

Wrap-up

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

NLG is Auto-regressive Task

- 현재의 최선의 선택이, 나중에 최선의 선택이 아닐 수 있음
 - 따라서 greedy search는 최선의 선택을 하지 못할 수 있음
 - Exhaustive Search는 현실적으로 불가능함
- Beam Search:
 - Greedy search를 좀 더 확장하여 진행하는 것
 - NLG 이외의 분야에서도 많이 활용됨
 - 모델의 성능을 한층 더 끌어올릴 수 있는 방법

Mini-batch Parallelized Beam Search

- Parallelization for single sample beam search:
 - $k(\text{beam-size})$ 만큼 inference가 반복됨
 - 병렬 처리를 위해서 k 개의 샘플인 것처럼 추론 수행
- Mini-batch Parallelization for multi sample beam search:
 - $k \times \text{batch_size}$ 만큼 inference가 반복됨
 - 병렬 처리를 위해서 $k \times \text{batch_size}$ 개의 샘플인 것처럼 추론 수행

Finally!

- Now, you can build your own machine translation system,
 - which is better than Google Machine Translation in early 2010s .

