A Note for those who are confused by the proof in our paper "Word Embedding Revisited: A New Representation Learning and Explicit Matrix Factorization Perspective"

Yitan Li
University of Science and Technology of China

November 21, 2015

Someone emailed me and told me that they are confused by a part of the **proof of theorem 1**[1]. I wrote this document to solve their confusion.

1 About the equation (7) in our paper

Some one doubt the correctness of equation (7) in our paper. Here I show correctness of (7) in a detailed way.

$$
\begin{aligned}
&\mathcal{L}_S(\mathbf{d}_w, \mathbf{C}^T\mathbf{w}) \\
&= -\log \frac{e^{\mathbf{d}_w^T \mathbf{C}^T \mathbf{w}}}{\sum_{\mathbf{d}'_w \in \mathcal{S}_w} e^{\mathbf{d}'_w{}^T \mathbf{C}^T \mathbf{w}}} \\
&\triangleq -\log P(\mathbf{d}_w | \mathbf{C}^T\mathbf{w}) \\
&= -\log \frac{e^{\sum_{c\in V_C} d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{\mathbf{d}'_w \in \mathcal{S}_w} e^{\sum_{c\in V_C} d'_{w,c} \mathbf{C}_c^T \mathbf{w}}} \\
&= -\log \frac{e^{\sum_{c\in V_C} d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{d'_{w,1}\in\mathcal{S}_{w,1}}, \dots, \sum_{d'_{w,|V_C|}\in\mathcal{S}_{w,|V_C|}} e^{\sum_{c\in V_C} d'_{w,c} \mathbf{C}_c^T \mathbf{w}}} \\
&= -\log \frac{\prod_{c\in V_C} e^{d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{d'_{w,1}\in\mathcal{S}_{w,1}}, \dots, \sum_{d'_{w,|V_C|}\in\mathcal{S}_{w,|V_C|}} \prod_{c\in V_C} e^{d'_{w,c} \mathbf{C}_c^T \mathbf{w}}} \\
&= -\log \frac{\prod_{c\in V_C} e^{d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{d'_{w,1}\in\mathcal{S}_{w,1}}, \dots, \sum_{d'_{w,|V_C|-1}\in\mathcal{S}_{w,|V_C|-1}}, \sum_{d'_{w,|V_C|}\in\mathcal{S}_{w,|V_C|}} \prod_{c\in V_C} e^{d'_{w,c} \mathbf{C}_c^T \mathbf{w}}} \\
&= -\log \frac{\prod_{c\in V_C} e^{d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{d'_{w,1}\in\mathcal{S}_{w,1}}, \dots, \sum_{d'_{w,|V_C|-1}\in\mathcal{S}_{w,|V_C|-1}}, \sum_{d'_{w,|V_C|}\in\mathcal{S}_{w,|V_C|}} e^{d'_{w,|V_C|} \mathbf{C}_{|V_C|}^T \mathbf{w}} \prod_{c\in V_C/\{|V_C|\}} e^{d'_{w,c} \mathbf{C}_c^T \mathbf{w}}} \\
&= -\log \frac{\prod_{c\in V_C} e^{d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{d'_{w,1}\in\mathcal{S}_{w,1}}, \dots, \sum_{d'_{w,|V_C|-1}\in\mathcal{S}_{w,|V_C|-1}} \prod_{c\in V_C/\{|V_C|\}} e^{d'_{w,c} \mathbf{C}_c^T \mathbf{w}} \left[\sum_{d'_{w,|V_C|}\in\mathcal{S}_{w,|V_C|}} e^{d'_{w,|V_C|} \mathbf{C}_{|V_C|}^T \mathbf{w}}\right]} \\
&= -\log \frac{\prod_{c\in V_C} e^{d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{d'_{w,1}\in\mathcal{S}_{w,1}}, \dots, \prod_{c\in V_C/\{|V_C|,|V_C|-1\}} e^{d'_{w,c} \mathbf{C}_c^T \mathbf{w}} \left[\sum_{d'_{w,|V_C|-1}\in\mathcal{S}_{w,|V_C|-1}} e^{d'_{w,|V_C|-1} \mathbf{C}_{|V_C|-1}^T \mathbf{w}}\right] \left[\sum_{d'_{w,|V_C|}\in\mathcal{S}_{w,|V_C|}} e^{d'_{w,|V_C|} \mathbf{C}_{|V_C|}^T \mathbf{w}}\right]} \\
&\quad \text{Through repeating the procedure as line 8 to 10, we can finally get the following} \\
&= -\log \frac{\prod_{c\in V_C} e^{d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\prod_{c\in V_C} \sum_{d'_{w,c}\in\mathcal{S}_{w,c}} e^{d'_{w,c} \mathbf{C}_c^T \mathbf{w}}} \\
&= -\log \prod_{c\in V_C} \frac{e^{d_{w,c} \mathbf{C}_c^T \mathbf{w}}}{\sum_{d'_{w,c}\in\mathcal{S}} e^{d'_{w,c} \mathbf{C}_c^T \mathbf{w}}} \\
&= -\sum_{c\in V_C} \log P(d_{w,c} | \mathbf{C}_c^T \mathbf{w})
\end{aligned} \tag{1}
$$

2 About the equation (9) in our paper

The equation (9) in "Word Embedding Revisited: A New Representation Learning and Explicit Matrix Factorization Perspective" is based on the following factorization:

$$\sum_{d'_{w,c}\in\mathcal{S}_{w,c}} e^{d'_{w,c}\mathbf{C}_c^T\mathbf{w}} = \prod_{q=1}^{Q_{w,c}} \sum_{d'_{w,c,q}\in\{0,1\}} e^{d'_{w,c,q}\mathbf{C}_c^T\mathbf{w}} \tag{2}$$

Someone thought that the left side of the equation is not equal to the right hand size. This is because that they were misleaded by their intuition, and they thought that

$$\sum_{d'_{w,c}\in\mathcal{S}_{w,c}} e^{d'_{w,c}\mathbf{C}_c^T\mathbf{w}} = \sum_{i=0}^{Q_{w,c}} e^{i\mathbf{C}_c^T\mathbf{w}} \tag{3}$$

Then they derive that the left hand size of the (2) is not equal to the right hand side.

Actually, what they thought is not correct, because

$$\sum_{d'_{w,c}\in\mathcal{S}_{w,c}} \neq \sum_{i=0}^{Q_{w,c}} \tag{4}$$

I have mentioned in the paper that "$\sum_{d'_{w,c}\in\mathcal{S}_{w,c}}$ is a summation in Hamming space that will be defined below in the proof of Theorem 1.". So $\sum_{d'_{w,c}\in\mathcal{S}_{w,c}}$ is not a normal summation and it is defined in the first equation in the **proof of theorem 1**:

$$\sum_{d'_{w,c}\in\mathcal{S}_{w,c}} = \sum_{d'_{w,c,1},d'_{w,c,2}\ldots,d'_{w,c,Q_{w,c}}\in\{0,1\}^{Q_{w,c}}} \tag{5}$$

Then,

$$\begin{aligned}
&\sum_{d'_{w,c}\in\mathcal{S}_{w,c}} e^{d'_{w,c}\mathbf{C}_c^T\mathbf{w}} \\
&= \sum_{d'_{w,c,1},d'_{w,c,2}\ldots,d'_{w,c,Q_{w,c}}\in\{0,1\}^{Q_{w,c}}} e^{\sum_{q=1}^{Q_{w,c}} d'_{w,c,q}\mathbf{C}_c^T\mathbf{w}} \\
&= \sum_{d'_{w,c,1},d'_{w,c,2}\ldots,d'_{w,c,Q_{w,c}}\in\{0,1\}^{Q_{w,c}}} \prod_{q=1}^{Q_{w,c}} e^{d'_{w,c,q}\mathbf{C}_c^T\mathbf{w}} \\
&= \prod_{q=1}^{Q_{w,c}} \sum_{d'_{w,c,q}\in\{0,1\}} e^{d'_{w,c,q}\mathbf{C}_c^T\mathbf{w}}
\end{aligned}$$

Finally, we have

$$\sum_{d'_{w,c}\in\mathcal{S}_{w,c}} e^{d'_{w,c}\mathbf{C}_c^T\mathbf{w}} = \prod_{q=1}^{Q_{w,c}} \sum_{d'_{w,c,q}\in\{0,1\}} e^{d'_{w,c,q}\mathbf{C}_c^T\mathbf{w}} \tag{6}$$

References

[1] Yitan Li, Linli Xu, Fei Tian, Liang Jiang, Xiaowei Zhong, and Enhong Chen. Word embedding revisited: A new representation learning and explicit matrix factorization perspective. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 3650–3656, 2015.