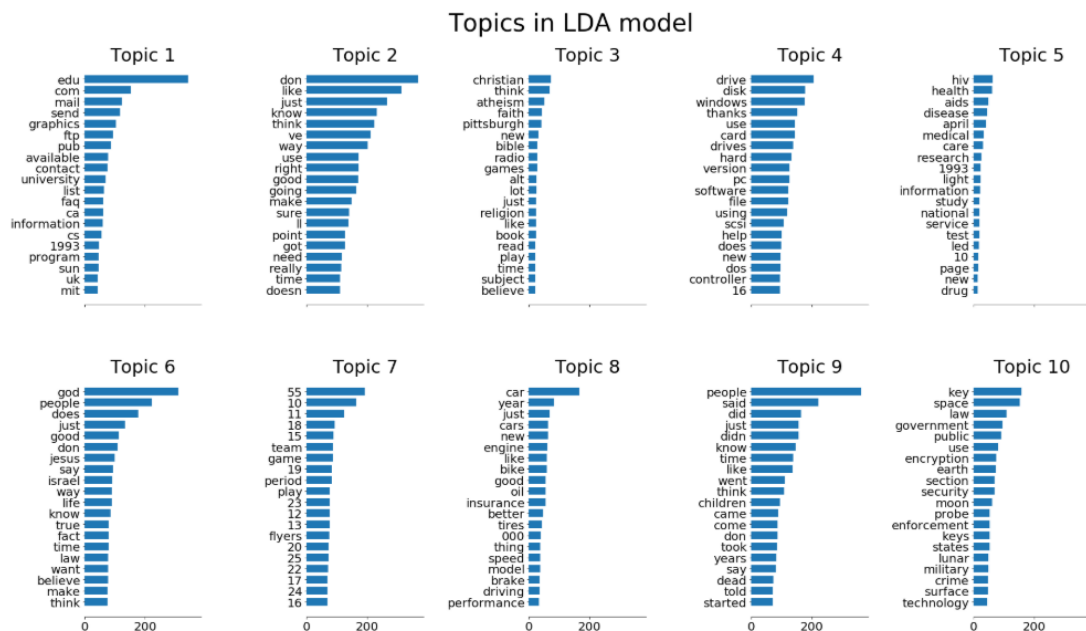


## 作业要求

(1) 从 Scikit-learn 库中导入新闻文本数据集(fetch\_20newsgroups), 利用 Scikit-learn 库中的 LDA 模型进行话题分析, 并绘制出各个主题下的高频词(类似下图所示)。



(2) 从交大新闻网主页新闻栏目(<http://news.xjtu.edu.cn/zyxw.htm>)下载最新的100 条新闻报道, 利用 Scikit-learn 库中的 LDA 模型进行话题分析, 并绘制出各个主题下的高频词, 以及各个主题的词云(类似下图所示)。

