

FastText 알고리즘을 이용한 사용자 지정 키워드 기반 동영상 요약 시스템

김경민^o, 박승민^{*}

^o동서대학교 소프트웨어학과,

^{*}동서대학교 소프트웨어학과

e-mail: sminpark@dongseo.ac.kr^{*}

Keywords-based Video Summary System using FastText Algorithm

Kyungmin Kim^o, Seungmin Park^{*}

^oDept. of Software, Dongseo University,

^{*}Dept. of Software, Dongseo University

● 요약 ●

본 논문에서는 FastText 알고리즘을 기반으로 한 사용자 지정 키워드 기반 동영상 요약 시스템을 제안한다. 사용자가 키워드를 입력하면 시스템은 해당 키워드와 관련된 단어들을 FastText를 통해 추출하며, 이를 STT (Speech-to-Text)로 변환된 동영상에서 타임 스탬프 기반으로 인식한다. 인식된 키워드와 관련된 내용은 클립 형식으로 요약되어 사용자에게 제공된다. 본 연구의 목적은 스포츠 콘텐츠 환경에서 효과적인 콘텐츠 추출 및 제공을 통해 사용자 경험과 정보 제공의 효율성을 향상시키기 위함이다. 제안된 시스템은 사용자 지정 키워드에 맞춰 다양한 동영상 플랫폼에서 효율적인 영상 요약을 제공함으로써 온라인 동영상 환경에서 큰 혁신을 이끌어낼 것으로 기대된다.

키워드: 자연어처리(Natural Language Processing), STT(Speech-To-Text), 신경망(Neural Network), FastText 알고리즘, 시계열 데이터 분석

I. Introduction

최근 디지털 기술의 발전과 함께 동영상 콘텐츠가 폭발적으로 증가하고 있다. 사용자가 동영상 콘텐츠를 시청할 때, 동영상에서 사용자가 원하지 않는 내용도 포함되어 있을 경우에는 정보 습득에 있어서 처음부터 끝까지 시청하는 것이 시간 대비 효율적이지 않다. 이와 같은 이유로 인해 사용자에게 원하는 정보를 빠르게 전달하기 위한 대안이 필요하다. 이에 본 논문에서는 FastText 알고리즘 기반 키워드 추출을 통한 동영상 요약 시스템을 제안한다. 본 논문에서 제안하는 동영상 요약 시스템은 FastText 알고리즘을 사용자가 원하는 키워드를 중심으로 효율적인 정보 접근을 제공한다.

여 단어를 표현할 수 있으며, 이를 통해서 효율적인 학습이 가능하다.

FastText의 핵심 중 하나는 부분 단어의 사용이다. 이를 통해 형태학적으로 유사한 단어간의 관계를 파악하고, 그에 따라 벡터 공간에서 이를 표현할 수 있다. 이러한 방식을 사용함으로써, 학습 데이터에 없더라도 효율적인 처리가 가능하다.

2.2 STT(Speech-To-Text)

음성 인식(STT: Speech-to-Text) 기술은 사람이 말하는 소리를 텍스트로 변환하는 기술이다. 동영상 콘텐츠에서 오디오 데이터를 이해하려면 먼저 음성 인식 기술을 통해 텍스트로 변환해야 한다. 이렇게 변환된 텍스트 데이터는 대화, 설명 내용 등의 정보를 포함한다.

II. Related works

2.1 FastText

FastText는 페이스북에서 개발한 자연어 처리를 위한 알고리즘으로, 기존의 Word2Vec과 비슷하지만, 문자 단위의 n-gram을 사용하

III. The Proposed Scheme

3.1 Research Environment

연구 환경에서 사용된 데이터셋은 AI Hub에서 제공하는 mp4 형식의 데이터셋을 사용하였으며, 그래픽카드로는 RTX 3070 Ti를 사용하였다.

3.2 Research Method

대략적인 연구 방법을 작성하면 아래와 같다.

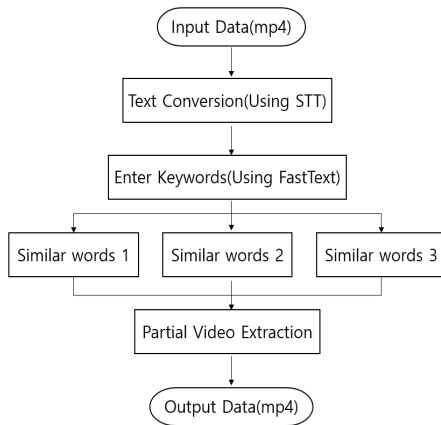


Fig. 1. Proposed System Diagram

STT 기술을 사용하여 동영상에서 음성 데이터를 텍스트로 변환한다. 또한 타임 스탬프를 사용하여 각 단어의 시작과 종료 시간을 기록하여 이후 처리에 활용한다. 그리고 FastText 알고리즘을 사용하여 사용자가 입력한 키워드와 연관된 단어들을 찾아낸다. 이 과정에서 FastText가 단어와 해당 단어의 연관 단어들을 벡터로 표현하고, 코사인 유사도를 기반으로 유사한 단어들을 찾는다. 예를 들면, ‘더위’라는 키워드를 입력했을 때, FastText를 사용하여 유사 단어(더움, 더운, 더위)를 탐색할 수 있다.

결과적으로 사용자가 입력한 키워드와 FastText 알고리즘으로 찾아낸 연관 단어들이 인식되어, 해당 키워드가 포함된 문장의 시작 시각과 종료 시각을 추출하고, 이를 바탕으로 클립 영상을 생성한다.

Video Keyword Search

포털 검색 news_onair_4127.mp4



Result



Fig. 2. The result of video summary using FastText Algorithm

IV. Conclusions

FastText 기반의 사용자 지정 키워드 기반 동영상 요약 시스템은 다양한 주제의 동영상에서 효과적인 콘텐츠 제공이 가능해진다. 이를 바탕으로 사용자 지정 키워드와 관련 키워드를 통한 다양한 주제의 콘텐츠를 손쉽게 요약 및 제공하면서, 온라인 동영상 환경에 더 큰 혁신을 선도할 것으로 기대된다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심 대학지원사업의 연구 결과로 수행되었음 (2019-0-01817)

REFERENCES

- [1] BOJANOWSKI, Piotr, et al. Enriching word vectors with subword information. Transactions of the association for computational linguistics, 2017, 5: 135-146.
- [2] FURUI, Sadaaki, et al. Speech-to-text and speech-to-speech summarization of spontaneous speech. IEEE Transactions on Speech and Audio Processing, 2004, 12.4: 401-408.
- [3] WU, Sun; MANBER, Udi. Fast text searching: allowing errors. Communications of the ACM, 1992, 35.10: 83-91.
- [4] POTAPOV, Danila, et al. Category-specific video summarization. In: Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13. Springer International Publishing, 2014. p. 540-555.