1)

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

지금부터 7조 발표 시작하겠습니다.

2)

발표는 다음과 같은 순서로 진행하겠습니다.

먼저 저희가 아이디어를 도출한 배경에 대해 설명드린 후, 저희가 활용하는 데이터와 수집방법에 대해 말씀드리겠습니다.

이후 데이터 처리방안과 분석기법에 대해 다룬 후 마지막으로 예상 활용방안 및 기대효과로 발표를 마무리 하도록 하겠습니다.

3)

먼저 배경입니다.

현재 유튜브는 사용시간이 꾸준히 증가하는 추세이며, 현재 타 매체보다 이용 시간 면에서 사용량이 월등히 높습니다.

이에 따라 뉴스기사 또한 타 매체보다 유튜브를 통해 접하는 사용자가 급증하여 유튜브 컨텐츠 분석의 필요성은 더욱 더 대두될 것입니다.

4)

실제로 유튜브 인기동영상의 많은 부분이 뉴스 컨텐츠로 구성되어 있음을 확인할 수 있습니다.

또한 영국 로이터저널리즘연구소 ‘디지털 뉴스 리포트 2020’에 따르면,

SNS로 뉴스를 접한다는 국내 응답자 비율은 44%(복수 응답)로 2019년 26%보다 큰 폭으로 증가했고

SNS를 통한 뉴스 이용 매체로 유튜브를 꼽은 응답자는 45%였습니다.

5)

하지만, 이렇게 많이 소비되는 뉴스 컨텐츠들에 대한 신뢰도는 매우 낮습니다.

앞서 언급한 리포트에서 한국은 조사대상국 40개국 중 언론 신뢰도 21%로 최하위를 기록했습니다.

특히 허위정보가 가장 우려되는 플랫폼이 한국에서는 유튜브가 31%로 1위를 기록했습니다.

만약, 뉴스 요약 정보를 통해 뉴스 컨텐츠별 빠른 비교를 가능하게 하면,

양질의 컨텐츠를 제공하기 더 좋아지지 않을까요?

6)

그럼, 타 플랫폼의 경우를 먼저 살펴보겠습니다. 네이버뉴스에는 요약봇, 다음뉴스에는 자동요약 서비스가 존재합니다.

7)

하지만, 유튜브에는 아직 이러한 요약 서비스가 존재하지 않습니다.

자막, 댓글 등 기사보다 더 많은 텍스트를 다뤄야 하기 때문에 분석으로 얻을 정보가 더 많을 수 있는 유튜브에서 뉴스 컨텐츠 요약 서비스가 제공된다면,

바쁜 현대인들이 빠르게 뉴스 내용과 그에 따른 반응을 습득할 수 있도록 도와줄 것입니다.

따라서, 저희는 유튜브 뉴스 컨텐츠를 요약하는 서비스를 만들고자 합니다.

8)

유튜브 요약 서비스를 구현하기 위해 수집한 데이터는 다음과 같습니다. 요약본이 이미 존재하고 있는 네이버뉴스와 다음뉴스에서 게시일자, 제목, 기사본문, 요약문을 크롤링해 모델 training 및 validation에 사용할 예정입니다. 최근 약 2년간의 카테고리별 상위 랭킹 뉴스를 수집했으며, 각 10만건으로 총 20만건을 수집합니다.

모델을 학습한 이후 유튜브 뉴스를 대상으로 test를 진행할 예정이며, YTN, JTBC, SBS 등과 같은 공식 뉴스 채널의 영상제목, 영상 자동생성자막(즉 내용), 동영상 정보와 댓글을 약 1000건 정도 수집해 예상 요약문을 생성할 예정입니다.

크롤링 라이브러리로는 beautifulSoup와 Selenium을 활용했습니다.

9)

전체적인 분석 flow는 다음과 같습니다. 앞서 언급한 바와 같이, train data로 네이버와 다음 뉴스기사를 활용할 예정이며, 게시일자, 제목, 기사본문을 x값으로, 요약문을 y값으로 하여 tokenizing 및 word embedding 을 하고 모델을 학습합니다.

test 데이터는 유튜브에서 수집한 모든 내용들이며 이 또한 tokenizing과 word embedding을 거친 후, 미리 학습된 모델을 적용시켜 최종적으로 yhat, 즉 예상 요약문을 생성합니다.

10)

활용 모델은 attention이 적용된 sequence to sequence 모델을 사용할 예정입니다. Attention이 적용된 Sequence to sequence 모델은 단어들에 가중치를 주며 학습되는 encoder와 decoder로 구성되어 있으며 이와같은 구조는 text summarization에 유용하여 가장 자주 쓰이는 모델로 알려져 있습니다.

11)

저희의 최종 목표는 다음과 같이 어플리케이션을 제작하는 것입니다.

뉴스 채널에서 업로드되는 영상들을 실시간으로 수집하여, 해당 영상에 미리 학습된 모델을 적용합니다.

영상 옆에 뉴스 영상의 핵심 키워드가 표시되며,

영상을 클릭하면 원본 영상이 있는 유튜브 페이지로 이동하고, 키워드를 클릭하면 해당 키워드가 나타난 구간으로 바로 이동합니다.

더불어 저희는 댓글 요약을 제공할 예정입니다. 이모티콘 등 구어체의 특수성을 반영해 유튜브 댓글을 두세줄로 요약하여 화면에 표시합니다.

12)

이런 서비스를 제공한다면 다음과 같은 기대효과가 있을 것으로 예상됩니다. 사용자는 정보의 홍수 속에서 양질의 정보 선택이 용이해지고, 정보 소비에 시간을 단축해서 더 효율적으로 컨텐츠를 소비할 수 있습니다.

언론사에서는 주어진 화면에서 영상에 대한 정보를 더 많이 제공하여 사용자들의 클릭을 적극적으로 유도할 수 있습니다.

‘’’

사용자: 빠른 시간 안에 양질의 뉴스 정보를 습득할 수 있다. 즉, 빠른 비교를 통해 올바른 정보를 습득할 수 있다.

각 언론사: 타 플랫폼과 달리 interactive. 관심이 있는 정보면 영상으로 더 자세한 정보를 볼 수 있도록.

뉴스 컨텐츠가 다양한 형식으로 제공되게 되면 유튜브 시청자들의 접근성을 높일 수 있다. (기존에 길어서 보려고 하지 않았던 시청자들의 유입도 가져올 수 있다.)

‘’’