一、LDA模型的概念

LDA模型实际上就是通过整个词库的词频数据进行构造各个主题的分布（即密度曲线）。在对每个文档构造词频密度曲线，并对各个主题进行一一比对，由此进行文本的聚类。

二、阶段成果

1、可以通过LDA模型进行文本降维，并获取各个主题的关键字和词频矩阵。

2、可以对新的句子进行预测。

三、存在问题

1、由于现阶段拟合分布的方法为MCMC（马尔科夫链蒙特卡洛模拟），因此每次训练结果存在一定的随机性。聚类结果不稳定（即存在对于单一文本，每次训练完后所对应的主题是不同的）

2、由于需要大量词语进行构造密度曲线，所以现阶段语料库不足。可能存在要反复训练的现象，且每次训练耗时为8-10s。