数据预处理

1. 提取合适的公司（招聘数量top1000）；
2. 对每个Post分词、去stop-words，得到每个post的向量表示，并累计公司所有post，获取其向量表示；
3. 对公司按招聘聚类，获取公司的近邻，获取近邻影响力向量表示（均值）；
4. 按照公司领域，获取每个公司的领域向量表示；

模型训练：

1. 初始化；
2. Collapsed Gibbs Sampling迭代；
   1. 采样z；
   2. 采样x；
3. 参数估计；
4. 模型输出；