

Chapter17 Unsupervised Learning – Word Embedding

1. 1-of-N Encoding to Word Embedding

- a) Generating Word Vector is unsupervised
- b) A word can be understood by its context

2. Context

a) Count-based (Glove Vector):

If two words w_i and w_j frequently co-occur, $V(w_i)$ and $V(w_j)$ would be close to each other

b) Prediction-based:

Sharing parameters:

$$z = W_1 x_{i-2} + W_2 x_{i-1}$$

The length of x_{i-1} and x_{i-2} are both $|V|$, the length of z is $|Z|$

The weight matrix W_1 and W_2 are both $|Z| * |V|$ matrices

$$W_1 = W_2 = W \Rightarrow z = W(x_{i-2} + x_{i-1})$$

Continuous bag of word (CBOW) model:

Predict the word given its context

Skip-gram:

Predict the context given a word

3. Multi-lingual Embedding

4. Multi-domain Embedding

5. Document embedding

- a) Word sequences with different length \rightarrow Vector with the same length
- b) Semantic Embedding

6. Beyond Bag-of-Word

- a) Paragraph Vector
- b) Seq2seq Auto-encoder
- c) Skip Thought