山东科技大学

本科毕业设计（论文）任务书

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

职称： 副教授

拟定题目： 面向特定商品的用户购买原因和用途识别

（任务起止日期： 2022 年 3 月 14 日至 2022 年 6 月 12 日）

|  |  |
| --- | --- |
| 设计（论文）的主要内容及任务要求：  本课题主要针对用户在电商平台上留下的评论数据，对其进行：   1. 压缩重复词、删除无用词等数据清洗。 2. [分词](https://so.csdn.net/so/search?q=%E5%88%86%E8%AF%8D&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/weixin_47922824/article/details/_blank)、词性标注和去除停用词等文本预处理。 3. 基于预处理后的数据进行情感分析。   4.使用LDA主题模型提取评论关键信息，了解用户的购买原因、用途，意见，以及产品的优缺点。 | |
| 设计（论文）进程安排：  第3-4周：针对电商平台的反爬机制，阅读相关文献，学习新的爬虫技术，获取更多文本评论数据以增加最终分析结果的可靠性。  第5-7周:学习Python-jieba库的相关知识，并用其对重复的非重要词汇进行压缩和适当删除，同时对经过初步数据清洗后获取到的文本数据合并同义词，初步统计词频。制作出词云图以便后续进一步的分析。  第8-10周：查阅相关文献，学习基于Python的对中文文本进行情感分析的常用方法（初步选定为利用snowNLP），了解当前研究现状。利用学习到的情感分析方法对前述的数据进行实际应用。  第11-13周：阅读国内外关于LDA主题模型的文献，学习LDA模型的相关实现思路和原理，学习其函数参数的意义及常用默认值。利用学习的此部分知识，对前述文本评论数据进行主题提取和分析,逐步完成毕业论文。  第14-15周：对整个过程进行完善，整理记录各过程中间结果，实现对特定商品的用户购买原因和用途的识别，完成毕业论文和答辩工作。 | |
| 主要参考文献：（指导教师指定）   1. Ian Sutherland,Youngseok Sim,Seul Ki Lee,Jaemun Byun,Kiattipoom Kiatkawsin. Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation[J]. Sustainability,2020,12(5). 2. JUNG Y J，SUH Y M． Mining the voice of employ- ees: A text mining approach to identifying and analyzing job satis- faction factors from online employee reviews［J］．Decision Sup- port Systems，2019，123: 113074． 3. SHARAN S，SUCHITHRA R ． Topic－based knowl- edge mining of online student reviews for strategic planning in u- niversities［J］． Computers ＆amp; Industrial Engineering，2019， 128: 974－984． [4]陈欢,黄勃,朱翌民,俞雷,余宇新.结合LDA与Self-Attention的短文本情感分类方法[J].计算机工程与应用,2020,56(18):165-170.   [5]吴琳.在线商品评论信息质量影响因素研究[D].黑龙江大学,2021.DOI:10.27123/d.cnki.ghlju.2021.001380.  [6]徐谢宁.在线用户群体评论行为及其影响研究[D].电子科技大学,2021.DOI:10.27005/d.cnki.gdzku.2021.001721.  [7]张芷欣.电商产品文本评论数据的情感购买决策分析[D].湘潭大学,2019.DOI:10.27426/d.cnki.gxtdu.2019.000915.  [8]周成骥. 基于机器学习的商品购买行为预测模型设计[D].广州大学,2018.  [9]李永海.一种使用在线评论信息的商品购买决策分析方法[J].运筹与管理,2018,27(02):32-37.  [10]杨瑞欣.电商空调产品的评论数据情感分析[D].山西大学,2017.  [11]崔宁,赵宗良,吴瑞雪.基于LDA主题模型和偏序集的在线商品评论研究[J].情报探索,2021(12):71-76. | |
| 指导教师签字：  年 月 日 | 系（教研室）主任签字：  年 月 日 |