山 东 科 技 大 学

本科毕业设计（论文）中期报告

题 目 基于特定商品的用户购买原因和用途识别

学院（系、部） 计算机科学与工程学院

专业班级 计算机科学与技术2018级2班

姓 名 于跃洲

学 号 201801060228

指 导 教 师 赵 华

填表时间： 2022 年 4 月 29 日

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 姓名 | 于跃洲 | | 学号 | 201801060228 | 指导教师 | 赵华 |
| 班级 | 2018级2班 | | 专业 | 计算机科学与技术 | 课题研究  有无变化 | 无 |
| 毕业设计（论文）题目 | | 基于特定商品的用户购买愿意和用途识别 | | | | |
| 1. 目前进展及成果   (1)完成了对特定商品评论的爬取、清洗。  (2)完成了对评论文本数据的分词、停用词过滤等操作。  (3)对预处理后的文本评论数据进行了初步的情感分析，将其按照情感倾向分为了正面评论数据（好评）和负面评论数据（差评）。  (4)分别对正、负面评价数据进行了LDA主题分析。  2.遇到的问题及解决措施  (1)利用Python对特定商品的评论数据进行爬取时触发了电商平台的反爬措施，导致爬取效率大幅下降。后通过更换新的爬取方法（Selenium）得以解决。  (2)对评论数据进行中文分词时，由于容易存在切分歧义和难以识别未登录词等问题，如何选择合适的分词方法成为新的难题。经查阅相关资料，决定使用Python的Jieba分词包进行中文分词操作。  (3)对评论文本进行主题提取分析时，由于先前未曾接触过相关技术，因此选用何种主题方法一时变为了难题。后通过积极查阅相关的资料，最终确定了使用潜在狄利克雷分配，即LDA主题模型。  3.下一步工作的计划  (1)对经过情感分析的文本评论数据的情感倾向进行进一步的修正，使其更加精确。  (2)继续寻找基于LDA模型的最优主题数，使最终主题分析结果更具有真实性和准确性。  (3)进一步分析结果，达到对评论数据进行用户购买原因和用途识别的最终目的。  (4)检查并改进整个毕设过程中的细节问题，并着手撰写毕业设计论文，为最终的答辩做准备。  学生签名：    年 月 日 | | | | | | |
| 指导教师意见：    指导教师签名：    年 月 日 | | | | | | |