1p)

안녕하세요. 저희는 YouTube 뉴스 컨텐츠 자동 요약 서비스 프로젝트를 진행하고 있는 고은경, 권지혜, 배형준이고 저는 발표자 배형준입니다.

7조 발표 시작하겠습니다.

2p)

발표는 서비스 기획 배경, 활용 데이터 정의 및 수집 방안, 데이터 처리와 분석 기법, 마지막으로 활용 방안과 기대효과 순서로 발표를 진행하도록 하겠습니다.

3p)

여러분들은 유튜브를 많이 이용하시나요? 저는 이동하면서, 자기전에 많이 이용하고 평균적으로 하루에 1시간에서 1시간 반정도 시청하고 있습니다. 통계를 한번 볼까요? 2018년 와이즈앱의 안드로이드 앱 사용시간 통계에 따르면 유튜브는 카카오, 네이버, 페이스북을 제치고 sns 어플 중 가장 많이 사용되고 있습니다. 추세를 보면, 현 시점에서는 더 많이 이용되고 있을 것입니다.

4p)

그럼 유튜브 내에서는 어떤 종류의 컨텐츠가 많이 이용되고 있을까요?

영국의 로이터 저널리즘 연구소의 2020년 디지털 뉴스 리포트에 따르면 sns로 뉴스를 접하는 비율이 44퍼센트이고 그 중 유튜브를 이용한다는 비율이 45%로 많은 비중을 차지하는 것을 확인할 수 있습니다.

실시간 이슈에 따라 다르겠지만, 유튜브 메인 인기 동영상 항목에 들어가보면 뉴스 동영상이 많이 배치되어 있는 것을 확인하실 수 있습니다.

5p)

오른쪽 표는 메인 언론사들의 sns 구독자 수입니다. 2019년 1월 기준 주요 22개 언론사의 sns 구독자 수 총합이 약 1600만 명이고, 해당 언론사들의 유튜브 구독자 수 총합은 2020년 8월 기준 약 1000만 명 정도로 뉴스 컨텐츠가 많이 소비되고 있음을 알 수 있습니다.

그런데! 뉴스 컨텐츠가 많이 소비되는 것이 저희같은 컨텐츠 소비자들에게 좋은 일일까요?

6p)

현대인들은 매우 바쁘게 살아갑니다. 끊임없이 제공되는 컨텐츠를 전부 소비하기에는 시간이 부족할 수 밖에 없습니다. 따라서 컨텐츠를 많이 소비하는 것 보다는 핵심만 골라서 소비하는 전략이 필요합니다. 그래서 저희 팀은 효율적으로 뉴스를 소비하려는 시장의 잠재 수요가 있다고 판단하여 뉴스 영상 요약 서비스를 기획하였습니다.

아래의 기사와 같이 요약 서비스 산업의 시장은 점점 커지고 있고, 거기에 인공지능 기술을 접목시켜 더 효율적으로 자동화 요약 서비스를 제공하려고 합니다.

7p)

서비스를 만들기 위해 수집한 데이터는 다음과 같습니다.

네이버와 다음에서 최근 2년간 카테고리별 상위 랭킹 뉴스를 포털별로 각 10만건씩 수집하였습니다. 해당 데이터를 이용하여 요약 모델을 학습시킨 후 유튜브 영상 정보를 실시간으로 가져와 요약 서비스를 제공하려고 합니다.

8p)

전체적인 work flow는 다음과 같습니다. train data로 네이버와 다음 뉴스기사를 활용하여, 기사본문을 x값으로, 요약문을 y값으로 하여 tokenizing 및 word embedding 을 하고 모델 학습을 위한 전처리를 진행합니다. 각 작업에는 koNLPy와 koBERT를 사용할 예정입니다.

요약에 활용할 모델은 attention이 적용된 sequence to sequence 모델입니다. 해당 모델은 단어들에 가중치를 주며 학습되는 encoder와 decoder로 구성되어 있으며 이와같은 구조는 텍스트 요약에 유용합니다.

test 데이터는 유튜브에서 수집한 내용들이며 같은 전처리 과정을 거진 후 학습된 모델을 통과하여 최종적으로 예상 요약문을 생성합니다.

9p)

서비스는 다음과 같이 제공할 예정입니다. 유튜브를 가장 많이 소비하는 방식이 어플이므로 저희도 어플리케이션으로 구현하려고 합니다. 오른쪽은 예시 이미지인데요, 기존 유튜브와 같이 썸네일과 제목을 제공하고 우측 또는 하단에 영상 요약과 댓글 요약, 영상의 주요 키워드를 추가하려고 합니다. 영상을 클릭하면 원본 영상이 있는 유튜브 페이지로 이동하여 영상을 시청할 수 있게 만들 예정입니다.

10p)

마지막으로 저희 서비스의 기대효과입니다.

컨텐츠 소비자는 수많은 정보 속에서 영상 선택을 위한 추가 정보가 제공되기 때문에 자신에게 필요한 정보를 선택하기 쉬워지고, 효율적으로 컨텐츠를 소비할 수 있게 됩니다.

영상을 클릭하기 전 시청자들에게 더 많은 정보가 제공되어 시청자의 컨텐츠 결정권이 강해지므로, 언론사들은 경쟁력을 확보하기 위해 정확한 정보 전달에 더 힘쓸 것입니다. 또한 긴 영상을 보지 않는 시청자에게 요약을 제공함으로써 클릭하게 만들어 신규 시청자 유입을 유도할 수 있습니다.

사회적으로도 긍정적 효과를 기대할 수 있는데요. 컨텐츠 소비자들이 유튜브 알고리즘에 의해 선택받은 영상만 시청하지 않고 컨텐츠를 자기가 결정할 수 있도록 도와줌으로써 여론의 확증 편향을 방지할 수 있을 것입니다. 또한 타임 푸어족의 뉴스 소비를 촉진시킴으로써 사회 구성원들간의 정보 격차를 해소할 수 있을 것이라 기대합니다.

이상으로 발표 마치도록 하겠습니다.

1. 배경
2. 진행사항 / 역할분담
3. 크롤링
   1. 네이버, 다음, 유튜브
4. Cleaning&Preprocessing
   1. konlpy/kobert
   2. 네이버, 다음, 유튜브
5. Modeling

3. TODO / 역할분담

1. Train Model
2. Hyperparameter Tuning
3. Apply Test Data(YouTube data)
4. Build Application?/Web?
   1. Web이라면 flask, django
   2. Application이라면 react native

4. 기대효과

5. 의문사항?자문?