

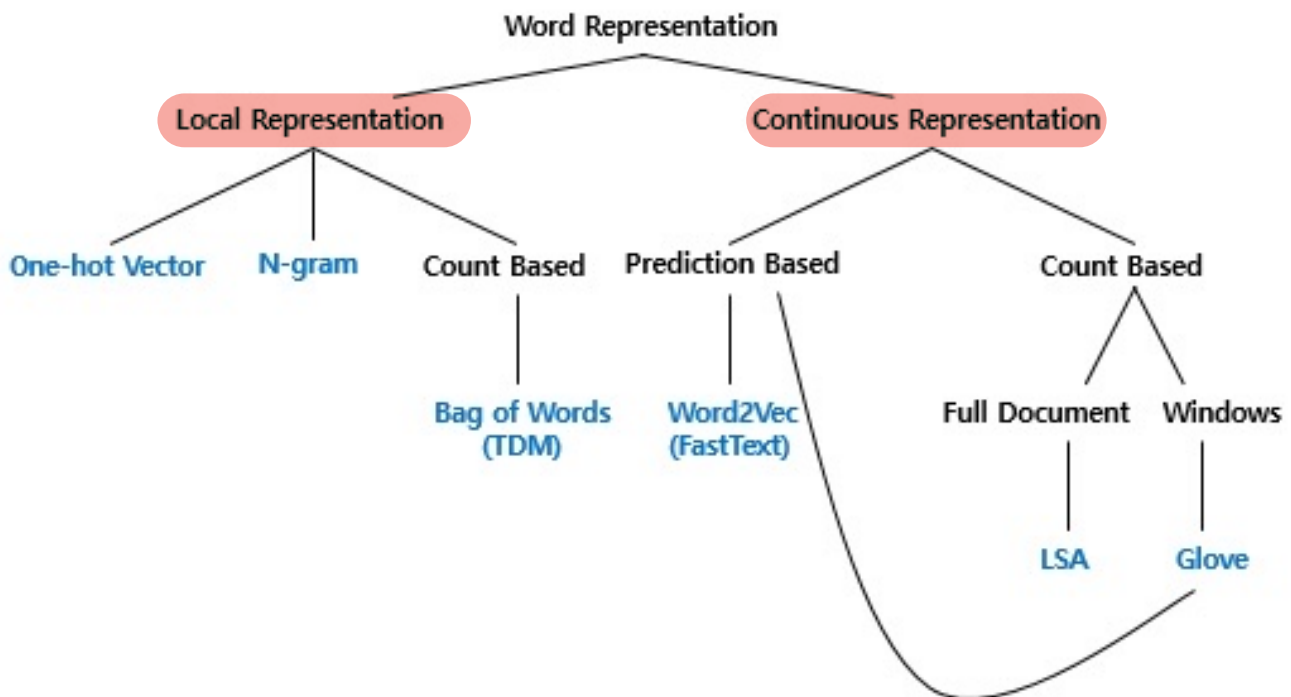
다양한 단어의 표현 방법

이번 챕터의 제목은 카운트 기반의 단어 표현 방법이지만, 카운트 기반의 단어 표현 방법은 다음 챕터인 **Bag of Words** 챕터에서 배우게 된다. 지금은 카운트 기반의 단어 표현 방법 외에도 다양한 단어의 표현 방법에는 어떤 것이 있으며, 앞으로 “딥 러닝을 이용한 자연어 처리 입문”이라는 책에서는 어떤 순서로 단어 표현 방법을 학습하게 될 것인지에 대해서 먼저 설명한다.

1. 단어의 표현 방법

단어의 표현 방법은 크게 ① **국소 표현 (Local representation)** 방법과 ② **분산 표현 (Distributed Representation)** 방법으로 나뉜다. ① 국소 표현 방법은 해당 단어 그 자체만 보고, “특정 값을 맵핑”하여 단어를 표현하는 방법이며, ② 분산 표현 방법은 그 단어를 표현하고자 주변을 참고하여 단어를 표현하는 방법이다.

ex) 등장한 횟수를 맵핑.



예를 들어 puppy(강아지), cute(귀여운), lovely(사랑스러운)라는 단어가 있을 때 각 단어에 1번, 2번, 3번 등과 같은 숫자를 맵핑 (Mapping)하여 부여한다면 이는 ① 국소 표현 방법에 해당된다.

반면, ② 분산 표현 방법의 예를 하나 들어보면 해당 단어를 표현하기 위해 주변 단어를 참고한다. puppy(강아지)라는 단어 근처에는 주로 cute(귀여운), lovely(사랑스러운)이라는 단어가 자주 등장하므로, puppy라는 단어는 cute, lovely한 느낌이다로 단어를 정의한다. 이렇게 되면 이 두 방법의 차이는 국소 표현 방법은 단어의 의미, 뉘앙스를 표현할 수 없지만, 분산 표현 방법은 단어의 뉘앙스를 표현할 수 있게 된다.

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