Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.27, NO.6, Dec. 2017 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2017.27.6.1431

악성코드의 이미지 시각화 탐지 기법을 적용한 온라인 게임상에서의 이탈 유저 탐지 모델*

임 하 빈,† 김 휘 강, 김 승 주‡
고려대학교 정보보호대학원

Using Image Visualization Based Malware Detection Techniques for Customer Churn Prediction in Online Games*

Ha-bin Yim,† Huy-kang Kim, Seung-joo Kim‡
Center for Information Security Technologies(CIST), Korea University

요 약

보안 분야에서 악성코드나 이상 행위를 탐지하기 위한 보안 로그의 분석은 매우 중요하며, 악성코드를 탐지하기 위한 이미지 시각화 분석 기술은 많은 선행 연구를 통해 논의되어져 왔다. 이러한 분석 기술은 온라인 게임에도 적용될 수 있다. 최근 온라인 게임에서 악성코드나 게임 봇, 매크로 도구 등의 악용 사례가 증가하므로 인해 정상적으로 게임을 이용하려는 유저들의 이탈이 늘어나는 추세로 서비스의 운영자가 제시간에 필요한 조치를 하지 않을 경우 게임 산업 자체가 무너질 수 있다. 본 논문에서는 분석의 효율성을 향상시키기 위해 로그 파일을 PNG 이미지로 변환하는 방식을 사용한 새로운 이탈 예측 모델을 제안한다. 제안하는 모델은 이미지 변환을 통해 기존의 로그 크기에 비해 52,849배 경량화된 분석이 가능하며 특성 분석이 별도로 필요하지 않은 방식으로 분석에 소요되는 시간을 단축시켰다. 모델의 유효성 검증을 위해서 엔씨소프트의 블레이드 앤 소울 게임의 실제 데이터를 사용하였고, 분석 결과 97%의 높은 정확도로 잠재적인 이탈 유저를 예측할 수 있었다.

ABSTRACT

In the security field, log analysis is important to detect malware or abnormal behavior. Recently, image visualization techniques for malware dectection becomes to a major part of security. These techniques can also be used in online games. Users can leave a game when they felt bad experience from game bot, automatic hunting programs, malicious code, etc. This churning can damage online game's profit and longevity of service if game operators cannot detect this kind of events in time. In this paper, we propose a new technique of PNG image conversion based churn prediction to improve the efficiency of data analysis for the first. By using this log compression technique, we can reduce the size of log files by 52,849 times smaller and increase the analysis speed without features analysis. Second, we apply data mining technique to predict user's churn with a real dataset from Blade & Soul developed by NCSoft. As a result, we can identify potential churners with a high accuracy of 97%.

Keywords: Log Analysis, Visualization, Online Game Data Mining, Churn Prediction

Received(10. 26. 2017), Modified(11. 27. 2017), Accepted(12. 04. 2017)

* 본 연구는 미래창조과학부 및 한국인터넷진흥원의 "고용계약형 정보보호 석사과정 지원사업"의 연구결과로 수행되었음 (과제번호 H2101-17-1001)

† 주저자, habin103@korea.ac.kr
‡ 교신저자, skim71@korea.ac.kr(Corresponding author)

I. 서 론

사물인터넷의 등장으로 인해 위협을 발생시키는 공격의 범위가 확장됨에 따라 공격이 보안 사고로 진행된 이후에 대응할 경우 피해의 규모가 커지게 되었다. 따라서 사고가 발생하기 이전에 적정한 임계치를 정하고 이를 기준으로 이상 징후를 탐지하여 식별된 위험을 예방하는 것은 중요하다. 정보 시스템을 위협하는 공격이 확장됨으로 인해 악성 행위를 추적하기 위한 로그데이터의 양이 증가하고 있으며, 방대한 규모의 로그데이터를 분석하기 위해 악성코드의 탐지 방법으로 이미지 시각화 기법에 대한 연구가 이루어져왔다. 또한, 이미지 시각화 기법을 활용한 로그 분석은 디지털 포렌식 분야에서도 중요시되는데 대표적인 예로 은퇴하려는 사람들을 사전에 예측하여 정보 유출의 피해를 최소화하는 목적으로 사용된다.

이러한 탐지 기법은 온라인 게임에도 적용이 가능하다. 최근 온라인 게임 환경에서 게임 봇, 자동 사냥 프로그램, 계정 도용 등의 악성행위가 빈번하게 발생함에 따라 정상적으로 게임 서비스를 이용하는 유저들은 악성 유저들로부터 상대적인 박탈감, 게임 내 시장에서의 금전적인 피해 등으로 인해 게임의 이용을 중단하게 된다[1]. 온라인 게임은 플레이 방식의 특성상 각각의 유저마다 행동, 유저들과의 상호작용, 게임 내 재화거래 등에 따른 대량의 로그를 실시간으로 발생시키는 환경이기 때문에 악성행위의 발생에 따라 유저들의 로그데이터가 변화하게 된다. 본 논문에서는 유저들의 게임 내의 활동을 반영하고 있는 정보인 로그데이터를 악성코드의 탐지에 사용되는 이미지 시각화 방식을 통해 이탈할 유저를 예측한다. 서비스의 운영자는 탐지된 이탈 유저를 대상으로 악성 행위의 발생을 발견하고 사전에 방지할 수 있게 된다.

온라인 게임 유저들의 이탈을 예측하는 데에는 대표적으로 두 가지의 어려움이 있다. 첫째, 유저들이 대규모로 동시에 접속하여 게임을 진행하는 MMORPG 장르의 게임의 경우 매 초당 대량의 이벤트를 발생시키기 때문에 대량의 로그파일이 생성된다. 이러한 로그를 분석하여 이탈을 적시에 예측하기 위해서는 로그파일 저장과 가공 및 분석을 위한 대용량의 스토리지와 컴퓨팅 파워가 요구된다. 둘째, 각 게임마다 게임 디자인 (예: 플레이 방식, 길드 형성 구조, 아이템 거래 방식 등)이 상이하기 때문에, 분석할 게임별로 장르와 유저들의 플레이 패턴을 고려하여 게임 플레이에 영향을 줄 수 있는 특성(feature)을 파악하기 위한 분석과정이 필요하다.

앞선 제약사항으로 인해 기존 이탈 예측에 대한 연구로는 다양한 방법이 있지만 대부분 플레이 시간, 게임 캐릭터의 레벨, 게임 내 구매도 등과 같은 일부 특성만을 활용하여 분석하였다. 본 논문의 예측 모델은 온라인 게임에서 사용자마다 발생하는 텍스트 형식의 게임 로그를 이미지 파일 형식으로 변환하는 방법을 사용하여 게임 환경에 따른 별도의 특성 분석 없이도 이탈 유저를 예측할 수 있는 이탈 유저 탐지 모델을 제안한다.

II. 관련 연구

2.1 악성코드의 시각화 탐지 기법

개인정보 유출, 기업의 내부 기밀유출, 금전적인 피해와 같은 보안 사고를 유발하는 변종 악성코드 수의 기하급수적으로 증가로 인해 악성코드 분석가들의 수동 분석에 의존하는 시그니처 기반의 탐지와 같은 전통적인 악성코드 공격을 대응하기가 어려워졌다. 따라서 대량의 악성코드를 분석하기 위해 기존 방식에 비해 자동화되고 지능화된 형태의 악성코드 시각화 탐지 기법 연구에 대한 이루어져왔다.

Nataraj 외 [2]는 악성코드의 경우 대부분 코드 재사용으로 인해 구성되는 바이너리의 형태가 반복된다는 점에 착안하여 이미지 시각화 기술로 9,458개의 악성코드 샘플 중에서 총 25개의 악성코드 패밀리를 분류하였다. 악성코드의 바이너리를 Gray Scale 이미지로 변환하여 동일한 악성코드 패밀리의 경우 레이아웃과 텍스처가 시각적으로 유사하게 구성되는 점을 이용한 분류 방식을 수행한다.

Trinius 외 [3]는 허니팟으로 수집한 대량 악성코드 데이터베이스를 샌드박스 환경에서 동적 분석을 수행한 정보를 시각화한 분석 방법을 제안하였다. 해당 연구에서는 두 가지 시각화 방식을 사용하는데, 악성코드의 수행 과정과 개별 스레드의 동작을 트리 그래프와 스레드 그래프로 시각화한다. 결과적으로 악성코드의 시각적인 분류로 세부 보고서를 통한 분류보다 직관적인 분류와 예측이 가능함을 보였다.

Daniel 외 [4]는 시각화 도구로 악성코드의 역공학적 분석을 위한 전반적인 프로세스를 단축시키는 방법을 제안하였으며 디스크 및 메모리에서 실행될 수 있는 코드, 높은 엔트로피를 가지는 코드, 실행된

명령을 시각화하였다. 그래프로 표현된 실행에 따른 메모리 사용량, 코드들의 관계를 분석가가 직관적으로 판단할 수 있게 지원하는 도구를 통해 효율적인 악성코드 분석에 활용할 수 있도록 하였다.

Syed 외 [5]는 악성코드의 동작을 시각화하는 방식으로 변종 악성코드를 높은 정확성으로 식별하는 기법을 제안하였다. 해당 연구의 실험에서 악성코드를 VM에서 실행한 후에 동작을 캡처하고 악의적인 행동을 유발하는 API의 색상의 지정한 후 악성코드 행동에 따른 이미지를 생성한 방식을 사용한다. 지정 API calls를 기반으로 RGB 색상을 매칭시켜 악성코드의 패밀리를 구분하였으며, 결과적으로 악성코드의 패밀리에 따라 나타나는 변환된 이미지가 유사함을 이용하여 변종 악성코드를 탐지할 수 있음을 보였다.

2.2 온라인게임에서의 이탈자 예측

정보보안 리서치 기관인 Ponemon Institute에 따르면 처음 보안사고 발생 시 34%의 고객이 이탈하게 되며, 이후 보안사고가 발생할 경우엔 나머지 고객의 45%가 이탈한다고 조사되었다[6]. 이탈자에 대한 예측은 텔레커뮤니케이션[7], 구독 서비스[8], 인터넷 공급업체[9] 등 지속적인 유저의 서비스 이용이 유지되어야하는 분야에서 연구되고 있다. 본 논문의 분석 대상인 온라인 게임의 이탈 고객을 예측하기 위해 아래와 같은 연구가 이루어져왔다.

Zoheb 외 [10]는 MMORPG 상에서 플레이어 활동을 분석하기 위해 게임 로그를 분석하여 라이프 사이클 기반의 방식으로 일반 플레이어와 이탈이 예측되는 플레이어 사이의 활동 특성을 비교하였다. 특히 거리 기반의 분류 체계로 두 집단 사이의 서로 다른 행동 프로파일을 통해 분류하는 모델을 제안하였다.

Julian 외 [11]은 캐주얼 소셜 게임을 대상으로 4가지 공통 분류 알고리즘에 대한 예측을 비교하였으며, 분류 방법으로는 시간 역학에 따른 마르코프 모델로 구현하였다. 예측에 따른 비즈니스 영향을 평가하기 위한 A/B 테스트의 수행으로 주요 플레이어가 게임 내의 대량의 화폐를 획득할 경우에 게임 이용에 영향을 미치지 않고 간단한 인센티브로 주요 유저를 유지할 수 없음을 분석하였다.

Christian 외 [12]은 액션 어드벤처와 슈팅게임 장르의 25만명 이상의 플레이어를 가지는 5가지 게임 데이터를 대상으로 시간의 지남에 따른 플레이어의 참여 변화 양상에 대해서 분석하였다. 해당 연구에서는 Weibull 분류로 총 플레이시간의 경험적 분포에 대한 분석 기법을 제안하였다.

위와 같은 기존의 이탈유저에 대한 연구의 경우에는 플레이시간이나 행동과 같은 유저의 지속적인 서비스 이용에 영향을 미치는 일부 특성만을 활용하여 이탈을 예측하기 때문에 게임 봇이나 오토 프로그램의 사용 등과 같은 악성 행위에 따른 유저의 로그데이터에 대한 특성은 고려되지 않는다는 한계가 있다. 본 논문에서 제안하는 모델은 각 유저마다 발생하는 방대한 양의 로그데이터를 이미지로 변환시켜 수동으로 분석하기 어려운 특성 분석 작업에 소요되는 시간을 줄이고, 유저의 게임 내 모든 활동을 고려한 탐지가 가능하다.

2.3 이미지 분류 기법

이미지 분류의 일반적인 과정[13]은 Fig. 1.과 같다. 먼저 이미지의 전처리 과정에서 불필요한 노이즈 데이터, 누락된 데이터, 불안정한 데이터 등의 처리 과정을 거쳐 머신러닝에 사용되는 데이터의 집합을 효과적으로 사용할 수 있도록 한다. 이후 전처리된 데이터에서 특성을 추출하고 분류에 사용될 주요 특성

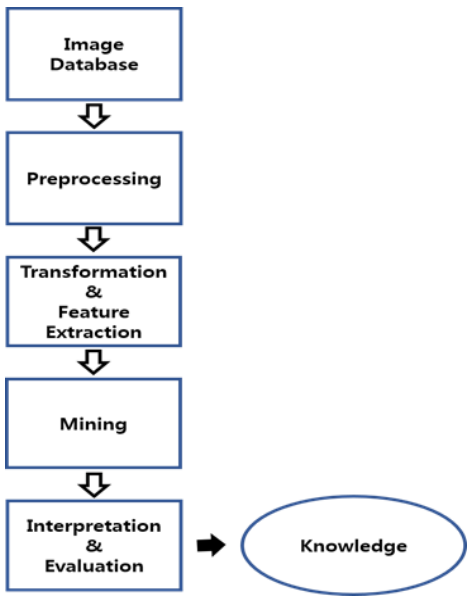

Fig. 1. image mining process

Fig. 2. PNG image based Churn Prediction Model

을 선정하고, 주요 특성을 통해 패턴을 발견하기 위한 과정을 수행한다. 최종적으로 발견된 패턴의 분석과 평가를 통해 의미 있는 결과를 얻을 수 있게 된다.

2.4 PNG 포맷 형식

Portable Network Graphics(PNG)는 무손실 그래픽 파일 포맷으로 1977년에 Jacob Ziv 등에 의해 새로운 종류의 무손실 데이터 압축 알고리즘에 대한 논문으로부터 고안되었다. 일반적인 그래픽 포맷에 비해 압축률이 더 높으며, 8비트 알파 채널을 이용해 다양한 투명층을 지원한다. 트루컬러의 지원으로 24비트의 색상을 표현할 수 있으며 16,777,216개의 색상을 사용할 수 있다[14].

III. 이탈유저 탐지 모델

온라인 게임 환경의 특성상 각 사용자에 대한 로그 데이터를 분석에 사용 가능한 특성의 수가 다양하기 때문에 모든 특성을 고려해서 분석하는 작업에 대한 볼륨이 매우 크다.

따라서 우리는 로그 데이터를 PNG로 변환하고, 변환된 이미지를 분류하는 알고리즘을 사용하여 이탈 고객을 예측하는 모델을 제안한다. 모델에서 분석에 사용된 로그 데이터는 레이블(label)정보가 포함된 2,000명의 데이터로 Table. 1.에 샘플링 비율에 따른 결과를 비교하였다. 실험에 사용된 데이터는 샘플링 결과에 따라 가장 높은 정확도를 보이는 9:1비율의 1,800명의 비이탈자 데이터와 200명의 이탈자 데이터로 구성하였다. 제안하는 모델에서 이탈 고객의 예측까지 Fig. 2.와 같은 3단계의 과정을 거친다.

Table. 1. An Result of sampling according to the rate

Rate (+ : -)	Precision	Recall	Accuracy	F-measure
9 : 1	1.00	0.94	0.97	0.97
8 : 2	0.85	0.83	0.84	0.84
7 : 3	0.90	0.89	0.90	0.89
6 : 4	0.87	0.90	0.89	0.89
5 : 5	0.91	0.97	0.94	0.94
4 : 6	0.81	0.89	0.84	0.85
3 : 7	0.82	0.88	0.85	0.85
2 : 8	0.82	0.95	0.87	0.88
1 : 9	0.90	1.00	0.95	0.95

3.1 실험 데이터

이탈 고객의 분석을 위해 사용하는 데이터는 엔씨소프트의 대표 MMORPG 서비스인 블레이드 앤 소울의 로그 데이터이다. 실험에 사용되는 데이터는 IEEE CIG 2017에서 제공하는 데이터 셋[15]을 대상으로 한다. 로그 데이터는 2016년 3월 16일부터 2016년 5월 11일 사이에 발생하였으며, 총 77개의 필드로 구성되어 있다. 각각의 로그 데이터는 캐릭터의 행동, 상태, 이벤트를 나타낸다.

본 논문의 분석대상인 로그데이터의 필드는 Common, Actor, Object, Target 으로 구성된다.

- Common : 모든 로그에 공통적으로 포함되는 정보로 행동이 발생한 시간(TIME), 캐릭터의 행동에 대한 ID(Log ID) 등이 포함된다.
- Actor : 게임 내의 행동 주체에 대한 정보로 캐릭터 ID(Actor_ID), 계정 ID(Actor_Account_ID), 게임 내의 캐릭터 직업(Actor_Job), 캐릭터 레벨(Actor_Level) 등이 포함된다.
- Object : 게임 내의 행동에 동반되는 정보로 게임 내의 아이템 ID(Entity_ID), 등급(Entity_Grade) 등이 포함된다.
- Target : 게임 내 캐릭터 행동에 영향을 받는 대상에 대한 정보로 대상 캐릭터 ID(Target_I

D), 계정 ID(Target_Account_ID), 게임 내의 캐릭터 직업(Target_Job), 캐릭터 레벨(Target_Level) 등이 포함된다.

로그데이터 중에서도 플레이어의 주된 행동에 대한 필드인 LogID는 82개의 행동을 나타낸다. 행동은 Connection, Character, Item, Skill, Quest, Guild로 이루어지며, Fig. 3.과 같이 LogID에 따라 영향을 받는 다른 필드의 데이터가 변하게 된다.

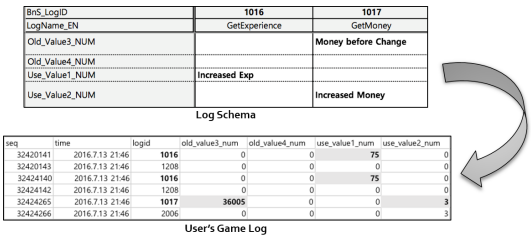

Fig. 3. An Example of changed User's Game Log according to LogID

3.2 로그데이터의 이미지

사용자 별로 생성된 각각의 CSV 파일 형태의 데이터는 Binary 형식의 파일로의 변환이 이루어지며, 이때 생성된 Binary 파일을 PNG 파일로 변환한다. 위 과정을 거친 변환 결과 Fig. 4.과 같은 이미지가 생성된다.

CSV에서 Binary로의 변환은 포맷에 따른 바이너리 형태로의 변형을 수행하며, Binary에서 PNG로의 변환은 python 스크립트[16]를 사용한다. 출력 이미지의 각 픽셀은 3Byte를 인코딩하는데, 각 바이트는 빨간색, 녹색, 파란색 채널에 인코딩된다. 모든 파일은 0으로 테일 패딩 되고 3배수인 바이트 수를 갖게 된다. 스크립트에서는 가능한 3배수에 가깝게 이미지 크기가 자동으로 계산되고 최소한의 패딩을 필요로 하는 여러 차원이 존재할 경우에 사각형에 가장 가까운 차원이 선택된다. 온라인 게임의 유저는 유저의 게임 내 활동 빈도에 따라 로그데이터의 양이 달라지므로 변환된 이미지의 사이즈를 일정하게 고정시켜주는 작업이 필요하다. 본 논문의 분석 대상인 데이터는 총 3,000명의 CSV 파일로 평균 크기는 17.01MB이며, 위의 방식을 통해 PNG로 변환할 경우에는 가장 높은 정확도를 보이는 20x20사이즈로 변환 시에 각 파일의 평균 크기는 337.49Byte로 기존의 파일에 비해 분석 대상의 크기가 현저하게 줄어들게 된다.

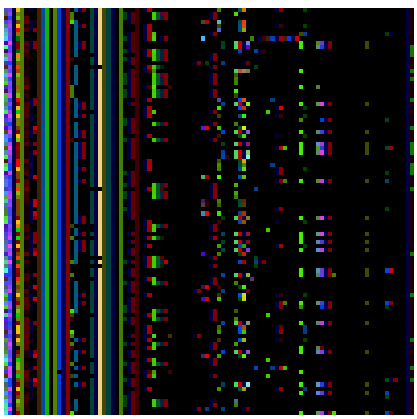

Fig. 4. An Example of transformed PNG(100x100) according to User's Game Log with total 92,808 records

3.3 모델에서 사용되는 이미지 분류

Bag of Word(BoW)는 텍스트를 분류하기 위해 사용되는 모델로 다양한 분야에서 연구[17-19]되었다. 본 논문에서 사용하는 Bag of Feature(BoF) 방식은 BoW 토대로 한 방식을 이용한다. 반직관적인 형태의 BoW는 스팸 필터링 및 주제 모델링에서 뛰어난 성능을 보이며, 각 단어를 독립적으로 가정한다. BoF 모델은 이미지 내의 극소 특징 벡터의 출현 빈도를 기준으로 이미지를 근사 표현하는 방식을 사용한다. BoW에서 분류에 사용하는 단어로 Visual Word가 사용되며 BoF에서도 특징 벡터가 양자화 과정을 거쳐서 출현 빈도를 계산하는 방식의 Visual Word가 사용된다. 그러므로 BoF 모델의 경우에는 학습 과정을 거침에 따라 탐지의 기준이 되는 특징점이 계속해서 변하기 때문에 온라인 게임 내 로그데이터의 변화에 따른 학습이 가능하다. BoF의 분류 과정은 Fig. 5.[20]와 같이 동작한다. 우리는 BoF 모델의 적용을 위해 학습에 필요한 데이터를 분류하는 작업을 수행한다. 전체 데이터에서 Train Data를 랜덤으로 추출하여 학습 데이터로 사용하고, 나머지 데이터를 Test Data로 사용한다. 실험에서는 BoF 모델에 이미지를 입력 받고 Train Data에서 특성을 추출하였다. 이후 k-means 클러스터링 방식을 통해 500개의 word visual

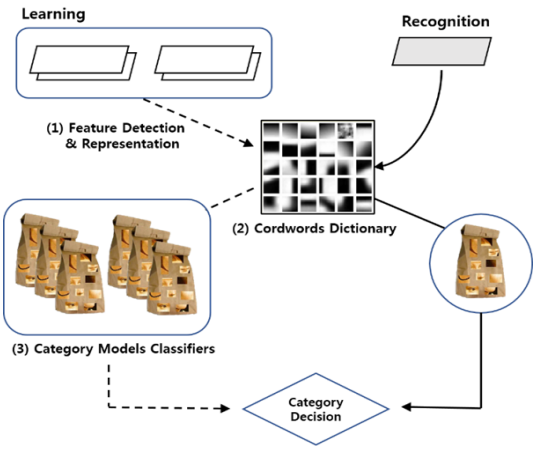

Fig. 5. Bag of Feature Classification Process

vocabulary를 생성하면 특성이 선정된다. 선정된 특성을 통해 Train Data로 카테고리에 따라 분류하고, Test Data로 분류기를 평가한다. 실험에서는 이미지 변환 과정에서 대상 데이터를 최대 100x100 사이즈에서 최소 10x10 사이즈로 변환한 이미지를 대상으로 학습에 사용되어지는 데이터의 비율에 따른 정확도를 평가하였다. 이미지의 사이즈는 줄어들수록 변환된 이미지에 포함되는 로그데이터가 손실될 수 있으나 이미지의 분류과정에서 불필요한 정보가 제거되어 정확도(Accuracy)에 영향을 준다. Fig. 6.는 제안하는 모델의 정확도를 측정한 결과이며 실험 결과 20x20 사이즈일 때 97%로 가장 높은 정확도를 보인다.

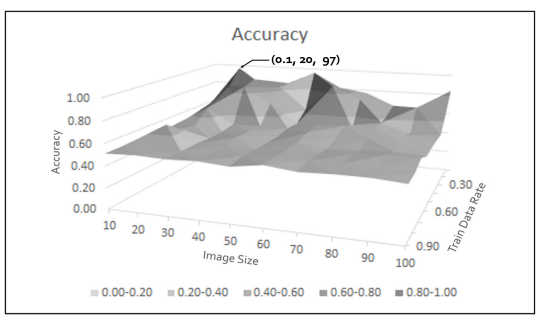

Fig. 6. An Analysis of the Change of accuracy according to the rate of used train data and image size

IV. 이탈유저 예측 모델의 성능 평가

본 논문에서 제안하는 예측 모델의 성능을 평가하기 위해서 Recall, Precision, F-measure[21]를 측정하여 정량적 비교평가를 수행한다. 평가에서 False Positive는 이탈자 데이터를 비이탈자 데이터로 분류, False Negative는 비이탈자 데이터를 이탈자 데이터로 분류한 것을 나타낸다. 성능평가에 사용된 성능 지표에 대한 수식은 Table. 2.와 같다.

Recall은 실제의 이탈자 중에서 탐지된 이탈자의 비율을 나타내며, 1에 가까운 높은 Recall 값을 가질수록 미탐율 (False Negative Ratio)이 적다. 제안하는 탐지 모델의 Recall은 Fig. 7.와 같으며, 20x20 사이즈일 때 94%로 가장 높은 Recall 값을 갖는다.

Precision은 분류된 이탈자 중에서 실제의 이탈자의 비율을 나타내며, 1에 가까운 높은 Precision을 가질수록 오탐율(False-Positive Ratio)이 적다. 제안하는 탐지 모델의 Precision은 Fig. 8.과 같으며, 최대 100%에 해당하는 Precision 값을 갖는다.

Recall과 Precision은 상호보완적 관계에 있기

Table. 2. Formulas of Accuracy, Precision, Recall, and F-measure

Evaluation Method	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TN}{TP+FN}$
F-meaure	$2*\frac{\mathrm{Precision}*Recall}{\mathrm{Precision}+Recall}$

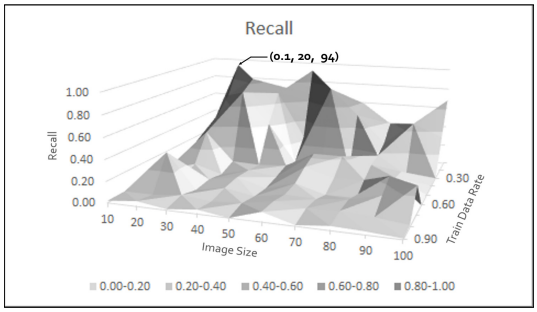

Fig. 7. An Analysis of the Change of Recall according to the rate of used train data and image size

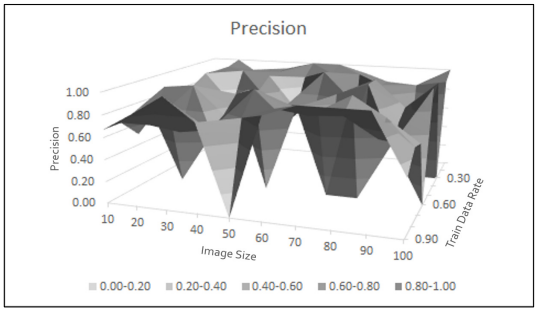

Fig. 8. An Analysis of the Change of Precision according to the rate of used train data and image size

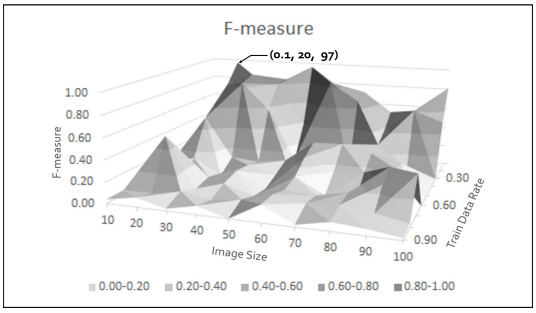

Fig. 9. An Analysis of the Change of F-measure according to the rate of used train data and image size

때문에 이를 보완하기 위해서 F-measure를 활용한다. F-measure는 Recall과 Precision을 결합한 방식의 조화 평균값을 나타내며, 1에 가까울수록 탐지 모델의 예측율이 높음을 의미한다. 제안하는 탐지 모델의 F-measure은 Fig. 9.과 같으며, 20x20 사이즈일 때 97%로 가장 높은 F-measure 값을 갖는다.

기존 연구들의 이탈 예측 정확도를 살펴보면 Zoheb 외 [10]의 논문은 앙상블 분류기를 사용하여 최대 77.25%의 F-measure 값을 보인다. Thomas 외 [22]는 콘텐츠와 유저의 관계를 분석한 방법으로 최대 90%의 recall 값을 가지지만 Precision 값이 28%로 다소 낮다. 최근 이탈 예측 연구로 Paul 외 [23]는 일일 로그인 횟수, 게임 내의 구매, 플레이 시간 등을 특성으로 조건부 추론 생존 앙상블 분류를 적용하여 이탈 가능성을 예측하고 AUC(area under the ROC curve)로 평가하였다. 예측 결과의 최대 AUC값은 0.960으로 본 논문에서는 가장 정확도가 높은 20x20 사이즈일 때

Table. 3. Comparison of AUC value

Research	Algorithm	AUC
PNG Image Conversion Based Churn Prediction	Bag of Feature	0.969
Paul etc.[23]	Survival Ensemble	0.960
	Support Vector Machines	0.940
	Naive Bayesian	0.900
	Decision Tree	0.934

AUC값이 0.969로 더 높은 수치를 갖는다. 제안하는 이탈 예측 모델의 경우 별도의 특성에 대한 분석이 필요가 없다는 측면에서 효율적이면서도 높은 탐지율을 보인다.

V. 결 론

온라인 게임 환경에서는 게임 봇, 오토 프로그램, 독소 행위(toxic-behaviors) 등의 악성 행위가 빈번하게 발생함에 따라 정상적으로 게임을 이용하고자 하는 유저들은 상대적인 박탈감을 느끼고 게임 서비스를 이탈하게 된다. 따라서 이탈이 발생하기 전에 사전에 유저의 이탈을 예측할 수 있어야하며, 보안 담당자는 예측 정보를 활용하여 악성 행위에 대응할 수 있어야한다. 기존의 이탈유저에 대한 예측방식은 플레이시간, 게임 내 참여도 등 일부 특성만을 기준으로 이탈을 분석하였다. 온라인 게임은 플레이 환경에서 실시간으로 상호작용이 활발하게 이루어지고 유저의 모든 기록이 로그데이터에 기록되기 때문에 악성 행위에 따라 일반 유저의 로그데이터는 영향을 받는다. 본 논문에서는 로그데이터의 분석을 위해 악성코드의 시각화 탐지기법을 사용하여 별도의 특성분석에 필요한 시간이 소요되지 않는 효율적인 이탈 탐지기법을 제안한다. 최종적으로 이탈 유저에 대한 예측의 정확도는 최대 97%로 나타났다.

References

[1] Woo, Jiyoung, and Huy Kang Kim. “Survey and research direction on online game security.” Proceedings of the Workshop at SIGGRAPH Asia. ACM, 2012.

[2] Nataraj, Lakshmanan, et al. “Malware

images: visualization and automatic classification." Proceedings of the 8th international symposium on visualization for cyber security. ACM, 2011.

[3] Trinius, Philipp, et al. "Visual analysis of malware behavior using treemaps and thread graphs." Visualization for Cyber Security, 2009. VizSec 2009. 6th International Workshop on. IEEE, 2009.

[4] Quist, Daniel A., and Lorie M. Liebrock. "Visualizing compiled executables for malware analysis." Visualization for Cyber Security, 2009. VizSec 2009. 6th International Workshop on. IEEE, 2009.

[5] Shaid, Syed Zainudeen Mohd, and Mohd Aizaini Maarof. "Malware behavior image for malware variant identification." Biometrics and Security Technologies (ISBAST), 2014 International Symposium on. IEEE, 2014.

[6] Ponemon Institute - Measuring Trust In Privacy And Security. https://www.ponemon.org/

[7] Wei, Chih-Ping, and I-Tang Chiu. "Turning telecommunications call details to churn prediction: a data mining approach." Expert systems with applications 23.2 pp. 103-112. 2002.

[8] Coussement, Kristof, and Dirk Van den Poel. "Churn pre-diction in subscription services: An application of support vector machines while comparing two parameter-selection techniques." Expert systems with applications 34.1 pp. 313-327. 2008

[9] Khan, Afaq Alam, Sanjay Jamwal, and M. M. Sepehri. "Applying data mining to customer churn prediction in an internet service provider." International Journal of Computer Applications 9.7 pp. 8-14. 2010. G. O. Young, "Synthetic structure of industrial plastics," in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill. pp. 15-64. 1964.

[10] Borbora, Zoheb H., and Jaideep Srivastava. "User behavior modelling approach for churn prediction in online games." Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). IEEE, 2012.

[11] Runge, Julian, et al. "Churn prediction for high-value play-ers in casual social games." Computational Intelligence and Games (CIG), 2014 IEEE Conference on. IEEE, 2014.

[12] Bauckhage, Christian, et al. "How players lose interest in playing a game: An empirical study based on distributions of total playing times." Computational Intelligence and Games (CIG), 2012 IEEE conference on. IEEE, 2012.

[13] Fatma, Shaikh Nikhat. "Image mining method and frameworks." International Journal of Computational Engineering Research (Ijceronline. Com) pp. 135-145. 2012

[14] PNG (Portable Network Graphics) Specif ication, Version 1.1 http://www.libpng.org/pub/png/spec/1.1/PNG-Rationale.html#R.PNG-file-signature

[15] CIG 2017. Game Data Mining Competition. https://cilab.sejong.ac.kr/gdmc2017/

[16] bin2png. https://github.com/ESultanik/bin2png

[17] Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." Machine learning: ECML-98 pp. 137-142. 1998.

[18] Lodhi, Huma, et al. "Text classification using string kernels." Journal of Machine Learning Research. pp. 419-444. Feb. 2002

[19] Cristianini, Nello, John Shawe-Taylor, and Huma Lodhi. "Latent semantic kernels." Journal of Intelligent Information Systems 18.2 pp. 127-152. 2002.

〔20〕 Visual Object Detection, Recognition & Tracking (without Deep Learning) https//www.slideshare.net/yuhuang/visual-object-detection-recognition-tracking-without-deep-learning
〔21〕 Powers. David M. W, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," Journal of Machine Learning Technologies, No.2, pp. 37-63. 2011.
〔22〕 Debeauvais, Thomas, et al. "Retention and progression: Seven months in World of Warcraft." FDG. 2014.
〔23〕 Bertens, Paul, Anna Guitart, and África Periáñez. "Games and Big Data: A Scalable Multi-Dimensional Churn Prediction Model." arXiv preprint arXiv:1710.02262. 2017.

〈저자소개〉

임 하 빈 (Ha-bin Yim) 학생회원
2016년 2월: 순천향대학교 정보보호학과 졸업
2016년 3월~현재: 고려대학교 정보보호대학원 석사과정
〈관심분야〉 온라인 게임 보안, 시스템 보안, 데이터 마이닝

김 휘 강 (Huy-kang Kim) 종신회원
1998년 2월: KAIST 산업경영학과 학사
2000년 2월: KAIST 산업공학과 석사
2009년 2월: KAIST 산업및시스템공학과 박사
2004년 5월~2010년 2월: 엔씨소프트 정보보안실장, Technical Director
2010년 3월~2014년 12월: 고려대학교 정보보호대학원 조교수
2015년 1월~현재: 고려대학교 정보보호대학원 부교수
〈관심분야〉 온라인게임 보안, 네트워크 보안, 네트워크 포렌식

김 승 주 (Seung-joo Kim) 종신회원
1994년~1999년: 성균관대학교 정보공학과 (학사, 석사, 박사)
1998년 12월~2004년 2월: KISA(舊 한국정보보호진흥원) 팀장
2002년~현재: 한국정보통신기술협회(TTA) IT 국제표준화전문가
2004년 3월~2011년 2월: 성균관대학교 정보통신공학부 조교수, 부교수
2011년 3월~현재: 고려대학교 사이버국방학과/정보보호대학원 정교수
2004년~현재: 한국정보보호학회 이사
2005년~2006년: 교육인적자원부 유해정보 차단 자문위원
2007년: 국가정보원장 국가사이버안전업무 유공자 표창
2007년~2009년: 전자 정부 서비스 보안 위원회 사이버 침해사고대응 실무위원회 위원
2010년: 방송통신위원회 정보통신망 침해사고 민관합동조사단 위원
2012년 3월~2012년 6월: 선관위 디도스 특별검사팀 자문위원
2013년 4월~2013년 12월: IT보안인증사무국 자문위원
2013년 9월~현재: 중앙선거관리위원회 자문위원
2014년 3월~현재: 헌법재판소 자문위원
2014년 12월~현재: 카카오 자문위원
2016년 1월~현재: 한국정보화진흥원 자문위원
〈관심분야〉 보안공학, 암호이론, 정보보증, 정보보호제품 보안성 평가, Usable security