

정보성 동영상 요약 및 키워드 기반 영상검색 시스템

김기훈^o, 문미경^{*}

^o동서대학교 소프트웨어학과,

^{*}동서대학교 소프트웨어학과

e-mail: 20181468@dongseo.ac.kr^o, mkmoon@dongseo.ac.kr^{*}

Information Video Summarization and Keyword-based Video Tracking System

Gihun Kim^o, Mikyeong Moon^{*}

^oDept. of Software, Dongseo University,

^{*}Dept. of Software, Dongseo University

● 요약 ●

비대면 교육이 증가함에 따라 강의, 특강과 같은 정보성 동영상의 수가 급격히 많아지고 있다. 이러한 정보성 동영상을 보아야 하는 학습자들은 자원과 시간을 효율적으로 활용할 수 있는 동영상 이해 및 학습 시스템이 필요하다. 본 논문에서는 GPT-3 모델과 KoNLPy 사용하여 동영상 요약을 수행하고 키워드 기반 해당 영상 프레임으로 바로 갈 수 있는 시스템의 개발내용에 대해 기술한다. 이를 통해 동영상 콘텐츠를 효과적으로 활용하여 학습자들의 학습 효율성을 향상시킬 수 있을 것으로 기대한다.

키워드: 정보성 동영상(informational video), 동영상 요약(video summarization), 질문 생성(question generation), GPT 모델 (Generative Pre-trained Transformer)

I. Introduction

정보성 동영상이란 전문 지식, 뉴스, 교육 등과 같이 다양한 주제에 대한 활용가치가 있는 정보를 제공하는 동영상이다. 인터넷과 스마트폰의 확산으로 사람들은 점점 더 많은 시간을 이러한 동영상 콘텐츠 시청에 사용하고 있다. 그러나 정보성 동영상은 내용이 방대하고 어려운 경우가 많아서 콘텐츠의 내용을 집중해서 받아들이기가 힘든 경우가 많다. 최근에는 긴 문장을 자동 요약해주는 시스템에 대한 개발이 많이 이루어지고 있다. 정보성 동영상도 핵심 내용을 자동 요약해 줄 수 있으면 그 내용을 학습하는 데 도움이 될 것이며, 학습 중 핵심 키워드에 대한 설명 부분을 바로 찾아갈 수 있다면 학습효과를 높일 수 있을 것이다.

본 논문에서는 GPT-3 모델과 KoNLPy 사용하여 동영상 요약을 수행하고 키워드 기반 해당 영상 프레임으로 바로 갈 수 있는 시스템의 개발내용에 대해 기술한다.

II. Related works

연구 [1]에서는 한국사 질의응답과 문제 자동 생성 시스템에 대한 연구가 진행되었다. 해당 연구는 Transformer 기반의 인코더인 BERT와 디코더인 GPT2를 활용한 시스템을 제안하였다. 연구 [2]에서는

BERT 모델을 요약 추출 및 요약 생성에 활용하여 성능을 향상시키는 방법을 소개하고 있다. 이 모델은 신경망 기반 문장 추출 모델을 사용하여 핵심 문장을 추출한 후, 추출된 문장을 BERT 모델의 입력으로 활용하여 요약문을 생성한다.

본 연구에서는 BERT 모델이 아닌 GPT-3 모델을 사용하여 좀 더 정확한 요약을 해주고 KoNLPy를 사용하여 키워드를 추출하고 검색된 키워드 기반으로 해당 영상 프레임으로 갈 수 있게 한다.

III. The Proposed Scheme

3-1. 동영상 자막 데이터 수집

본 연구에서는 동영상 자막을 추출하기 위해 YouTube API를 활용한다. YouTube API를 통해 동영상의 자막 데이터를 시간별 및 분별로 수집한 후, JSON 형태로 변환하여 저장한다. 수집된 데이터는 동영상에서의 내용과 정보를 확보하기 위해 사용되며, 이를 분석에 활용하기 전에 데이터 튜닝 작업을 거친다. Fig. 1의 왼쪽은 시간별로 수집한 자막 데이터이며 이중 summary의 문장을 재구성하여 전체 자막 데이터를 만든다. (Fig. 1의 오른쪽)

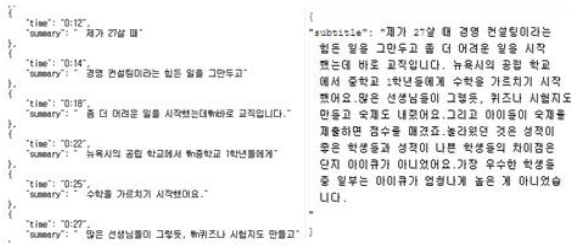


Fig. 1. 동영상 자막 데이터

3-2. GPT 요약 및 키워드 추출

본 연구에서는 추출된 자막을 기반으로 GPT-3 모델을 활용하여 내용을 요약한다. 요약은 요약정도에 따라 세 단계 수준으로 구성된다. 요약된 텍스트를 KoNLPy의 Mecab 라이브러리를 통해 문장의 형태소를 분리하고, 자연어 처리 과정에서 품사 태깅과 불용어 제거하는 과정을 거친다. 이후, 정제된 텍스트를 바탕으로 TF-IDF 모델을 구축한다. 각 단어의 빈도수에 대한 가중치를 계산하며, 가중치가 특정 기준을 넘어서는 경우 그 단어를 키워드로 식별하여 이를 요약문에서 강조시킨다.

3-3. 키워드 기반 영상검색

본 연구에서는 사용자가 특정 키워드에 대한 영상 장면 검색기능을 제공한다. 사용자가 키워드를 입력하면 영상에서 해당 키워드와 관련 있는 시간대를 GPT-3 모델을 통해 추적하여, 시청자가 원하는 영상 장면으로 바로 이동할 수 있도록 한다. 이러한 기능은 사용자가 불필요한 영상 부분을 검색할 필요 없이 원하는 내용을 효과적으로 찾을 수 있게 해준다.

3-4. 개발 결과

Fig. 2는 제안한 방법을 구현한 결과화면이다. 왼쪽에는 06분 13초 분량의 "Grit: the power of passion and perseverance"에 대한 동영상이 나오고 있으며, 오른쪽에는 본 동영상에 대한 요약을 1수준에서 보여주고 있다. 또한 '아이들', '아이큐', '성적'이 중요한 키워드로 추출이 되어 빨간색으로 강조되어 표시되고 있다. 이 중, '똑똑함'으로 키워드를 검색했을 때 영상의 1분 35초로 바로 이동할 수 있음도 확인하였다.

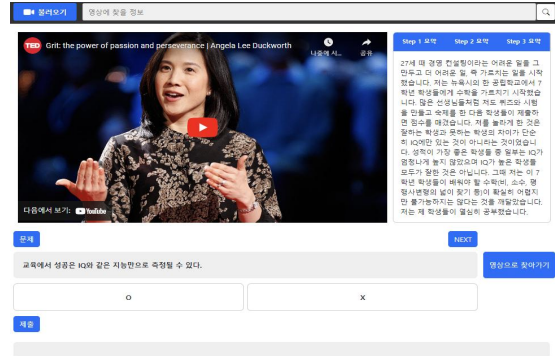


Fig. 2. 결과화면

IV. Conclusions

본 논문에서는 정보성 동영상 요약을 수행하고 키워드 기반 해당 영상 프레임으로 바로 갈 수 있는 시스템의 개발내용에 대해 기술하였다. 해당 시스템은 향후 웹 확장프로그램으로 추기하여 정보성 동영상 시청 시 바로 활용할 수 있도록 기능을 확장해 나갈 것이다.

ACKNOWLEDGEMENT

본 연구는 2023년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2019-0-01817)

REFERENCES

- [1] W. Cho, H. Oh, J. Lee, H. Lee. "Korean history Q&A and Quiz system using BERT and GPT2" , Proceedings of KIIT Conference, pp.508-511, 2021.
- [2] E.K. Kim, W. Son, "Subtitle Automatic Generation System using Speech to Text", IEEE Trans. on Multimedia, Vol. 16, No. 1, pp.81-88, 2021.