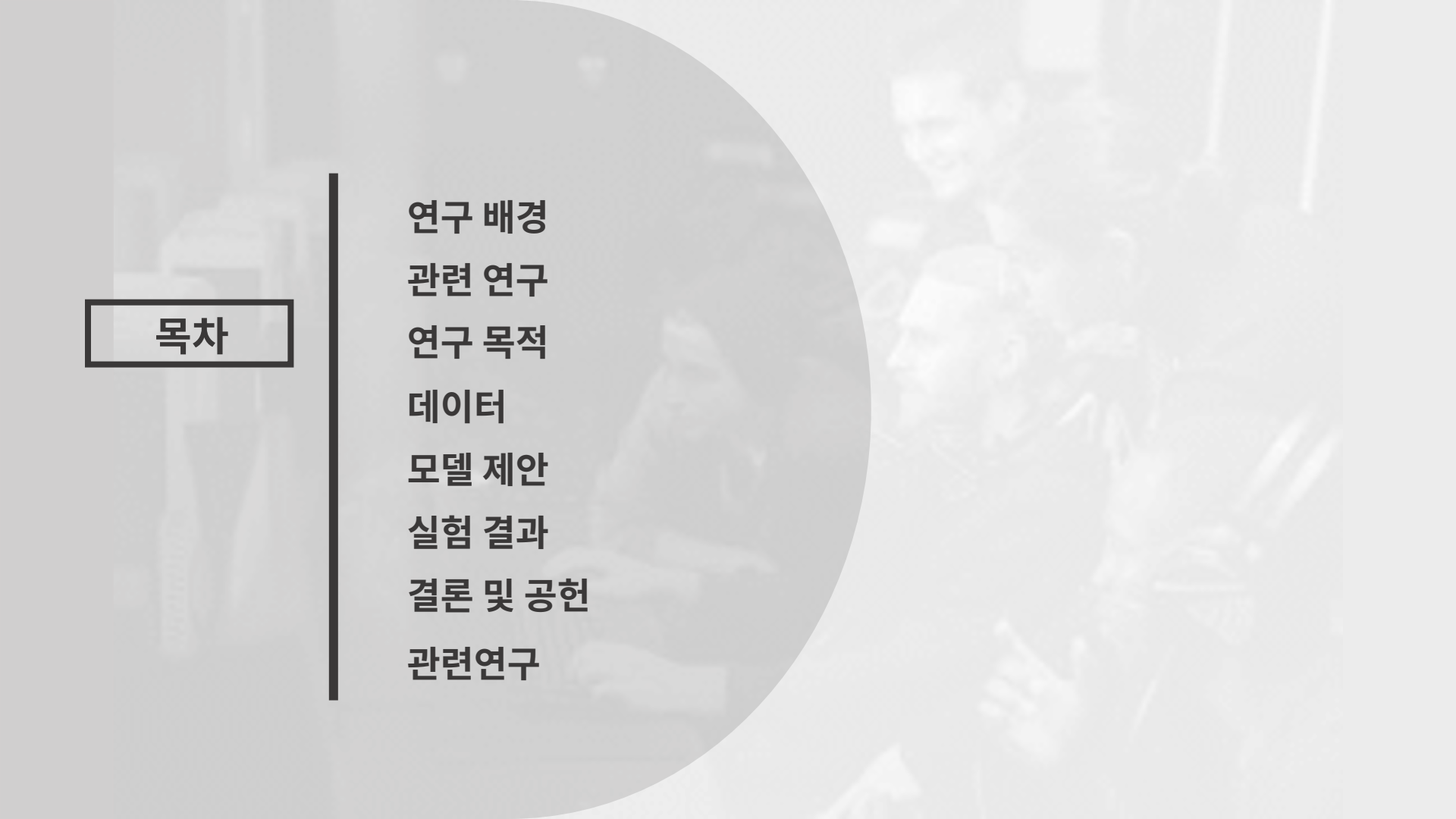


멀티모달 학습을 이용한 E-sports 하이라이트 자동 추출 방법

박강민,현혜인



목차

연구 배경

관련 연구

연구 목적

데이터

모델 제안

실험 결과

결론 및 공헌

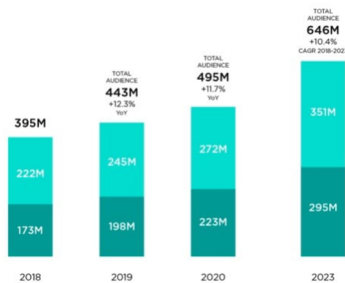
관련연구

E-sports의 시장 규모

연 평균 10% 성장

2018년 3억 9500만명

2020년 4억 9500만명



(출처 : 한국콘텐츠진흥원)

E-sports 하이라이트 시청자 수

평균 10만뷰 시청

2019 롤 월드컵 하이라이트 영상 평균 10만명 시청

롤 전문 하이라이트 제공 유튜브 채널 25만명 구독

노동 집약적인 하이라이트 영상 제작

영상 길이 대비 최소 10배 시간 소요

평균 35분 39초 게임 중계 영상에 대해 10분 내외 길이 제작

5분 내외의 영상 편집에 5만원 이상 비용발생

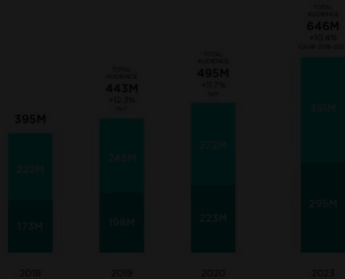
영상 제작시, 영상 길이 대비 최소 10배 시간 소요

E-스포츠의 시장 규모

연 평균 10% 성장

2018년 3억 9500만명

2020년 4억 9500만명



(출처 : 한국콘텐츠진흥원)

E-스포츠 하이라이트 시청자 수 AI 기술을 활용한

평균 10만뷰 시청

2019 롤 월드컵 하이라이트 영상 10만명 시청

롤 전문 하이라이트 제공 유튜브 채널 25만명 구독

하이라이트 자동 추출 필요

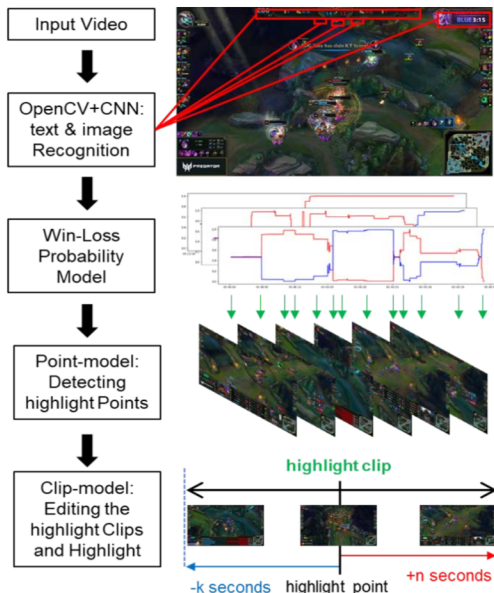
노동 집약적인 하이라이트 영상 제작

평균 영상 제작 X 시간 소요

평균 35분 39초 게임 중계 영상에 대해 10분 내외 길이 제작

5분 내외의 영상 편집에 5만원 이상 비용발생

10분 내외 영상 제작 평균 X 시간 소요



데이터

- 2년간 개최된 공식 경기의 **최종 스코어 및 승,패 데이터**를 Riot developer API를 사용해 수집
- 경기중 화면에 표시되는 스코어 데이터를 OCR과 CNN을 통해 수집
- 경기 하이라이트 통계 데이터 활용

모델

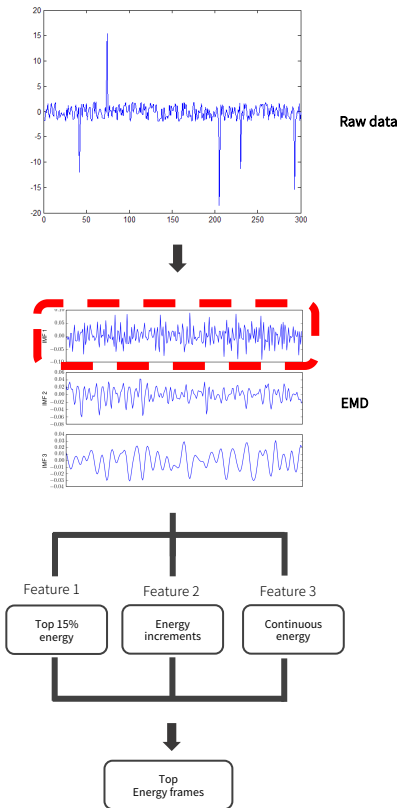
- 최종 스코어 및 승,패 데이터를 이용해 경기 전구간 **승률 예측 모델**을 생성
- 승률 예측 모델을 통한 승률의 **변화율**을 사용해 하이라이트 포인트를 추출
- 하이라이트 포인트에 **통계기반 하이라이트 길이**를 더해 하이라이트를 추출

특징

- **승률이 급격하게 변하는 지점**을 하이라이트 지점으로 간주
- 승률 모델을 통해 전체 하이라이트가 아닌 **하이라이트 포인트를 추출**
- 하이라이트 구간의 길이를 보충하기 위해 하이라이트 **통계 데이터 사용**

한계

- 승률의 변화없이 한쪽이 우세한 **단조로운 경기**의 경우 **하이라이트 추출 불가**
- 승패에 직접적인 영향이 아닌 **간접적인 영향**을 주는 장면(snowballing)은 하이라이트 예측 불가
- 하이라이트 길이는 예측된 하이라이트 포인트 기준 앞,뒤 통계에 기반한 길이



데이터

- World Cup, Laliga 축구 4경기의 음성 데이터를 사용

모델

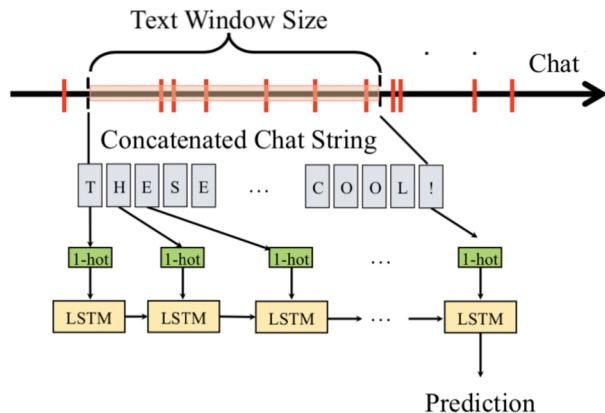
- Raw data에 EMD(Empirical mode decomposition)를 적용 후 IMF(Implicit mode function) 중 1번째 선택(노이즈 제거)
- Energy함수를 정의 후 3가지 features를 생성
- $E(n) = \sum |X(i)|^2; i = i_1 \sim i_2$
- Feature1: 상위 15% 에너지의 프레임
- Feature2: 평균 에너지가 2배 이상 증가한 샷(초)
- Feature3: 상위 15% 평균 에너지의 샷
- Feature1,2,3을 통해 하이라이트 샷 선택

특징

- EMD를 통해 음성데이터에서 노이즈를 제거
- 음성데이터를 가공한 에너지 feature를 제안
- 3가지 feature를 정의하여 음성데이터의 전역적, 지역적 특징을 잡아냄

한계

- 실제 연구의 평가지표는 하이라이트 순간(골, 반칙 등)을 탐지했는지를 확인하기 때문에
- 정확한 구간의 하이라이트 탐지 불가
- 경기내용에 대한 데이터 없이 관중 및 해설자의 반응 데이터로 하이라이트를 추출
- 경기내용, 해설자, 음질상태에 따라 정확도 편차가 심함



데이터

- 트위치 LOL 경기의 **유저 채팅 로그**를 사용

모델

- 문자별로 1-hot 벡터 임베딩**하여 사용
- 윈도우 사이즈 7
- 3 layer LSTM

특징

- 문자 별 임베딩을 진행하여, 인터넷 용어, 축약어, 오타 등의 노이즈에 영향을 받지 않음
ex) LOL : Laugh Out Loud
- 유저 반응 중 하이라이트에 직결되는 **강조, 흥분**을 표현하기 위해 반복사용하는 문자등을 파악할 수 있음
ex) !!!, happppy

한계

- 경기마다 채팅 수 등이 달라 예측 **정확도 편차가 심함**
- 유저의 악성 도배 글 등의 **노이즈에 취약**

연구목적

Twitch E-sports 중계방송
League of legends (LOL)



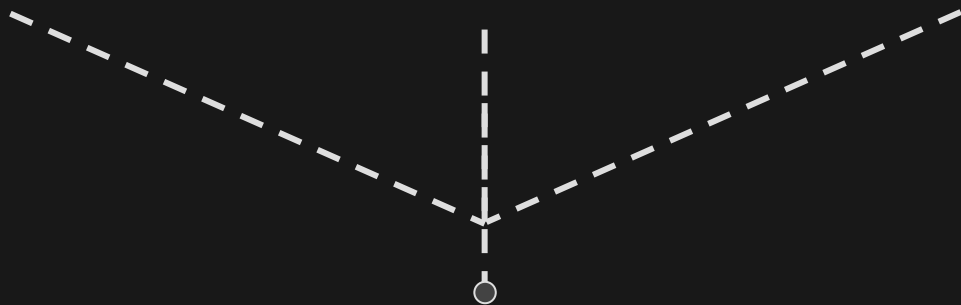
영상 데이터



음성 데이터



채팅 데이터



멀티모달 학습을 이용한
E-sports **하이라이트 자동 추출 방법**

베이스 모델 vs 멀티모달 모델

데이터

2019년 개최된 LoL 월드챔피언십 112경기 | 데이터 수집

이미지

• 데이터 수집 방법



DEVELOPER

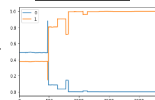
Riot API로 경기 데이터 수집

• 데이터 Feature 추출

API game score



승률 데이터



Blue, Red 팀 승률 변화율

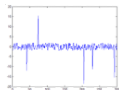


• 데이터 특성

하이라이트 구간이 아닌 승률의 변화로 인한 하이라이트 포인트를 예측하기에 용이한 데이터

음성

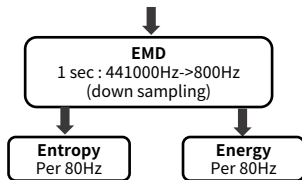
• 데이터 수집 방법



경기영상에서 음성데이터 추출

• 데이터 Feature 추출

Raw 음성 데이터



$$p(t) = \frac{|x(t)|}{\max|x|}$$
$$H(t) = - \sum_{k=-l/2}^{l/2} p(k) * \log(p(k))$$
$$E(n) = \sum |X(i)|^2; i = i_1 \sim i_2$$

엔트로피 데이터

에너지 데이터

• 데이터 특성

음성데이터의 진폭을 함수화 한 에너지데이터, 구간별 변화 정도가 반영된 엔트로피데이터

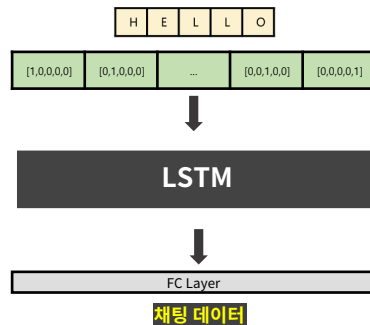
채팅

• 데이터 수집 방법



트위치 API를 통해 채팅로그 수집

• 데이터 Feature 추출



• 데이터 특성

Feature당 해당 프레임부터 7초간의 채팅 정보를 가짐

하이라이트

• 데이터 수집 방법



Tesseract OCR

Tesseract API 를 통해 화면상 데이터 추출



error handle

[00:01,

...

07:50]

하이라이트 라벨링

모델제안

단순 피쳐 퓨전을 통한 **멀티모달 모델**

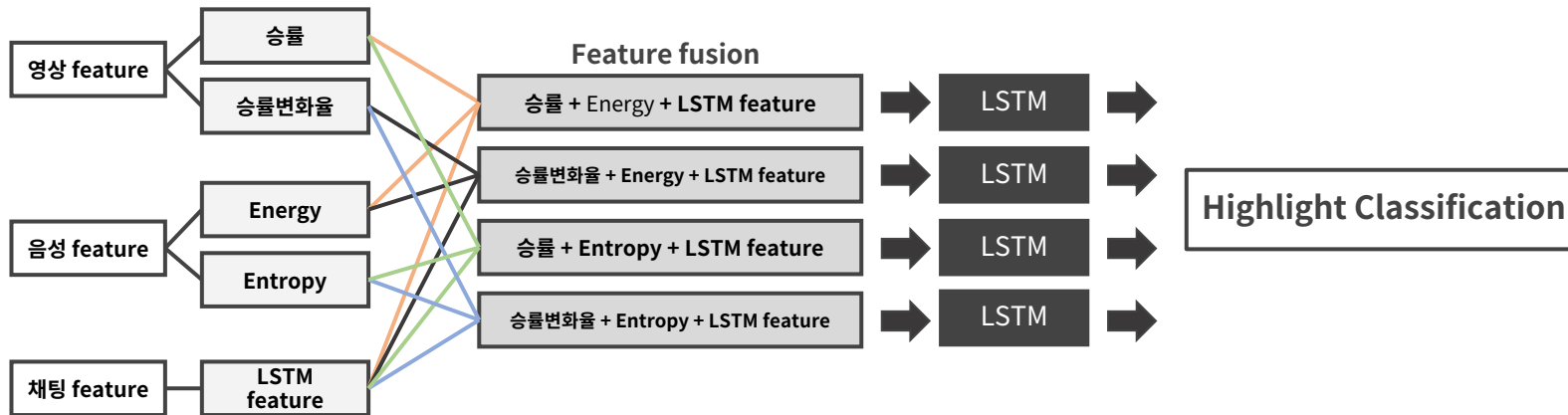
LSTM

3 simple layer LSTM (window size = 23)
SGD optimizer (momentum=0.9)
Learning rate 0.01 ([0,1,0.0001])

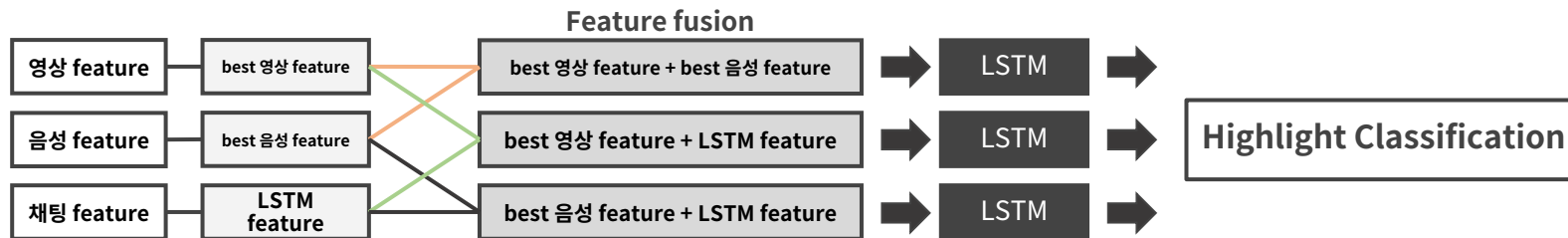
Feature Fusion

영상, 음성, 채팅 데이터 Concatenation

모델 1 3 features 멀티모달



모델 2 2 features 멀티모달



실험결과

단순 피쳐 퓨전을 통한 멀티모달 모델

Base 모델 실험

데이터	모델	Precision	Recall	F1
영상 feature	Base model	0.661	0.673	0.658
음성 feature	Base model	0.219	0.427	0.286
채팅 feature	Base model	0.318	0.390	0.343

모델 1 실험 결과

데이터	모델	Precision	Recall	F1
승률 + Energy + LSTM feature	모델1	0.682	0.409	0.500
승률변화율 + Energy + LSTM feature	모델1	0.814	0.228	0.343
승률 + Entropy + LSTM feature	모델1	0.696	0.414	0.508
승률변화율 + Entropy + LSTM feature	모델1	0.612	0.351	0.432

모든 경우가
Baseline보다 성능이 낮음

모델 2 실험 결과

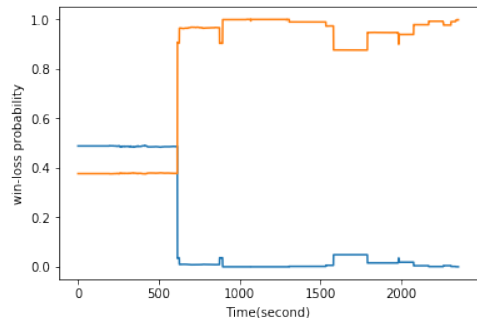
데이터	모델	Precision	Recall	F1
best 영상 feature(승률) + best 음성 feature	모델2	0.253	0.764	0.372
best 영상 feature(승률) + LSTM feature	모델2	0.453	0.740	0.555
best 음성 feature(Entropy) + LSTM feature	모델2	0.503	0.650	0.559

영상 feature와 퓨전한 경우,
모든 조합에서 영상 feature baseline보다 성능이 낮음

음성 feature와 채팅 feature를 퓨전하였을 때,
각 baseline보다 성능이 좋아짐

실험결과

데이터 특성에 대한 | **피처 엔지니어링**



HKA vs. ISG | Play-In Knockouts Day 2
승률 그래프

팀간의 승률 차이가 없는 **단조로운 경기인 경우,**



영상 데이터

Precision	Recall	F1
0.148	0.24	0.183

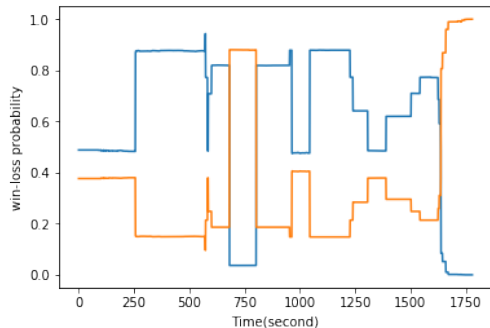


음성 데이터

채팅 데이터

Precision	Recall	F1
0.732	0.396	0.514

경기 스코어를 기반으로 하는
영상데이터는 성능이 낮고,
관중의 반응을 기반으로 하는
음성+채팅데이터는 성능이 높음



CG vs. RNG | Groups Day 7
승률 그래프

팀간의 승률 차이가 심한 **치열한 경기인 경우,**



영상 데이터

Precision	Recall	F1
0.608	0.952	0.742



채팅 데이터

Precision	Recall	F1
0.643	0.501	0.564

경기 스코어를 기반으로 하는
영상데이터가 성능이 높음

데이터 특성에 따라 경기별 하이라이트를 탐지하는 구간이 다름
따라서, **데이터 특성을 고려한 멀티모달 모델이 필요**

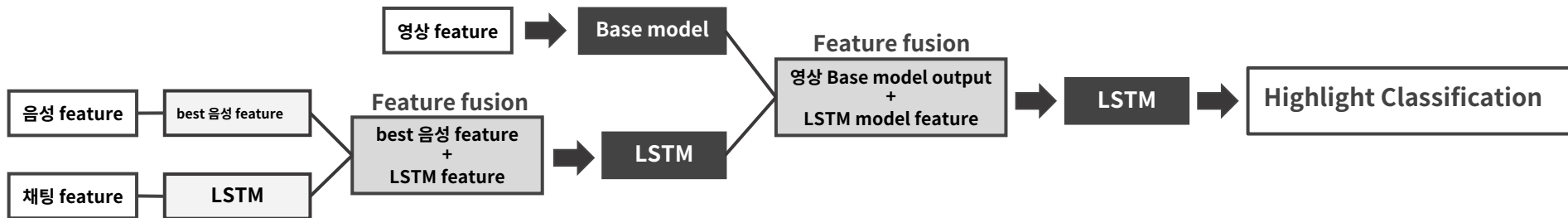
데이터 특성을 고려한 | 멀티모달 모델

LSTM

3 simple layer LSTM (window size = 23)
SGD optimizer (momentum=0.9)
Learning rate 0.01 ([0,1,0.0001])

Feature Fusion

음성, 채팅 데이터 Concatenation
영상 Base model output, LSTM model output Concatenation



영상 Base model output + 채팅, 음성 LSTM model output	모델3	0.663	0.700	0.670
--	-----	-------	-------	-------

데이터 특성을 고려한 **2Step 3feature** 멀티모달이
베이스라인보다 **좋은 성능**을 보임.

노동 집약적인 E-sports 하이라이트 시장에 자동화 추출 기술 필요

E-sports 중계 방송으로 얻을 수 있는 **영상, 음성 채팅**을 이용한 멀티모달 모델 제안

데이터 특성을 반영해 경기 정보 및 관중 반응을 효과적으로 활용한

2step 멀티모달 모델 제안

베이스 라인 모델 및 단순 피쳐 퓨전 멀티모달 모델보다 **좋은 성능을 보임**

관련 연구

- [1] An E-sports Video Highlight Generator Using Win-Loss Probability Model (2019)
- [2] Sports highlights generation using decomposed audio information (2019)
- [3] Video Highlight Prediction Using Audience Chat Reactions (2017)
- [4] AENet- Learning Deep Audio Features for Video Analysis (2018)
- [5] Video Highlight Prediction Using Multiple Time-Interval Information of Chat and Audio (2019)
- [6] A Mathematical Theory of Communication (2001)