**캡스톤디자인 주간 보고서**

|  |  |
| --- | --- |
| 팀원 | 이예준, 박기홍 |
| 주제 | 미리내 – ML based Korean Educational Reconjugator using seq2seq |
| 기간 | 2020. 11. 09. ~ 2020. 11. 15 |
| 진  행  내  용 | 1. **화요일 회의 (11.10)**  * 형태소를 합치는 method를 직접 개발하기로 하고, 관련 자료를 받고 method에 대한 의견을 제안받았다. 그 결과 Merger v1.0를 개발하고 테스트를 진행하였다. 지난주에 이어 n gram의 n 파라미터에 대해 조금씩 수정해보고 dimention도 바꿔보면서 stable한지 테스트를 해보라는 CTO분의 의견이 있었다.  1. **Morpheme merger 개발 (11.10 ~ 11.11)**  * 이예준 : seq2seq model을 거쳐 나온 일련의 형태소 배열의 결과물들의 형태소를 다시 이어 붙혀 완전한 문장을 만들 필요가 있었다. 그래서 한국어의 여러 규칙 및 불규칙들을 조사하였고, 이를 여러 개의 함들로 나누어 큰 뼈대를 만들었다. 설계를 바탕으로 Morpheme Merger v1.0을 개발하였고 테스트 결과 한국어가 stable하게 잘 합쳐지는것을 확인하였다.  1. **Apply 90% for training & 10% for tesing (11.13 ~ 11.14)**  * 박기홍 : 현재 만들어진 seq2seq과 fastText 기법을 결합한 Reconjugator를 가지고 테스트를 진행하였다. 90%의 sentence pairs만 seq2seq model을 통과시켰고, 10%의 pairs는 테스트를 위해 통과시키지 않았다. Sentence pairs를 랜덤하게 구성하여 모델을 10회 구성하고 테스트를 했는데, 올바르게 변환된 것은 전체 표본의 59%에 불과하였다. 그러나 main verb를 제외하고 tense 자체가 옳게 변환된 경우는 전체 표본의 90%로 나타났기 때문에, main verb가 변하는 현상을 줄이는 방향으로 개선해 나간다면 좋은 결과를 보일 수 있을 것으로 예상되었다.  1. **Attention 및 테스트 결과값 visualizing method 탐색 및 구상**  * 이예준 : 모델을 개선한 이후의 테스트 결과값을 분석하고 시각화하면 이후 모델을 개선할 방향을 모색하는데 좀더 도움이 될것이라고 판단되어 결과값을 visualizing할 필요를 느끼고 R&D 및 구상을 하였다. 다음주까지 결과값을 visualizing하는 방법에 대해 방향을 정하고 method를 구현할 예정이다. |