

Machine Translation, Seq2Seq, Attention

"Machine Translation, which is a major use-case of sequence-to-sequence, which is improved by attention."

Machine Translation 기계번역

Source → Target

Statistical Machine Translation (SMT)

- 흐름 모델
- Translation Model + Language Model
- 많은 양의 parallel data 필요
- alignment
- Decoding 계산량 많음 \Rightarrow Dynamic programming

Neural Machine Translation (NMT)

- end-to-end 학습 가능, subcomponents 별 따로 최적화 필요 X
- 2개 RNN으로 구성된 seq2seq 모델
- Human effort 강도 (feature engineering X, 모든 source-target 동일 method O)
- Interpretability 부족, Rule: guideline 적용 어려움 (-)

Sequence-to-sequence

- 인코더 RNN & 디코더 RNN 두로 구성
 - ↳ 양방향 Source sentence 인코딩
 - ↳ 인코딩을 condition으로 target sentence 생성
- 인코더의 마지막 cell output이 디코더의 초기 hidden state로 사용,
디코더의 각 timestep의 output이 다음 단어 예측
 $t-1$ 의 output이 t 의 입력으로 사용
- summarization, dialogue, parsing, code generation, time series ...

Multi-layer RNNs

- timestep 진행과 함께 layer 추가, 좀 더 higher-level feature

Greedy decoding

- step별 argmax word 선택 (exhaustive search 보다 효율적)
- 최적화를 보장하지 않으므로 부자연스러운 문장 생성 가능

Beam search decoding

- k 개의 best translation 유지하며 decoding (k: beamsize)
- 최적화를 보장하지 않으나 exhaustive 보단 효율적, greedy 보다 자연스러움

Evaluation for MT \rightarrow BLEU score

Attention

기존 NMT는 encoder의 마지막 cell만 decoder의 초기 hidden state로 전달

\Rightarrow Bottleneck prob. 발생

\therefore Decoder의 각 step에서 source sentence의 특정 부분에 집중하도록 connection 추가

① 'Encoder의 step별 encoding' & 'Decoder hidden state' 사이
similarity 계산 \Rightarrow attention score & weight 계산

② Softmax로 구한 attention distribution을 사용해 encoder hidden state를
가중합해 attention output 계산

③ Attention output과 decoder hidden state를 concat하고
이후 step의 word를 생성

\rightarrow NMT 성능 향상 : decoder가 source sentence 특정 영역에 집중
bottleneck prob. 해결
gradient vanishing 완화