



분석 과제 - Data Scientist

주제

(1) 레드 팀 정글러 챔피언 '리신'에 대해서 1분 30초부터 3분 30초까지의 초반 정글 동선을 분석하고자 합니다. 몇 가지 동선으로 분류할 수 있을까요? 제공된 챔피언 위치 데이터를 사용하여, '리신'의 초반 정글 동선이 유사한 경기들을 분류해 주세요.

분석이 어려운 영상의 경우에는 분석에 직접적으로 활용하지 않으셔도 되지만, 활용하기에 어려웠던 이유를 같이 남겨주시기 바랍니다.

(2) 제공된 인게임 타임라인 데이터를 사용하여, (1)에서 수행한 방법을 업그레이드하려고 합니다.

어떤 데이터를 함께 사용하면 위치 데이터 분석을 고도화할 수 있을까요?

어떤 데이터를 사용할지, 어떤 방식으로 할 수 있을지, 결과는 어떻게 나올지 정리해 주세요.

과제 진행 방법

- 과제 분석 언어
 - 개인이 잘할 수 있는 분석언어를 선택하여 데이터 분석을 진행해주시면 됩니다.
- 제출 형태
 - 분석 과정에서의 내용들은 최대한 자세히 문서화하여 제출해주시면 됩니다.
 - 문서의 형태는 PPT 혹은 노션 (PDF) 으로 제출 가능합니다.
 - 결과물은 압축된 하나의 파일 형태로 제출해주세요.
 - 코드를 함께 제출할 필요는 없으나, 기술 면접이 진행될 때 코드 리뷰가 있을 수 있습니다.
- 마감 기한

- 과제 마감은 메일이 도착한 시점으로부터 7일 뒤 24시(자정)까지입니다.

활용 데이터

데이터 개요

- 분석 대상 게임은 'League of Legends' 입니다.
- 19 경기에 대한 데이터를 제공합니다.
 - 데이터는 각 경기마다 3가지 형태로 주어집니다.
 - 메타데이터
 - 챔피언 위치 데이터
 - 게임 결과 데이터 & 인게임 타임라인 데이터
- 모든 경기에서 챔피언 '리신'이 '레드 팀'의 '정글러' 포지션으로 플레이 되었습니다.
- 챔피언 위치 데이터의 경우, 모든 경기의 데이터가 0분 0초부터 시작하지는 않습니다.

메타 데이터

분석에 사용하는 19 경기의 메타 정보를 정리한 값 입니다.

- **Name**: 해당 경기의 영상 이름
- **MatchID**: 해당 경기의 고유 ID
- **StartTime**: 인게임에서 해당 영상의 녹화가 시작된 시간
 - ex) 1:20, 인게임 시간 기준 1:20 부터 녹화된 영상입니다.
- **EnemyJungler**: 적 정글러 챔피언의 이름
 - ex) Diana, 상대 팀의 정글러 챔피언은 Diana입니다.
- **Winner**: 승리한 팀
- **Patch**: 해당 경기의 패치 버전

메타 데이터

Name	MatchID	Video Url	StartTime	EnemyJungler	Winner	Patch
bidqBtK0Hd8	5344643717	https://www.youtube.com/watch?v=bidqBtK0Hd8	1:20	Diana	Blue	11.15
oRMVvIriewU	5334576251	https://www.youtube.com/watch?v=oRMVvIriewU	1:22	Graves	Blue	11.14

Name	MatchID	Video Url	StartTime	EnemyJungler	Winner	Patch
dAGL5xrydUo	5326569918	https://www.youtube.com/watch?v=dAGL5xrydUo	1:07	Olaf	Red	11.14
3EzTJ4tmXKA	5326750258	https://www.youtube.com/watch?v=3EzTJ4tmXKA	1:23	Viego	Red	11.14
KbTmkrALaLM	5307156651	https://www.youtube.com/watch?v=KbTmkrALaLM	1:04	Gwen	Red	11.13
ZhoROIQluXw	5291299018	https://www.youtube.com/watch?v=ZhoROIQluXw	0:06	Nidalee	Red	11.13
m_jhvtqmoE	5357852875	https://www.youtube.com/watch?v=m_jhvtqmoE	1:14	Diana	Red	11.15
0l-1-U7bcgY	5353602405	https://www.youtube.com/watch?v=0l-1-U7bcgY	1:12	XinZhao	Red	11.15
NPL3AENowDo	5353719037	https://www.youtube.com/watch?v=NPL3AENowDo	1:15	Viego	Red	11.15
jOUK8efu2EM	5344240480	https://www.youtube.com/watch?v=jOUK8efu2EM	1:15	Nidalee	Red	11.15
SV5_cPym_7o	5359855377	https://www.youtube.com/watch?v=SV5_cPym_7o	0:08	Gwen	Red	11.15
Sodn83u_kFQ	5369042922	https://www.youtube.com/watch?v=Sodn83u_kFQ	0:02	XinZhao	Red	11.15
uk4g0gt-NQE	5362506298	https://www.youtube.com/watch?v=uk4g0gt-NQE	0:02	Viego	Red	11.15
uu03GFr1wU	5360191762	https://www.youtube.com/watch?v=uu03GFr1wU	0:02	Olaf	Red	11.15
sNW-KwhVqZs	5357406162	https://www.youtube.com/watch?v=sNW-KwhVqZs	0:02	Rumble	Red	11.15
wDE_YLycT4M	5357748665	https://www.youtube.com/watch?v=wDE_YLycT4M	0:02	XinZhao	Red	11.15
VYnijgikia4	5353629623	https://www.youtube.com/watch?v=VYnijgikia4	0:02	Sejuani	Red	11.15
OZpVbV-CK2s	5369856875	https://www.youtube.com/watch?v=OZpVbV-CK2s	1:01	Viego	Blue	11.15
3CW1nm5C1w	5368955793	https://www.youtube.com/watch?v=3CW1nm5C1w	1:04	Nidalee	Red	11.15

챔피언 위치 데이터

- LOL 영상에 딥러닝 모델을 사용해서 1초마다 추출한 챔피언들의 위치 데이터입니다.
 - 최대 10명 (블루 팀 5명, 레드 팀 5명)의 챔피언에 대한 위치 데이터가 존재할 수 있습니다.
 - 모델이 챔피언을 감지하지 못해서 위치 데이터가 누락된 경우가 있을 수 있습니다.
- **opgg_data_project.tar.gz**
 - 19개 영상에 대한 위치 데이터와 챔피언 검출 결과 영상이 포함되어있습니다.

- 데이터 구조
 - opgg_data_project / {Name} / labels / {챔피언 위치 데이터}_crop_second.txt
 - opgg_data_project / {Name} / {Name}_crop.mp4
- 예시
 - opgg_data_project / bidqBtK0Hd8 / labels /bidqBtK0Hd8_crop_1.txt 파일을 예시로 데이터를 설명드리겠습니다.
 - 행마다 6개의 값이 있습니다.
 - class id, (바운딩 박스 중심의 좌표 x, y), (바운딩 박스의 크기 width, height), confidence

```
53 0.426259 0.311151 0.0899281 0.0899281 0.825161
17 0.127698 0.266187 0.0899281 0.0935252 0.877195
27 0.636691 0.794964 0.0935252 0.0935252 0.883645
46 0.516187 0.609712 0.0899281 0.0899281 0.89579
51 0.419065 0.830935 0.0899281 0.0935252 0.897473
128 0.874101 0.419065 0.0935252 0.0971223 0.927462
```

- class id는 챔피언의 이름과 1:1로 매핑되는 값이며, champion_class_id.yaml을 참고해주세요.
 - ex) 53 → Kassadin
- 바운딩 박스 중심의 좌표 x, y는 이미지 크기 288에 대해서 0~1까지 normalization 되었습니다.
 - ex) 0.426259 0.311151 → 122.762592 89.611488
- 바운딩 박스의 크기는 이미지 크기 288에 대해서 0~1까지 normalization 되었습니다.
 - ex) 0.0899281 0.0899281 → 25.8992928 25.8992928
- confidence는 딥러닝 모델이 미니맵에서 챔피언의 위치를 검출할 때 나온 확률 값이며, 0~1까지 normalization 되었습니다.
- LOL 영상의 StartTime을 기준으로 데이터가 저장되었습니다.
 - ex) bidqBtK0Hd8의 경우 StartTime 1:20부터 위치 데이터가 저장된 것이며, bidqBtK0Hd8_crop_1.txt는 인게임 시간 1:20일 때 챔피언들의 위치 데이터를 나타내고, bidqBtK0Hd8_crop_2.txt는 인게임 시간 1:21일 때 챔피언들의 위치 데이터를 나타냅니다.
- https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/champion_coordinate_data/opgg_data_project.tar.gz
- **champion_class_id.yaml**
 - 챔피언에 대한 class id를 나타냅니다.
 - ex) Aatrox = 0 ~ Zyra = 158
 - ex) LeeSin = 67
 - https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/champion_coordinate_data/champion_class_id.yaml

게임 결과 데이터 & 인게임 타임라인 데이터

- 게임 결과 데이터
 - 게임의 결과 값이 포함된 데이터입니다.
 - 예시 :
 - <https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/processed/{MatchID}.json>
 - 원하는 영상의 MatchID 값을 링크에 넣어주세요.
 - ex) <https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/processed/5291299018.json>
- 인게임 타임라인 데이터
 - 개별 경기를 1분 단위로 스냅샷을 찍은 형태의 인게임 데이터가 포함된 json 파일입니다. 타임라인 json 파일은 이름에 (1) 이 붙어있습니다.
 - 예시 :
 - {MatchID} (1).json
 - 타임라인 json 파일의 경우에는 timestamp를 이용해서 인게임 데이터를 매핑할 수 있습니다.
 - 밀리세컨드 기준이므로, 초 단위로 사용하시려면 `timestamp/1000/60` 으로 환산해주세요.
 - [https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/processed/{MatchID}+\(1\).json](https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/processed/{MatchID}+(1).json)
 - 원하는 영상의 MatchID 값을 링크에 넣어주세요.
 - ex) [https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/processed/5291299018+\(1\).json](https://opgg-data-analytics.s3.ap-northeast-2.amazonaws.com/recruit/202108/processed/5291299018+(1).json)