

Appendix: Sentence Embedding Overview

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Sentence Embedding via Word Embedding

- 흔히 시도되는 방법은 문장 내 단어 임베딩 벡터들을 모두 더하는 것

$$v_{sent} = \sum_{i=1}^n w_i$$

- 하지만 어순을 무시한 채 평균을 구하는 것이므로 detail이 뭉개질 수 있다.
 - I did **not** say that I am idiot.
 - I did say that I am **not** idiot.

Context Embedding

- 단어는 문맥(또는 문장 내 위치)에 따라서 의미가 변화
 - 따라서 문맥에 따른 임베딩이 필요함
- 문맥을 고려하기 위해서는 주변 단어들의 쓰임새를 살펴야 함
 - 이후 다룰 <텍스트 분류>, <자연어 생성> 주제들의 기본 원리

RNN

Seq2seq