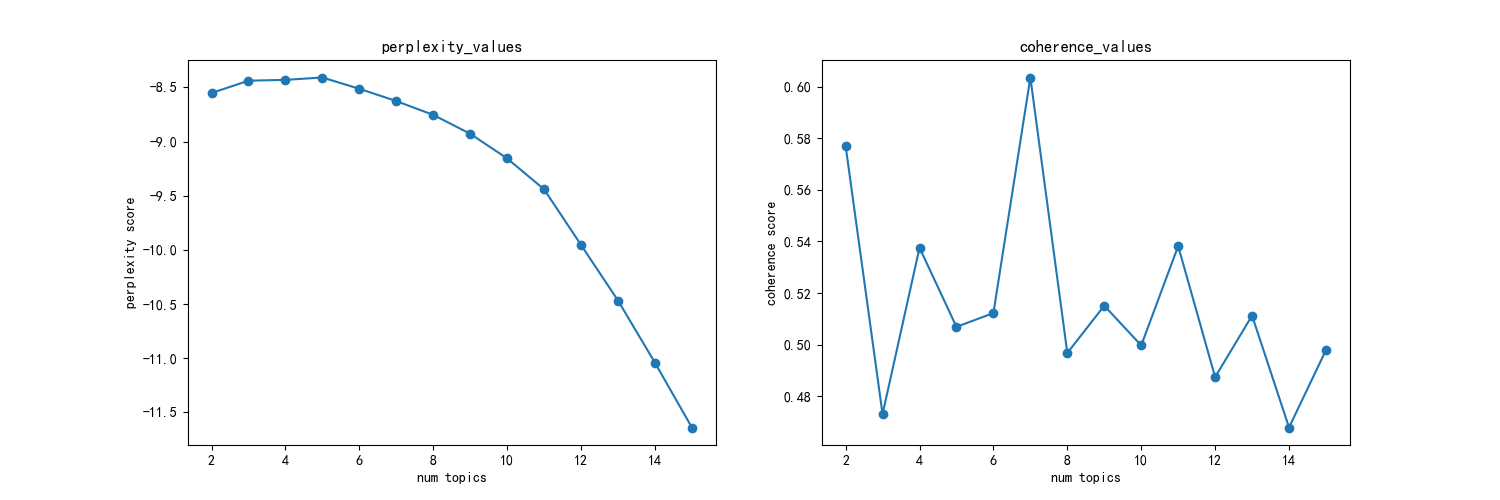
# 事件维度说明文档

首先根据前面以及处理好的文档，根据处理好的分词内容，这边去做LDA建模

把分词去构建成字典的格式，为后续建模做准备

接着通过构建好的字典去计算它的困惑度和一致性，通过困惑度和一致性来找到最优的主题数

困惑度越接近0越好，一致性越大越好，根据如下，这边的一致性比较突出，故此选择主题数为7



接着根据确定好的主题数，进行主题建模

得到lda.html、主题词分布表.csv、特征词.csv文件，可以知道每个主题，对应的词汇内容是什么，他们对应的权重是什么，特征词是什么，方便我们对该主题进行自定义

接着根据lda\_data.csv文件，去对每个主题进行词云可视化，来查看每个主题，他们的用户画像是怎么样的，让我们对主题有个更加直观的认识，从而对事情整体有个大概的分类情况

接着去计算相关的指标，根据数据可视化.py文件去进行可视化分析

根据转发，评论，点赞，阅读数，对这四个指标进行加权处理，采用StandardScaler（），进行标准化，处理后使得每个指标的数值都代表同一个指标，得到得出事件综合热度的大小。

接着再去根据时间按照月进行求总，从而得出每个月热度变化的趋势图

再接着，根据前面采用snownlp去计算出来的情感得分，我们根据时间维度，按照月去计算每个月，情感得分的均值，用折线图的方式来进行展示。

最后根据 信源地域列和sentiment score列，去查看每个地区情感得分的均值，接着把属于中国的地理坐标保留下来，其他的去掉，采用pyecharts进行地图可视化

采用pyecharts里面的Geo()函数，来展示中国地图，把数据按照3个阶段进行划分，65-75 75-85 85-100 这三个维度，这个数值是情感得分，越接近1说明情感越积极，越接近0说明情感越消极，当然这只是机器进行打分，每个人对于情感理解不同，打分也不同，更何况机器打分 所以仅作为参考而已