### GloVe(Global Vector)

目录

[1. 官方 2](#_Toc27747)

[2. 背景 2](#_Toc3981)

[3. 共现矩阵 2](#_Toc18924)

[4. 模型 3](#_Toc31403)

[5. 模型公式 4](#_Toc9868)

[6. 模型推导 4](#_Toc3969)

[7. 结论 7](#_Toc23047)

# **官方**

文档：<https://nlp.stanford.edu/projects/glove/>

代码：<https://github.com/stanfordnlp/GloVe>

# **背景**

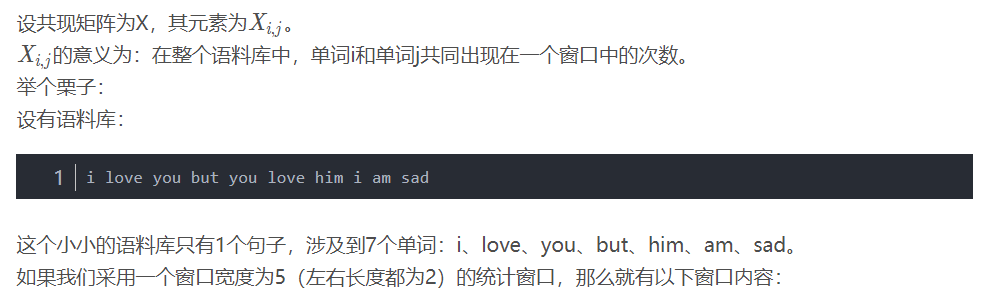
Cbow 或者Skip-Gram取得了很大的成功，不管在训练效率还是在词向量的计算效果上面。

但是考虑到Cbow/Skip-Gram 是一个local context window的方法，比如使用NS来训练，缺乏了整体的词和词的关系，负样本采用sample的方式会缺失词的关系信息。

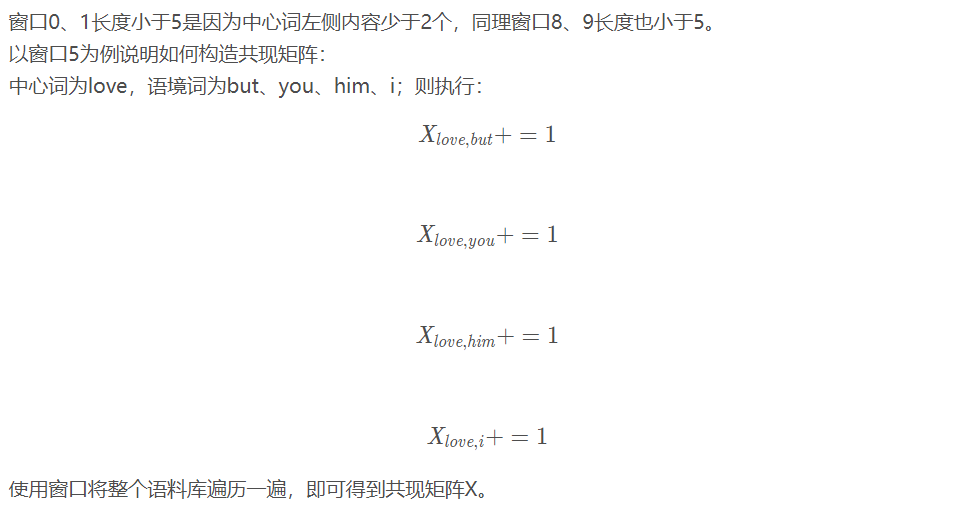
另外，直接训练Skip-Gram类型的算法，很容易使得高曝光词汇得到过多的权重。

Global Vector融合了矩阵分解Latent Semantic Analysis (LSA)的全局统计信息和local context window优势。融入全局的先验统计信息，可以加快模型的训练速度，又可以控制词的相对权重。（skip-gram、CBOW每次都是用一个窗口中的信息更新出词向量，但是Glove则是用了全局的信息（共线矩阵），也就是多个窗口进行更新）

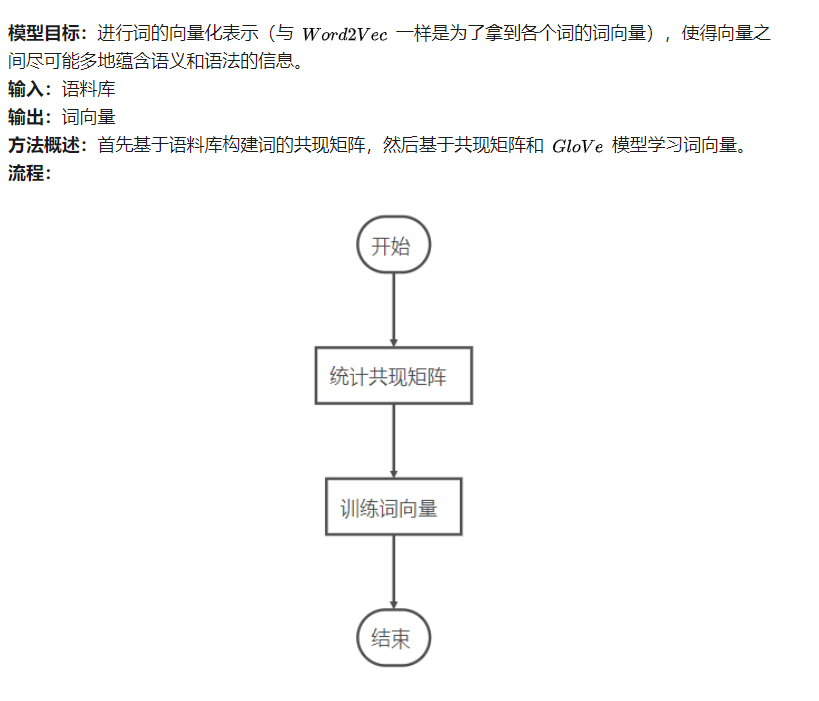
# **共现矩阵**



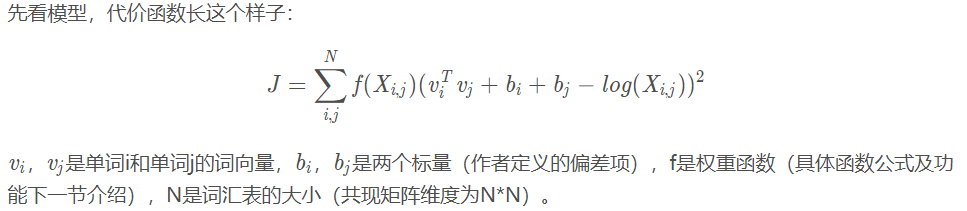




# **模型**

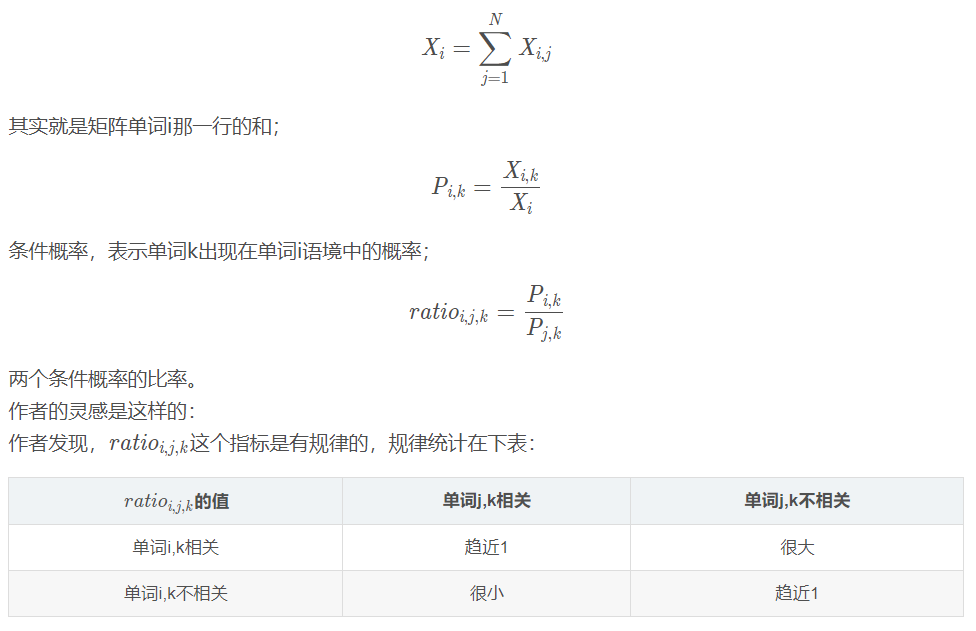


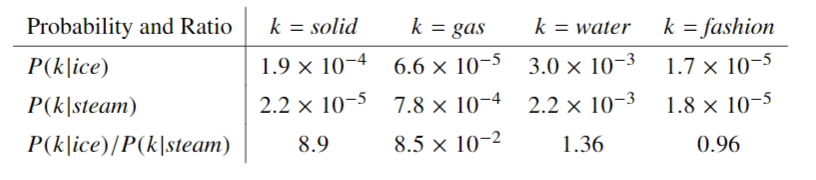
# **模型公式**

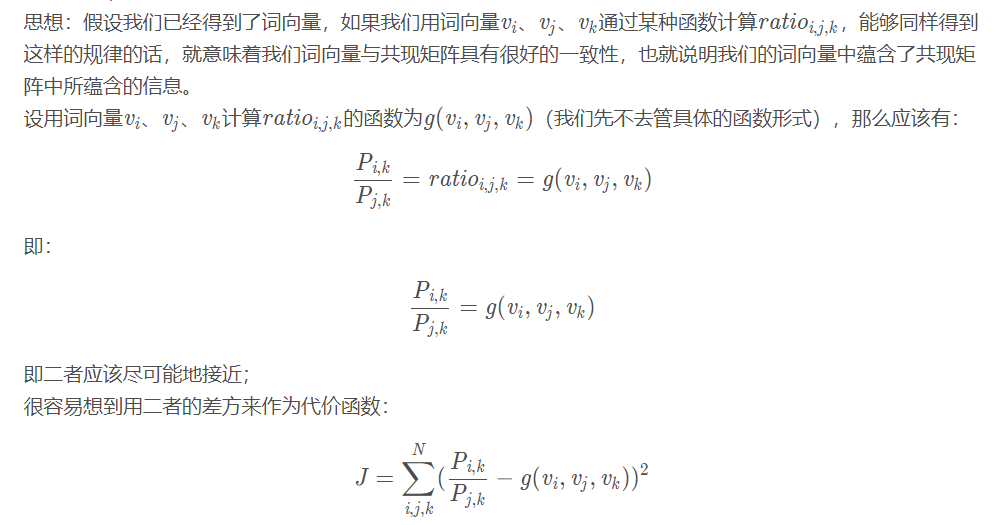


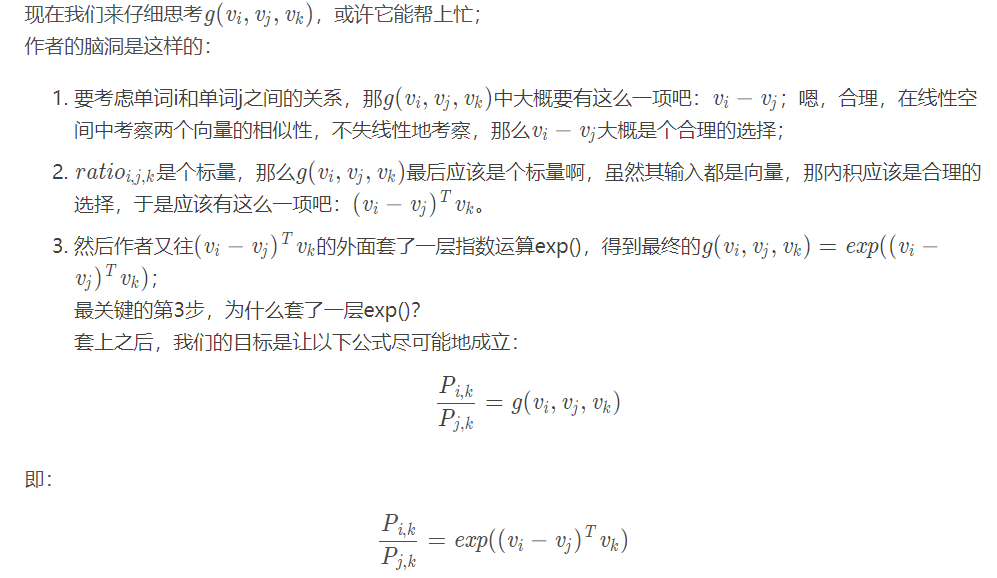
GloVe模型没有使用神经网络。

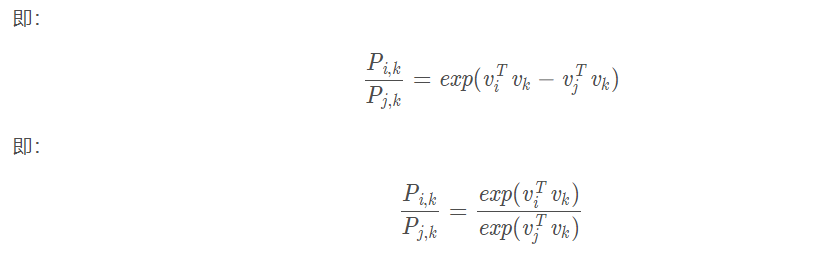
# **模型推导**

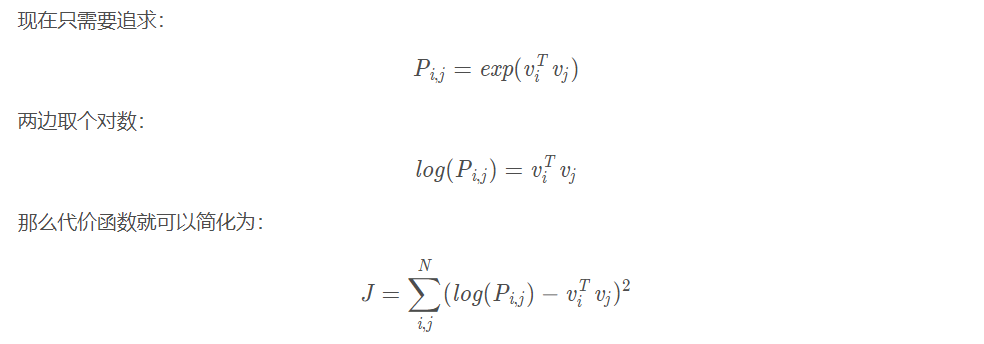




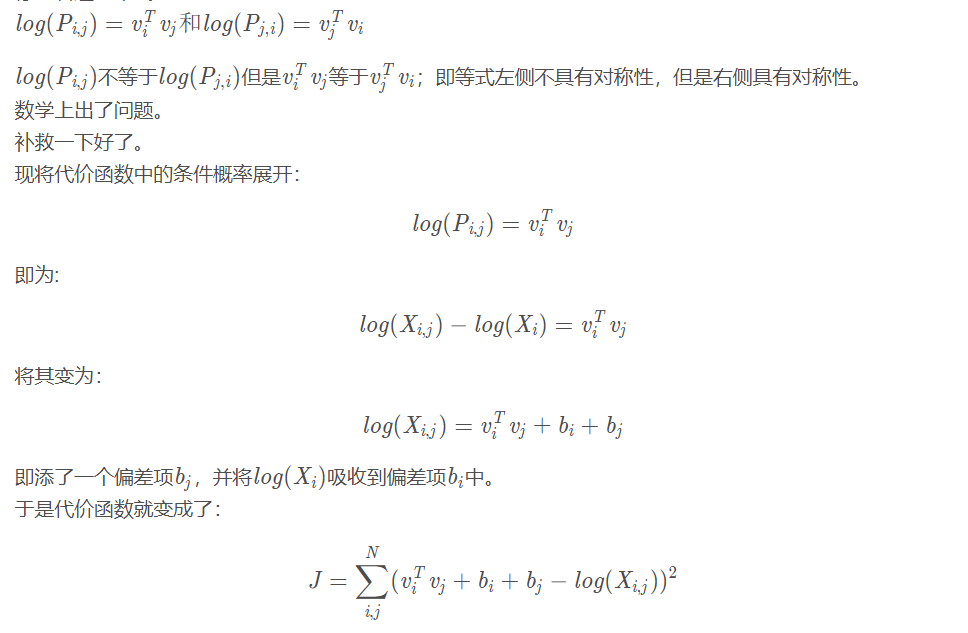


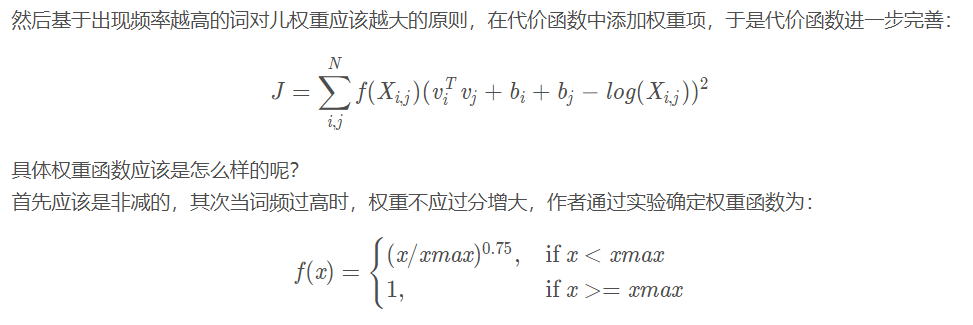




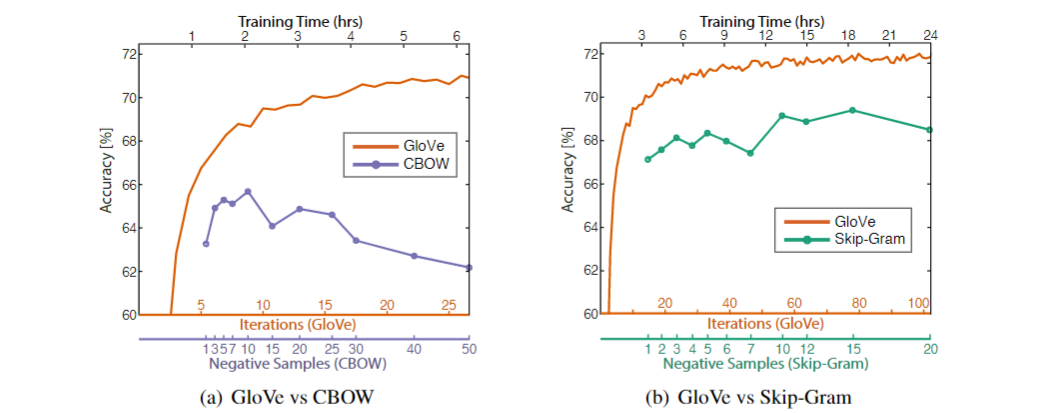


因为在共现矩阵中，中心词的角色和上下文词是可以随时互换的，所以式子需要具有对称性





# **结论**



类比任务的整体准确性与训练时间的关系，取决于GloVe的迭代次数以及CBOW（a）和skip-gram（b）的否定样本数量。 在所有情况下，我们都在具有相同400,000词表的相同6B令牌语料库（Wikipedia 2014 + Gigaword 5）上训练300维向量，并使用大小为10的对称上下文窗口。

结果GloVe是一种用于单词表示的无监督学习的新的全局对数-双线性回归模型，在词类比，词相似性和命名实体识别任务方面优于其他模型。