Word Embeddings

Machine Learning

Word Embeddings

- Word Embeddings: An NLP technique of representing words as vectors

- Several methods exist to do this kind of translation. Popular methods include

  - Dimension reduction techniques, like SVD

  - word2Vec (Neural networks)

  - GloVe

Context ≈ Meaning

- The purpose of word embeddings is to use closer vectors to represent similar words.

- That is, it detects similarities mathematically.

- More specifically it aims to capture similarities between words based on their <u>distributional properties</u> in large samples of language data.

- "a word is characterized by the company it keeps"

- Given enough data, usage and contexts, word embeddings can make highly accurate guesses about a word's meaning.

Similar words and similar vectors

- Two words with similar contexts mean similar things

  - Eg: Red & Blue, Cat & Dog

- Cosine Similarity - measuring distance between word vectors

- Embedding can (surprisingly) build even more meaning into the vectors

  - Eg: 'King'-'Man'+'8Woman' ≈ 'Queen'

Applications

- Analyzing any form of text for insights: reviews, survey responses, tweets, comments

- NLP

  - Text classification

  - Sentiment analysis

  - Translation

- Recommendation systems

Singular Value Decomposition

- Begins with a text corpus (Eg. All the tweets in say a yr)

- Assemble a word co-occurrence matrix: M

- Find a lower dimensional word embedding matrix W such that, $M = W * W^T$.

  - This step be done using Singular Value Decomposition (SVD)

- The W matrix now contains the vector representation for each word.

6

word2Vec

- Uses a simple neural network with an input layer, a hidden layer and an output layer

- Train the neural network to either

    - Predict context given a word (skip gram model) or

    - Predict word given a context (continuous bag of words model)

- I went to the cinema on Sunday

- I went to the beach on Monday

- My favorite pet is a cat

- I like walking my pet dog

- I bought a red Honda

- I bought a red car

- I bought a blue Toyota

I will have oats and eggs for breakfast