**캡스톤디자인 주간 보고서**

|  |  |
| --- | --- |
| 팀원 | 이예준, 박기홍 |
| 주제 | 미리내 – ML based Korean Educational Reconjugator using seq2seq |
| 기간 | 2020. 11. 02. ~ 2020. 11. 08 |
| 진  행  내  용 | 1. **수요일 회의 (11.04)** - 지난 주 형태소 단위의 Morpheme words를 pre-trained embedding vector로 사용하여 seq2seq model에 학습을 시킨 것에 대한 결과가 상당히 좋게 나왔지만 미세하게 틀리는 것이 뒤의 conjugating하는 부분은 괜찮았지만 맨 앞 verb-phrase들의 실수는 문장의 의미 전달에 있어서 영향이 크기 때문에 pre-training 하는 부분의 n-gram의 n 파라미터에 대해 조금씩 수정하면서 context를 학습시키는 것이 어떠냐는 CTO분의 제안이 있었고, 아무래도 형태소 단위로 쪼개어 학습을 진행했기 때문에 결과물도 형태소들의 배열로 나오는데 이것을 다시 합쳐서 완전한 문장을 만들어 줘야 하기 때문에 이를 도와주는 method를 직접 개발을 할 것인지 아니면 또 다른 학습 모델을 만들어야 할지에 대한 회의를 진행하였다. 2. **Morpheme merger 개발 (11.05 ~ 11.07)**  - 이예준 : seq2seq model을 거쳐서 나온 일련의 형태소 배열의 결과물들의 형태소를 다시 이어서 완전한 문장을 만들기 위해 한국어의 여러 규칙 및 불규칙들을 조사했고 이를 여러 개의 함수들로 나누어서 큰 뼈대를 만들었다. 다음주중으로 v1.0의 개발 완성을 목표로 하고 있다. 3. **Apply 90% for training & 10% for tesing (11.08)**   - 박기홍 : 기존에 하던 테스트는 단순히 10개 정도의 random한 choice에 의해 진행되었지만, 제대로 학습이 되었는지에 대해서 확인하기 위해서 원래의 Morpheme words는 vocab-dictionary에 등록을 하되, 90%의 sentence pairs만 seq2seq model을 통과시켰고 10%의 seq2seq model을 거치지 않은 데이터들을 통해 테스트를 진행했다. 이에 대한 결과로 학습이 잘 진행되고 있다는 것을 재확인했다. |