

게임봇(Game Bot) 탐지 트랙

데이터셋 설명

2019. 10. 1.

목차

| | | |
|-------|------------------------|---|
| 1 | 챌린지 문제 | 2 |
| 2 | AION 데이터셋 | 2 |
| 2.1 | 기본정보..... | 2 |
| 2.1.1 | 파일 정보..... | 2 |
| 2.1.2 | 로그 아이디(log_id)..... | 2 |
| 2.1.3 | 용어 및 컬럼 참고사항 | 2 |
| 2.2 | 유저 선별 | 3 |
| 2.3 | 분석용/제출용 유저 계정 목록 | 3 |
| 2.4 | 개인정보 등 데이터 수정사항 | 3 |

1 챌린지 문제

주어진 AION 데이터셋을 분석하여 게임봇을 탐지할 수 있는 알고리즘을 제안하는 것이 본 트랙의 주요 목표입니다. 알고리즘은 룰, 머신러닝, AI 기반 등 형태에 제한은 없으나, 프로그램으로 구현이 가능해야 합니다. (즉, 자동화가 가능해야 합니다.)

2 AION 데이터셋

2.1 기본정보

2.1.1 파일 정보

데이터셋은 일자별 로그 파일(log_20100416.csv ~ log_20100430.csv)로 나누어져 총 15개 파일입니다. 압축 해제 시 약 112GB의 용량입니다. 문자 인코딩은 UTF-8입니다.

2.1.2 로그 아이디(log_id)

유저의 행위 등 게임 내에서 발생하는 이벤트들은 로그 아이디를 기준으로 세부적으로 나누어 집니다. 예를 들어 유저의 월드 입장 이벤트는 103번, 유저의 스킬 습득은 401번의 로그 아이디 값을 가집니다.

로그 아이디 값에 따라 데이터의 각 컬럼에 기록되는 정보의 종류가 달라지므로 유의하시기 바랍니다. 로그 아이디별 기록되는 내용은 로그정의서 [AION datasheet for challenge.xlsx] 파일에서 살펴보실 수 있습니다.

2.1.3 용어 및 컬럼 참고사항

- **PC:** Player character로 유저가 직접 플레이하는 캐릭터를 의미합니다.
- **NPC:** Non-player character로 유저에 의해 컨트롤되지 않는 캐릭터를 의미합니다. 상점 NPC, 몬스터 등이 포함됩니다.
- **big_log_id 컬럼:** log_id의 백자리 수를 나타내며 로그 이벤트의 큰 카테고리를 의미합니다.
- **천족/마족:** AION에서 플레이 가능한 종족은 천족과 마족으로 나누어지며, 서로 대립하는 구도입니다.
- **레기온(Legion):** AION에서 길드를 나타내는 용어입니다.
- **어비스(Abyss):** 천족과 마족 간의 대규모 전투 및 요새전 콘텐츠를 의미합니다.

2.2 유저 선별

제공된 로그에는 더 많은 유저가 포함되어 있으나, 로그 기간 중 3시간 이상 플레이한 유저만을 분석 대상으로 합니다. 상세한 선정 기준은 아래와 같습니다.

- ① 접속 시 5분마다 주기적으로 발생하는 160번 로그를 카운트하여 캐릭터별 예상 접속시간 계산
- ② 예상 접속시간이 3시간 이상인 경우 캐릭터 선별하여, 해당 캐릭터를 소유한 계정을 대상으로 선정
(일반적으로 유저는 하나의 게임 계정을 보유하고 있으며, 계정마다 여러 캐릭터를 생성하여 플레이할 수 있습니다.)

2.3 분석용/제출용 유저 계정 목록

데이터셋을 다운로드 받으면 2가지 유저 계정 목록이 포함되어 있습니다. 각 파일의 목적은 다음과 같습니다.

- **[labeled_account.csv]** 분석 편의를 위해 클래스가 구분된 8,100개의 계정
(일반유저 0, 게임봇 1)
- **[test_account.csv]** 클래스가 구분되지 않은 계정 목록.
해당 계정에 대해 로그를 분석하여 게임봇 탐지 결과를 제출해야 합니다.

[labeled_account.csv]에서 게임봇으로 분류된 계정은, 실제로 당시에 봇(자동화 툴) 활동이 관찰되어 제재한 계정입니다. 이를 참고하여 [test_account.csv]의 계정 중 제재해야 할 게임봇 계정을 찾아주시기 바랍니다. 여기서 계정 식별자는 로그정의서에 "계정UID"로 기록된 값을 의미하며, 주로 actor_account 컬럼에 위치합니다.

2.4 개인정보 등 데이터 수정사항

실제 로그에서 개인정보에 해당하는 사항은 삭제하였습니다.

- 삭제 대상: 캐릭터명, 계정명, 레기온명 등 유저가 임의로 기재 가능한 텍스트.
- 유저의 이름이나 다른 개인정보가 포함되어 있을 가능성이 존재하여 일괄 삭제하였습니다. 로그정의서 상에는 노란색 셀로 표시되어 있습니다.

실제 로그에서 캐릭터UID, 계정UID, IP 주소에 해당하는 사항은 다른 값으로 변환하였습니다.

- 튜토리얼에서 소개하는 AION 데이터를 분석한 기존 연구에서, 웹사이트 상에 탐지결과를 일부 오픈하고 있습니다. 캐릭터UID와 계정UID 값으로 이와 대조할 경우 정답이 노출될 여지가 있기 때문에 분석 대상 캐릭터 및 계정은 다른 값으로 변환하였습니다.
- IP 주소의 경우 개인정보이나 작업장의 게임봇 대량 운영 등에 초점을 맞추어 분석할 시 필요한 정보일 수 있기에, 다른 값으로 변환하여 제공합니다. 단, IP 주소의 표준 형태를

따르지는 않으며 임의 정수값으로 변환하였습니다.

- 로그정의서 상에는 주황색 셀로 표시되어 있습니다.