

## 基于SVD降维的表示方法

- 词-文档矩阵

我们建立一个词组文档矩阵 $X$ ，具体是这么做的：遍历海量的文件，每次词组 $i$ 出现在文件 $j$ 中时，将 $X_{ij}$ 的值加1。不过大家可想而知，这会是个很大的矩阵 $R^{V \times M}$ ，而且矩阵大小还和文档个数 $M$ 有关系。

- 基于窗口的共现矩阵 $X$

把矩阵 $X$ 记录的词频变成一个相关性矩阵。我们先规定一个固定大小的窗口，然后统计每个词出现在窗口中次数，这个计数是针对整个语料集做的。

## GloVe结合了SVD与Word2Vec的优势

查阅相关资料学习了损失函数的由来，后面的评估只是了解了一下，这部分掌握的不太好。