

행동 데이터 분석 인공지능 AI 경진대회

human_learning

18등(상위 20%)

주최: (사)한국인공지능협회 / 주관: DACON

2020.03.01 ~ 2020.04.15

1. 대회 개요

- 목적
 - 행동 데이터(스타크래프트의 경기 기록 데이터)를 이용하여 경기 승자와 패자 예측
- 채점 방식
 - AUC
- Public Score & Private Score
 - Public Score – 대회 중 test 데이터의 30%로 채점
 - Private Score – 대회 이후 Private Score 채점에 쓰이지 않은 test
로 채점

2. 주어진 데이터

- Train
 - 학습 세트
 - Column: game_id, time, winner, player, event, event_contents
- Test
 - 예측 대상이 되는 세트
 - Column: winner 컬럼을 제외한 Train 컬럼과 동일
- Sample Submission
 - 채점을 위해 업로드하는 파일의 예시

3. Columns

- Game_id - 경기 구분 기호
 - Ex) 1, 2
- Winner - 선수
 - 0: player 0, 1: player 1
- Time - 경기 시간
 - ex) 2.24 = 2분 24초
- Player - 선수
 - 0: player 0, 1: player 1
- Species - 종족
 - T: 테란, P: 프로토스, Z: 저그

- Event - 행동 종류
 - Ability : 생산, 공격 등 선수의 주요 행동
 - AddToControlGroup : 부대에 추가
 - Camera : 시점 선택
 - ControlGroup : 부대 행동
 - GetControlGroup : 부대 불러오기
 - Right Click : 마우스 우클릭
 - Selection : 객체 선택
 - SetControlGroup : 부대 지정
- Event_Contents - 행동 상세

데이터 분석 설명

1. 주요 가설들 (Ability외의 항목)

1. Column Time

- 경기의 **Running Time**이 승패에 영향을 줄 것이다.
- 종족의 상성에 따라 초반에 유리한 종족과 후반에 유리한 종족이 존재

2. Column Event – Value: Camara

- 카메라를 움직인 사람이 승자일 확률이 높을 것이다.
- **카메라 좌표의 분산**으로 피쳐 생성

3. Column Event – Value: Right Click

- 오른쪽 클릭은 이동이나 공격을 의미
- **Right Click행동의 횟수**가 많은 플레이어가 승리할 확률이 높을 것이다.

1. 주요 가설들 (Ability)

4. Column Event – Value: Ability가 가장 중요하다.
 - Ability외의 Event에는 행동 상세에 큰 정보가 없다.
 - **Ability의 Event_content에는 여러 정보가 줄글로 존재한다.**
 - 따라서 이것을 주된 피쳐로 사용해야한다.

5. Column Event – Value: Ability – 건물 건설
 - 어떤 건물을 건설하는지에 따라 빌드가 달라짐.
 - 따라서 어떤 건물을 한 게임 안에 건설하는지가 중요
 - 더 나아가 건물에는 지을 수 있는 순서가 존재하는데
어느 건물을 얼마나 건설했는지는 중요한 변수

1. 주요 가설들 (Ability)

6. Column Event – Value: Ability – 유닛 생성

- 같은 건물을 건설하더라도 생산할 수 있는 유닛은 여러 가지
- 따라서 어떤 유닛을 얼마나 생산했는지도 중요
- 더 나아가 유닛도 상위의 유닛과 하위의 유닛이 존재하는데
어느 위치에 존재하는 유닛을 얼마나 생산했는지는 중요한 변수

7. Column Event – Value: Ability – 업그레이드

- 유닛을 뽑더라도 이 유닛을 잠깐의 방어를 위해 뽑았는지, 아니면 주력으로 사용할 것인지는 업그레이드 여부로 판단할 수 있음
- 따라서 어떤 업그레이드를 했는 가는 중요
- **공격에 대한 업그레이드, 방어에 대한 업그레이드로 분류하여 Count**

2. 전처리 및 피쳐 엔지니어링

1. 전처리 - Event 컬럼이 Ability인 행의 Event_contents
 - 정보들이 줄글로 적혀 있기 때문에, 어떠한 글들이 적혀있는지 추출
 - 그 중에서 앞서 제시한 가설의 내용을 포함한 행들을 추출하여 집계
 - 피쳐 생성
2. 피쳐 엔지니어링 - 상대적인 피쳐 생성
 - 게임은 두 플레이어가 하는 것
 - 따라서 두 사람의 행동 차이가 게임의 승패를 결정한다고 판단
 - 이를 반영하기 위해 피쳐들간의 차이를 피쳐로 생성

2. 전처리 및 피쳐 엔지니어링

3. 피쳐 엔지니어링 – 피쳐 폭발

- 가설에 해당하는 모든 피쳐를 추가하고 성능향상에 한계
- 이 성능의 한계를 해결하기 위해 피쳐 폭발 사용
- 피쳐들 간의 모든 조합으로 차이를 구하여 피쳐로 추가

4. 피쳐 엔지니어링 – 피쳐 선택

- 피쳐 폭발로 느려진 학습 속도 개선 목적
- 피쳐 폭발로 생긴 피쳐들 중에 성능에 악영향을 주는 피쳐 제거 목적
- Feature Permutation 사용

3. 모델링

1. Light Gradient Boosting Model




- 입력 데이터의 피쳐의 수를 고려할 때, 학습 속도가 빠른 모델이 필요
- (데이터 수, 피쳐 선택을 하기 전의 피쳐 수) = (38872, 1288)
- 트리기반 모델 중에서 제일 성능이 준수

2. Bayesian Optimization




- 파라미터 튜닝을 위해 사용
- 10번의 초기 그리드 탐색과 100번의 베이지안 최적화 과정을 통해 최고 성능 모델로 선택

4. 최종성능 및 순위

Private Score

#	팀		최종점수	제출수	등록일
18	human_learning	  	0.7027	15	2년 전

Public Score

#	팀		점수	제출수	등록일
27	human_learning	  	0.70395	15	2년 전