

1. Paperswithcode에서 NLG extractive summarization task에 대해서 본인 블로그에 정리해보세요. 아래 3가지 항목에 대해서 정리하세요. (각 항목 고려 사항 참고)

- 문제 정의

- task가 해결하고자 하는 문제가 무엇인가?

- 데이터셋 소개(대표적인 데이터셋 1개)

- task를 해결하기 위해 사용할 수 있는데 데이터셋이 무엇인가?

- 데이터 구조는 어떻게 생겼는가?

- SOTA 모델 소개(대표적인 모델 최소 2개 이상)

- task의 SOTA 모델은 무엇인가?

- 해당 모델 논문의 요약에서 주요 키워드들에 대한 설명

- <https://blog.naver.com/ctk456/222655856826>

## 추출적 요약 (Extractive summarization)

아기별 · 방금 전

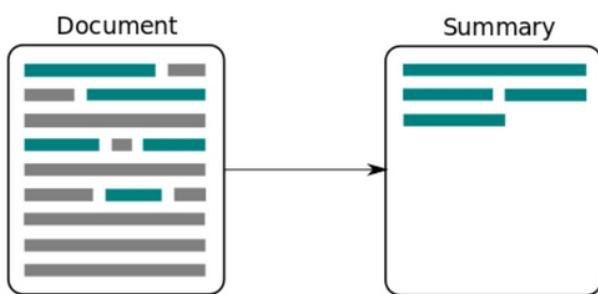
URL 복사

통계

:

NLG(Natural Language)

<Text summarization>



출처 <https://ceuity.tistory.com/33>

텍스트 요약이란 위 그림과 같이 긴 길이의 문서원문(Document)을 핵심 주제만으로 구성된 짧은 요약(Summary)문장으로 변환하는 것을 말한다. 상대적으로 큰 텍스트인 뉴스 기사로 작은 텍스트인 뉴스제목을 만들어내는 것이 텍스트 요약의 대표적인 예로 볼 수 있다. 이때 중요한 것은 요약 전후에 정보 손실 발생이 최소화되어야 한다는 점이다. 요약문은 문서 원문이 담고 있는 정보를 최대한 보존하고 있어야 한다. 이것은 원문의 길이가 길수록 만만치 않은 어려운 작업이 된다.

텍스트 요약은 크게 추출적 요약(Extractive Summarization)과 추상적 요약(Abstractive Summarization)의 두 가지 접근으로 나누어볼 수 있다.

- 추상적요약(Abstractive) : 문단을 새로운 문장으로 정리하는 요약하는 방식

- 추출적요약(Extractive) : 주어진 문장내에서 추출하는 요약방식

텍스트 요약의 성능평가는 ROUGE score를 이용한다.

ROUGE 평가지표는 텍스트 요약 뿐만아니라 기계번역의 성능을 평가할 때도 쓰인다.

### ROUGE(Recall-Oriented Understudy for Gisting Evaluation)

자연어생성(NLG)의 성능을 평가하기 위한 지표이며, 모델이 생성한 요약본 혹은 번역본을 사람이 미리 만들어 놓은 참조본과 대조에 성능 점수를 계산한다.