

1. Machine Translation

machine translation이란 특정 언어의 문장을 다른 언어의 문장으로 번역하는 것을 말한다.

- input(x)-source language
- output(x)-target language

Rule-based 방식 사용 → 미리 각 언어의 단어에 맞는 의미를 구축한 후 대응되는 단어를 찾아 번역하는 방식.

$\operatorname{argmax}_y P(x|y)P(y)$

- source language, target language 모두 표현되는 데이터인 parallel 데이터를 통해 확률을 계산한다.

$P(x|y)$ 값의 학습

$P(x|y)$ 란 y문장이 주어졌을 때, x가 정렬되어 있을 확률

*alignment*는 서로 다른 언어 두 문장 사이에 특정 단어 쌍들의 대응을 의미한다. 서로 다른 언어 중 대응되는 단어도 있지만 대응되지 않는 단어도 존재한다. 이런 단어를 *spurious word*라고 한다.

*fertile word*는 *spurious word*와 달리 하나의 단어가 여러가지 단어와 대응되는 단어를 의미한다.

Heuristic 알고리즘 → 너무 낮은 확률을 버리고 계산하는 알고리즘

SMT의 단점

- 성능에 비해 구조가 복잡하다.
- 미리 구축된 단어의 의미를 사용하므로 추가적인 자료가 필요하다.
- 유지보수를 위해 사람의 노력이 많이 필요하다.

2. Sequence to sequence

Neural Machine Translation NMT → seq2seq 구조

-두 개의 RNN 포함

Encoder

- 번역을 진행할 문장 입력
- 각 단어의 임베딩 벡터가 각 시점마다 입력값으로 사용됨
- 소수 문장을 마지막에 인코딩시켜 디코더에 넘김

Decoder

- 인코더의 마지막 hidden state에서 넘어온 번역할 문장과 문장의 시작을 의미하는 start 토큰을 입력 받아 다음에 나올 단어의 확률분포 계산
- decoder의 출력값은 다음 decoder의 입력값이 된다.
- end 토큰을 출력하면 종료

BLEU 지표

Machine Translation을 평가하는 지표.

$$BLEU = \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

위 식은 candidate의 문장이 짧으면 점수가 높게 나오는 문제 발생.

$$BLEU = BP \times \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

따라서 brevity penalty 적용.

Machine translation의 한계

- 학습 데이터에 존재하지 않는 단어가 입력되면 목표 단어 생성 불가
- 긴 문장의 문맥을 유지하기 힘들다
- MNT 학습을 위한 많은 양의 병렬 corpus 구축 어려움.

3. Attention

seq2seq는 incoder의 마지막 hidden state에 번역하고자 하는 문장의 모든 정보가 담겨 있어야 한다. 따라서 마지막 hidden state에서 인코딩되어 정보가 물리는 정보병목현상이 발생한다.

Attention은 decoder의 각 단계에서 incoder를 직접 연결하여 source 문장의 특정 부분마다 집중함으로써 정보병목현상을 해결한다.

attention의 장점

- decoder가 source 문장의 특정 부분에 집중하기때문에 NMT의 성능을 향상
- 문장의 길이가 길수록 기울기 loss가 발생하는데 attention은 전체 문장 중 확률이 높은 단어에 더 집중하므로 기울기 loss 문제를 해결