

Part. 01

실전 데이터 분석 프로젝트

텍스트마이닝을 활용한 카카오톡 대화분석

FASTCAMPUS

ONLINE

머신러닝과 데이터분석A-Z

강사. 김용담

Life Changing Education

1워드 임베딩 역사



(Neural Probabilistic Language Model) (Skip-Gram with Negative Sampling)

- 처음으로 제안된 dense representation model
- Neural Network를 이용하여 주변 단어 •
 의 단어 등장 확률을 예측함
- NPLM에서 높은 계산량을 요구하는 문제점을 획기적으로 해결
- 본격적인 word embedding 시대 개막

fastText

(Subword SGNS)

- Word2vec에서의 OOV(Out-of-Vocabulary) 문제를 해결
- 학습 단위가 subword로 변경

→ ELMo

(Embeddings from Language Model)

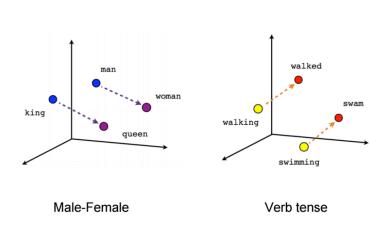
HERE

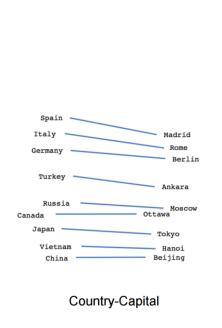
- Pretrained Language Model을 제안
- Bi-directional Language Model을 제안하 여 문맥을 반영한 워드 임베딩 기법 제시
- NLP에서 transfer learning이 확산됨

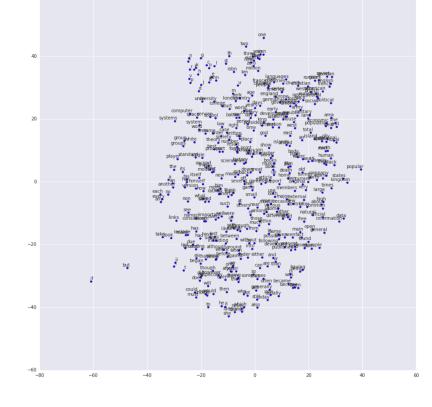
FAST CAMPUS ONLINE



I 워드 임베딩 공간의 특징







Properties of word2vec embedding space

Reference: https://tensorflowkorea.gitbooks.io/tensorflow-kr/content/g3doc/tutorials/word2vec/

FAST CAMPUS ONLINE

김용담 강사.





Part. 01 실전 데이터 분석 프로젝트

1감사합니다.

FASTCAMPUS

ONLINE

강사. 김용담

머신러닝과 데이터분석A-Z