MAS: BART 와 WebRTC 라이브러리를 이용한 실시간 회의 스크립트화 및 요약 서비스

권기준, 고건준, 주영환, 지정희
건국대학교 컴퓨터공학부
kijune@konkuk.ac.kr, koeunjun7@konkuk.ac.kr, wndudghks@konkuk.ac.kr, jhchi@konkuk.ac.kr

MAS: Real-time Meeting Scripting and Summarization Service using BART and WebRTC library

Ki-Jun Kwon, Geon-Jun Ko, Yeong-Hwan Joo, Jeong-hee Chi
Konkuk University, Computer Science & Engineering

요 약

COVID-19 사태의 지속화로 재택근무 및 화상 수업의 수요가 증가함에 따라, 화상 회의 서비스에 대한 수요 또한 증가하고 있다. 본 논문은 회의 내용의 텍스트화 및 요약 회의록 생성에 관한 연구를 통해 보다 효율적인 화상 회의 서비스를 제공하고자 한다. WebRTC를 기반으로 화상 회의 서비스를 제공하며, WebSpeech API 를 활용하여 회의 내용을 스크립트화 한다. 회의 스크립트는 BART를 통해 요약본으로 재생성되며, 회의 스크립트와 요약본은 언제든지 열람 및 다운로드가 가능하다. 본 논문은 회의 요약 기능을 제공하는 화상 회의 서비스 MAS (Meeting Auto Summarization)를 제안하며, MAS 의 설계 및 구현 방법을 소개한다.

1. 서론

COVID-19 사태의 오랜 지속으로 사회적 흐름이 급격히 변화했고, 이를 배경으로 한 언택트(Untact) 시대의 등장과 함께 온택트(Ontact) 서비스가 많은 수요를 이루게 되었다. 온라인 수업, 재택근무, 화상 회의 등 대부분의 업무가 온택트 서비스 기반의 비대면 방식으로 진행되고 있다. 코로나 19 종식 이후에도 이러한 흐름은 계속될 것이다. 전문가들은 대기업이나 IT 관련 중견 규모 이상의 기업은 재택근무, 화상회의 상시화의 길을 걸을 것이라는 전망을 하고 있다[1]. 우리는 이러한 온택트 업무 환경의 편의성과 효율성을 극대화 시킬 수 있는 화상회의 서비스를 제공하고자 한다.

줌(ZOOM), 시스코 웹엑스 미팅 (CISCO WEBEX MEETINGS), 구글 미팅 (GOOGLE MEET)등과 같은 대부분의 기존 화상 회의 서비스는 발화 내용에 대한 스크립트를 제공하지 않으며, 회의 내용 (스크립트) 확인, 회의 요약본과 같은 회의 정보 파악에 용이한 추가 기능을 제공하지 않는다.

이를 개선하고자 우리는 기존의 대면 회의와 달리 오디오 스트림을 지속적으로 다룰 수 있는 화상 회의의 이점을 적극적으로 활용하고자 한다. 화상 회의에서 발생하는 발화 내용에 대한 스크립트를 제공하고, 그 이점을 더욱 극대화하고자 스크립트 기반 회의 요약 기능을 제공하고자 한다.

본 논문은 회의 요약 기능이 포함된 화상회의 서비스 MAS (Meeting Auto Summarization)를 구축한 방법 및 사용된 기술에 대해 설명하고 있다. MAS 는 화상회의에서 발생하는 발화를 텍스트로 변환하여 실시간으로 제공하며, 이를 토대로 회의 요약본을 생성하는 핵심 기능을 바탕으로 한다. 이 외에도 사용자 정보 관리, 회의 내용 저장 및 열람 등의 서비스 요소를 제공하는 종합 화상회의 플랫폼으로 기획 및 구현되었다.

2. 관련 연구

2.1 WebRTC, socket.io, Speech Recognition

WebRTC[2] 프로토콜은 웹 브라우저 간에 플러그인 없이 서로 실시간으로 오디오나 영상 등의 데이터를 교환할 수 있도록 하는 기술로 정의된다. socket.io 양방향 통신 구현을 위한 라이브러리이다. 통신 방식으로는 websocket 방식과 COMET (Polling, Long Polling, Streaming 등)을 사용한다.

Speech Recognition 에는 Web 브라우저 환경에서 음성인식 기능을 제공하는 API 중 하나인 Web Speech API[3]를 사용한다.

본 실험에서는 실시간 회의 구현에 WebRTC, socket.io, Web Speech API 의 조합으로 P2P 연결방식의 실시간 화상회의와 음성인식 기반의 채팅을 구현한다.

2.2 한국어 생성 모델

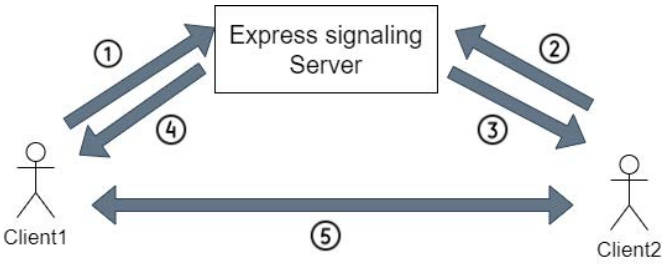
대표적인 한국어 생성 모델로는 SKT 에서 발표한 BART[4] 기반의 한국어 모델 KoBART 와 GPT2[5] 기반의 한국어 모델 KoGPT2, Facebook 에서 제안한 다국어 생성 모델 mBART-50[6] 등이 있다.

딥러닝 모델을 통한 대화 요약은 문어체에 비해 생략이나 변형이 많은 점, 대화의 문맥을 고려해야 하는 점, 화자가 여러명인 점 등의 특수성이 존재함에 따라 그 어려움이 강조되며, 성능 개선을 위해 꾸준히 연구가 이루어지는 분야 중 하나이다.

본 실험에서는 해당 3 개의 모델에 대해 동일한 데이터, 동일한 기법을 통한 미세 조정을 통해 모델을 생성하였다.

3. 제안 구조 및 방법론

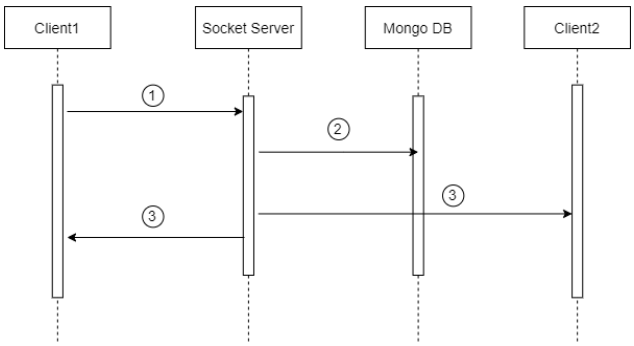
3.1 Signaling server 와 화상회의 프로세스



(그림 1) WebRTC P2P 연결 프로세스

전체적인 화상회의 연결 프로세스이다. 본 연구에서 채택한 P2P 연결의 결과는 client 간의 direct 연결로, 이를 중계하는 역할을 하는 서버가 signaling server 이다. Signaling server 는 각 Peer client(client1, client2)의 SDP 와 iceCandidate[7] 정보를 중계하여 다른 client 에게 전달하는 역할(그림 1 의 ①, ②, ③, ④)을 한다. 각 client 이 정보를 바탕으로 WebRTC 를 이용해 P2P 연결을 맺고 media 정보를 서로 교환하여 화상회의 연결을 성립(그림 1 의 ⑤)시킨다.

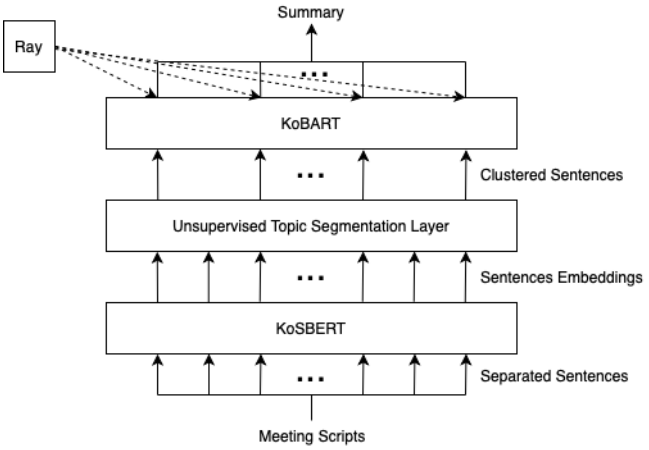
3.2 실시간 STT 프로세스



(그림 2) 실시간 STT 시퀀스 다이어그램

Web Speech API 를 통해 client(그림 2 의 client1)에서 STT 가 진행된다. 진행된 STT 결과는 socket 통신을 통해 client 와 Socket Server 가 주고받는다. 먼저 client1 에서 server 에 STT 결과를 전달(그림 2 의 ①)한다. Server 내부에서는 client 에게 전달받은 STT 결과를 발화자, 발화 시간 정보와 함께 database 에 저장하고(그림 2 의 ②) 같은 회의실에 있는 모든 참여자(client1, client2)에게 결과 스크립트를 전달(그림 2 의 ③)한다.

3.3 회의 요약 알고리즘 설계



(그림 3) 회의 요약 알고리즘 설계 구조도

본 연구에서는 (그림 3)과 같은 구조를 통한 회의 요약 방법론을 제안한다.

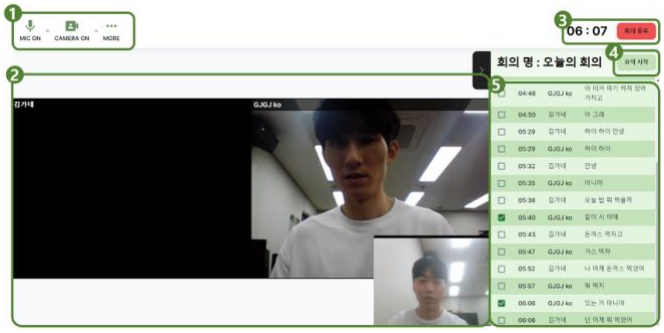
먼저, 전체 회의 스크립트를 문단으로 나누는 과정을 거친다. 이를 위해 Unsupervised Topic Segmentation[8] 이라는 방식을 사용하였으며, 문장 임베딩 추출을 위해 SBERT[9] 기반의 한국어 모델 KoSBERT 를 사용했다.

주제 별로 나뉜 각 문단을 seq2seq 기반 모델에 입력하여 요약본을 생성하였다. 본 연구에서는 BART[4]의 구조를 사용하였으며, AIHUB 의 한국어 대화 요약 데이터 셋[9]을 통해 미세조정을 진행하였다. 이 때, Ray[10]를 통해 각 문단의 요약을 분산 처리하였다.

모든 요약이 완료되면 각 요약 문장을 종합하여 최종 요약본을 생성한다.

4. 구현 결과

본 논문에서 제안된 회의 요약 기능을 제공하는 MAS 시스템의 구현 결과는 그림 4, 5, 6 을 통해 확인할 수 있다.



(그림 4) 회의 진행 페이지

회의 진행 페이지는 사용자의 캠 화면과 음성 및 실시간으로 문자화되는 회의 스크립트 등을 제공한다.

(그림 5) 회의 요약 범위 지정 페이지

회의가 종료되면 회의 요약을 위해 다음과 같은 세가지 단계를 거쳐야 한다.

1) 회의 스크립트를 수정 또는 삭제 2) 요약본 제목 및 요약본 형식을 지정 3) 각 제목에 해당하는 요약 범위 지정

(그림 6) 회의 결과 페이지

생성한 요약본과 스크립트를 확인할 수 있는 페이지로 이동하게 된다. 회의 요약본과 스크립트는 txt, docx 파일로 저장할 수 있다.

5. 결론

본 논문은 효율적이고 간편하며, 회의 내용을 직관적으로 정리할 수 있는 화상 회의 서비스 구축에 대해 연구하고 있다. 연구 목표를 달성하기 위해 STT 와 요약 모델을 통한 회의록 생성이라는 컨셉을 고안했으며, 이러한 컨셉이 유용한 서비스로 성장하기 위해선 STT 와 대화 요약 성능의 고도화가 필수적이다. 본 논문에서는 다양한 사전 학습 모델과 방법론을 비교하여 선정된 모델을 적용하였다.

현재 회의록 생성은 사용자가 제목을 지정하고 각 제목에 해당하는 요약 범위를 지정해야 하는 단계가 존재한다. 추후 요약 범위를 자동으로 지정하여 각 범위에 해당하는 제목을 생성하는, 군집화 및 제목 생성 기술을 적용하기 위한 연구를 진행하여 사용성을 보완할 수 있다.

Acknowledgement

본 연구는 2022 년도 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학지원사업의 결과로 수행되었음(No.2018-0-00213)

참고문헌

[1] “코로나 19 가 가져온 언택트 문화, 포스트 코로나 시대를 준비하며”, 복지타임즈, 2020 년 06 월 05 일 수정, 2022 년 9 월 9 일 접속, https://www.bokjitimes.com/news/articleView.html?idxno=23242

[2] H. Alvestrand, RFC 8825 - Overview : Real-Time Protocols for Browser-Based Applications, 2021.

[3] Julius Adorf, "Web Speech API", 2013.

[4] Mike Lewis et al., BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, 2019

[5] Alec Radford et al., Language Models are Unsupervised Multitask Learners, 2018

[6] Yinhan Liu et al., Multilingual Denoising Pre-training for Neural Machine Translation, 2020

[7] M. Petit-Huguenin et al., RFC 8839 - Session Description Protocol (SDP) Offer/Answer Procedures for Interactive Connectivity Establishment (ICE), 2021.

[8] Alessandro Solbiati et al., Unsupervised Topic Segmentation, 2021

[9] Nils Reimers et al., Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, 2019

[10] Philipp Moritz et al., Ray: A Distributed Framework for Emerging AI Applications, 2017