文献综述

经过阅读国内外的相关研究文献，获悉了用户的在线评论(Online Reviews)是消费者网上购物消费后生成的重要信息,蕴含着消费者们对商品的使用质量、感受、功能描述等多维度的感受，具有较强的主观意愿，以文本的形式直观的体现了用户购买并使用了某个产品之后对其的评价。在线评论可以帮助更多的消费者更好的选择符合他们个人使用需求的商品，同时还能够使商家更好的掌握消费者需求从而有方向的改进营销策略以提高销量。

目前，国外有许多研究人员已经利用了隐含狄利克雷分布，即LDA模型（Latent Dirichlet Allocation）从用户的在线评论文本中提取出用户对商品或服务在某些层面的影响因素。例如: 文献［1］利用 LDA 主题模型对旅社的用户在线评论进行在线留宿评论主题建模，并得出这些因素对旅社的星级评定具有重要影响。通过使用直接来自消费者本身的大型、非结构化文本，验证并扩展了当前关于酒店行业中影响客户的维度的理论。更准确地说，在文献中所看到的重要维度，如服务质量、位置、价值等，都被LDA方法所证实。并根据所提取的主题在不同住宿特征中的比例进行评估，发现竞争力和独特性是相对突出的主题;文献［2］针对相关平台官网上职工产生的在线评论文本，基于 LDA 主题模型提取出影响职工工作满意度的因子，如假期长度、管理制度、薪酬水平等，然后根据这些因素进行了一系列的分析，并测量了行业、公司、团体和时间层次上的每个工作满意度因素的情绪和重要性。同时还分析考察了每个星级工作满意度因素对整体工作满意度的相对重要性。最终计算各工作满意度因子与整体工作满意度之间的关联强度。该研究的结果将为企业管理者在各方面管理员工的工作满意度提供深刻的见解。文献［3］使用LDA 模型分析了某高等院校学生的在线评论文本，由此提取出了影响学生对学校印象的因素，如学校环境、教学能力等因素。

与此同时，国内也有学者进行了相关的研究。文献[4]针对在对短文本进行情感分类时,由于文本长度过短引起数据稀疏,进而使得分类任务准确率下降的现象提出了一种基于LDA主题模型与Self-Attention技术的短文本情感分类方法。利用这些评论文本实现了在高维向量空间中对同一主题的聚类,并且通过Self-Attention进行动态权重分配进一步进行了分类的完善。文献[5]把在线商品评论信息质量的影响因素分为传播者因素、信息因素、信息媒介因素、信息反馈因素四个方面，结合信息采纳模型及已确定的的影响因素,建立在线商品评论信息质量影响模型,最后以问卷调查的方式验证了模型合理。文献[6]使用幂律分布的尺度参数刻画了用户群体评论行为的时间特性,扩展了评论行为相关研究内容。提出了新的刻画群体用户评论内容特征的指标,包括主题的分布和话题多样性。同时,除了关注评论内容中产品相关的信息,还研究关注了评论内容中的服务信息,扩展了评论内容特征抽取的研究。文献[7]将文本评论数据划为正面（1）、中立面（0）以及负面（-1）三部分,对其中的正面和负面评论进行LDA主题模型分析,挖掘相应的主题以及具有高频频率特征的单词,进而对商品做出了评价。文献[8]设计并实现了一种基于时间序列的数据预处理方法，并且最终证实了SSP算法可以有效筛选特征,并能在一定程度上使模型的复杂度得到简化,从而提升分类精确度。文献[9]通过对用户评论的情感分析生成了关于各候选产品的商品购买决策矩阵，并由此提出出一种基于随机TOPSIS的排序思路，最终完成了对候选商品的排序。文献[10]对产品评论数据用tf/idf的方法来提取商品属性，最终挖掘出文本评论数据中有价值的内容。文献[11]利用LDA主题模型从在线产品评论文本中挖掘出影响消费者消费决策的影响因子，并基于偏序集理论对通过调查问卷方式获取的数据完成了商品排序。

综合上述学者的研究结果，对文本数据的情感分析是对具有较强烈主观情感色彩的文本进行预处理、主题提取、主题分析、归纳总结的过程。值得注意的是，大部分研究仅仅只是提取和分析评论文本数据中的主题或情感。而随着电子商务模式发展的日益成熟，消费者们对电商平台上产品评价的聚焦点早已不只是限制在主题分析或情绪好坏。因此如何在情感分析过程中更加高效准确地识别分析产品特征成为产品在线评论文本挖掘中急需解决的问题。同时，由于评论内容中的评论评级、评论语义深度，以及评论目标特性方面的商品类型等评论的实用性较高; 并且评论的质量及产品的整体先后排名也会对评论的有用性产生影响，进而对用户的选择产生辐射影响; 除此之外，用户的先前的购买经验以及产品的品牌效应也会在很大程度上影响到在线评论的实用性。基于电商平台上消费者产生的在线评论的独立和可信特征，从海量的在线评论文本中提取有利用价值的评论已然成为消费者、商家共同关注的热点问题，越来越多的研究人员开始研究在线商品评论文本。

参考文献

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.