Glove 公式推导

为了捕捉作者提出的条件概率比例，作者提出

$$F(w_i, w_j, w_k) = \frac{P_{ik}}{P_{jk}}$$

，其中F函数未知

作者开始发挥想象力：

① $\frac{P_{ik}}{P_{jk}}$ 表示词对之间的关系，向量空间是线性，最直接就是用两个向量的差值来表示相似性，即有 $F(w_i - w_j, w_k) = \frac{P_{ik}}{P_{jk}}$

② 右式为标量，所以对左式做点乘：$F((w_i - w_j)^T w_k) = \frac{P_{ik}}{P_{jk}}$

③ $F(w_i^T w_k - w_j^T w_k) = \frac{P_{ik}}{P_{jk}}$，左式为差，右式为商，于是想到 $F = exp(\ )$，可以把两者结合起来：$exp(w_i^T w_k - w_j^T w_k) = \frac{exp(w_i^T w_k)}{exp(w_j^T w_k)} = \frac{P_{ik}}{P_{jk}}$

④ 于是让 $P_{ik} = exp(w_i^T w_k)$，$P_{jk} = exp(w_j^T w_k)$

⑤ $exp(w_i^T w_k) = P_{ik} = \frac{X_{ik}}{X_i}$ （X_{ik}：w_i, w_k共现次数；$X_i = \sum_{k=1}^{|V|} X_{ik}$）

$$w_i^T w_k = \log\left(\frac{X_{ik}}{X_i}\right) = \log X_{ik} - \log X_i$$

⑥ $\log X_{ik} = w_i^T w_k + \log X_i$，交换i,k顺序，因为$\log X_i$这一项导致模型不对称，于是引入$b_i$和$b_k$偏置项(bias)

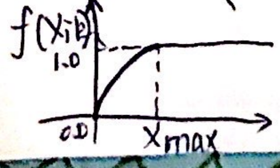
$$\log X_{ik} = w_i^T w_k + b_i + b_k$$

⑦ 用最小二乘法计算loss $J = \sum_{ik} (w_i^T w_k + b_i + b_k - \log X_{ik})^2$

⑧ 根据⑦，每个共现关系对J的贡献是相等的，作者引入$f(X_{ik})$作为权重

$$J = \sum_{ik} f(X_{ik}) (w_i^T w_k + b_i + b_k - \log X_{ik})^2$$

最后确定 $f(x_{ik}) = \begin{cases} (x_{ik}/x_{max})^{\alpha}, & if\ x_{ik} < x_{max} \\ 1, & otherwise \end{cases}$ （$\alpha = 3/4$，$x_{max} = 100$，基于经验）



（共现次数越多，权重越高/或不变，表示其越重要）