivation Through Avatar Identification in Digital Games," *CHI Conference on Human Factors in Computing Systems*, pp.2982-2995, 2016.
- [4] F. Hadji, R. Sifa, A. Drachen, C. Thureau, K. Kersting,

- and C. Bauckhage, "Predicting player churn in the wild," *IEEE Conference on Computational Intelligence and Games*, pp. 1-8, 2014.
- [5] A. Drachen, C. Thurau, R. Sifa, and C. Bauckhage, "A comparison of methods for player clustering via behavioral telemetry," arXiv preprint arXiv:1407.3950, 2014.
- [6] A. Drachen, N. Ross, J. Runge, and R. Sifa, "Stylized Facts for Mobile Game Analytics," *IEEE Computational Intelligence and Games (CIG)*, 2016.
- [7] kocca, "GLOBAL GAME INDUSTRY TREND," kocca, 2017.
- [8] Jayoung Jeong, "2016 Google Play Game Category Final Report," igaworks, 2016.
- [9] A. Clerico, C. Chamberland, M. Parent, P.-E. Michon, S. Tremblay, T. H. Falk, J.-C. Gagnon, P. Jackson, "Biometrics and Classifier Fusion to Predict the Fun-Factor in Video Gaming," *IEEE Conference on Computational Intelligence and Games*, 2016.
- [10] S. Ertekin, J. Huang, L. Bottou, L. Giles, "Learning on the border: active learning in imbalanced data classification," *Proc. of ACM conference on Conference on Information and Knowledge Management*, pp. 127-136, 2007.
- [11] 2015 Google Play Game Category Final Report, 2015



전 지 훈

2016년 세종대학교 컴퓨터공학과 졸업 (학사). 2016년~현재 세종대학교 컴퓨터공학과 석사과정. 관심분야는 인공지능, 기계학습, 데이터분석



양 성 일

1998년 연세대학교 컴퓨터과학과(박사수료). 1998년~2000년 아시아나항공SW연구소 주임연구원. 2000년~현재 한국전자통신연구원 책임연구원. 관심분야는 기계학습, 게임 인공지능, 자연어처리 등



김 경 중

2007년 연세대학교 컴퓨터과학과 졸업 (박사). 2007년~2009년 코벨대학교 박사후연구원. 2009년~2019년 세종대학교 컴퓨터공학과 조교수, 부교수. 2019년~현재 광주과학기술원 융합기술학제학부 부교수. 관심분야는 게임 지능, 강화학습,

자율주행차