

Wrap-up

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Enhancing NLG using Monolingual Corpus

- 자연어 생성 연구의 꽃 (개인 의견)
 - Parallel corpus는 매우 한정적이고 비싼 자원
 - 꼭 기계번역 task가 아니더라도 labeling corpus도 포함됨
 - 따라서 적은 양의 코퍼스로 학습한 decoder의 언어모델은 약해질 수 밖에 없음
 - 이에 반해 monolingual corpus는 온 세상에 널려 있음
 - 그러므로 monolingual corpus를 활용하여 decoder의 성능을 높일 수 있다면 좋을 것
- Back Translation
 - 반대쪽 번역 모델을 활용하여 pseudo corpus를 만들어, 모델의 성능을 향상시키는 방법
 - 매우 간단하고 직관적인 방법

RL tries to solve:

Training with better objective

- PPL은 실제 번역의 품질을 반영 못함
- BLEU와 같은 더 좋은 metric을 통해 학습할 순 없을까?
 - e.g. GAN: 진짜 같은 이미지 생성
- 하지만 BLEU는 미분 불가능

Remove Teacher Forcing

- Auto-regressive task이므로 보통 teacher forcing을 통해 학습
- 따라서 training과 inference 사이의 괴리(discrepancy) 발생
- RL을 통해 샘플링 기반의 학습 수행

However, RL has limits

- MLE에 비해 훨씬 비효율적인 학습 구조
 - 샘플링 방식에 의존
 - Reward는 scalar 값이므로 최대화 하는 방향을 알 수 없음
- MLE 방식 위에서 regularization을 통해 해결하고자 하는 시도들이 연구됨

