

⚙ Word2Vec의 특징

- 단어를 벡터로 취급함
- 단어 사이의 거리와 방향까지 알 수 있음
- 단어 임베딩의 size를 정함
- 많은 정보를 보여줌
- 더 확장하여 텍스트 간의 거리를 구하는 방법이 연구됨
- Tf-idf나 BagOfWords의 한계를 극복함

⚙ Skip Gram 알고리즘

- 단어의 주변에 나타나는 단어가 무엇인지 예측하는 신경망
- 최종적으로 나타나는 가중치 행렬은 해당 단어를 나타내는 임베딩 벡터의 모임

⚙ Skip Gram 변수

- V : 사전의 크기로서 전체 단어들의 개수
- N : 단어를 표현할 임베딩 벡터의 크기로서 신경망에서 hidden layer의 size
- Window Size : 주변에 나타나는 단어를 선정할 때 반경