Distributed representations of words and phrases and their compositionality

Wednesday, June 28, 2023 12:35

- Skip gram model - efficient method for learning high quality distributed vector representations that capture a large number of precise syntactic and semantic word relationships
- Distributed representations of words in director space help learning algorithms to achieve better performance in NLP by grouping similar words
- Unlike other models used for training, training of the skip gram model does not involve dense matrix multiplications
- More efficient training
- An optimized machine can train on more than 100 billion words in one day
- Learned vectors explicitly encode many linguistic regularities and patterns
- Word representations are limited by their inability to represent idiomatic phrases that are not combinations of the individual words (Boston globe newspaper)
- Extension from word based to phrase based is relatively simple.
- non-obvious degree of language understanding can be obtained by using basic mathematical operations on word vector representations
- skip-gram model
    - Find word representations that are useful for predicting tee surrounding words in a sentence or a document
    - Given a sequence of training words, objective of the model is to maximize the average log probability where larger training sizes results in more training examples = higher accuracy
    - Cost of completing can be high
- Hierarchical softmax
    - softmax - a mathematical function that converts a vector of numbers into a vector of probabilities were probabilities of each value are proportional to the relative scale of each value in the vector
    - Uses a binary tree representation of the output layer with the W words as leaves and for each node explicitly represents the relative probabilities of its child nodes
    - The structure of the tree used by the hierarchical softmax has a considerable effect on the performance

.

    - Grouping words together by their frequency helps speed up
- Negative sampling
    - Noise constructive