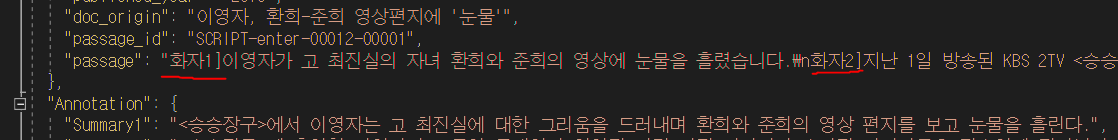
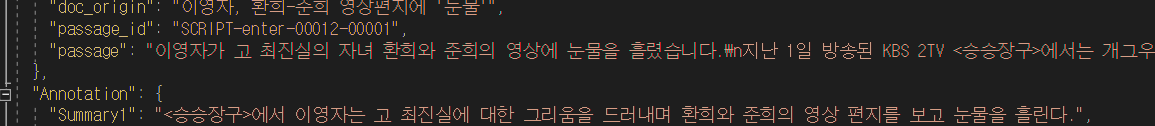
두 시스템 사이의 정확한 비교를 위해 두 모델을 유사한 데이터로 학습을 시켰습니다. 데이터는 ai hub의 `방송 콘텐츠 대본 요약 데이터`를 사용하여 kobart를 사용하여 학습하였는데, 한 모델은 기존 데이터를 그대로 사용하고, 한 데이터는 화자 정보를 지운 다음 학습을 진행하였습니다.



<원본 데이터>

<수정 후 데이터>

요약 모델은 kobart로, 파라미터는 Batch size: 4, Max length: 512, Optimizer: adam, learning rate: 0.00003, epochs: 20로 통일한 뒤 성능 평가를 진행하였습니다.

먼저 두 모델의 성능은 다음과 같이 측정되었습니다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | rouge-1 | rouge-2 | rouge-ㅣ | meteor | Blanc |
| 화자 분리X | 0.354 | 0.257 | 0.313 | 0.334 | 0.288 |
| 화자 분리O | 0.557 | 0.299 | 0.399 | 0.532 | 0.307 |

모든 지표가 화자 분리를 한 모델에서 더 좋은 성능을 보였습니다.

다음은 해당 모델을 적용한 두 시스템에 대한 평가입니다.

데이터는 총 60개의 뉴스 음성을 사용하였습니다. 뉴스에서 여러명이 말하는 multi speaker와 single speaker을 각각 30개씩 사용하였고, 음성에 대한 참조 요약문이 없기 때문에 blanc 평가 지표를 사용하여 시스템의 성능을 평가하였습니다. 그 결과는 다음과 같았습니다.

|  |  |  |
| --- | --- | --- |
|  | Multi speaker | Single speaker |
| 화자 분리X | 0.278 | 0.263 |
| 화자 분리O | 0.253 | 0.222 |

화자 분리를 적용하지 않은 모델에 비해 화자 분리를 적용한 모델이 화자가 여러명인 경우에는 더 좋은 성능을 보였습니다. 하지만 화자가 한명인 경우에는 오히려 요약문의 품질이 낮아진 것을 확인할 수 있었습니다.

해당 결과가 나타난 이유에 대해서 생각을 해 보았는데, 화자 분리를 적용한 시스템의 경우 화자분리, 화자 인식의 딥러닝 모델을 사용하는데, 이 부분에서 모델이 틀린 결과를 내는 경우에 만들어지는 대본에서 오류가 발생하고, 이러한 요인이 결과에 영향을 미친 것으로 생각됩니다. 다만, 화자가 여러명인 경우에는 이러한 오차에도 불구하고 성능 향상이 나타났는데, 이는 대본에서 화자 정보를 제공해 주는 것이 요약문 품질 향상에 어느정도 도움이 된다는 것으로 생각할 수 있습니다.