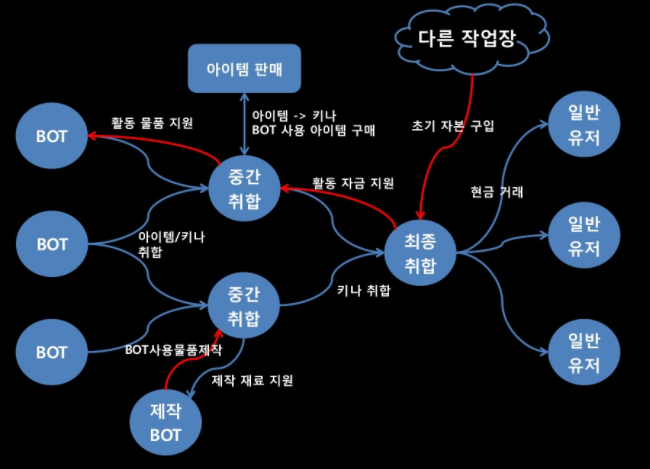
**R 기반 GFIS 게임 내 부정행위 탐지**

201402308 이상인

1. 서두

컴퓨터 및 모바일 환경이 비약적으로 발전하면서 전세계적으로 이를 이용한 게임 산업이 빠르게 발전하고 있다. 국내 앱(스마트폰 응용프로그램) 분석 업체인 와이즈랩에서 안드로이드 폰 사용자들을 대상으로 한 분석(2017.08)에서 사용자들이 하루 평균 10.2회씩 게임을 플레이 하는 것으로 조사됐다. 그 중 NCSOFT(이하 ‘NC’)의 ‘리니지M’이라는 게임은 한 달 동안 약 28억분이라는, 모바일 게임 중 최장시간을 기록했고 매출은 2256원으로 모바일 게임 업계 최고 매출을 기록했다. 이 모바일 게임과 이 게임의 원조인 ‘리니지1’ ‘리지니2’는 30-40대 직장인을 겨냥한 게임으로 아이템 하나에 몇 천 만원씩 오가고 최고가 아이템이 5억원에 달하는 등 돈이 없으면 할 수 없는 대표적인 게임들이다. 이렇다 보니 게임 내에서 계정 도용(다른 사람의 계정을 도용해서 보유한 게임 아이템/자산을 몰래 처분)이나 게임 어뷰징(게임 버그를 악용하여 게임 아이템 무한 복사, 보스 몹 무한 사냥 등을 수행)이 타 게임에 비해 예민할 수 밖에 없다. 이에 NC는 Game Fraud Investigation Service를 통해 자기 유사성(Self-Similarity)을 이용한 이상 탐지 및 네트워크 분석을 통한 작업장 배후 탐지를 하둡을 기반한 R로 수행하였다.

[그림 1] ‘리니지’ 작업장 구조



2. 과거 실태 와 R 적용

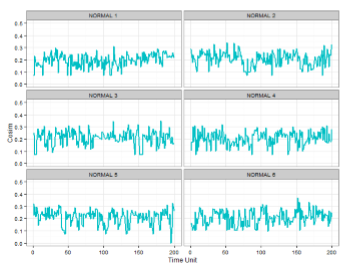
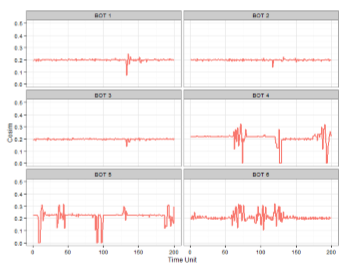
현 작업장의 구조를 보면 [그림 1]과 같다. 다수의 활동 BOT을 매크로를 통해 조종해서 아이템과 키나(화폐 단위)를 획득한다. 중간 취합 캐릭터는 BOT가 생성한 아이템과 키나를 취합하고 BOT이 아이템을 생성할 때 필요한 재료와 키나를 지원하거나 일부 아이템을 판매하여 모두 키나로 환전 한다. 최종 취합 캐릭터는 중간 취합 캐릭터들이 가지고 있는 키나를 모두 취합한 후 일반 유저와 현금 거래를 한다. 이러한 ‘작업장’은 이전까지 신고를 확인하고 캐릭터를 차단하는 방식으로 차단하였다.

하지만 하둡과 R의 도입으로 다양한 알고리즘을 적용해서 이를 해결 할 수 있는 기반이 마련되었다. 게임 로그 raw data를 하둡에 Pig를 이용해 분산처리(MapReduce)하여 저장하였고 R은 높은 유연성, 하둡과의 쉬운 연동, 다양한 패키지의 장점이 있어 어뷰징, 계정도용, 작업장을 탐지할 수 있는 모델의 사전 분석 및 모델링에 이용하였다. 여기서는 작업장을 탐지하는 방법에 대해 집중하여 다루고자 한다.

3. 이상 탐지

이상 탐지란 게임 내에서 버그를 활용하는 플레이를 탐지하는 기법이다. 아이템을 무한으로 복사한다든지, 보스 몹을 무한 사냥한다든지 하는 행동들을 탐지할 수 있으며 작업장의 BOT은 (일반 캐릭터와 비교하여) 오직 효율적인 동작만 하기 때문에 캐릭터의 패턴을 분석하여 이상 탐지를 수행한다. 이 탐지의 기본 가정은 다음과 같다. 첫 째, BOT은 미리 정해진 패턴에 따라 반복적 플레이를 수행한다. BOT이 행동 할 매크로를 만들 때, 뚜렷한 목적을 위해 몇 가지 행동들을 반복하므로 실제 이용자들이 할 수 있는 불필요한 행동들을 BOT은 하지 않는다. 둘 째, 실제 이용자들이 비슷한 패턴을 보일 수는 있으나 오래 행동하지는 못한다. 실제 이용자들은 같은 행동을 계속 반복한다 하더라도 지루함을 느끼는 등 다양한 이유로 동일한 행위를 오랫동안 하지 않는다.

[그림2] BOT(위)와 HUMAN(아래)의   
자기 유사성 그래프



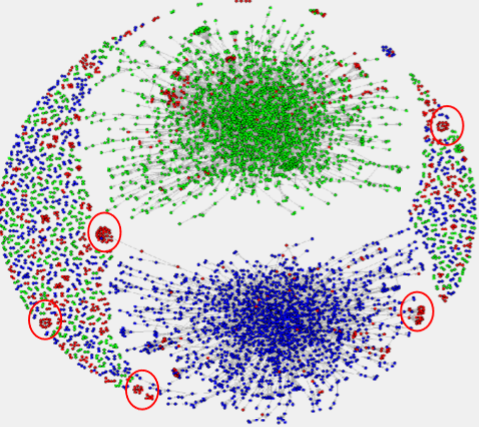
자기 유사성을 이용한 로그 탐지를 하기 위해서는 각 캐릭터 별 자기 반복적인 경향을 정량화 한 후 이 정량화된 수치가 높은 캐릭터들을 탐지한다. 정량화를 하기에 앞서 분 별로 어떠한 캐릭터가 어떠한 행동을 했는지 로그 기록을 벡터화한다. 예를 들어 12시 1분에 ‘앉기’가 0번, ‘엎드리기’ 1번, ‘사다리타기’ 1번, ‘휴식’이 3번 이루어지면 (0,1,1,3)으로 기록을 벡터화한다. 분 별로 변환된 벡터를 을 이용해서 정량화한다. 이를 time index의 그래프로 나타내어 일정한 간격으로 동일한 행동을 하는 BOT인지 아닌지 볼 수 있고 Cosine similarity에 대한 표준 편차 계산 후 자기 유사도 값으로 변화시켜 box plot으로 볼 수 있다. [그림2]는 BOT과 인간의 자기 유사도를 그래프로 표현하였는데 BOT의 경우 일정 시간마다 같은 값으로 행동의 변화가 나타나는 반면 인간은 매 분마다 다양한 행동을 하여 BOT에 비해 그래프가 산발적임을 알 수 있다. 이를 학습 집합을 이용하여 자기유사도 값을 BOT일 확률로 전환시킨다. 이렇게 캐릭터 별 자기 유사도 변화 추이를 분석하여 오토를 비 상시적으로 이용하는 BOT을 탐지할 수 있다. 오토를 비 상시적으로 사용하는 BOT의 경우 행동 패턴 변화에 의한 탐지율 저하 방지 방안이 필요하므로 지수가중 이동평균모형을 통해 과거 오토 확률과의 상관계수 변화량을 추적하여 정해진 범위를 넘어선 변화 발생 시 모델 수정 혹은 학습 집합을 재 구축 하였다.

4. 작업장 배후 탐지

이상 탐지처럼 캐릭터가 BOT인지 아닌지 판별하는 분석이 있는가 하면 작업장의 머리와 몸통인 중간 취합 캐릭터(이하 BankerA)와 최종 취합 캐릭터(이하 BankerB)를 잡는 기법인 작업장 배후 탐지가 있다. 기존 BOT과 BankerA의 거래내역을 추적하기에는 BOT의 수가 많았기 때문에 BOT보다는 수가 적고 탐지하였을 때 아이템과 키나를 회수할 수 있는 양이 월등히 많은 BankerA와 Banker B의 거래 추적이 효과적이다. 작업장 배후 탐지는 다음과 같은 가정을 따른다. 첫 째, BankerA는 BankerB에게 아이템과 키나를 무상으로 준다. BankerB가 자금을 지원해 주는 경우도 있지만 BankerA가 무상으로 물품을 건네주는 횟수에 비해서 미미한 횟수이다. 둘 째, 작업장은 일반유저의 보다 가 많다. BankerA만 하더라도 BOT과의 거래, BankerB와의 거래로 한 작업당 수십 단위씩 거래가 이루어지므로 일반 유저의 거래보다 훨씬 많지만 거래 한 단위당 유저 간 거래의 최대치를 넣어 거래를 하므로 일반 유저의 것보다 많다. 이를 추적하기 위해 네트워크 분석을 통한 시각화를 진행하고 일반 유저를 작업장으로 판단하는 것(False Positive)을 막고 BankerB의 추적을 Contagion Model을 통해 진행한다.

작업장 배후 계정을 탐지하기 위해 유저 데이터 분석 및 작업장 의심 캐릭터를 추출한다. 이 캐릭터들의 사회 활동을 추출하기 위해 관계 네트워크를 구축을 하여 시각화를 하는데 비슷한 특성의 캐릭터끼리 모이는 경향을 확인할 수 있다. 이러한 특징을 확인한 후 그래프 클러스터링을 통해 작업장의 클러스터링을 진행하며 가중치 값은 거래 횟수를 사용한다. 거래량을 가중치 값으로 사용하기에는 수천가지의 아이템 가치가 정량화되기 어렵고 일반 유저는 값비싼 아이템을 거래하는 한편 중간 취합 캐릭터는 값이 싼 일반적인 아이템을 대량으로 거래하기 때문이다. [그림 3]을 보면 거래량에 따라서 유저간 뭉치는 모습이 다른 것을 확인할 수 있는데 독립적으로 많이 뭉치는 그룹(빨간 원 표시)을 작업장으로 의심할 수 있다. 오토가 많이 있는 작업장을 구분하였으면 Banker 캐릭터의 탐지를 진행한다. 캐릭터 간 관계 및 불량성을 정량화하여 감염 로직을 적용하기 위해 을 계산한다. 이를 통해 주로 키나를 취합하는 BankerB가 BankerA 캐릭터들과의 거래를 통해 재화량이 갑자기 높아짐을 확인하여 작업장 자체를 막을 수 있다.

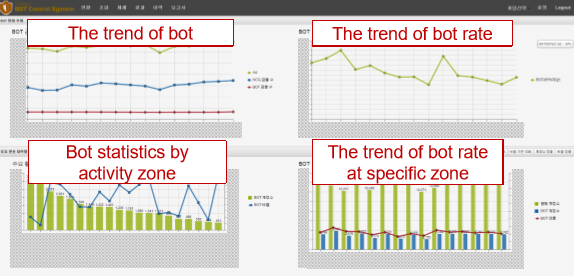
[그림 3] 거래 모형의 네트워크 시각화



4. 결과

NC는 위의 모델을 기반으로 업데이트를 진행하였고 Banker 캐릭터들의 제재를 통한 자산 압류로 작업장에 실질적인 경제적 타격을 주었다. 기존에는 서버 내 키나 량 대폭 증가에 따른 인플레이션이 심하였는데 첫 모델을 서버에 적용한 이후 1383개의 계정을 탐지하였고 약 1조 6천억 키나, 당시 현금 가치로 약 1.5억원을 환수하는 결과를 내었다. 하지만 제재가 계속되고 탐지 범위가 확대되자 902개의 계정이 False Positive로 분석 오류를 내어 다수의 고객 항의를 받기도 하였다. 이후 분석 결과에 대한 증거 자료 보강, GM팀의 자료 검증 절차 추가와 매주 정기적 분석 및 제재 활동 수행을 통해 2016년 기준 총 92차례 BOT 배후 계정을 분석하였고 약 3만개의 BOT 배후 계정 제재 및 20조 키나, 현금 가치 약 25억원을 회수할 수 있었다.

[그림 4] BOT Detection System Dashboard



5. 느낌

필자는 다양한 게임을 하면서 BOT들을 많이 보았었기 때문에 이러한 BOT제재를 GM들이 직접 하나씩 하는가에 의문을 가지게 되었고 이번 조사를 통해 BOT을 기반으로 한 작업장의 실체를 알아볼 수 있었다. 자료를 찾아보기도 하고 직접 관계자에게 메일을 보내 답변을 받아 보고서를 쓰면서 다양한 모델과 알고리즘들이 실제로 어떻게 적용되는지 알 수 있어서 매우 흥미로웠다. BOT의 경우에는 아이디 자체가 의미없는 문자와 숫자들의 나열로 지정되는 경우가 많으므로 이를 분석하여 1차적으로 BOT을 판별하는 알고리즘을 만들면 유용할 것 같다고 느꼈다. 특히 R이 학생들한테는 많이 사용되고 있지만 실제로 직접 사용하는 사례가 기존에 알고있던 바로는 간단한 시각화 정도였고 그 부분도 요즘에는 tableau가 대체하고 있는 시점에서 R의 유용성을 다시 확인할 수 있는 좋은 시간을 가질 수 있었다.

[출처]

이지나 외 2명, Detecting the bank character in MMORPGs by analysis of a clustered network, NCSOFT

이은주 외 4명, You are a Game Bot: Uncovering Game Bots in MMORPGs via Self-similarity in the Wild, NCSOFT (사내 발표자료)

이은주, Data Analysis for Game Fraud Detection, Nexon Developers Conference 2014(컨퍼런스 발표 자료)

윤서호. (2017). [IGC2017] "게임은 애들만 즐기는 것이 아니다" 통계로 본 한국 모바일 게임 시장 분석, INVEN, http://www.inven.co.kr/webzine/news/?news=184916