WORD EMBEDDINGS CRASH COURSE

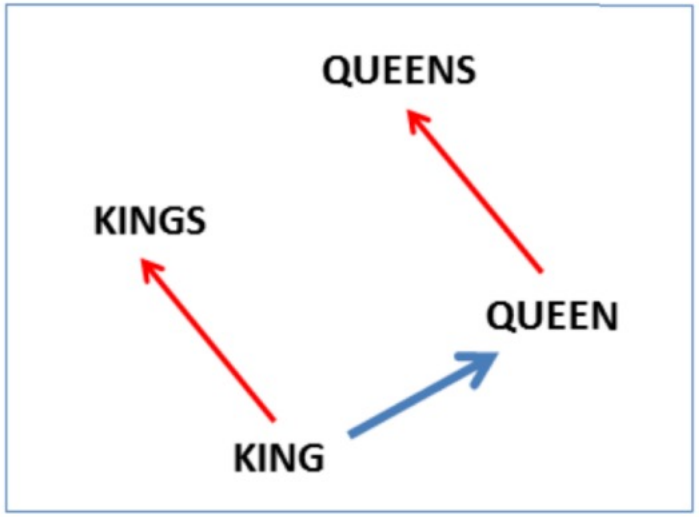
ALEX MEI | FALL 2021 | TEACHING IN CS

OVERVIEW

- What are Word Embeddings?
- Why do we need Word Embeddings?
- Word Embeddings Methods

CONTEXT

- Word Embeddings: translations of text into alternative representations
- Why? alternative representations are easier to manipulate and quantify
- Example Task: plagiarism checker for text similarity
- Example Solution: cosine distance between two vectorized representations



BAG OF WORDS MODEL

- Corpus (C): set of unique words in vocabulary under consideration
- Bag of Words: vector of length C with each element i denoting count of word i
- One Hot Vector: instead of count of word i, use 1 if word in text else 0
- Problems? hard to compare words with similar meanings

$$\underbrace{[0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ \ldots\ 0]}_{\text{car}}\ \text{AND}\ \underbrace{[0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ \ldots\ 0]}_{\text{motorcycle}} = 0$$

WINDOW-BASED MODELS

- Incorporate advantage of surrounding context for similarity comparison
- Co-occurrence Matrix: M[i][j] stores count of word j after word i
- N-gram Model: use window of n-word blocks
- Problems? High-dimensional and sparse
- Expensive Solution: singular value decomposition

similarity > 0

Counts	I	love	enjoy	UCSB	deep	learning
I	0	2	1	0	0	0
love	2	0	0	1	1	0
enjoy	1	0	0	0	0	1
UCSB	0	1	0	0	0	0
deep	0	1	0	0	0	1
learning	0	0	1	0	1	0

WORD2VEC (SKIPGRAM)

- Learning Task: predict surrounding context words given a target word
- Objective Function: maximize P(context word | target word)
- Model optimized using standard stochastic gradient descent
- Benefits: efficient and dynamic with new texts; distance related to similarity
- Continuous Bag of Words Variation: predict target word given context words

GLOVE

- Unlike Word2Vec which uses local context, GloVe uses global word context
- Training Objective: log-bilinear model with weighted least-squares based on the co-occurrence matrix
- Has explicitly defined locally linear contexts (i.e., word analogies)

PRACTICE I

- Why are word embeddings useful?
- What is the primary difference between the Word2Vec and GloVe models?
- What is one significant problem of the Bag of Words model that both Word2Vec and GloVe is able to solve?

PRACTICE II

- Compute the co-occurrence matrix (using the right neighbor) of the following sentences:
  - “Loose lips sink ships”
  - “I am jumping ship”
  - “This is a sinking ship”

- Use cosine distance to compute the similarity between “jumping” and “sinking”

SOURCES

- Many slides are adapted from Professor William Wang (UCSB)
- nlp.stanford.edu/projects/glove/
- code.google.com/archive/p/word2vec/