연구과제명: 화자분리와 대본을 이용한 뉴스 음성 요약

|  |  |  |  |
| --- | --- | --- | --- |
| 분류 별 meteor점수와 rouge점수는 위와 같았다. 위 점수는 각 분류별로 화자 별 참조 요약문을 만들고, 모델을 통해 생성된 결과와 비교한 값인데, 각 분류별로 3개의 참조 요약문을 만들었다. 참고를 위한 타 요약모델의 rouge및 meteor점수는 참조 자료1에 나와있다.    하지만, 해당 점수는 참조 요약문의 성향과도 관계가 있고, 3개는 해당 모델의 성능을 평가하기에 대표성이 부족하다는 단점이 있다. 해당 지표에 대해서 모델의 성능을 신뢰할 수 있게 하기 위해서는 많은 평가 데이터가 필요한데, 그러면 화자 별 요약의 참조 요약문의 수를 늘려야 한다. 그러나 이는 시간적, 비용적 한계에 의해 현실적으로 힘들기 때문에 다른 평가 방식에 대해서 고려를 해 보았다. 참조 자료2에서는 참조 요약문 없이 요약문의 성능을 평가하는 방식인 BLANC를 제안했다. 위 평가 지표에 대한 간단한 설명은 다음과 같다. 모델이 생성한 요약문을 특정 기준으로 분리하고, random하게 특정 부분을 blanc처리한다. 이렇게 blanc처리된 요약문을 언어 모델을 사용하여 원문으로 복구하고, 복구된 원문과 원문 사이의 유사도를 기반으로 점수를 평가한다. 이 방법을 사용하면 참조 요약문 없이 모델을 평가할 수 있고, 참조 자료1의 data를 통해 타 모델과의 성능 비교 또한 가능하다. 고로 해당 평가 지표를 사용하여 각 분류당 100개의 data를 평가하여 성능을 측정할 예정이다. 화자 별 요약의 평가 흐름도는 아래와 같이 나타낼 수 있다.    위의 방법으로 화자 별 요약의 성능을 측정하고 기존 모델과의 성능을 비교한다.  참조자료 1: Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R., & Radev, D. (2020). SummEval: Re-evaluating Summarization Evaluation. arXiv preprint arXiv:2007.12626.  참조자료 2: Vasilyev, O., Dharnidharka, V., & Bohannon, J. (2020). Fill in the BLANC: Human-free quality estimation of document summaries. Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, 11-20. arXiv preprint arXiv:2002.09836. | | | |
| **작성자** | **일자 2023-05-26** | **확인자** | **일자 2023-05-26** |
| **서명 신원철** | **서명 김유성** |