

Architecture

- Bidirectional LSTM + Language Model
- **Bidirectional Language Model** 구조를 가짐.

초기 임베딩

- N-gram 방식으로 단어를 부분으로 쪼개어 Conv 연산 후 Max Pooling
- Highway Network를 통해 불필요한 정보를 억제하고 중요한 정보를 보존.

Bidirectional Language Model

- Forward LM + Backward LM
 - Forward : 현재 토큰을 통해 다음 토큰을 학습
 - Backward : 현재 토큰을 통해 이전 토큰을 학습
- 두 방향의 Log likelihood를 최대화하는 방식으로 학습.

ELMo Embedding

1. 각 층에서 양방향 출력값을 Concat
2. 각 층마다 가중치를 부여
3. 가중치가 부여된 벡터를 더해줌.
4. Scale Factor r 을 곱해 최종 출력 크기 조정.

Evaluation

- SQuAD, SNLI, SRL, Coref, NER, SST-5의 태스크를 수행.
- ELMo를 추가하는 것으로 모든 task에서 SOTA 달성
- BiLM의 모든 층을 사용하는 것이 더 높은 성능을 보임.
- Input, Output 중 어느 층에 ELMo를 적용함에 따라 task에 따라 성능 향상을 보임.