

Week2

本笔记聚焦在 Glove。

最早的词的表示是用词袋 (Bag Of Words) 方法，然此方法未考虑词顺序。之后引入了词向量的方法：Word2vec 和 Glove，可把词映射到对应向量中，用多维向量表示一个词。

Word2vec 是使用临近关系的词去构建浅层神经网络，然后把神经网络中间层的特定词的输出变成对应的词向量。主要有 CBOW 和 skip-gram 两种方法。但此方法只关注全文的局部信息，没有考虑全文语言特性，如词在全局出现次数等表述信息。

1.Glove (Global Vectors for Word Representation)

GLOVE 是一种新的词矩阵生成方法，使用了全局统计信息和局部统计信息生成词向量，在模型的嵌入层使用。此方法在运用时，可用已融入的全局先验统计信息，加快模型的训练速度和控制词的相对权重。

1.1 Glove 模型概览

首先基于语料库构建词的共现矩阵，然后基于共现矩阵和 Glove 模型学习词向量。通过构建语料库所有词对的共现矩阵（见下图），我们知道若词对 i, k (ice, solid) 的共现概率很大（相关），而词对 j, k (steam, solid) 的共现概率很小（不相关），则它们相除的值（比值）很大（相比 steam 来讲，在 ice 的上下文中，更容易想到 solid），反之很小。若两对词对均不相关或相关，那么它们比值接近 1。

因此，我们可通过词对的共现概率，去描述词之间相关性。进而通过词向量，描述词的相关性。而它们的比值可捕捉我们想要的某些特征。

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k ice)/P(k steam)$	8.9	8.5×10^{-2}	1.36	0.96

$P(k ice)/p(k steam)$ 的值	单词 i, k 相关	单词 i, k 不相关
单词 i, k 相关	接近 1	很大
单词 i, k 不相关	很小	接近 1

通过算法设计推导，推导出词对向量 (w_i 和 w_k) 和词对的共现词频 (X_{ik}) 关系，如下：

$$w_i^T \tilde{w}_k + \tilde{b}_i + \tilde{b}_k = \log(X_{ik})$$

可统计

然后运用随机梯度下降方法，最小化代价函数 J ，求出词向量。

代价函数：

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

词频重要性
最小化，随机梯度下降

若需运用通过 Glove 方法获得词向量，可通过预训练模型获取。网址如下：

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