

Wrap-up

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Language Model 이란

- 실제 우리가 사용하는 (or 타킷 도메인) 언어의 분포를 확률 모델로 모델링 한 것
 - Chain rule에 의해서 문장의 확률을 모델링하는 것은
단어들이 주어졌을 때, 다음 단어의 확률을 모델링하는 것과 같음

$$\begin{aligned}P(x_{1:n}) &= P(x_1, \dots, x_n) \\&= P(x_n | x_1, \dots, x_{n-1}) \cdots P(x_2 | x_1) P(x_1) \\&= \prod_{i=1}^n P(x_i | x_{<i})\end{aligned}$$

$$\log P(x_{1:n}) = \sum_{i=1}^N \log P(x_i | x_{<i})$$

- 언어 모델을 통해 우리는 아래의 task를 수행할 수 있음
 - 1) 주어진 문장들 중에서 가장 fluent한 문장을 골라낼 수 있다.
 - 2) 단어들이 주어졌을 때, 다음 단어를 확률적으로 예측할 수 있다.

Perplexity

- 매 time-step 마다 모델이 동등하게 헛갈리고 있는 평균 단어 수
 - 헛갈리는 단어가 적을수록 좋은 것 == lower is better
- 문장의 확률의 역수에 단어 수 만큼 기하 평균을 취한 것
 - 문장의 likelihood가 높을수록 좋은 것 == lower is better
- Cross Entropy에 exponential을 취한 것
 - GT 분포와 모델의 분포가 비슷할 수록 좋은 것 == lower is better

n-gram and Neural Network Language Model

n-gram

- 단어를 discrete symbol로 인식
 - Exact matching에 대해서만 count
- 학습 코퍼스에 word sequence가 존재해야만 확률 값을 추정 가능
 - Markov Assumption 도입
- 쉽고 직관적인 구현
 - 학습(counting) 후, 추론(table look-up)
 - Scalable하며, 저렴한 계산 비용

NNLM

- 단어를 continuous vector로 변환
 - Unseen word sequence에 대처 가능
 - Generalization에 강점
- 비싸고 느린 연산 추론 과정
- Generation task에 굉장히 강함