WORD2VEC

- description : WORD2VEC
- author : 오션
- email : shlim@repia.com
- lastupdate : 2020-11-09

WORD2VEC

- Word to Vector [워드 투 벡터]

개념

- “ 자연어 처리(NLP)” 방식 중 하나로 수식을 통해 단어를 다차원 공간에 매칭하고, 벡터(숫자)로 표현해주는 머신러닝 알고리즘
- 같은 문맥에서 많이 등장하는 단어들은 기본적으로 비슷한 의미를 지닌다고 가정
- 또한 주위에 등장하는 단어들을 가지고 같이 나타날 수 있는 단어를 컴퓨터가 스스로 유추하고 학습하게 하는 알고리즘(비지도학습)

특징

- 인공지능은 이 단어들을 벡터(숫자)로 변환하여 좌표값을 통해 단어간 거리 계산 가능
- 거리가 가까워질수록 단어 간 유사도가 높아지면, 멀어질수록 유사도가 낮아짐
- 단어가 벡터(숫자)로 표현되었기 때문에 각 단어들의 벡터값을 통해 덧셈, 뺄셈, 평균 등 연산이 가능
- 이를 통해 단어 추론, 연관어 계산, 복잡한 개념 등의 표현이 가능함

의미

- 텍스트 기반의 모델 만들기는 텍스트를 숫자로 변경하여 알고리즘에 넣고 계산을 한 후 결과값을 낼 수 있다.
- 그 일환으로 텍스트를 숫자로 바꾸는 것으로 단어를 벡터로 바꾸는 것이다.
- 이때 벡터에 단어의 의미를 추가하여 단어와 단어간의 관계를알아내기 위함이다.
- 단어를 벡터로 바꾸는 모델을 단어 임베딩 모델(word embedding model)이라고 하며, word2vec는 단어 임베딩 모델들 중 대표적인 모델이다.

Ref

쉽게 씌어진 word2vec

딥 러닝을 이용한 자연어 처리 입문

youtube word2vec

첨부

오션, WORD2VEC, 워드투벡터

From:
http://125.132.25.164/dokuwiki/ - **문제를 잘 정의하는 것은 문제를 절반 해결한 것이다.** - 2023.12

Permanent link:
http://125.132.25.164/dokuwiki/doku.php?id=wiki:miscellaneous:word2vec&rev=1604889864

Last update: **2022/03/10 19:52**