연구과제명: 화자분리와 대본을 이용한 뉴스 음성 요약

|  |  |  |  |
| --- | --- | --- | --- |
| 분리된 segment에서 segment내에 포함된 화자의 수 별로 화자 분리를 진행하였다. 이제 분리된 segment들 사이에 같은 화자끼리 묶어야 하므로 이 작업을 위해 화자 인식 기술을 사용하였다.  화자 인식 task를 수행하기 위해 사용할 모델은 spkrec-ecapa-voxceleb으로 위와 같은 EER을 보인다. 위 모델을 voxceleb1과 voxceleb2로 trained되었다. segment별로 분리된 화자를 해당 모델을 사용하여 같은 등록된 화자들과 비교하고, 일치하는 화자와 같은 화자 이름을 가지게 된다. 이렇게 되면 각 segment별로 어떤 화자가 포함되어 있는 지 확인할 수 있고, 이를 VAD알고리즘을 다시 사용하여 순서별로 구분한다.    전체적인 흐름을 나타내면 왼쪽 그림과 같다. 다만 음성 내에 화자가 몇 명인지를 파악하는데 CountNet이 적절하지 않았으므로 직접 명시하였는데, 이 부분을 해결하기 위해 speaker diarization에 대한 연구가 추가적으로 필요하다. 해당 시스템을 통하여 뉴스 음성을 넣어 결과를 확인하였다.  1.    첫 오류는 발음 또는 음질에 따른 STT모델의 오류일 수 있지만 두번째 오류의 음성을 확인해 보니 segment를 만들 때 10초의 길이의 제한에 의해 중간에 잘린 사실을 확인할 수 있었고, 이로 인해서 STT모델이 부정확한 결과를 산출한 것으로 생각되었다.  2.  2번째 뉴스의 요약을 확인해 보니 위와 같이 전체적으로 정확한 결과가 나타났다. 하지만 chatGPT를 ‘채 GPT’또는 ‘채축 PT’와 같이 부분적으로 오류가 발생하였다. 기본적으로 STT모델의 음성 input은 길이가 길수록 문맥 정보를 활용하여 성능이 향상되지만 짧은 segment를 넣어 문맥을 정확하게 파악하지 못한다고 생각하였다.  이러한 문제를 해결하기 위하여 화자 분리된 segment들을 각각 STT모델에 넣지 않고, 화자 별로 하나의 음성으로 합친 후에 STT모델에 넣으면 정확성이 향상될 것으로 판단하였다. | | | |
| **작성자** | **일자 2023-05-04** | **확인자** | **일자 2023-05-05** |
| **서명 신원철** | **서명 김유성** |