World Embedding Evaluation on Clinical Notes

Agenda

Word Embeddings

Introduction

Word2Vec Model

About word2vec model

BERT

About Bert representation

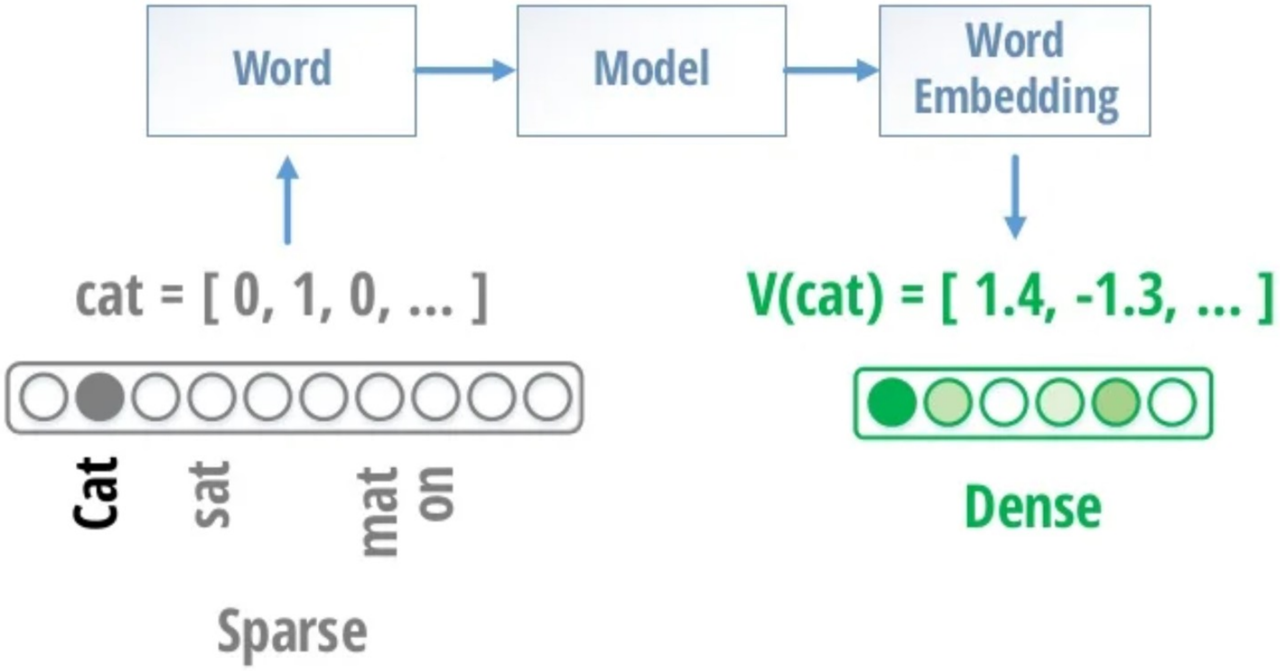
Experimentation

Evaluation on different models

Result

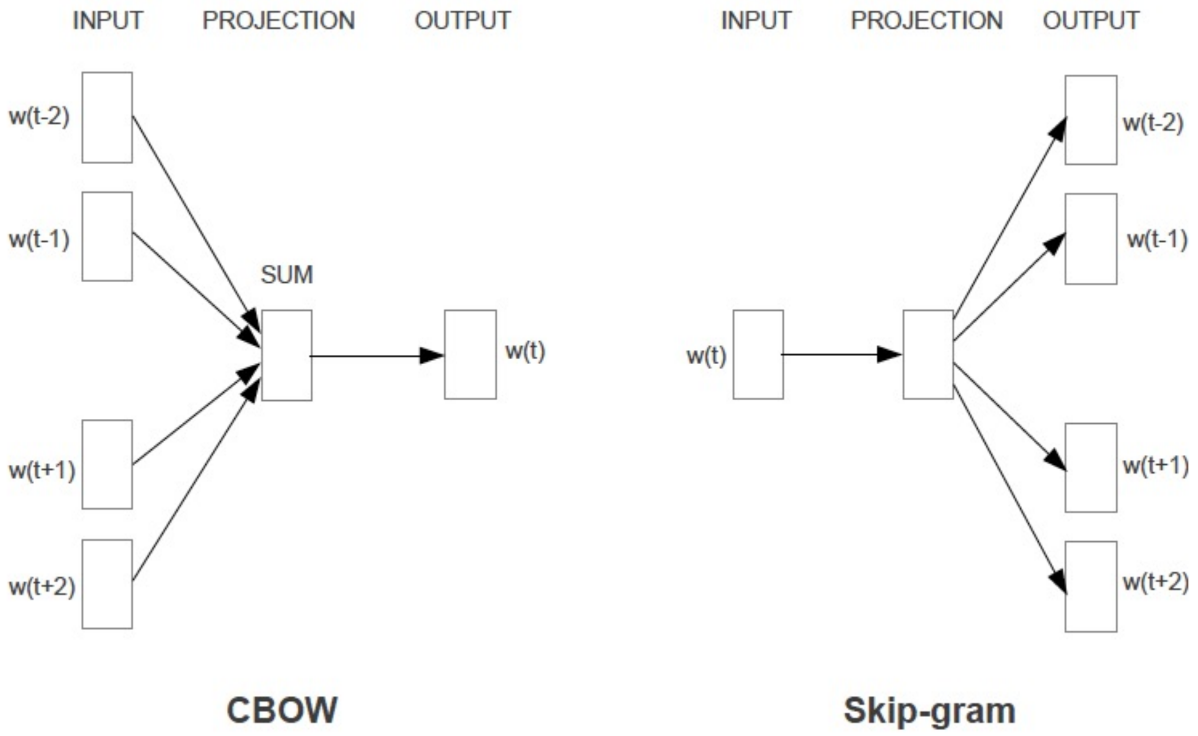
Final result

Word Embeddings



- Word embeddings are a type of word representation that allows words with similar meaning to have a similar representation.
- Word embedding methods learn a real-valued vector representation for a predefined fixed sized vocabulary from a corpus of text.
- Word embeddings give us a way to use an efficient, dense representation in which similar words have a similar encoding.
- From a sparse represntation (usually one hot encoding) to a dense representation

Word2Vec Model

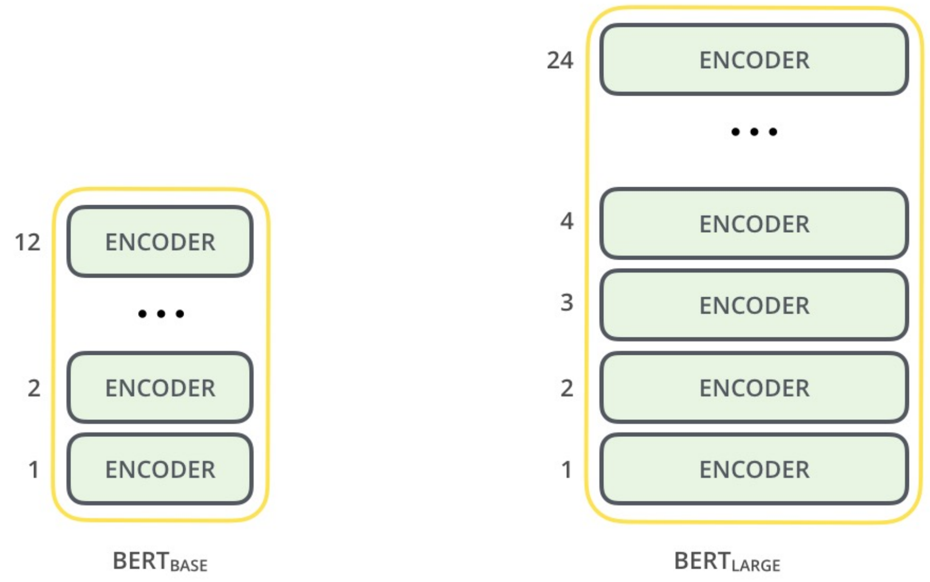


- Word embedding is one of the most popular representation of document vocabulary. It is capable of capturing context of a word in a document, semantic and syntactic similarity, relation with other words, etc.
- Word2Vec is one of the most popular technique to learn word embeddings using shallow neural network.
- Word2Vec is a method to construct such an embedding. It can be obtained using two methods (both involving Neural Networks): Skip Gram and Common Bag Of Words (CBOW)

CBOW Model: This method takes the context of each word as the input and tries to predict the word corresponding to the context.

Skip-Gram model: This looks like multiple-context CBOW model just got flipped. In the skip-gram model, given a target (centre) word, the context words are predicted.

Bidirectional Encoder Representations from Transformers (BERT)



- BERT makes use of Transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text.
- It is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of NLP tasks.”
- BERT is pre-trained on a large corpus of unlabelled text including the entire Wikipedia(that’s 2,500 million words!) and Book Corpus (800 million words).

The BERT architecture builds on top of Transformer. We currently have two variants available:

- BERT Base: 12 layers (transformer blocks), 12 attention heads, and 110 million parameters
- BERT Large: 24 layers (transformer blocks), 16 attention heads and, 340 million parameters
- BERT is pre-trained on two NLP tasks:
  - Masked Language Modeling
  - Next Sentence Prediction

Experimentation

1. **PubMed Word2vec:** Word vectors were induced from PubMed and PMC texts and their combination using the word2vec tool. The word vectors are provided in the word2vec binary format.

2. **Bert Large Model:** BERT is a transformers model pretrained on a large corpus of English data in a self-supervised fashion.

3. **Bio Bert model:** BioBERT (Bidirectional Encoder Representations from Transformers for Biomedical Text Mining), which is a domain-specific language representation model pre-trained on large-scale biomedical corpora.

4. **Bio Clinical Bert model:** The Bio_ClinicalBERT model was trained on all notes from MIMIC III, a database containing electronic health records from ICU patients at the Beth Israel Hospital in Boston, MA. For more details on MIMIC, see here. All notes from the NOTEEVENTS table were included (~880M words).

5. **PubMed Bert model:** PubMedBERT is pretrained from scratch using abstracts from PubMed and full-text articles from PubMedCentral. This model achieves state-of-the-art performance on many biomedical NLP tasks, and currently holds the top score on the Biomedical Language Understanding and Reasoning Benchmark.

Result

Medical Concept result:

No	Term 1	Term 2	PubMed Word2vec	Bert Large	Bio Bert	Bio Clinical Bert	PubMed Bert
1	Glaucoma	Fibrillation	0.384217	0.894104	0.872067	0.941967	**0.950928**
2	Carbatrol	Dilantin	0.547312	0.921481	0.877101	0.926522	**0.954106**
3	Cardiomyopathy	Tylenol	-0.002026	0.879889	0.778894	0.901180	**0.926579**
4	Herpes	Hyperthyroidism	0.227363	0.883667	0.872658	0.909672	**0.956872**
5	Seasickness	Nausea	0.240243	0.844041	0.858088	0.827762	**0.944510**

Clinical Notes result:

No	Model	Algorithm	Test Accuracy
1	PubMed Word2vec	Linear SVC	**94.51**
2	Bert Large	Linear SVC	72.56
3	Bio Bert	Linear SVC	89.02
4	Bio Clinical Bert	Linear SVC	90.24
5	PubMed Bert	Linear SVC	93.29

Thank You