词向量（比较经典的模型）(word2vec)：

CBOW、Skip-Gram、Glove

Glove（既考虑了skip-gram的local信息，也考虑了matrix factorization的global信息）

删除出现次数比较少的单词：

阈值的选择可以通过可视化的方法

倒排表：

先去掉没有任何交集的问题。速度会大大提升。

词向量--》句子向量：

将句子中所有的单词向量做平均。