西安交通大學
XI'AN JIAOTONG UNIVERSITY

Natural language processing
with deep learning

Language Model & Distributed Representation (3)

Chen Li

cli@xjtu.edu.cn

2024-10

Outlines

1. NNLM
2. CBOW
3. Skip-gram
4. Hierarchical softmax
   & Negative sampling
5. GloVe

GloVe

- GloVe: Global Vectors for Word Representation

1. Constructing co-ocurrence matrix from corpus

   with decreasing weighting
   (The further the two words are, the less weight they have in the total count)

$$decay = 1/d$$

2. Constructing approximations between word vector and co-occurrence matrix

$$w_i^T \tilde{w}_j + b_i + \tilde{b}_j = \log(X_{ij})$$

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36	0.96

Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

GloVe

- GloVe: Global Vectors for Word Representation

1. Constructing co-ocurrence matrix from corpus

   with decreasing weighting
   (The further the two words are, the less weight they have in the total count)

$$decay = 1/d$$

2. Constructing approximations between word vector and co-occurrence matrix

$$w_i^T \tilde{w}_j + b_i + \tilde{b}_j = \log(X_{ij})$$

3. Constructing Loss function and training:

Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

GloVe

- GloVe: Global Vectors for Word Representation

1. Constructing co-ocurrence matrix from corpus

   with decreasing weighting
   (The further the two words are, the less weight they have in the total count)

$$decay = 1/d$$

2. Constructing approximations between word vector and co-occurrence matrix

$$w_i^T \tilde{w}_j + b_i + \tilde{b}_j = \log(X_{ij})$$

3. Constructing Loss function and training:

$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log(X_{ij}))^2$$

Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

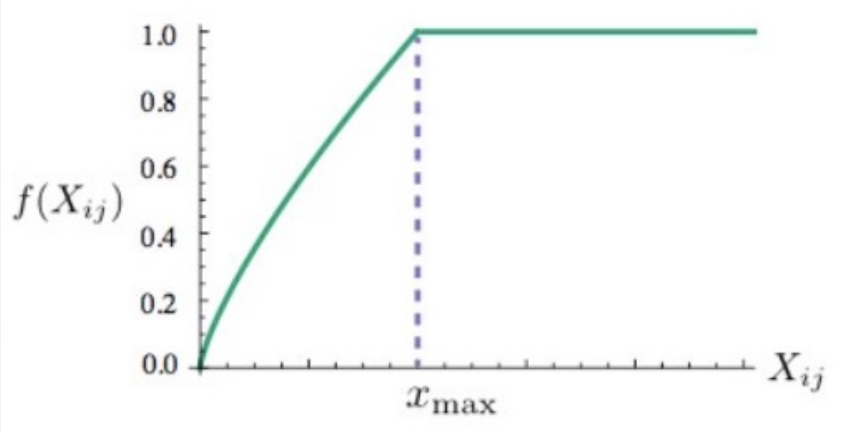
GloVe

- GloVe: Global Vectors for Word Representation

  3. Constructing Loss function and training:

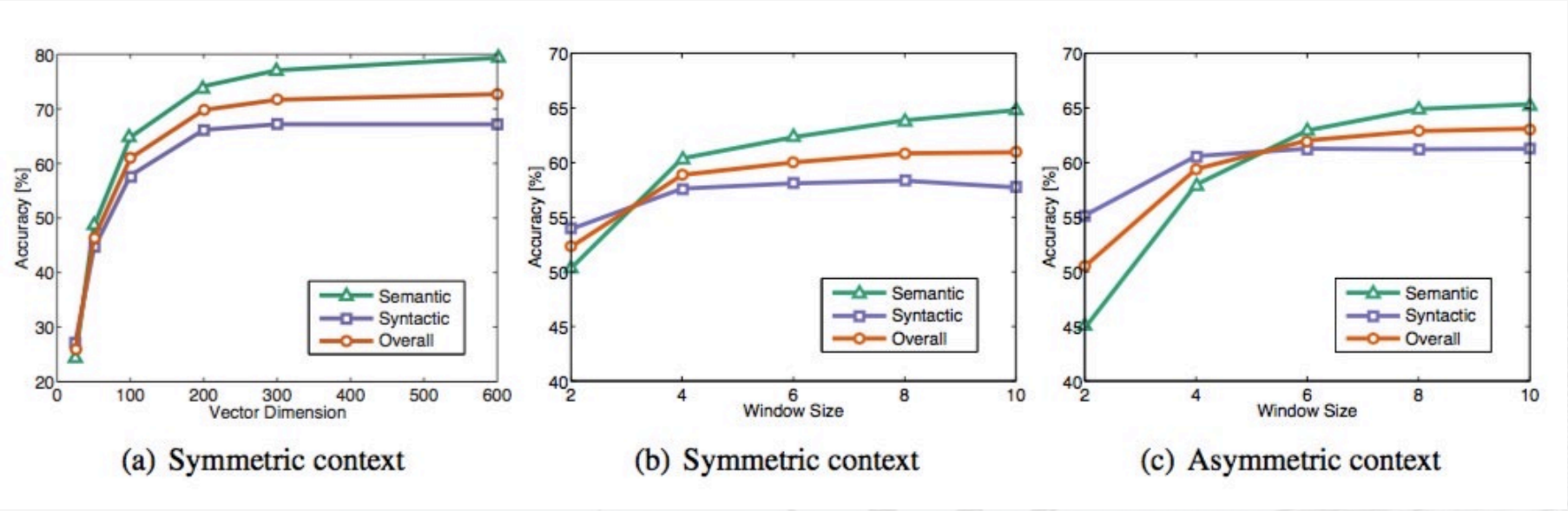
$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b_j} - \log(X_{ij}))^2$$

where $f(X_{ij}) = \begin{cases} (x / x_{\max})^{\alpha} & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases}$



Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

GloVe



- 共采用了三个指标：语义准确度，语法准确度以及总体准确度；
- Vector Dimension在300时能达到最佳；
- 而context Windows size大致在6到10之间。

Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

西安交通大学
XI'AN JIAOTONG UNIVERSITY

Q & A