

一 什么是 GloVe

GloVe 的全称是 Global Vectors for Word Representation，它是一个基于全局词频统计的词表征工具

二 GloVe 论文: <https://www.aclweb.org/anthology/D14-1162>

三 GloVe 实现

步骤 1: 构造共现矩阵(co-occurrence Matric)，根据语料库构建，矩阵中的每一个元素 X_{ij} 代表单词 i 和上下文单词 j 在特定大小的上下文窗口内共同出现的次数。

Co-occurrence: 协同出现指的是两个单词 w_1 和 w_2 在一个 Context Window 范围内共同出现的次数

Context Window: 指的是某个单词 w 的上下文范围的大小，也就是前后多少个单词以内才算是上下文

一般而言:

对于 GloVe，它根据两个单词在上下文窗口的距离 d ;

提出了一个衰减函数(decreasing weighting): $\text{decay}=1/d$ 用于计算权重，也就是说距离越远的两个单词占总计数(total count)的权重越小。

例: He is not lazy. He is intelligent. He is smart. Context Window=2

	He	is	not	lazy	intelligent	smart
He	0	4	2	1	2	1
is	4	0	1	2	2	1
not	2	1	0	1	0	0
lazy	1	2	1	0	0	0
intelligent	2	2	0	0	0	0
smart	1	1	0	0	0	0

步骤 2: 构建词向量和共现矩阵(co-occurrence Matric)之间的近似关系

公式 1: $w_i^T \bar{w}_j + b_i + \bar{b}_j = \log(X_{ij})$, w_i^T 是最终要求解的词向量

构造损失函数

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

对于公式 2，在一个语料库中，肯定存在很多单词，它们在一起出现的次数很多，那么:

1. 这些单词的权重要大于那些很少在一起出现的单词
2. 但我们也不希望这个权重过大，当达到一定程度之后应该不再增加
3. 如果两个单词没有在一起出现，也就是 $X_{ij}=0$ ，那么它们应该不参与损失函数的计算中，也就是 $f(X_{ij})=0$

步骤 3: 训练

GloVe 虽然是无监督的学习方式（不需要标记 label），但 GloVe 是有 label 的，label 就是 $\log(X_{ij})$ ，所以

GloVe 的训练和有监督的训练方法类似。

论文中，GloVe 采用了 AdaGrad 的梯度下降算法，对矩阵 X 中的所有非零元素进行随机采样，学习率设为 0.05，在 vec_size 小于 300 的情况下迭代了 50 次，其他大小的迭代了 100 次，直到收敛。

最终学到的是两个 vec 是 w 和 \tilde{w} ，因为 X 是对称的，所以从原理上讲 w 和 \tilde{w} 也是对称的，唯一区别就是初

始值不一样而导致最终的值也不一样。并选取 $w + \tilde{w}$ 作为最终的 vec。

四 GloVe 实现

mxnet 安装（目前支持 GloVe 最好的库）：`pip install mxnet`

GloVe 应用

[GloVe.html](#)