

Wrap-up: n-gram

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Conclusion (n-gram)

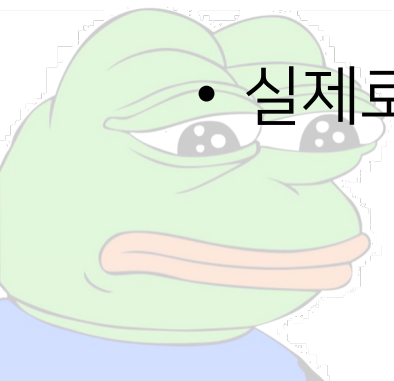
- Pros

- Scalable: 쉽게 (large vocabulary 등의) 대형 시스템에 적용 가능
- n-gram 훈련 및 추론 방식이 굉장히 **쉽고 간편**

- Cons

- **Poor generalization:** 등장하지 않은 단어 조합에 대처 미흡
 - 단어를 discrete symbol로 취급
 - 따라서, 비슷한 단어에 대한 확률을 이용(leverage, exploit)하지 못함
 - Smoothing과 Back-off 방식을 통해서 단점을 보완하려 했으나, 근본적인 해결책이 아님
- Poor with long dependency: 멀리 있는 단어에 대해 대처 불가
- n이 커질수록 용량도 커짐

- 실제로 어플리케이션 적용(ASR, SMT)에 있어서 큰 과제



Conclusion (n-gram)

- Pros

- Scalable: 쉽게 (large vocabulary 등의) 대형 시스템에 적용 가능
- n-gram 훈련 및 추론 방식이 굉장히 **쉽고 간편**

- Cons

- **Poor generalization:** 등장하지 않은 단어 조합에 대처 미흡
 - 단어를 discrete symbol로 취급
 - 따라서, 비슷한 단어에 대한 활용을 이용(leverage, exploit)하지 못함
 - Smoothing과 Back-off 방식을 통해서 단점을 보완하려 했으나, 근본적인 해결책이 아님
- Poor with long dependency: 멀리 있는 단어에 대해 대처 불가
- n이 커질수록 용량도 커짐

- 실제로 어플리케이션 적용(ASR, SMT)에 있어서 큰 과제

