

단어의 다양한 표현 방법

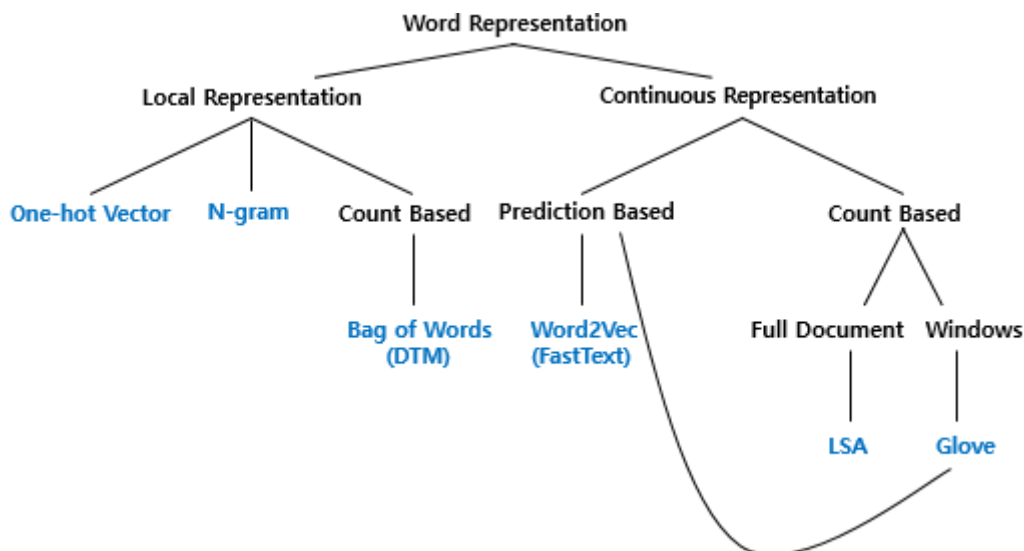
단어의 표현 방법

1. 국소 표현(Local Representation) 방법 : **해당 단어 그 자체만 보고**, 특정값을 맵핑하여 단어를 표현하는 방법
example : puppy(강아지) , cute(귀여운), lovely(사랑스러운)
각 단어에 1, 2, 3과 같은 숫자를 맵핑하여 단어를 표현
2. 분산 표현(Distributed Representation) 방법 : 그 단어를 표현하고자 **주변 단어를 참고**하여 단어를 표현하는 방법
example : puppy(강아지) , cute(귀여운), lovely(사랑스러운)
puppy(강아지)라는 단어 근처에 주로 등장하는 cute, lovely 단어를 이용. 즉, "puppy는 cute, lovely한 느낌이다"로 단어를 정의.

이 두 방법의 차이는 Local Representation은 단어의 의미, 뉘앙스를 표현할 수 없지만, Distributed Representation은 단어의 뉘앙스를 표현할 수 있음

비슷한 의미로 국소 표현 방법(Local Representation)을 이산 표현(Discrete Representation)이라고도 하며, 분산 표현(Distributed Representation)을 연속 표현(Continuous Representation)이라고도 함.

단어 표현의 카테고리화



※ N-gram이 Local Representation인 이유의 개인적인 생각 :

example : puppy is scared

n-gram을 사용하면 puppy와 scared가 상관관계가 있다는 것을 알 수는 있지만 "puppy가 cute, lovely, small한 느낌이다"라는 것을 알 수는 없음. (vector representation의 초점을 맞추어)

출처

- <https://wikidocs.net/31767>

