山东科技大学毕业设计（论文）开题审核表

|  |  |  |  |
| --- | --- | --- | --- |
| 学生姓名 | 于跃洲 | 学 号 | 201801060228 |
| 专 业 | 计算机科学与技术 | 开题日期 | 2022年3月27日 |
| 指导教师 | 赵华 | 职 称 | 副教授 |
| 论文（设计）题目 | 面向特定商品的用户购买原因和用途识别 | | |
| 指导教师评语：  签字：  年 月 日 | | | |
| 开题答辩记录：  1.问：如何快速获取所需的数据？  答：利用爬虫技术，模拟浏览器发送网络请求并且接收请求响应，按照事先设定好的规则自动地到相关网站获取本课题所需的相关数据。  2.问：为何要对获得的数据进行预处理？  答：获取到的文本评论数据中存在大量价值含量很低甚至甚至没有价值含量的条目，如果将这些评论数据也引入进行分词、词频统计乃至情感分析等，必然会对分析造成很大的影响，得到的结果的质量也必然是存在问题的。因此，在正式使用到这些文本评论数据之前就必须先进行文本预处理、把大量的此类无价值含量的评论去除。  3.问：为何要选择LDA模型进行主题分析？  答：LDA主题模型是一种无监督学习模型，只需要提供训练文档，它就可以自动训练出各种主题概率，无需任何人工标注过程，节省大量人力及时间。相对于其他主题模型，其引入了狄利克雷先验知识。因此，模型的泛化能力较强，不易出现过拟合现象。  （审核内容：1.选题是否合适；2.方案论证是否可行；3.工作量是否适宜；4.文献是否充足；5.学生对课题内容和要求的理解是否深入；6.进度是否得当；7.基础条件是否具备等。） | | | |
| 是否通过开题： 通过 不通过 | | | |