

기업과제4_10팀_권예환_개인보고서

TRUE SUMMARY 생성 방법

- 카카오브레인에서 제공하는 pororo 모델을 이용해 제공받은 sports_news 데이터를 Extractive summary하였고, 이를 Ground Truth로 하였다.

Metric 함수 선정 및 근거

- ROUGE score를 사용하였다.
- 두 summary간의 precision, recall을 기반으로 f1 score를 계산하는 방식이다.
- ROUGE가 Extractive summary 모델을 사용했을 때 둘의 유사도를 비교하기 적절하다고 생각했고, Abstractive summary 모델에 대해서도 활용 가능하다고 판단하였다.

모델 선정

- KoBertSum, KoBart, MatchSum을 사용가능한 모델로 판단하고 조사하였다. 이 중 MatchSum은 구현하는 데 어려움이 있어 KoBertSum, KoBart를 사용하였다.

성능 평가

- Ground Truth에 대해 KoBart 모델은 ROUGE score가 0.531, KoBertSum 모델은 0.625가 나왔다.
- 이로 인해 둘 중 score가 더 높은 KoBertSum 모델을 최종적으로 사용하였다.

담당한 역할

- MatchSum 논문 리서치 후 학습에 적용을 시도했으나 실패
- 이 후 평가 metric으로 ROUGE score를 조사하고 모델 output으로 ROUGE score를 도출하는 코드 구현