* 依据评分，将评论分为积极评论和负面评论进行分类分析，4-5分为积极评论，1-2分为负面评论。
* 对评论进行分词，并通过词性标注筛选。
* 对所有的评论进行转换
  + 所有句子中的词都映射为（id, frequency），frequency表示词在句子中出现的频率
  + 进而所有评论变成了一个语料库，作为LDA模型的corpus的输入
* 人工选择主题的数量
* 最终输出对应的主题的评论
  + 词与id的映射关系作为输出的转换
  + 每个主题里按照主题的相关性排序，越靠前的表示该主题越突出