**캡스톤디자인 주간 보고서**

|  |  |
| --- | --- |
| 팀원 | 이예준, 박기홍 |
| 주제 | 미리내 – ML based Korean Educational Reconjugator using seq2seq |
| 기간 | 2020. 10. 19. ~ 2020. 10. 25 |
| 진  행  내  용 | 1. **월요일 정기 회의 (10.19)** - CTO분께서 Mirinae pipeline에 AI hub 등에서 가져온 여러 corpus들을 가지고 seq2seq model에서 학습시킬 수 있도록 training-pair를 추출해주는 기능을 새롭게 추가해주셨다. 우선 jamo-level까지 분해하지 않고 기존의 모델로 학습시켜보자는 의겨이 나왔다. 2. **New training-pairs (10.21)**  - 이번에 제공해주신 새로운 training-pairs를 가지고 jamo-level이 아닌 기존의 모델로 학습을 시켜보았다. 이는 이전보다 유의미한 시도였다. 이전엔 되게 적은 수의 training set이었기 때문에 1064개의 pairs뿐이었지만 이번엔 822842개의 training-pairs였다.   Text  Description automatically generated  따라서 indexing된 word의 수도 확연히 차이가 났고, 이전엔 over-fitting이 의심될 정도의 결과값을 보였다면 이번엔 인풋에 대하여 정확하게 맞추진 못했지만 학습한대로 결과값을 유추하려는 모습을 보였다. 현재는 인코더나 디코더의 내부구조를 조금씩 바꿔보거나 계속해서 jamo-level의 wild-card rule을 정하거나 학습시킬 데이터를 조금 더 형태소에 가깝게 분해시켜보는 설계를 계획중이다. |