

分类号: _____

密级: _____

UDC: _____

编号: 201721702011

河北工业大学硕士学位论文

基于在线评论有用性的商品评价模型研究

论文作者: 王然

学生类别: 全日制

学科门类: 管理学

学科专业: 管理科学与工程

指导教师: 耿立校

职 称: 副教授

Dissertation Submitted to
Hebei University of Technology
for
The Master Degree of
Management Science and Engineering

RESEARCH ON COMMODITY EVALUATION
MODEL BASED ON THE HELPFULNESS OF
ONLINE REVIEWS

by
Wang Ran

Supervisor: Prof. Geng Lixiao

December 2019

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本学位论文不包含任何他人或集体已经发表的作品内容，也不包含本人为获得其他学位而使用过的材料。对本论文所涉及的研究工作做出贡献的其他个人或集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：王然

日期：2019.12.9

关于学位论文版权使用授权的说明

本人完全了解河北工业大学关于收集、保存、使用学位论文的以下规定：学校有权采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供本学位论文全文或者部分内容的阅览服务；学校有权将学位论文的全部或部分内容编入有关数据库进行检索、交流；学校有权向国家有关部门或者机构送交论文的复印件和电子版。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：王然

日期：2019.12.9

导师签名：王然

日期：2019.12.9

摘 要

随着网购的迅速普及，越来越多的用户选择线上购买商品及服务的方式满足自己个性化的需求，随之而来的是用户在线评论数量的急剧增加。已有研究表明，在线评论会影响用户的购买决策。因此，帮助用户从海量评论信息中提取商品相关信息，在考虑已购买用户体验的基础上对商品做出评价，辅助用户做出决策，是十分重要且必要的。

本文在对商品评价、在线评论有用性、商品特征提取、情感分析等多个主题进行综述后，选取了在线评论全面性与相关性、情感强度、评论文本长度、用户打分、评论距今时间跨度等指标，并使用商品特征提取、情感分析等方法对指标进行量化。然后，结合层次分析法和 TOPSIS 排序算法，计算在线评论有用性量化值。在对商品评价时，通过计算每条评论在商品各个特征维度的情感值，并使用该条评论的有用性量化值对情感值进行修正，以区分其对于商品评价过程的参考价值。最终，本文立足于全体数据集，对每个商品 id 建立模糊评价矩阵，并计算商品评价得分。通过对比商品评价值的相对大小，为用户决策提供可参考角度，进而实现了在充分考虑用户体验以及用户发声的基础上，完成了对商品的评价。

本文的创新点有三点：(1) 以往在进行商品评价的研究中，尽管已有研究在逐步完善，但其评价并未对数据本身做出处理，存在将时间较为久远等评论参与到了商品评价过程中。为此，本文在构建模糊评价矩阵时，使用了评论有用性对矩阵进行修正，更具有合理性；(2) 对在线评论有用性指标中的商品特征个数和情感词个数指标进行了优化。区别于以往计算在线评论有用性时使用的商品特征个数和情感词个数指标，本文分别将其改进为综合考虑商品特征个数和商品特征重要性、以及使用情感分析计算情感强度的方法，这种改进不仅将其指标影响更加细化，同时将情感度量落地。这对于指标的完善和进一步的探索打下了基础；(3) 本文使用情感词典以及定义规则的方法计算情感强度值，并使用 F1 值作为情感分类结果衡量的指标，通过使用随机森林结合不同向量进行情感分类，其所得结果的 F1 值与本文情感分析方法进行比较，证明了本文方法的准确性与可靠性。

关键词：评论有用性；全面性与相关性；情感分析；TOPSIS；模糊综合评价法

ABSTRACT

With the rapid spread of online shopping, more and more users choose to purchase goods and services online to meet their individual needs. What has changed is that the number of online reviews has increased dramatically. Studies have shown that online reviews can influence a user's purchasing decision. Therefore, it is very important and necessary to help users to extract product-related information from massive commentary information to help users in making decisions.

After reviewing the literature on commodities evaluation, the helpfulness of online reviews, commodity feature extraction, sentiment analysis, and other topics, the comprehensiveness and relevance of online reviews, sentiment intensity, review text length, user scores, and time span of reviews are selected, and the indicators are quantified using commodities feature extraction and sentiment analysis. Then, combined with the analytic hierarchy process and TOPSIS ranking algorithm, the quantitative value of the helpfulness of online reviews is calculated. In the evaluation of a commodity, the sentiment value of each review in each characteristic dimension of the commodity is calculated, and the sentiment value is corrected by using the helpfulness quantitative value of the review to distinguish its reference value for the process of commodity evaluation. Finally, based on the entire data set, this paper establishes a fuzzy evaluation matrix for each commodity id and calculates the commodity evaluation score. By comparing the relative sizes of the commodity evaluation values, a reference angle can be provided for the user's decision-making, and then the evaluation of the commodity is completed on the basis of fully considering the user experience and the user's voice.

There are three innovations in this paper: (1) In the past research on commodity evaluation, although the existing research is gradually improving, the evaluation did not deal with the data itself, and there were comments that took a long time to participate in the evaluation process. For this reason, in constructing the fuzzy evaluation matrix, this paper uses the helpfulness of the review to modify the matrix, which is more reasonable; (2) The indicators of the number of commodity features and the number of sentiment words in the helpfulness index of online reviews are optimized. Different from the index of the number

of commodity features and the number of sentiment words used to calculate the helpfulness of online reviews in the past, this paper improves them into a method that comprehensively considers the number of commodity features and the importance of commodity features, and calculates the intensity of sentiment using sentiment analysis. This improvement not only refines the impact of its indicators, but also grounds sentiment measurement. This lays the foundation for the improvement and further exploration of the indicators; (3) This paper uses the sentiment dictionary and the method of defining rules to calculate the sentiment intensity value, and uses the F1 value as an index for the sentiment classification result, by using a random forest combined with different vectors. The sentiment classification, the F1 value of the result is compared with the sentiment analysis method in this paper, which proves the accuracy and reliability of the method in this paper.

Key words: review helpfulness; comprehensiveness and relevance; sentiment analysis; TOPSIS; fuzzy comprehensive evaluation

目 录

第一章 绪论	- 1 -
1.1 研究背景、目的和意义	- 1 -
1.1.1 研究背景	- 1 -
1.1.2 研究目的	- 2 -
1.1.3 研究意义	- 3 -
1.2 研究内容	- 3 -
1.3 研究方法和技术路线	- 4 -
1.3.1 研究方法	- 4 -
1.3.2 技术路线	- 5 -
1.4 文章创新点	- 6 -
1.5 本章小结	- 7 -
第二章 文献综述	- 9 -
2.1 商品评价	- 9 -
2.2 商品评价要素	- 10 -
2.2.1 在线评论有用性含义	- 10 -
2.2.2 在线评论有用性影响因素研究	- 10 -
2.2.3 在线评论有用性排序研究	- 11 -
2.3 商品评价要素挖掘方法	- 12 -
2.3.1 商品特征提取	- 12 -
2.3.2 情感分析	- 14 -
2.4 研究述评	- 15 -
2.5 本章小结	- 16 -
第三章 评论有用性度量研究	- 17 -
3.1 要素选取	- 17 -
3.2 要素量化	- 19 -
3.2.1 评论全面性与相关性量化	- 19 -
3.2.2 评论情感强度量化	- 23 -
3.2.3 其他要素量化	- 26 -

3.3 本章小结	- 27 -
第四章 在线商品评价模型构建研究	- 29 -
4.1 商品评价指标权重确定	- 29 -
4.1.1 评论有用性指标权重赋值	- 29 -
4.1.2 加权 TOPSIS 排序算法	- 31 -
4.1.3 评论有用性量化	- 32 -
4.2 商品评价模型构建	- 34 -
4.3 本章小结	- 36 -
第五章 实证研究	- 37 -
5.1 数据获取与预处理	- 37 -
5.1.1 数据获取	- 37 -
5.1.2 数据预处理	- 42 -
5.2 在线评论有用性计算	- 43 -
5.2.1 指标量化	- 43 -
5.2.2 指标权重计算	- 50 -
5.2.3 评论有用性量化结果展示	- 52 -
5.3 商品模糊综合评价	- 53 -
5.3.1 模糊评价矩阵构建	- 54 -
5.3.2 商品评价得分	- 55 -
5.3.3 商品模糊综合评价结果展示	- 56 -
5.3.4 商品评价模型验证与分析	- 57 -
5.4 研究启示	- 59 -
5.5 本章小结	- 61 -
结 论	- 63 -
参考文献	- 65 -
附录 A 数据源部分展示	- 71 -
附录 B Doc2vec 指标表现	- 75 -
附录 C 情感向量指标表现	- 79 -
附录 D 均衡数据集指标表现	- 83 -
附录 E 调查问卷	- 87 -
攻读硕士学位期间承担的科研任务与主要成果	- 89 -
致 谢	- 91 -

第一章 绪论

随着网购规模的扩大，商品在线评论积累速度加快，这使得消费者从用户评论中挖掘商品有关信息变得更加困难。因此，有必要从用户的角度对商品进行综合评价，以辅助用户高效便捷地选择商品。本章针对本文的研究目标，明确本文的研究内容、研究方法和研究框架，为全文研究制定总体研究思路。

1.1 研究背景、目的和意义

1.1.1 研究背景

2019年6月，中国社会科学院发布了“2019中国电商半年报”，报告显示，2019年上半年网上零售额持续增长，近半的新增消费支出通过网络消费实现。报告显示，截止至2018年12月，我国网购用户数量达到6.10亿，平均每个家庭已有超过1.3人使用电商购物。随着网购在大众生活中的普及，网上购物人群的扩大，人均购买次数的增多，可供选择的商品越来越多，因此如何在网上购买到合适的商品值得关注。电子商务快速发展的同时，用户在线评论数据迅速沉淀积累。在线评论不仅是反映用户体验的第一手资料，也是潜在消费者做出购买决策的重要依据^[1]。早前已有研究表明，相较于商家的广告宣传，在线评论更具有实时性和说服力，因而备受关注^[2]。线上购物区别于线下购物时可参考实物本身，线上购物过程中，在线评论对用户购买决策有重要的影响^[3]，也就是说，在线评论已经成为大多数消费者购买决策中使用的新信息来源^[4]。然而，由于在线评论数量巨大，用户难以逐条阅读所有评论数据，这就导致用户难以整合海量评论中对商品情况的表达。因此，使用大数据的相关技术手段，基于用户高有用性感知的在线评论，对商品做出客观且反映大众视角的评价具有重要的现实意义。

从学术研究的角度来看，以往帮助潜在购买者选择商品的模型有：一类是基于产品参数，该方法将产品重要性放在第一位，但是并没有考虑用户的体验；另一类是基于问卷调查，该方法考虑了用户体验，但是问卷的设计、发放、回收、处理费时费力^[5]。而用户评论不仅可以反映用户体验，还相对较容易获取，因此，基于用户评论的商品评价对潜在消费者而言是具有参考价值的，一方面因为评论在很大程度上反映了

已购买用户对商品的实际体验，另一方面对商品评论整体情感倾向的分析也可帮助分析商品及服务的整体情况如何。因此，基于在线评论的商品评价对电子商务平台进行商品列表排序及其他服务而言，是极具研究价值的。除了商品选择的相关研究，目前关于推荐系统算法的研究，大多集中于如何深入且准确地挖掘用户喜好、如何提高算法效率等方面，缺乏对推荐列表商品的筛查和过滤。因此，为了更好地提高商品转化率，推荐满足用户喜好商品的同时，还应当考虑商品及服务的情况。

从企业和用户的角度来看，目前电子商务平台正在使用的一些商品排序方法，如在京东搜索页面，可以按照价格、评价数等方法对商品进行排序，这些排序的依据过于单一，且不能反映商品真实的综合面貌。而电子商务平台拥有的用户在线评论数据量已经达到了大数据规模，因此，我们应当借助大数据的优势，充分考虑用户发声，从用户的角度解读商品，尽可能客观、全面、真实的反映商品本质。从点击商品开始，用户逐渐了解和熟悉商品的过程，可以看作是学习的过程，用户通过浏览商品首页、商品详情页来获得商品的感官初步认识。然而由于商家的包装效果和宣传效果的夸大，用户会阅读商品评论对商品进行深层次的判断。在用户阅读评论的过程中，会对评论的有用性有一定的感知，进而在线评论有用性会影响用户的购买决策^[6]，这说明用户对不同有用性的评论有不同的参考价值，因此，有必要在商品评价模型中将评论有用性指标加入其中，已符合大众消费者的认知习惯。

综合上述观点，无论从学术研究的角度还是企业、用户的角度，使用科学的方法，整合现有的资源，对商品做出符合大众认知习惯的评价，这为用户提供一个良好的购物环境、增强用户体验和保持用户粘性打下了坚实的基础。

1.1.2 研究目的

结合商品多维度特征以及在线评论中对于商品特征的情感表达，为商品做出评价，有利于用户对商品进行评判，缩短用户搜索时间，降低购物过程的不确定性，提高商品转化效率。本文以电子商务在线商品为研究对象，用户在线评论为研究数据源，通过对在线评论的文本挖掘实现对在线商品做出相对优劣的评价。与此同时，由于商品评价模型可以适用于多种场景，例如商品推荐、商品搜索页排序等，因此，本文希望通过用户在线评论来评价商品，可以实现将商品不仅推荐给合适的人，更能为合适的用户推荐相对优质的商品。

此外，本文引入在线评论有用性进行商品评价可以看作是模拟用户进行商品选择，用户通过高有用性的评论获取商品信息以辅助其做出购买决策，因此，本文以评论有用性作为区别在线评论参考价值大小的依据参与到商品评价中，以更好地站在用户的

角度、按照用户的习惯评价商品。

1.1.3 研究意义

第一，理论意义上，关于商品评价的研究模式较为固定，其研究的重点和关注点忽略了对数据源即在线评论的区别。通过梳理相关文献，相关研究者认为不同的在线评论有用性不同。因此，基于在线评论来评价商品时，应当对在线评论的有用性加以区分。其次，在线评论的有用性方面，学者关于影响有用性因素的探索尚未形成统一的结论，因此，通过梳理学者提出的有用性影响因素，本文发现学者使用的部分指标是可以进一步细化和落地的。通过文本挖掘的方法细化指标，使得研究结果更具有合理性。此外，通过基于在线评论有用性对商品进行多维度的综合考量，结合用户消费后的反馈，以实现商品的评价。由于本文的研究工作具有基础性，希望通过本文提出的商品评价模型，为商品推荐等相关领域研究提供另一种思考。

第二，实践意义上，商品广告宣传和包装带来的商品质量、效果等多方面信息的夸大，导致用户不再单纯依赖商家提供的信息，而是更多依赖已经购买该商品的用户提供的在线评论。然而，随着在线评论的积累，用户陷于海量的评论中，难以对商品做出判别。因此，本文基于在线评论辅助用户评价商品的同时，对在线评论的有用性即可参考价值进行了区分，加强了相对高有用性评论的参考价值，更加合理地评价商品，快速辅助用户做出决策，提高决策效率。

1.2 研究内容

文章主体部分研究流程图如图 1.1 所示。本文基于在线评论有用性对商品评价进行了实证研究，主要包括如下几个方面的研究内容：

第一，明确商品评价使用的对象。目前，商品评价研究较少。基于已有的研究现状，结合对商品评价逻辑以及评价模型构建过程的梳理，发现商品评价过程中存在的可创新点。本文发现商品评价所使用的数据源即在线评论的有用性是参差不齐的，因此，商品评价可以从在线评论本身作为切入点进行后续研究。

第二，在线评论有用性量化研究。在线评论有用性是在线评论相关研究的热点话题，但是其影响因素至今尚未统一。本文通过梳理在线评论有用性量化研究中所使用的影响因素，找出本文量化在线评论有用性的指标维度，并进行指标量化方法的优化。

第三，商品评价研究。基于在线评论有用性，本文通过构建指标集、评价集、模糊评价矩阵使用模糊综合评价法对商品进行综合评价。

第四，实证研究。本文以“羽绒服”评论数据集为例，按照上述的方法、规则、逻辑在真实数据集上进行实证研究。通过实证分析，研究基于在线评论有用性的商品评价，证明其具有可行性与有效性。

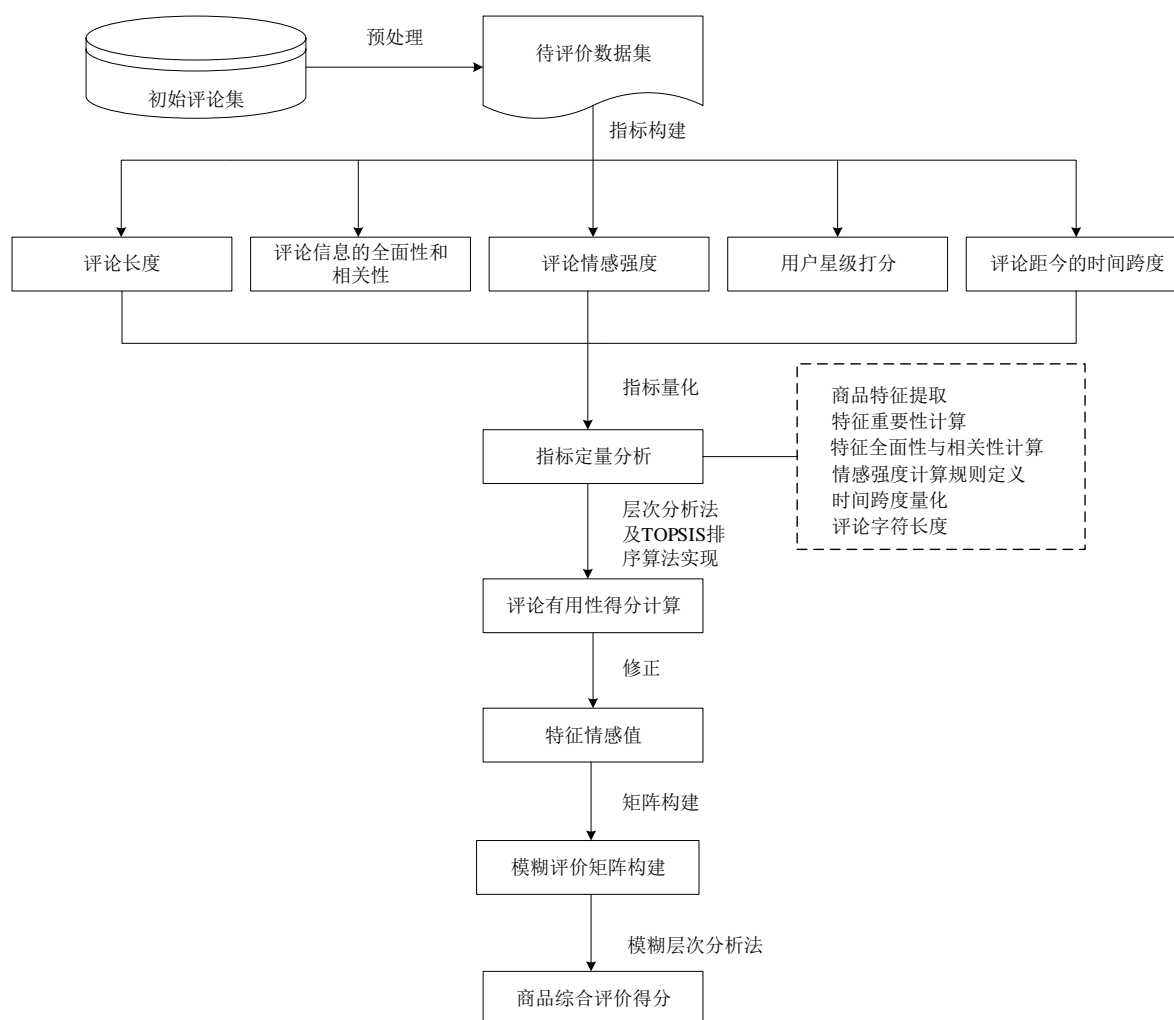


图 1.1 研究框架图

Figure 1.1 Research framework

1.3 研究方法和技术路线

1.3.1 研究方法

(1) 文献研究法

文献研究有助于研究者更深入地了解研究前沿，更好地把握研究问题。首先，查阅大量相关文献，对商品评价和评论有用性进行文献研究，找出可研究并且值得研究的角度，然后查阅大量有关商品特征提取、情感分析等方面的文献，理清研究脉络和

思路。

(2) 问卷调查法

问卷调查法可以让研究者接触到消费者，了解消费者的想法，并将消费者的需求转化成数据，目前，问卷调查法是一种重要的实证研究方法。本文为了确定在线评论有用性指标权重，采用问卷调查的方法获得相关数据。

(3) 文本挖掘

文本挖掘是对评论文本挖掘其深层次信息的方法。通过文本预处理、分词、去除停用词等工作构建词向量，为后续研究工作奠定基础。

(4) 情感分析

情感分析包括两个方面，一个是情感倾向的分类，一个是情感强度的计算。情感强度作为在线评论有用性的指标，需要情感分析方法实现。

1.3.2 技术路线

图 1.2 为本文的技术路线。首先，通过对研究目标的文献梳理，明确研究的背景及意义，在此基础上，对商品评价、在线评论有用性、商品特征提取、情感分析等主题进行深入的文献研究；其次，对在线评论有用性指标选取、指标量化方法以及有用性量化进行探讨；然后，建立在线评论有用性修正后的商品评价研究模型，并进行实证研究；最后，对本文的研究进行总结，并对未来研究方向进行展望。

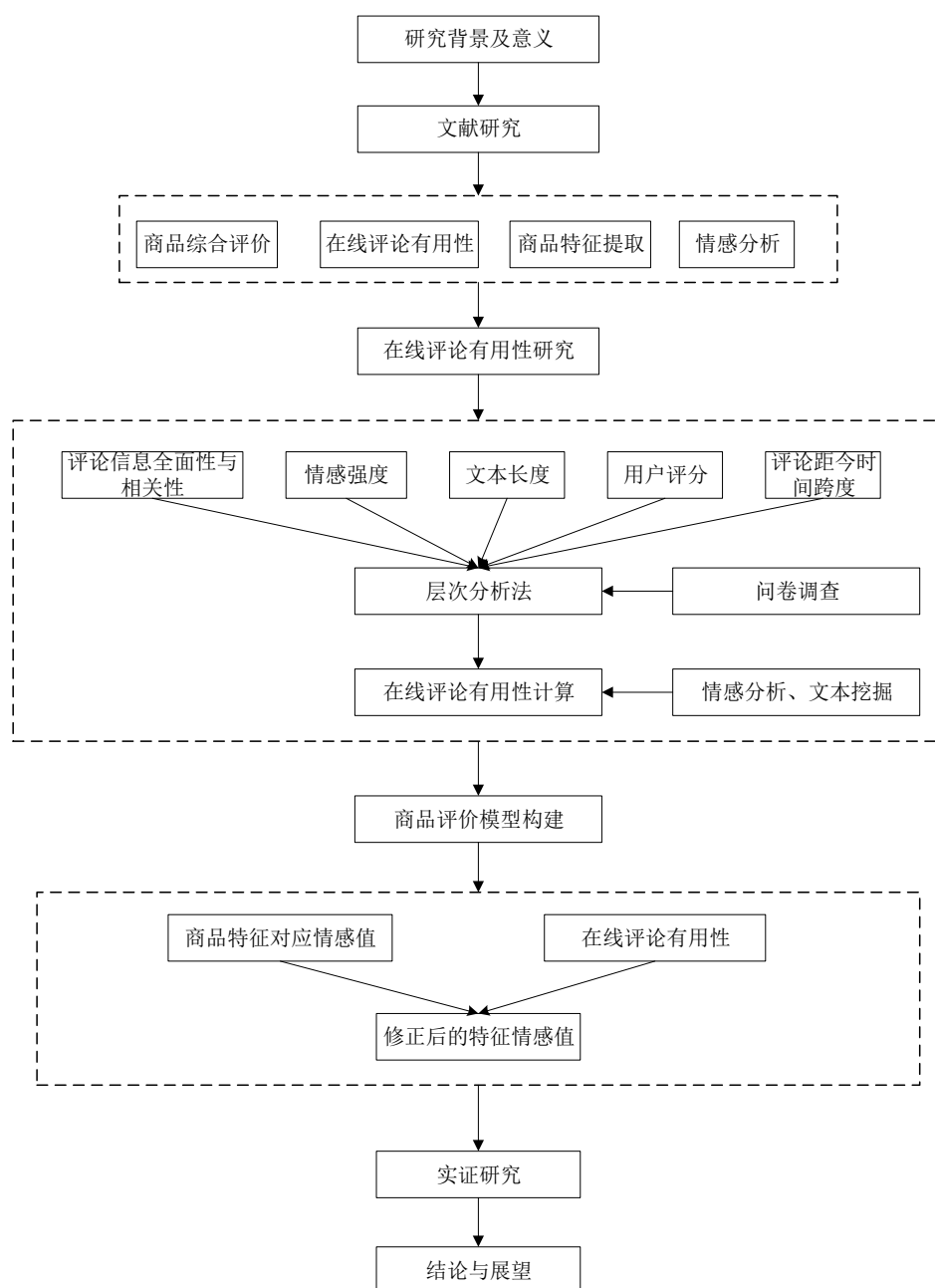


图 1.2 技术路线图

Figure 1.2 Technical roadmap

1.4 文章创新点

本文的创新点有三点：

第一，在线评论有用性指标量化优化。以往研究在进行在线评论有用性量化过程中，考虑了评论提及特征个数。但是对于不同的商品特征，其重要性是不同的。本文使用商品特征被用户提及的频率作为特征重要性加以考虑。相较于传统计算方法而言，

更具有合理性；此外，本文对在线评论有用性指标中的情感词个数指标转换为情感强度作为情感指标的量化值，但是鉴于不同情感词表达的情感强度不同，因此，本文使用情感分析方法将情感强度计算真正落地，并且构造了羽绒服品类的领域情感词典。

第二，本文使用情感词典和定义规则相结合的方法，计算评论情感强度，通过使用随机森林结合不同向量进行情感分类，其所得结果的 F1 值与本文情感分析方法进行比较，证明了本文方法的准确性与可靠性

第三，在构建模糊评价矩阵时，以往研究使用各个商品特征的情感强度构建，但是本文结合在线评论有用性对模糊评价矩阵进行了修正，使商品评价过程更具有合理性。

1.5 本章小结

本章探讨了商品评价的相关研究背景、目的及意义，明确了本文研究主题的重要性和必要性。根据文献梳理，明确了研究目标、研究内容以及研究方法，并对其进行整合，制定了本文研究的技术路线。通过阅读文献，梳理研究脉络，明确研究目标与模块，选取合适的研究方法，为本文后续的研究提供了理论基础和支撑。

第二章 文献综述

本章将从商品评价、商品评价要素、商品评价要素的挖掘方法等三个模块进行文献综述。首先，梳理商品评价的相关研究，明确研究内容和可改进之处；由于商品评价时引入了关键要素即在线评论有用性，因此在商品评价要素部分对在线评论有用性相关研究进行了综述，明确商品评价要素研究现状，明晰其研究范围、内容以及普遍认知的评价指标。商品评价要素挖掘方法对商品特征提取、情感分析等方法进行了文献综述，通过文献阅读，了解可以适用本文的要素提取及情感表达的量化方法。

2.1 商品评价

商品评价方法可以大致分为两种，一是根据指标通过权重连接，计算商品数值得分，二是使用模糊综合评价方法对商品进行评价。

商品评价方法之一是根据指标值，定义商品评价规则，设定权重计算商品得分值。Li 等人^[7]认为商品评价可以根据负面评论句子数的占比作为指标，因此其构造商品通过权重连接情感值与负面句子占比值计算商品得分。王乾和王晓耘^[8-9]采用用户文本评分和数值评分不一致的角度计算商品评分。

模糊综合评价法利用模糊数学的思想将定性评价转化为定量评价，即根据各指标对应的等级构建模糊评价矩阵，结合各指标的权重和各等级的得分赋值，计算得到商品得分。1994 年，李梵蓓^[10]首次将模糊综合评价法应用于商品评价，作者使用商品耐用程度、价格、颜色样式等六个指标对商品质量进行刻画，并将商品质量归结为四个等级，其指标权重由咨询法平均估值设定，且模糊矩阵是根据参与商品评价的 20 位专家对各指标等级的投票占比决定的，因此商品评价过程中主观性较强。且由通过商品多维度特征的情感表达，结合特征权重，将情感表达结合起来，并计算商品综合得分。王涛^[11]从电子商务层面构造出电子商务产品指标体系，应用模糊数学综合判断方法，对电子商务企业产品进行整体分析和综合评判，评价出产品优劣。其不足是采用指标两两进行评价方法确定权重，主观性太强，对商品的综合评价结果说服力不够强。杨静^[12]着眼于电子商务网站售后商品评价环节存在的问题，构建的评价模型考虑了评价的公正性和易于理解性，但是对于模糊评价矩阵是用调查统计的方法得到，此过程需要更多的人力物力。并且在各指标权重的判断过程中，采用了专家经验判断，主观性较强。徐勇等人^[13]将商品特征作为商品评价指标，商品特征被用户提及的频率作为

指标权重，模糊矩阵是用户对指标情感表达的情感值，这在很大程度上避免了主观因素，并且充分利用了用户反馈的评论信息。

商品评价研究相对较少，且研究过程中并没有对数据源即在线评论的有用性大小加以区别，因此，本文将在线评论有用性引入商品评价模型，对评价模型进行改进，使得商品评价研究更加合理。

2.2 商品评价要素

2.2.1 在线评论有用性含义

在线评论有用性由 Chatterjee^[14]于 2001 年提出，作者认为在线评论有用性是指信息使用的影响程度。后续关于在线评论有用性研究使用较多的是由 Mudambi 和 Schuff^[15]于 2012 年提出的观点，即消费者在决策过程中对在线评论是否有帮助的主观感知价值，该定义最早从感知价值角度定义了在线评论有用性。因此，本文关于在线评论有用性的研究是基于用户感知的角度进行的，这和许多研究的出发点是一致的。

2.2.2 在线评论有用性影响因素研究

在线评论有用性影响因素研究是该领域的关键问题。目前，学者对于有用性影响因素尚未形成统一的结论。刘伟和徐鹏涛^[16]认为信息可读性、信息丰富性以及点评负面性对评论有用性有显著正向影响。李中梅等人^[17]认为产品属性特征词、情感特征词正向影响在线评论质量，在线评论质量正向影响在线评论可信度，评论者的情感越正向，评论有用性越高。同样的，Simon 等人^[18]认为正向情感强度对有用性有显著影响。吴江等人^[19]认为以亚马逊平台作为研究对象，其认为是否含有“已购买”标签对评论有用性有显著影响。王晗等人^[20]基于 word2vec，采用深度学习技术开发的语义解析学习工具对在线评论有用性进行预测。田依林等人^[21]研究了不同类型的在线商品的负面评论有用性影响因素。郝媛媛等人^[22]认为评论句子的长度、正向情感倾向对在线评论有用性呈现正向影响。Han 等人^[23]使用国内外电影影评数据集进行评论有用性的对比研究，作者认为评论文本以及情感强度与评论有用性在两个数据集上均成正相关关系。Zhou 和 Yang^[24]指出评论文本长度和数值打分对有用性有显著影响。Singh 等^[25]根据香农提出的自然语言信息熵的计算方法，用信息熵表示每条评论中平均每个单词产生的信息量，信息熵越大，对评论有用性影响越大。同样的，艾时钟和曾鑫^[26]等人指出情感总量对在线评论有正向影响，且信息熵对体验型商品在线评论有用性有正向影响。

综合在线评论有用性相关文献的阅读，关于有用性影响因素部分总结如表 2.1 所示。

表 2.1 有用性评价指标研究

Table 2.1 Helpfulness evaluation index research

有用性评价指标	文献
文本长度	[15]、[27]、[28]
情感强度	[29]、[30]、[31]、[32]
用户评分	[28]、[33]、[15]、[34]、[35]
产品词汇量	[27]、[36]、[37]
修饰词数量	[6]、[36]、[37]
时间跨度	[27]、[28]、[36]、[38]

综合影响因素文献阅读，在线评论有用性影响因素研究较为丰富，但其指标研究范围较为固定，文本长度、情感强度、特征词等部分指标得到了较多学者的证实。

2.2.3 在线评论有用性排序研究

在线评论有用性排序研究是基于有用性量化的方法实现的。其采用的方法集中为以调查问卷、专家研究等方法为关键指标赋权重值，结合各个指标量化结果，计算评论有用性量化值。陈市等人^[27]将文本长度、商品特征词数量、负面情感词数量、评价者等级、追加评论天数、时间进度等六个指标量化值，结合模糊层次分析法赋权，计算有用性的量化值。蔡晓珍等人^[29]采用人工标注 1000 条评论数据的评论质量等级，以该值作为因变量，以评论涉及产品词汇量、评论长度、修饰词数量、情感强度四个指标进行多元线性回归方程，并根据所得权重计算剩余文本的质量等级。王倩倩^[39]等人通过计算评论一致性、评论时间、评论投票数、评论长度等因素使用专家赋权的方法计算评论有用性。毛郁欣等人^[40]将商品在线评论有用性投票作为标签，结合 SVM 对评论有用性进行分类，其概率值作为评论有用性得分。郭顺利等人^[36]使用加权灰色关联度的方法计算评论有用性。张梦莹等人^[6]采用模糊层次分析法确定指标权重，并结合灰色关联度对有用性进行排序。

综合有关文献的阅读，从有用性排序的方法上来说，可以分为三步：一是指标选取；二是指标量化；三是权重确定。其中，第三步的权重确定有两种方法：一是通过人工标注有用性等级，根据回归方程求解权重；二是采用传统指标赋值方法，如问卷调查、专家赋权等。

2.3 商品评价要素挖掘方法

2.3.1 商品特征提取

已有文献通常采用机器学习、根据句法依存关系等方法提取商品特征。Brody 和 Elhadad^[41]首次使用 LDA 提取的主题作为商品特征。Santosh 等人^[42]使用特征本体树结合 LDA 特征聚类提取商品特征。彭云等人^[43]使用词义相似度和上下文相关度相结合的词聚类度量算法,构建了一种基于词聚类先验知识的潜在狄利克雷分配的商品主题特征提取模型。Liu 等人^[44]使用条件随机场提取显式商品特征。Wu 等人^[45]结合词性以及句法依存关系制定规则,并结合词语间相似度和无向图制定聚类规则,实现对商品特征的提取;Devasia 和 Sheik^[46]使用英文单词两两词语间语法依赖关系作为规则进行商品特征提取。Wai 和 Aun^[47]利用观点词与特征词之间的关系建立双向传播提取规则;Rodrigues 和 Chiplunkar^[48]使用按词性抽取词语并使用关联规则提取商品特征。邱云飞等人^[49]根据句法依存关系构建句法树进行商品特征提取。

本文借鉴已有文献^[50-51]提取商品特征的方法,即 Word2Vec 和聚类分析方法相结合的方法提取商品特征。该方法的好处是其构建的词向量为后续工作打好了基础。此外,本文使用 TextRank 算法提取前 200 个关键词,并从中筛选词性满足商品特征要求的词语,随后对少量词语进行人工筛选,加入商品特征词中。

Word2Vec 是 Google 公司在 2013 年使用深度学习技术开发的一款开源学习工具。由于该工具可以将文本嵌入为空间词向量,并且词向量中是基于上下文学习而得,因此该工具既可以简化计算,且可以保证信息不会过度丢失。Word2Vec 构建出的词向量可以应用于很多方面,如词语相似度计算,词语分类等。

Word2Vec 包括两个模型^[52-53],即 CBOW 和 Skip-gram。两个模型的区别在于:CBOW 模型是基于上下文计算当前词向量,而 Skip-gram 是基于当前词预测上下文向量。但是二者结构相同,均为三层结构:即输入层、映射层和输出层,两个模型的结构如图 2.1 和 2.2 所示。CBOW 词向量相较于 Skip-gram 模型而言,训练速度较快,但是 Skip-gram 在识别生僻字上具有优势。

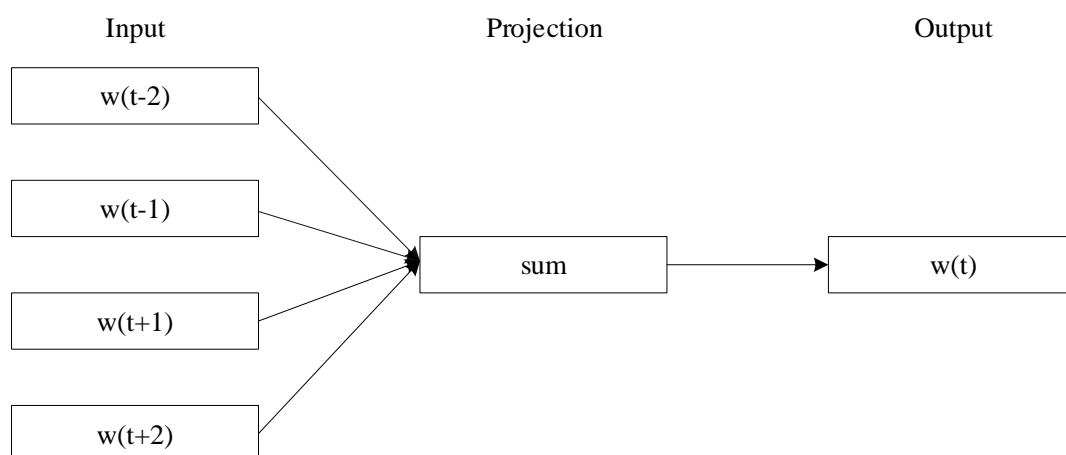


图 2.1 CBOW 模型结构图

Figure 2.1 CBOW model structure diagram

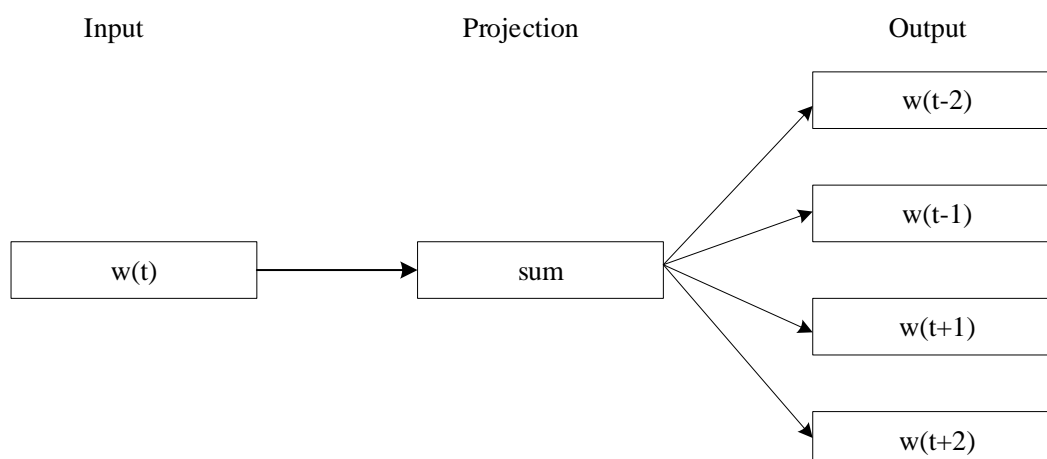


图 2.2 Skip-gram 模型结构图

Figure 2.2 Skip-gram model structure diagram

TextRank 算法常用于文本关键词提取以及自动摘要。该算法来自于 Google 公司的 PageRank 算法^[54], PageRank 算法根据网页被链接的多少以及链接网页的质量综合考虑, 以进行搜索结果排序。利用其思想, 亦可以用于文本重要性计算。TextRank 是一种基于图排序的算法。使用 TextRank 算法对关键词排序的流程如图 2.3 所示, 其基本思想是: 如果一个词和其他多个词语都有联系, 则该词被认为更重要, 因此评分就高。图顶点的权重均分给每一个有连接关系的其他顶点, 每一个顶点的评分不仅仅和自己的属性有关, 还和其相关联的顶点有关系。算法通过不断迭代计算各个顶点的评分, 直到最终满足一个收敛的阈值。TextRank 算法对各个顶点词的评分公式如公式(2.1)所示。

$$Score(w_i) = (1 - d) + d \times \sum_{w_j \in In(w_i)} Score(w_j) \times w_{ij} / \sum_{w_k \in Out(w_j)} w_{jk} \quad (2.1)$$

其中, d 为阻尼因子, 一般取值为 0.85, 表示候选词 i 跳转到 j 的概率; $In(w_i)$ 表示候选词 i 的顶点集合; $Out(w_i)$ 表示候选词 j 所指出的顶点集合。 w_{ij} 表示候选词 i 和 j 的边权重, 表示各个词语之间的影响关系。

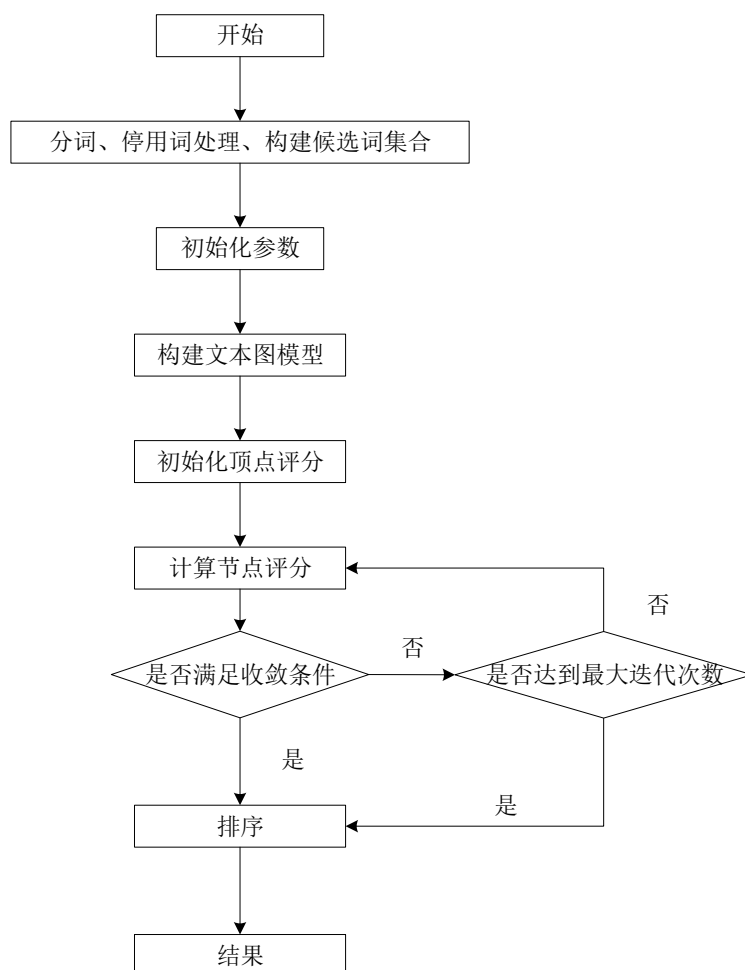


图 2.3 TextRank 算法流程图^[55]

Figure 2.3 TextRank algorithm flow chart

2.3.2 情感分析

情感分析的研究路径包括使用机器学习或深度学习和使用情感词典两种。前者多应用于句子级粗粒度情感分析, 后者可以进行不同粒度的情感分析。情感分析的实验效果一般通过 F1 值或者准确率(Accuracy)来衡量。

使用机器学习进行情感分析, 通常是基于商品评论文本构建特征向量, 进行有监督或者无监督的学习。颜端武等人^[56]通过句法依存关系构建产品特征树, 结合 LSTM 神经网络算法进行情感分类, 其正面评论的分类效果 F1 指标为 83.16%, 负面评论的分类效果 F1 指标为 84.67%。Aleebrahim 等人^[57]分别提取商品特征词语和形容词作为

情感词，并根据特征词语及其对应的情感词极性构建向量，使用 SVM 作为文本情感二分类器，其准确率为 83.6%。Sharaff 和 Soni^[58]使用 TF-IDF 权值向量表示评论文本并结合多种分类器进行实验，其中随机森林分类结果最好，其 F1 值为 78.2%，准确率为 85.9%。Bijari 等人^[59]使用基于图表示文本向量，结合了不同的机器学习算法并对结果进行比较，准确率和 F1 值最大均为 98.73%。

使用情感词典进行文章分类时进行已有文献通常基于台湾大学 NTUSD-简体中文情感极性词典或 HowNet 情感词典或大连理工大学构造的情感词汇本体作为知识资源建立情感词典，但是由于研究领域的多样性，上述情感词典中的词汇并不能完全满足研究的需要，因此，在研究情感分析相关工作时，往往需要构建领域专用情感词典。领域情感词典的构建可以使用基础情感词典，配合人工增加词汇完成^[60-61]，或者采用较为自动化的方法完成^[62-63]。由于非英文词典资源较为匮乏，已有学者^[64]提出了不依赖词典进行情感分析的方法，其 F1 值分别为 74%，71%；杜嘉忠等人^[61]通过构建特征-情感词对进行细粒度情感分析，其 F1 值为 80.14%；赵志滨等人^[65]对不同的商品属性赋予不同的权重，情感计算时针对含有至少商品属性维度的评论为有效评论计算情感值，属性权重越大，计算得到的情感值就越大，其准确率 81%，F1 值 82%；樊振等人^[62]将领域情感词典和机器学习这两类情感分析方法在电影数据集上做了对比，其准确率指标均小于 80%。

除了上述正负情感分类等相关研究以外，部分学者^[66-68]将情感细分为情绪，对情绪分类进行了深入的研究。

综合 2.3.1 和 2.3.2 小节可知，商品特征提取和情感分析的方法思路类似，即这两方面的研究方法可以分为两条路径：一是通过定义规则可以实现上述内容；二是通过机器学习或者深度学习实现上述内容。

由于本文有商品特征提取的工作，而商品特征抽取是细粒度情感分析的基础^[69]。因此，本文采用细粒度情感分析并通过实际数据集的情感分类结果以及与随机森林算法不同角度的对比，证明本文细粒度情感分析方法的有效性。

2.4 研究述评

通过文献阅读，本文发现模糊综合评价法应用于商品评价时，缺少对评论参考价值即评论有用性的修正。因此，本文以此为出发点，进行在线评论有用性的研究。

目前关于评论有用性的研究，主要集中于指标选取和量化两个方面。然而，在使用在线评论的情感色彩等相关特征作为有用性指标时，大多采用情感词的个数等方法衡量在线评论的情感，即没有将情感分析真正落地，不能相对精确的得出在线评论的情感强度。故本文将使用定义情感强度计算规则的方法，以得出最终的情感得分，并

通过与随机森林结合不同向量训练模型所得预测结果进行对比，证明了本文情感强度计算方法的有效性和准确性。相较于原有衡量在线评论情感表达的方法，本文的方法更具有合理性和参考性。此外，在考虑在线评论丰富性和充分性时，已有文献仅考虑了商品特征的个数，并没有考虑商品特征的重要性。而计算在线评论有用性时，在线评论中所包含的特征越重要，显然对于用户而言，从该条在线评论中所感知到的有用性会越高，同时在对商品做出评价时，该评论所述的各个商品特征的观点和态度，也会因为其有用性高而得到重点关注，因此，商品特征重要性对于在线评论有用性和商品评价十分重要。故本文使用用户提及商品特征的频率作为商品特征重要性，相较于原有仅考虑商品特征个数的方法，本文的方法避免了研究人员的主观判断，体现了客观性的同时，也保证了充分考虑用户发声的初衷。

综合上述考虑，本文使用在线评论有用性量化的主流指标，结合改进后的评价指标量化方法，将其量化结果应用于模糊综合评价法，并对评价模型进行修正，以区分相对高有用性评论和相对低有用性评论的参考价值，进而使得商品评价更能考虑已购买商品用户的发声，因此，无论从理论研究上还是实践的角度，本文所做出的改进均具有创新性。

2.5 本章小结

本章回顾了前人在以商品评价为研究主题所做出的贡献。在总结归纳前人研究时，发现可以改进研究的角度。通过对研究目标即商品评价的文献阅读，以及评价要素和评价要素挖掘方法的总结和梳理，发现在商品评价流程中，缺少对在线评论有用性因素的考量，因此，本文在商品评价过程中加入了在线评论有用性，使得研究更具有合理性和解释性。针对可以改进的角度，继续追溯相关研究的现状。其次，结合已有数据源和资料，选取研究在线评论有用性时可借鉴的角度、指标和方法。最终，在前人研究的基础上，实现了本文商品评价模型的搭建。

第三章 评论有用性度量研究

本章主要介绍在线评论有用性度量时采用的要素指标的选择以及要素指标的量化。基于文献研究，本章选取评论文本长度、评论的全面性与相关性、评论的情感强度、用户星级打分以及评论距今的时间跨度等五个指标，针对各个指标采用不同的量化方法，包括商品特征提取、情感分析等方法对在线评论文本进行深入的分析与研究，以量化在线评论有用性的各个要素，为后续研究内容打下基础。

3.1 要素选取

(1) 评论文本长度

评论文本长度是指评论文本的总字数^[29]。字符长度较长的评论，所包含的信息越有可能丰富、全面，越有可能降低消费者对于商品的不确定性^[15,17]。并且越长的评论，说明买家评论态度认真可能性越大，越有可能具有较大的参考价值^[39]。综合来说，较长的文本反映的信息越可能充分，同时反映出有用信息的可能性越大。因此，本文选取句子长度作为影响评论质量的指标之一，即评论长度越长，评论有用性越高。

本文将评论文本的有效字符长度作为该指标的量化值，其中，有效字符长度指去除标点以及停止词之后剩余的文本长度。

(2) 评论的全面性与相关性

评论长度这一个指标只是涉及了信息的完整性，无法反映评论信息的相关性和全面性。部分文献^[36,70-71]中，将评论中所含的关键词的个数作为为了在线评论长度的补充，但是本文认为商品特征词数量可以反映评论内容对商品描述的详尽程度，但是仍然无法反映当前商品类型特征的重要程度。即评论中涉及的商品特征丰富且重要时，用户感知到的评论有用性更高。因此本文结合产品特征词数量与特征词重要程度作为评论全面性与相关性(C-R, Comprehensiveness and Relevance)的量化值。

评论的全面性与相关性涉及三个步骤：一是商品特征提取；二是商品特征重要性计算；三是评论全面性与相关性计算即前面二者的结合。计算公式如公式(3.1)所示，具体量化过程见 3.2.1 节所示：

$$C-R = \sum_{i=1}^n -p_i \ln(p_i) \quad (3.1)$$

上式中， n 表示评论中所涉及的商品特征个数， p_i 表示商品特征词语的重要度。该值的求和项，反映了评论谈及商品特征的个数；求和公式中的每一项反映了商品特

征的重要程度。 $C-R$ 值越大,说明评论中涉及的商品特征越丰富,描述的商品特征越重要。

(3) 评论的情感强度

评论的情感强度反映了用户对商品的满意或者不满意程度。当用户对于商品满意时,评论中会表达出较高的情感强度;反之,评论中会表达出较低的情感强度。一般来说,在计算情感强度时,会将评论情感强度大于 0 的认为是正向情感表达,反之,会认为是负向情感表达。消费者通过阅读带有感情色彩的评论,可以感知评论者对于商品的情感倾向,评论者的情感强度的表达会影响评论阅读者的购买决策^[36]。在线评论情感强度早在 2000 年^[72]就被证实其对有用性感知有影响。郝媛媛等人^[22]认为,正向情感强度与长文本对有用性有正向影响。游俊等人^[73]从类似的出发点指出情感表达越频繁,所含特征词越多,评论越具有参考价值。李启庚等人^[74]考虑了时间跨度的问题,认为正负情感的评论对有用性感知有影响,并且近距离时间下,主观感受型评论文本的感知有用性更高。

和评论全面性与相关性类似,部分文献^[37,75]通过情感词的个数作为情感强度的评分。但是并没有考虑否定词、不同情感词语的情感强度以及情感词语所描述的是何种商品特征,而这些因素也是反映评论情感强度的重要因素。因此本文将结合商品情感词对应的特征权重、情感词极性、以及否定词等方面共同计算评论的情感强度。相较于单纯选取情感词个数评价在线评论的有用性,更具有合理性和可靠性。

根据文献阅读,以往关于情感强度作为有用性影响因素的研究大多采用情感词个数以及修饰词个数的方法进行量化,但是本文使用算法计算出在线评论的情感强度,相较于使用个数计算替代情感强度而言,更准确且情感强度指标量化实现了真正落地计算。

评论情感强度的计算分为两个部分进行,一是领域情感词语构建,二是根据语义结构计算情感强度。具体计算步骤如 3.2.2 节所示。

(4) 用户星级打分

用户的星级打分是指电子商务平台的评论者对商品给出以星级表达的评价。一般采用一到五星的评价方式。一星表示最不满意,五星表示最满意,三星表示中立态度。星级打分是很多文献都会探讨的一个影响在线评论有用性的重要因素^[76]。用户打分反映了用户对商品和服务等多维度分析之后得到的整体感知^[71]。由于数值打分的简洁性,其相较于评论文本而言,用户的数值打分真实性较高。因此本文将评论星级作为评论有用性的指标之一。

该指标的量化值以实际数据集中用户的真实打分数据为指标值。

(5) 评论发布距今时间跨度

评论的时效性可以使用评论发布距今的时间跨度来表示。评论发表时间是测量评

论质量常用的客观指标^[17]。时效性高的评论，可以体现商品和服务的最新信息，有助于消费者更好地了解商品及服务，具有更高的参考价值^[70]。

本文对于评论发布距今的时间跨度采用时间差的天数表达形式，即计算时间差后，统一转化为天数的形式，保证了单位的一致性。

3.2 要素量化

3.2.1 评论全面性与相关性量化

(1) 商品特征提取

本文借鉴已有文献^[50-51]提取商品特征的方法，即 Word2Vec 和聚类分析方法相结合的方法提取商品特征，为了保证词语的全面性，本文增加 TextRank 算法共同选取特征词语。

已有文献表明^[77]，名词和名词短语是商品特征对应的词性表述。根据 TF-IDF 筛选出重要词语作为词语库，词语库中词语满足上述词性要求时作为候选商品特征词语。并将 TextRank 算法所得的前 200 个词语经过人工筛选后，加入候选商品特征词语中。为了构建出反映上下文的词向量，本文使用 Word2Vec 学习在线评论的上下文，并根据上下文计算候选特征商品特征的词向量。由于 Word2Vec 是根据上下文学习得到的词向量，因此空间距离较近的词向量更有可能表述类似的商品特征。基于此，本文构建词向量后，对候选商品特征向量使用 KMeans 聚类算法，在聚类过程中，通过不断删除不属于商品特征的类别和调整类别个数，确定最终商品特征类别及其对应的商品特征词语。具体算法如表 3.1 所示。

表 3.1 商品特征提取

Table 3.1 Commodity feature extraction

Algorithm 1 商品特征提取

输入：评论文本

输出：候选特征分类结果

1. 初始化词语频次统计字典
2. 进行 tf-idf 计算，并按照阈值筛选候选特征
3. 从候选特征中筛选名词和动名词作为有效候选特征
4. 将评论文本写入“comment.txt”文件
5. for each comment:

 将词性标注的文本选取“n”，“vn”词性的词语，加入词语频次统计列表

6. 计算 `numpy.percentile(频次统计字典.values(), [25, 50, 75])`，取上四位数为阈值
-

表 3.1 (续)

Table 3.1 (Continuation)

Algorithm 1 商品特征提取

```

7. tr4w = TextRank4Keyword()
tr4w.analyze(text=text, lower=True, window=2)
for item in tr4w.get_keywords(200, word_min_len=2):
    有效候选特征.append(item.word)
7. 进行 Word2Vec 读入 comment.txt 并训练，不断调整参数，找到最优组合
8. 计算第四步中保留的词语词向量并进行 KMeans 聚类
9. 不断调整 KMeans 的 k 值，并删除不是商品特征的类别
10. 确定最终聚类个数
11. 输出分类结果
    
```

本文使用 Word2Vec 的 CBOW 模型构建词向量，通过使用评论数据集训练模型，并得到词向量。词语聚类是基于计算词向量之间的距离实现的。本文采用 KMeans 算法将距离相近的词语聚为一簇。KMeans 算法将词向量聚类的步骤如图 3.1 所示。

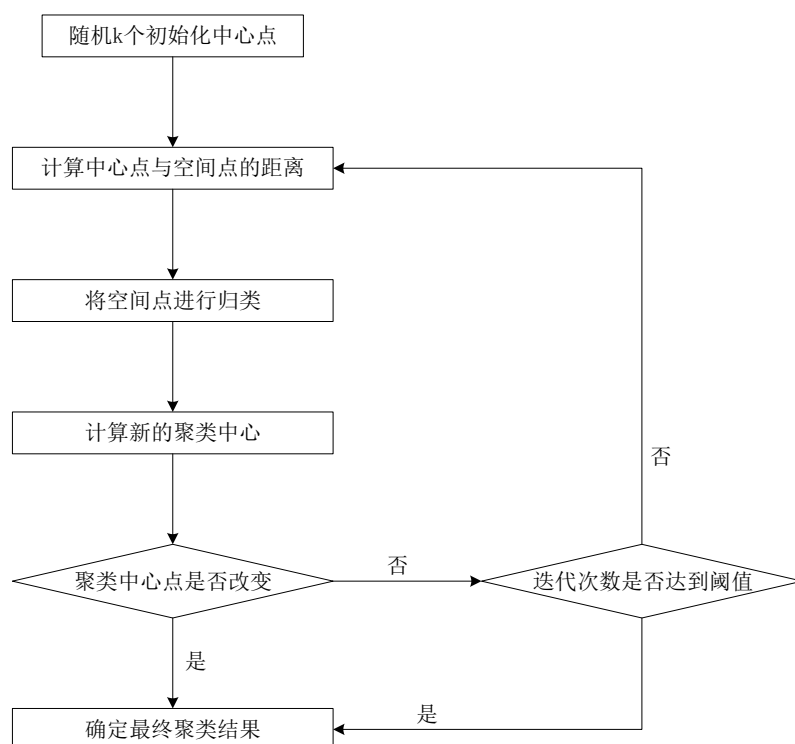


图 3.1 KMeans 聚类流程图

Figure 3.1 KMeans clustering flow chart

具体来说，聚词为簇的过程为：

1) 随机选取 k 个聚类中心，k 的个数取决于目标聚类数量，例如，研究目标是

聚为两类，则 $k=2$ ；

- 2) 计算空间点距离聚类中心的距离，并将其与距离最近的聚类中心归为一类。这里空间点与聚类中心点的距离按照公式(3.2)计算。

$$distance_i = ((x_{i1} - \sigma_{i1})^2 + (x_{i2} - \sigma_{i2})^2 + \dots + (x_{in} - \sigma_{in})^2)^{1/2} \quad (3.2)$$

其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ 表示空间样本点， $\sigma_i = (\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{in})$ 表示类中心点。

取点和类中心距离最小的类别为该点的类别，聚类后的点记为 x_i^k ，表示 x_i 被归为第 k 类；

- 3) 以各个类别中点的均值作为该类别的新聚类中心，即新的聚类中心 $\pi_k =$

$$(\sum_{i=1}^m x_{i1}^k / m, \sum_{i=1}^m x_{i2}^k / m, \dots, \sum_{i=1}^m x_{in}^k / m), m \text{ 表示类别 } k \text{ 中样本点的个数；}$$

- 4) 重复计算 2) 和 3)，直至聚类中心不再变化。

结合上述过程，可以看出，KMeans 的操作流程简单易懂，并且聚类结果具有较强的解释性。但是聚类个数 k 较难确定，需要根据人工进行不断删减聚类类别。

其次，为了使商品特征词库更加完善，本文结合 TextRank 算法，评论数据集中关键词进行提取，并按照词语词性进行筛选，最终对剩下的少量特征词语进行人工挑选，并将挑选出的商品特征词语加入商品特征词库。至此，完成商品特征词语的提取部分，商品特征词库已经较为完善和丰富。

(2) 商品特征重要性计算

评论的 C-R(Comprehensiveness and Relevance)值是综合评论所提及特征的丰富性以及所提及的特征的重要性两方面所得值。

计算该值的出发点在于 C-R 值的大小，可以反映评论中所提及商品特征的多少以及商品特征重要性的大小。因此在计算该值时，主要包括两点：即特征丰富性计算和特征重要性计算。本文假设每个词最多只可以反映一种商品特征。

对于评论中所提及的商品特征重要性：尽管不同消费者关注商品的特征维度是不同的，但是从总体上讲，是服从统计规律的^[65]。因此在计算特征重要性时，本文采用各个类别在用户评论中的提及频率作为商品特征重要性的衡量。

本文从词性及相似度两个方面评估潜在商品特征词语的质量，具体来说，采用计算词语相似性，并结合阈值挖掘未在商品特征列表中列出的商品特征词语。例如，“款式”和“样式”的相似度为 0.82，“面料”和“布料”相似度为 0.86。通过词语间相似度计算，可关联出相似的商品特征词语表达。为了避免某一个潜在特征词语被多个类别重复计数，本文会从大于阈值的相似度值中选取最大值所属的类别作为该潜在特征词所描述的商品特征类别。如若有多个相等的最大值，则任选一个即可。计算公式如公式(3.3)-(3.5)所示。具体算法如表 3.2 所示。

$$S_i = \max_{1 \leq i \leq c, 0 \leq j \leq k_i} \text{similarity}(\text{word}, x_{ij}) \quad (3.3)$$

$$f_i(S_i) = \begin{cases} 0 & S_i \leq threshold \\ 1 & S_i > threshold \end{cases} \quad (3.4)$$

$$importance_i = \sum_{word} f_i / \sum_i \sum_{word} f_i \quad (3.5)$$

上式中, *word*表示在线评论文本中分词后的词语; x_{ij} 表示商品特征词语; *c*表示特征词语类别; k_i 表示第*i*个商品特征类别所包含的词语; S_i 表示词语间相似度值; *threshold*表示在根据相似度可记为商品特征词语的相似度阈值; $f_i(S_i)$ 表示当 S_i 满足阈值要求, 第*i*类商品特征被提及; $importance_i$ 表示第*i*类商品特征对提及的总次数占所有商品特征类别提及总次数的比值。

表 3.2 特征重要性计算

Table 3.2 Feature importance calculation

Algorithm 2 特征重要性计算
输入: 经过词性标注的评论
输出: 各类别被提及次数统计结果
初始化字典 <code>category_count = {}</code>
对每条词性标注的评论(列表格式)进行处理:
for i in 每条词性标注后的评论:
新建比较池列表 <code>compare_max = []</code>
if 该词词性 in ["n", "vn"]:
for j in 商品候选特征:
if <code>similarity(i,j) > 0.75</code> :
<code>compare_max</code> 中加入[j 所在类别, <code>similarity(i,j)</code>]
判断 <code>compare_max</code> 中最大值及其对应的类别
词频计数 <code>category_count[i] = category_count.get(i,0)+1</code>
最终生成 <code>category_count</code> 即为输出结果

(3) C-R 值计算

对于评论信息中所提及的商品特征丰富性: 本文借用信息熵的公式, 以相加的方式体现评论信息中所提及商品特征的丰富性, 即相加项越多, 说明本条评论所提及的特征越多, 即越丰富, 如公式(3.6)所示。

$$C-R = \sum -P(s_i) \log P(s_i) \quad (3.6)$$

综合两方面的考虑, 避免了单纯考虑特征多样性和丰富程度而忽略了特征重要性的不同。与此同时, 这样做的好处是, 根据 C-R 值的大小, 可以判断出评论的全面性与相关性, 以及评论中所提商品特征的重要性。具体算法如表 3.3 所示。

表 3.3 C_R 值计算

Table 3.3 C_R value calculation

Algorithm 3 C_R 计算

输入：经过词性标注的评论文本，特征词类别权重

输出：C-R 值

1. 初始化 $sum = 0$

2. for i in 评论文本词性标注后的词语:

 for j in 各个类别特征:

 if $model.wv.similarity(i,j) > 0.7$ and i 词性 in ["n", "vn"]:

 将 i 归属为相似度最大的特征词所属类别

$sum +=$ 计算该类别权重 $\times \log(\text{该类别权重})$

3. 返回每条评论的 C-R 值

(4) 实例计算

以“衣服保暖性很好，物流很差”为例，进行评论全面性与评论相关性的计算。

首先，将评论文本按照标点符号拆分为若干子句，该评论文本拆为两个子句。

其次，在第一个子句中存在商品特征词语“保暖性”，第二个子句中存在商品特征词语“物流”。

再次，记“保暖性”的商品特征权重为 w_1 ，“物流”的商品特征权重为 w_2 。

最后评论信息的全面性与相关性得分为 $C-R = -w_1 \times \log(w_1) - w_2 \times \log(w_2)$ 。

3.2.2 评论情感强度量化

(1) 领域情感词典构建

本文使用 Word2Vec 结合 KMeans 选取情感种子词语。情感词语分为多种词性，但是主要词性是形容词(如“不错”)及部分名词(如“差劲”)。

首先对出现次数大于阈值的形容词词语聚类，形成第一批种子词语。其次，对出现次数大于某阈值的名词词语聚类，形成第二批种子词语，两次生词的种子词语合并得到正负向情感种子词语。

接下来基于情感种子词语扩充情感词典。扩充情感词典分为两部分，一部分是通过分别计算 n, vn, d, v, a, z, l, c 词性的词语与正、负情感词之间的相似度，选取相似度最大值所属的情感类别，并且最大值大于阈值的词语作为情感词扩充至情感词典。另一部分是商品特征词语中，部分词语既表示商品特征也是商品的情感表达，这类词语也应加入情感词典。

由于词语情感强度不同，本文基于大连理工大学的情感词汇本体^[78]，对其进行扩

充。扩充规则为根据已有词语的词义及情感表达强度，关联词义相近词语，将情感强度赋给词义相近词语，其余词语逐个人工定义情感词语强度。

(2) 情感强度计算

情感得分包括四个部分：即情感词强度、否定词个数、程度词个数、情感词所描述商品特征的权重。首先，扩充的情感词汇本体中各个词对应的强度值，作为情感基础得分。其次，本文采用 HowNet 否定词典和程度词词典用于修正情感强度得分。即在每个子句中，定位否定词与情感词，若情感词前存在否定词，则情感强度值为其相反数；其次，若该子句中存在程度词，如果情感值为正，则情感强度+2，反之，如果情感值为负，则情感强度-2。若该子句中存在多个商品特征词，则特征词共享情感词，则情感词权重为商品特征权重的平均值。故得分计算综合考虑了商品特征权重、程度词、否定词、情感强度等因素。

在定义领域情感词典之后，需要进行情感强度的计算。本文采用定义规则的方式进行情感强度的计算。

具体来说，情感强度计算各步骤为：

Step1: 按照标点符号将评论文本划分为多个子句，记子句个数为 q 。由于评论中往往包含多个子句，为了后续计算情感词语对应的商品特征权重以修正情感词语，需要定位情感词所属的商品特征词，因此，当把评论分为若干子句时，情感词所属的商品特征较为容易对应；

Step2: 确定每个子句中是否含有商品特征词语。即看每个子句中分词后短语是否有和商品特征词库中词义相近的词语，若存在这样的词语，则认为该子句中有商品特征词语，否则，认为该子句中没有商品特征词语；

Step3: 若当前子句中没有商品特征词语，则确定该子句中是否有情感词语，若存在情感词语，则将该情感词语加入情感词的临时列表中，转至 **Step4**。即当子句中没有商品特征词语而含有商品情感词语时，将情感词语先暂存，以备后续处理。

若当前子句中有商品特征词语，则确定该子句中商品特征词语的个数 n ，并计算商品特征词语权重的均值，作为该子句情感的权重值，如公式(3.7)所示。

$$\bar{w} = \sum_{j=1}^n w_j / n \quad (3.7)$$

其中 w_j 表示子句中每个商品特征词语的权重值；

该权重的计算是为了对情感值进行修正，以便准确计算出评论文本的情感倾向。例如：对于“衣服保暖性好，物流有点慢”而言，如果“保暖性”对应的商品特征权重大于“物流”对应的商品特征权重，则情感倾向为正；反之，情感倾向为负。

Step4: 若 **Step3** 中存在商品特征词语，则将该子句的情感词 s_i 作为商品特征的描述词，其中， s_i 表示子句中第 i 个情感词， $i = 1, 2, 3, \dots, m$ ， m 表示情感词个数。观察情

感词临时列表中是否为空，若不为空，设情感词列表长度为 p ，则将情感词的临时列表的 p 个词语加入该子句的商品特征词语对应的情感词中，并清空情感词的临时列表。

Step5: 确定情感词 s_i 位置前，是否存在否定词以及否定词个数 f_i 、情感强度词以及情感强度词个数 v_i ，且每个情感程度词的值为 2，计算子句所有情感词对应的总得分，如公式(3.8)所示。

$$senti = \sum_{i=1}^{m+p} (-1)^{f_i} s_i + v_i \times 2 \quad (3.8)$$

Step6: 将 Step3 中的商品特征权重作为系数乘以 $senti$ ，如公式(3.9)所示。

$$score_{part} = \bar{w} \times senti \quad (3.9)$$

Step7: 将情感得分按照子句个数相加，得到评论文本情感总得分，如公式(3.10)所示。

$$score_{all} = \sum_{part=1}^q score_{part} \quad (3.10)$$

本文采用的情感强度的计算流程如图 3.2 所示。

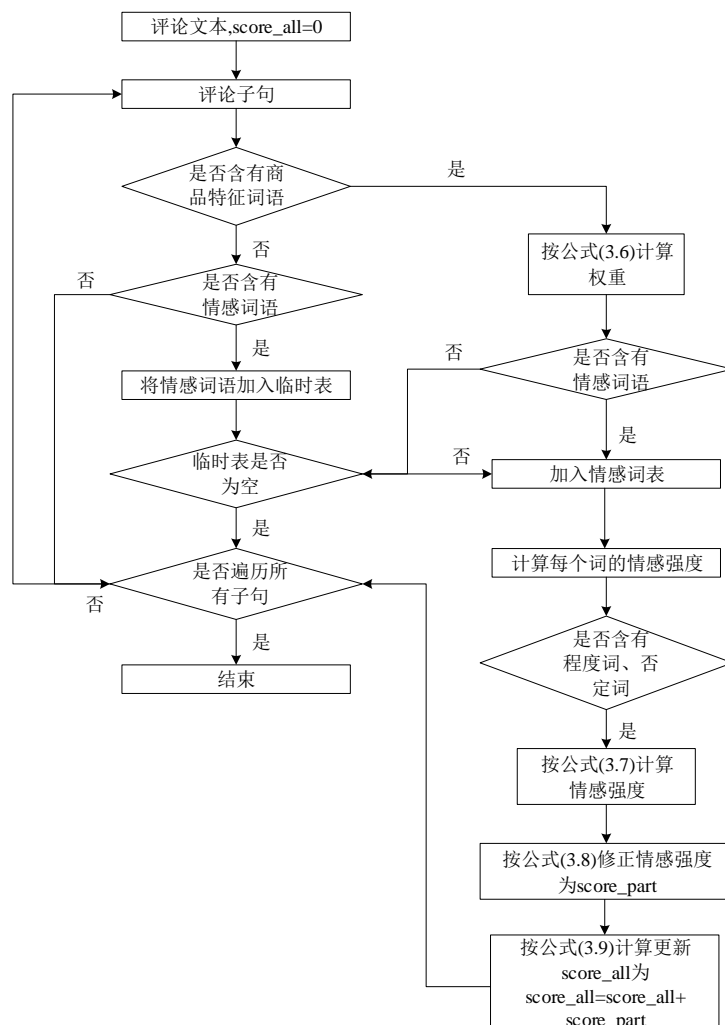


图 3.2 情感强度计算流程图

Figure 3.2 Sentiment intensity calculation flow chart

(3) 实例计算

以评论文本“衣服保暖性很好，物流很差”为例进行情感强度计算：

Step 1: 将评论文本按照标点符号划分为若干子句，即评论文本划分为“衣服保暖性很好”，“物流很差”，这里 $q = 2$ ，令第一个子句为子句 1，第二个子句为子句 2。

Step 2: 子句 1 中含有商品特征词语“保暖性”，子句 2 中含有商品特征词语“物流”，对于每个子句 n 均为 1。

Step 3: 子句 1 的特征权重为“保暖性”对应的权重 w_1 ，子句 2 的特征权重为“物流”对应的权重 w_2 。

Step 4: 子句 1 的情感词语为“好”，子句 2 的情感词语为“差”，对于每个子句， $i = 1$ 。

Step 5: 子句 1 和子句 2 的情感词语前均不存在否定词，即 $f_i = 0$ ，存在程度词“很”，即 $v_i = 1$ 。记“好”的情感强度值为 l_1 ，“差”的情感强度值为 l_2 ，第一个子句的 $senti_1 = (-1)^0 \times l_1 + v_1 \times 2 = l_1 + 2$ ，第二个子句的 $senti_2 = (-1)^0 \times l_2 + v_2 \times 2 = l_2 + 2$ 。

Step 6: 对于第一个子句而言， $score_{part1} = w_1 \times senti_1$ ，对于第二个子句而言， $score_{part2} = w_2 \times senti_2$ 。

Step 7: 对于评论文本整体而言， $score_{all} = score_{part1} + score_{part2} = w_1 \times senti_1 + w_2 \times senti_2$ 。

综合上述步骤可知，由于两个子句中，否定词个数和程度词个数均相等，因此，最终评论文本“衣服保暖性很好，物流很差”的情感倾向由情感词的情感强度值和特征权重共同决定。

3.2.3 其他要素量化

其他要素包括：在线评论长度、用户星级打分、时间跨度等三个要素。由于 TOPSIS 排序算法对分类变量的适用性较差。因此，本文量化这三个要素为连续型变量。量化情况如表 3.4 所示：

表 3.4 要素量化标准

Table 3.4 Factor quantization standard

要素	评分方式	得分值
评论长度	评论文本的有效长度 ^[29]	去除停止词、标点符号等剩余 评论文本长度
用户星级打分	数据获取时的用户打分信息	按用户打分原值计算 由于评论数据是爬取而得,所以 取爬取结束日期的下一天的时间
评论时间跨度	评论发表时间距今的时间差	作为截止时间,计算各个评论 日期距今的时间差,并统一转化 为天数。

3.3 本章小结

本章选取部分指标用于评价在线评论有用性,并逐一对各个指标的量化方法和逻辑进行了介绍,其中,着重介绍了在线评论全面性与相关性指标的量化以及情感强度量化。本章的创新点在于,区别于以往研究指标量化方法,本文使用情感分析定义规则代替前人情感词个数指标、使用评论全面性与相关性代替评论特征词的个数指标,较以往方法更合理、准确。本章各个指标的量化方法及其结果对后续研究而言具有基础性作用。

第四章 在线商品评价模型构建研究

在线商品评价模型构建研究包括两个主要内容，一是模糊评价矩阵权重确定，二是结合模糊综合评价法进行商品评价。对于模糊评价矩阵权重是指在线评论有用性的量化值，该值作为修正系数引入模糊评价矩阵构建过程中，调节不同的在线评论所表达的情感值，以表示不同评论其情感值的参考价值。

4.1 商品评价指标权重确定

鉴于在线评论的有用性是不同的，即在线评论的可参考程度是不同的。因此，对于在线评论所表达的情感强度的参考价值应当有所区分。对于每一条评论所述内容应当赋予可参考价值的权重，使用修正后的情感强度评价商品更为科学、合理。

考虑这种想法的原因在于：1、将所有的评论数据直接纳入数据源有些不妥。显然有些评论参考价值并不高，如时间较为久远的评论等，这些评论会随着商品的批次不同、流行迭代周期不同而导致评论有效性降低，因此其评论所表达情感的参考价值应当加以区分；2、由于评价商品的最终目的是为了提升商品的转化率，而用户在初次接触商品时，其了解商品的主要来源就是用户的评论，用户评论的有用性与否会影响用户对商品的认知，进而影响其购买决策，因此在对商品进行评价时，有必要将评论有用性考虑进去。

商品评价指标权重是指在构建模糊评价矩阵的过程中，商品特征对应的情感值作为矩阵的构成要素时，情感值将会乘以在线评论有用性的量化值，以表示当前特征对应的情感强度的可参考程度。即这是将有用性得分作为修正系数引入模糊评价矩阵构建的过程。从研究内容上看，应当首先计算在线评论有用性，再对构建模糊评价矩阵对商品进行模糊评价。因此，本部分将围绕在线评论有用性量化展开。

4.1.1 评论有用性指标权重赋值

本文采用层次分析法(AHP)计算指标权重。层次分析法由美国运筹学家 A.L.Saaty 于 20 世纪 70 年代初提出的一种多方案、多目标的决策方法，该方法模仿人的决策思维过程，将定性与定量相结合进行决策的方法，不仅实现了对非定量事件科学的定性分析，而且满足了精确的定量分析，是一种综合分析的方法。将层次分析法用于有用性评价指标权重的计算中，以确定各个指标对于在线评论有用性的重要程度。

层次分析法具体使用流程如下：

(1) 确定和分解目标层

针对某一个复杂的研究目标，可进行目标层-准则层-方案层逐层分解，如图 4.1 所示。其中准则层可以包括若干子因素，子因素可以继续向下拆解，共同构成方案层。

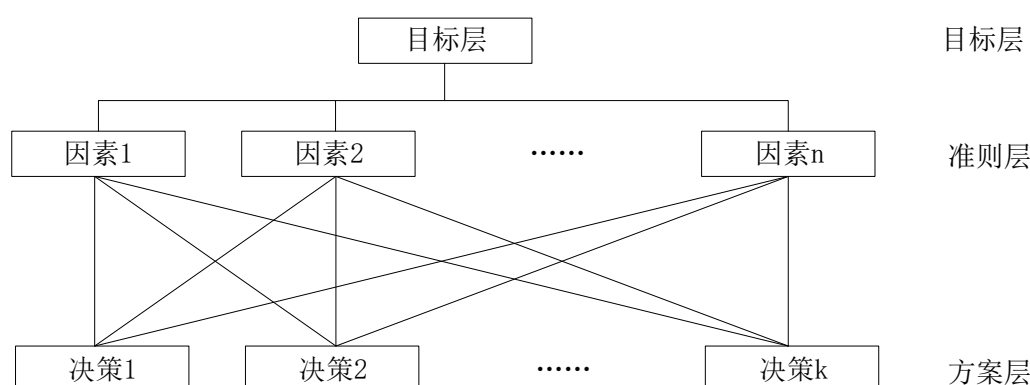


图 4.1 层次分析法目标分解图

Figure 4.1 Analytic hierarchy process target decomposition map

(2) 构建判断矩阵

记判断矩阵为 $Q = \begin{bmatrix} 1 & \cdots & q_{1l} \\ \vdots & \ddots & \vdots \\ q_{l1} & \cdots & 1 \end{bmatrix}$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, l$, l 为指标个数，其中

q_{ij} 表示指标 i 对 j 的相对重要程度。相对重要程度一般采用 1~9 标度法完成，各个数值表示如表 4.1 所示。判断矩阵对角线元素为 1，对角线之上元素与其下方元素成倒数关系，即 $q_{1i} \times q_{i1} = 1$ 。

表 4.1 各标度赋值含义

Table 4.1 Meaning of each scale assignment

标度赋值	定义	含义
1	同等重要	两个指标，同等重要
3	稍微重要	两个指标，前者比后者，稍重要
5	相当重要	两个指标，前者比后者，明显重要
7	明显重要	两个指标，前者比后者，强烈重要
9	绝对重要	两个指标，前者比后者，极端重要
2, 4, 6, 8		两个指标，重要度介于上述标度两者之间
倒数 1/n		两个指标，比之上述标度重要性相反

(3) 计算特征向量及特征值

得到上述判断矩阵之后, 判断矩阵 Q 按列进行归一化, 归一化后为 $Q' = \begin{bmatrix} q'_{11} & \cdots & q'_{1l} \\ \vdots & \ddots & \vdots \\ q'_{l1} & \cdots & q'_{ll} \end{bmatrix}$, 其中, $q'_{ij} = q_{ij} / \sum_{i=1}^l q_{ij}$ 。

(4) 对 Q' 按行求和得到向量 v , 即 $v = \left(\sum_{j=1}^l q'_{1j}, \cdots, \sum_{j=1}^l q'_{lj} \right)$, 将向量 v 进行归一化得到特征向量 v' 。

(5) 计算最大特征值, 即 $\lambda_{\max} = (\sum_{i=1}^l (Q'v')_i / v'_i) / l$ 。

(6) 一致性检验

为了确保分析的准确性, 判断矩阵需要通过一致性检验。一致性检验是否通过取决于 CR 系数, CR 计算公式如公式(4.1)所示。

$$CR = CI / RI \quad (4.1)$$

其中 $CI = (\lambda_{\max} - l) / (l - 1)$, RI 取值与指标个数有关, 具体取值如表 4.2 所示:

表 4.2 RI 取值表

Table 4.2 RI value table

l	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

当 $CR < 0.1$ 时, 通过一致性检验, 即认为矩阵的一致性是可以接受的; 当 $CR \geq 0.1$ 时, 应对矩阵进行修改。最终通过一致性判断后, 得到权重指标。

4.1.2 加权 TOPSIS 排序算法

加权 TOPSIS 排序算法, 其基本思想是将原始评分矩阵进行量化处理, 将得到的标准矩阵与指标权重结合, 即得到加权决策矩阵, 并根据正负理想解, 计算待评判对象到正负理想解的距离, 根据相对贴近程度即距离最优和最劣向量的相对值, 对指标因素进行排序^[79]。

加权 TOPSIS 排序算法具体步骤如下:

(1) 指标属性矩阵同向化

记原始标准属性矩阵为 $X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$, 其中 x_{ij} 表示第 i 个样本的第 j 个指标的

属性值, $i = 1, 2, \dots, m$, m 表示样本个数, $j = 1, 2, \dots, n$, n 表示指标个数。由于加权 TOPSIS 算法的目的在于根据距离尺度进行排序, 因此当指标维度中部分指标越大越好, 而另一部分指标越小越好时, 会造成尺度混乱。故需要进行指标属性矩阵同向化

处理。记同向化处理后的属性矩阵为 $X' = \begin{bmatrix} x'_{11} & \cdots & x'_{1n} \\ \vdots & \ddots & \vdots \\ x'_{m1} & \cdots & x'_{mn} \end{bmatrix}$ ，其中 x_{ij} 表示第 i 个样本的第 j 个指标同向化后的属性值。一般同向化处理向正向调整，即指标值越大越好。

(2) 构造加权规范矩阵

即对属性进行向量规范化，每一列元素除以当前列向量的范数，得到归一化处理的标准化矩阵 Z ：

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & \ddots & \vdots \\ z_{m1} & \cdots & z_{mn} \end{bmatrix}, \text{ 其中 } z_{ij} = x'_{ij} / (\sum_{i=1}^m x'_{ij}{}^2)^{1/2}, i = 1, 2, \dots, m, j = 1, 2, \dots, n,$$

m 为样本个数， n 为指标个数。

(3) 计算最优方案与最劣方案

记 Z^+ 表示最优方案， Z^- 表示最劣方案。当指标属性矩阵同向化处理按照正向调整时，由于此时矩阵每一列中，值越大越好，因此， Z^+ 由 Z 的列最大值组成。 $Z^+ = (\max(z_{11}, z_{21}, \dots, z_{m1}), \max(z_{12}, z_{22}, \dots, z_{m2}), \dots, \max(z_{1n}, z_{2n}, \dots, z_{mn}))$ ； Z^- 由 Z 的列最小值组成。 $Z^- = (\min(z_{11}, z_{21}, \dots, z_{m1}), \min(z_{12}, z_{22}, \dots, z_{m2}), \dots, \min(z_{1n}, z_{2n}, \dots, z_{mn}))$ 。

(4) 计算待排序属性向量与最优方案和最劣方案的贴近程度

贴近程度按照向量间距离来衡量。记 d_i^+ ， d_i^- 分别表示待评价对象距最优方案和最劣方案的距离，计算公式如公式(4.2)，(4.3)所示。

$$d_i^+ = (\sum_{j=1}^n w_j (Z_j^+ - z_{ij})^2)^{1/2} \quad (4.2)$$

$$d_i^- = (\sum_{j=1}^n w_j (Z_j^- - z_{ij})^2)^{1/2} \quad (4.3)$$

其中， w_j 表示第 j 个指标的权重。由于 d_i^+ 、 d_i^- 表示距离最优、最劣方案的距离，因此 d_i^+ 越小越好， d_i^- 越大越好。

(5) 根据贴近程度进行排序

相对贴近程度由待评价对象到最优方案距离 d_i^+ 和到最劣方案距离 d_i^- 共同决定。确定各个评价对象与最优值与最劣值的相对贴近程度 f_i ，计算公式如公式(4.4)所示，相对贴近程度介于 0~1 之间。

$$f_i = d_i^- / (d_i^+ + d_i^-) \quad (4.4)$$

当 f_i 越接近 1 时，说明 d_i^+ 趋近于 0，即待评价对象距最优方案越近，各指标表现越好，排名越靠前。当 f_i 越接近 0 时，说明 d_i^- 趋近于 0，即待评价对象距最劣方案越近，各指标表现越差，排名越靠后。

4.1.3 评论有用性量化

本文在线评论有用性量化过程如图 4.2 所示。

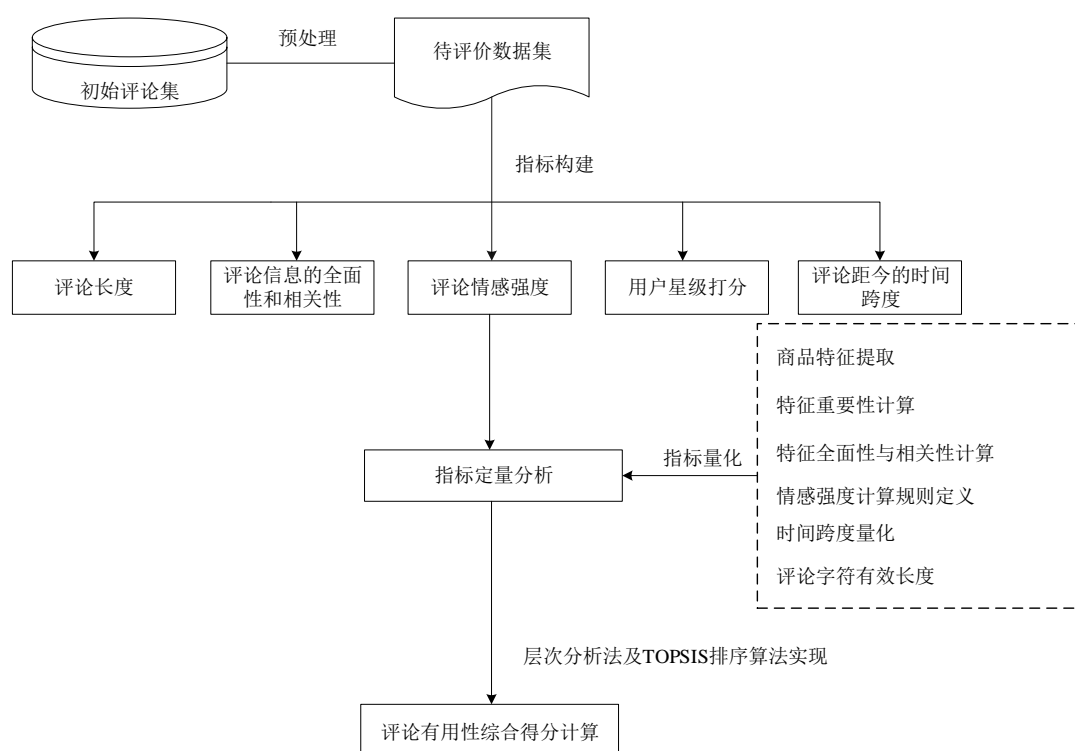


图 4.2 在线评论有用性评价过程模型

Figure 4.2 Online review helpfulness evaluation process model

本文选取评论信息的全面性与相关性、评论情感强度、评论文本长度、评论时间跨度、用户星级打分等五个指标进行在线评论有用性的量化。其中，有用性评价属性矩阵为 $X_{m \times 5}$ ，其中 m 为评论数据集样本个数，将构造的属性矩阵按照第三章所示步骤，对在线评论有用性的各个指标要素进行量化。

首先，同向化属性矩阵。通过文献综述中对各个指标的梳理，评论信息的全面性与相关性、评论情感强度、评论文本长度等四个指标与在线评论有用性呈正相关，即量化值越大表示有用性越高，而对于评论时间跨度表示发布时间距今的天数，根据文献可知，距今时间越短即时间跨度越小，有用性越高。此时，各个指标不符合矩阵同向的要求，因此，应当对指标做变换以使得矩阵具有同向性。具体做法是：将评论信息的全面性与相关性、评论情感强度、评论文本长度等四个指标的量化值保持不变，对于评论发布时间距今跨度指标，本文采用取倒数的方法将其转变为与有用性正向变化的指标。故此时的属性矩阵已经满足同向化的要求。

其次，计算最优、最劣方案。由于属性矩阵已经满足了同向化的要求，因此，矩阵的最优方案和最劣方案应当分别为属性矩阵的最大值集合和最小值集合。

最后，根据每个评价对象距最优、最劣方案的距离，计算相对贴近程度。并根据待评价方案的相对贴近程度，对评论数据集进行排序，排序结果即为在线评论有用性的排序结果。

加权 TOPSIS 算法表述如表 4.3 所示。

表 4.3 TOPSIS 排序算法

Table 4.3 TOPSIS sorting algorithm

Algorithm4 TOPSIS 排序算法
输入：各个特征值计算完成的 dataframe 格式数据，指标权重
输出：评论有用性排序结果
1. 将数据按照方向一致转化，如方向不一致按照 $1/x_i$ 转换
2. 数据归一化处理 $\text{topsis_data} / \text{np.sqrt}((\text{topsis_data} ** 2).sum(\text{axis}=0))$
3. 找出最优方案与最劣方案
<code>Standard_Data = pd.DataFrame([topsis_data.min(),topsis_data.max()],index=["负理想解","正理想解"])</code>
4. 计算评论指标与最优、最劣方案之间的距离
<code>d1 = np.sqrt((((topsis_data - Standard_Data.loc['负理想解']) ** 2 * weight).sum(axis=1))</code>
<code>d2 = np.sqrt((((topsis_data - Standard_Data.loc['正理想解']) ** 2 * weight).sum(axis=1))</code>
5. 计算各个评论的综合得分
<code>f = d1/(d1 + d2)</code>
6. 按照综合得分计算排序

4.2 商品评价模型构建

本文使用模糊综合评价法搭建商品评价模型。模糊综合评价法是一种将模糊数学理论引入综合评价的一种方法。即用模糊数学的方法对受到多因素制约的事物和对象做出相对客观的总体评价^[80]，其基本思想是根据多目标评价问题的性质和总目标，把问题本身按层次进行分解。模糊综合评价法步骤如图 4.3 所示：

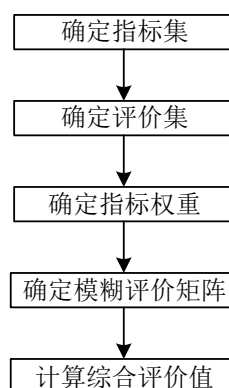


图 4.3 模糊综合评价法步骤

Figure 4.3 Fuzzy comprehensive evaluation flow chart

(1) 确定指标集

记商品评价指标集合为 $U = \{u_1, u_2, \dots, u_m\}$, 其中 u_i 表示第 i 个评价指标, $i = 1, 2, \dots, m$, m 为评价指标个数, m 由评价体系和评价目标共同决定。集合中 u_i 表示一级评价指标, 即表示评价指标按照属性划分为的某一类。一级指标仍可以继续下沉为二级指标, 以此类推。

由于本文按照商品特征对应的情感值进行商品评价, 因此, 本文的指标集为商品特征类别。

(2) 确定评价集

记评价集为 $L = \{l_1, l_2, \dots, l_n\}$, 其中 l_j 表示第 j 个评价结果, $j = 1, 2, \dots, n$, n 为评价等级个数, 一般 n 的范围是 3~5。

本文将商品评价等级分为 5 个等级, 即差、较差、中等、较优、优等五个级别。

(3) 确定指标权重

记指标权重为 $W = [w_1, w_2, \dots, w_p]$, 其中 w_i 表示第 i 个指标对应的权重, 其值表示指标重要程度, $i = 1, 2, \dots, p$, p 为赋权指标个数。当评价指标体系只有一级指标时, $p = m$, m 为指标集合元素个数。指标权重的计算方法, 一般有层次分析法、专家赋值法、德尔菲法等主观判断方法。

由于本文的指标集采用的是商品特征的类别, 因此, 指标权重是指商品特征类别的权重, 这和第三章计算评论全面性与相关性时使用的商品特征重要性权重是一致的, 因此, 本文采用商品特征的重要性作为指标集权重。

(4) 确定各个评价等级得分

将(3)中所划分的评价等级赋予一定的得分, 以更好地计算和区分商品综合得分。记评价等级得分为 $E = [e_1, e_2, \dots, e_p]$, 本文将按照情感值的分布情况进行评价等级得分的赋值。

(5) 确定模糊评价矩阵

记模糊评价矩阵为 $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pn} \end{bmatrix}$, 其中 a_{ij} 为第 i 个指标对第 j 个评价等级隶属度, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, n$, p 为赋权指标个数, n 为评价等级个数。

本文将各个商品特征对应的情感强度划分为五个等级, 并使用各个等级上评论数的占比作模糊评价矩阵的元素。但是由于本文引入了在线评论有用性的计算, 因此, 本文将优先对每条评论中提及的情感值乘以该评论的有用性量化值, 根据修正后的情感值分布进行等级划分, 并计算模糊评价矩阵中的各个元素。

(6) 计算综合评价值

模糊评价结果向量 $B = (b_1, b_2, \dots, b_n)$ 通过合成算子合成指标权重 W 和模糊评价矩阵 A 。常用的合成算子有: $M(\wedge, \vee)$, $M(\cdot, \vee)$, $M(\wedge, \oplus)$, $M(\cdot, \oplus)$ 。鉴于 $M(\cdot, \oplus)$ 体现权数明显、综合程度强, 利用模糊评价矩阵 A 充分的特点, 本文选择 $M(\cdot, \oplus)$ 合成算子合成 W 和 A , 如公式(4.5)所示。

$$B = WAE = [w_1, w_2, \dots, w_p] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pn} \end{bmatrix} [e_1, e_2, \dots, e_p]^T = (b_1, b_2, \dots, b_n) \quad (4.5)$$

4.3 本章小结

本文选取商品特征类别以及有用性指标, 结合模糊综合评价法构建商品评价模型。该部分的创新点在于将评论有用性作为一个重要因素在商品评价模型中加以考虑。由于在线评论是用户了解商品的一个重要途径, 且阅读有用评论有利于用户更好地感知商品, 而高有用性的评论有利于提高商品转化率^[6], 因此我们引入该因素是十分必要且重要的, 加入在线评论有用性因素有利于完善商品评价模型, 使模糊综合评价体系更好地贴近研究目标。

第五章 实证研究

本章基于前四章提出的研究目标、研究内容以及研究方法，将其应用到真实数据集上进行实证分析，以验证本文提出方法的有效性、可行性以及合理性。本章的数据集来源于京东(www.jd.com)电商平台，通过对在线评论的爬取、预处理等工作，结合情感分析、商品特征提取等方法量化评论有用性指标，并使用随机森林算法验证了本文情感分析方法的优劣。通过层次分析法和 TOPSIS 排序算法进行在线评论有用性计算，并将该量化值应用于模糊评价矩阵构建，通过模糊综合评价计算商品的得分。此外，本章构建了测试数据集，旨在说明模糊综合评价过程中，考虑引入在线评论有用性后对商品评价产生的差异，以证明本文在商品评价过程中考虑在线评论有用性的有效性及合理性。

5.1 数据获取与预处理

5.1.1 数据获取

本文实验数据来自于京东(www.jd.com)电商平台以“羽绒服”为搜索词所得的搜索页面商品的用户评论，取搜索页的前两页作为实验数据。本文使用 python 编写网络爬虫代码实现评论及其相关数据的爬取。部分数据源展示如附录 A 所示。经过页面解析，每页共有 60 件商品，前 30 件商品与后 30 件商品所在的 url 不同，因此本文采用评论爬取策略如下：

首先，分别请求每个搜索页面中前 30 件和后 30 件商品所在页面的 url，以获取商品 id，该部分代码如表 5.1 所示。其中前 30 件商品的 url 在网页源码中可以获取，后 30 件商品的 url 通过动态加载生成，使用谷歌浏览器的抓包工具将后 30 件商品 id 所在的 url 找到。通过请求上述 url，同时从返回的网页中结合 xpath 表达式和正则表达式获取相应页面的商品 id。同时观察商品 id 所在页面 url 的规律，构造出后续搜索页面的 url。

表 5.1 商品 id 获取

Table 5.1 Commodity id acquisition

商品 id 获取代码

```
import requests

import re

from lxml import etree

import time

def productid_Top30(num = 1): #前 30 个 url

    url="https://search.jd.com/Search?keyword=%E7%BE%BD%E7%BB%92%E6%9C%8D%E5%A5%B3&enc=utf-8&qrst=1&rt=1&stop=1&vt=2&suggest=1.his.0.0&psort=4&page={ }&s=1&click=0".format(str(num))

    response = requests.get(url, headers=headers).content.decode()

    html = etree.HTML(response)

    product_id = html.xpath('//*[@id="J_goodsList"]/ul/li/div/div[4]/a/@href')

    return [re.findall("//item.jd.com/(\\d+).html", i) for i in product_id]

def productid_Last30(num = 2): #后 30 个 url

    head = {"scheme": "https", "accept": "/*/*", "accept-encoding": "gzip, deflate, br",

            "referer": "https://search.jd.com/Search?keyword=%E7%BE%BD%E7%BB%92%E6%9C%8D%E5%A5%B3&enc=utf-8&qrst=1&rt=1&stop=1&vt=2&suggest=1.his.0.0&psort=4&click=0",

            "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/69.0.3497.92 Safari/537.36"}

    url="https://search.jd.com/s_new.php?keyword=%E7%BE%BD%E7%BB%92%E6%9C%8D%E5%A5%B3&enc=utf-8&qrst=1&rt=1&stop=1&vt=2&suggest=1.his.0.0&psort=4&page={ }&s=31&scrolling=y&log_id={ }".format(str(num), time.time())

    response = requests.get(url, headers=head).content.decode()

    html = etree.HTML(response)

    product_id = html.xpath("//div[@class='p-name p-name-type-2']/a/@href")

    return [re.findall("//item.jd.com/(\\d+).html", i) for i in product_id]

def final_productid(num1,num2):

    full_product_id = productid_Top30(num1) + productid_Last30(num2)

    return full_product_id
```

其次，构建 IP 池。具体代码实现如表 5.2 所示。由于爬取的数据量略大，因此本文采用构建 IP 代理池的方式缓解被京东网站反爬的问题，具体做法是：从 <https://www.xicidaili.com/> 网站上获取 IP，如图 5.1 所示。其页面内容中涉及 IP 的存活时间信息。从 IP 稳定性的角度考虑，本文选取存活时间大于 1 分钟且类型是 HTTP 或者 HTTPS 的 IP 进入本文的 IP 池。

表 5.2 IP 池构建

Table 5.2 IP pool construction

IP 池构建代码

```
import requests
import random
from lxml import etree
def IP():
    IP_url = "https://www.xicidaili.com/"
    response = requests.get(IP_url, headers=headers).content.decode()
    html = etree.HTML(response)
    ip_pre = html.xpath('//*[@id="ip_list"]/tr/td[6]/text()')
    ip_back = html.xpath('//*[@id="ip_list"]/tr/td[2]/text()')
    maintenance_time = html.xpath('//*[@id="ip_list"]/tr/td[7]/text()')
    IP = []
    for i in zip(ip_pre, ip_back, maintenance_time):
        if str(i[2]) != "1 分钟" and str(i[0]) != "socks4/5":
            IP.append(i[:2])
    IP1 = random.choice(IP)
    IP2 = random.choice(IP)
    IP_ava = {}
    IP_ava[IP1[0]] = IP1[1]
    IP_ava[IP2[0]] = IP2[1]
    print(IP_ava)
    return IP_ava
```

国内高匿代理IP							更多
国家	代理IP地址	端口	服务器地址	是否匿名	类型	存活时间	验证时间
	163.204.246.33	9999	广东	高匿	HTTP	1分钟	28分钟前
	27.43.191.252	9999	广东梅州	高匿	HTTP	1分钟	28分钟前
	163.204.241.194	9999	广东	高匿	HTTP	1分钟	28分钟前
	112.85.168.48	9999	江苏南通	高匿	HTTPS	85天	28分钟前
	120.83.106.80	9999	广东揭阳	高匿	HTTPS	87天	28分钟前

图 5.1 IP 来源网页

Figure 5.1 IP source page

最后，构建评论页面 url。在线评论页面仍为动态加载，因此，使用谷歌抓包工具获取在线评论的 url，并且发现，评论所在网页的实质格式是 json。通过对在线评论 url 的分析，发现其构成与商品 id 和在线评论页数有关，因此，本部分需要使用当前爬取的商品 id 以及在线评论页数进行 url 的拼接。在线评论页数的获取如表 5.3 所示。本文采用将 json 格式转化为字典格式并按照键-值取出相应字段的方式，取出本文的需要的字段，实现代码如表 5.4 所示；同时，将数据存储至 MongoDB 数据库。

表 5.3 评论页数获取

Table 5.3 Comment page number acquisition

评论页数获取代码

```
import requests
import json
import time
from pymongo import MongoClient

def num_of_page(productid): #page_number 表示评论的第几页
    url="https://sclub.jd.com/comment/productPageComments.action?productId=%s&score=0&sortType=5&page=0&pageSize=10&isShadowSku=0&fold=1"%(productid)
    try:
        response = requests.get(url, headers=headers).content.decode("gbk")
        response_dict = json.loads(response)
        page_number=int(int(response_dict["productCommentSummary"]["commentCount"])/10) + 1
    except:
        page_number = 0
    return page_number
```

表 5.4 评论字段获取

Table 5.4 Comment field acquisition

评论字段获取代码

```
def data_crawl(productid):
    ip = IP()
    page_num = 0
    page_number = num_of_page(productid)
    if page_number > 100:
        page_number = 100
    while page_num <= page_number:
        url="https://sclub.jd.com/comment/productPageComments.action?productId=%s&score=0
        &sortType=5&page=%s&pageSize=10&isShadowSku=0&fold=1" %
        (productid,str(page_num))
        response=requests.get(url,headers=headers,proxies=ip).content.decode("gbk").encode("utf
        -8")
        response_dict = json.loads(response)
        comment_count=response_dict["productCommentSummary"]["commentCount"]
        if response_dict["comments"]:
            for i in response_dict["comments"]:
                if len(i["content"]) == 0 and len(i["creationTime"]) == 0:
                    break
                nickname = i["nickname"]
                creationTime = i["creationTime"]
                isMobile = i["showOrderComment"]["isMobile"]
                content = i["content"]
                productColor = i["productColor"]
                productSize = i["productSize"]
                score = i["score"]
                collection.insert({"productid":productid,"nickname":nickname,"creationTime":cre
                ationTime,"isMobile":isMobile,"content":content,"productColor":productColor,"productSize":product
                Size,"score":score,"comment_count":comment_count})
            page_num += 1
            time.sleep(4)
```

最终共爬取 38049 条数据，涉及 60 个商品 id。本次采集的数据字段名称包括：商品 id、用户昵称、用户评论时间、品牌名称、商品颜色、商品尺码、用户评分、用户评论等八个字段。存储至 MongoDB 后的数据样式如下所示：

```
{"commodityid":"3313215","nickname":"j***8","creationTime":"2019-01-16
23:02:07","brand":"骆驼 (CAMEL)","commodityColor":"藏蓝男",
"commoditiesize":"L","score":5,"comment":"质量还不错，价格便宜，穿上很合身。"}
最终通过主程序运行，并实现评论数据的爬取与保存，主程序如表 5.5 所示。
```

表 5.5 主程序代码

Table 5.5 Main program code

主程序
<pre>if __name__ == "__main__": num1 = 1 #num 表示商品列表第几页 num2 = 2 client = MongoClient(ip, port) db = client.cloth collection = db.yurongfu while num1 <= 199 and num2 <=200: productid = final_productid(num1,num2) print(productid) for i in productid: data_crawl(i) num1 += 2 num2 += 2</pre>

5.1.2 数据预处理

数据预处理包括查看重复评论数据、去除评论中出现的数字和英文字符、分词、去除停用词及词性标注。

首先，在京东平台上，评论大多以匿名的身份发布。同时，经过对评论的分析，本文认为重复评论主要包括两种：匿名昵称不同但评论相同、匿名昵称相同且评论相同且评论时间不同。对于第一种，本文需要去除；对于第二种情况，由于无法确定匿名相同是否代表用户相同，因此本文对这类评论未去除。最终，第一种重复评论共去除 7411 条，剩余 30638 条在线评论。

其次，通过对评论文本的观察分析，部分评论中涉及数字及英文字符，这些字符

大多涉及价格、折扣、身高、体重、码数等，这些信息通过所在评论的其他描述可以提炼这些信息，因此，本文通过正则表达式将其去除。

最后，使用 python 的 jieba 分词器完成分词工作。并使用哈工大停用词表，在分词过程中直接去除停用词。与此同时，进行词性标注。

5.2 在线评论有用性计算

5.2.1 指标量化

(1) 评论信息的全面性与相关性量化

由于“轻薄”等类似形容词词语也是羽绒服商品特征的反映，因此结合 python 词性标注结果，本文将商品特征词性归结为：“n”，“vn”(如“服务”等)，“a”。按照 3.2.1 节所述的方法，对商品特征进行提取以及 Word2Vec 模型训练，通过提取的商品特征词语结合词义相似性，对评论信息的全面性与相关性进行计算，其中提取的商品特征词如表 5.6 所示。

表 5.6 商品特征词表

Table 5.6 Commodity feature words table

类别	商品特征词
上身效果	款式，颜色，版型，时尚，……
物流	物流，发货，速度
尺寸	尺码，标准，码数，尺寸
价格	价钱，价位，实惠，优惠，价格
质量	线头，异味，问题，瑕疵，质量，质感，……
与图片是否相符	实物，照片，图片，想象，实体店
服务态度	服务态度，服务，态度，热情，卖家，客服，店家，商家，老板

由于不同的商品特征的重要性不同，为了从用户的角度衡量商品特征的重要性，本文根据用户提及商品特征的频率作为商品特征重要性的衡量。商品特征重要性如表 5.7 所示。下表中频次之和不等于研究数据源的总量，这是因为一条在线评论中往往提及多种商品特征。

表 5.7 商品特征重要性

Table 5.7 Commodity feature importance table

类别	频次统计	频率统计	词语频次排序
上身效果	21222	0.313	款式, 外套, 颜色, 时尚, 效果, ...
物流	4207	0.062	物流, 速度
尺寸	2615	0.039	尺码, 标准
价格	4474	0.066	价钱, 实惠
质量	27796	0.410	质量, 面料, 袖子, 色差, 线头, ...
与图片是否相符	2672	0.039	照片, 实体店, 想象, 实物
服务态度	4846	0.071	卖家, 服务态度
统计项	67832	1.0	

本文使用 F1 值作为评价指标。分类结果矩阵如表 5.8 所示, 记 TP 表示真正例的个数, 即真实情况是正例, 且预测结果也为正例的总个数; FP 表示假正例的个数, 即真实结果是反例, 而预测结果是正例的总个数; FN 表示假反例的个数, 即真实结果是正例, 预测结果是反例的总个数; TN 表示真反例的个数, 即真实结果是反例, 预测结果也是反例的总个数。

表 5.8 分类结果矩阵

Table 5.8 Classification result matrix

真实情况	预测结果	
	正例	反例
正例	TP	FN
反例	FP	TN

Precision 表示精准率, 其计算公式为 $Precision = TP / (TP + FP)$ 。召回率 Recall 表示查全率, 其计算公式为 $Recall = TP / (TP + FN)$ 。F1 是 Precision(精准率)和 Recall(召回率)的调和平均, 其计算公式为 $F1 = 2 \times Precision \times Recall / (Precision + Recall)$ 。

本文采取人工打标的方法, 验证本文商品特征提取的效果。本文通过人工打标 1000 条评论文本中每条评论文本所提及的商品特征, 再对应至商品特征提取的结果, 从两个角度进行二者对比, 即计算出未识别出的商品特征个数, 以及误识别的商品特征个数。经统计, 本文使用的商品特征提取方法, 在 1000 条文本中, 共提取出 3401 个商品特征, 有 246 个商品特征词语识别错误, 107 个商品特征词语未识别出。即识别为商品特征且识别正确的个数(TP)为 3155 个, 识别为商品特征且识别错误的个数(FP)为 246 个, 实际为商品特征且未识别出的特征词语个数(FN)为 107 个。各指标计

算结果如表 5.9 所示。

表 5.9 商品特征提取衡量指标

Table 5.9 Commodity feature extraction metric

TP	FP	FN	P	R	F
3155	246	107	92.8%	96.7%	94.71%

(2) 情感强度计算

按照 3.2.2 节所示的领域情感词典建立的方法,构建领域情感词典如表 5.10 所示。其中包括褒义词 268 个,贬义词 169 个。

表 5.10 情感种子词典

Table 5.10 Sentiment seed dictionary

情感倾向	情感种子词语
正向	不错, 轻便, 合适, 舒服, 舒适, 便宜, 轻薄, 快, 厚, 暖和, 正合适,
负向	差, 臃肿, 差劲, 肥, 贵, 偏差, 毛病, 太差, 重, 色差, 异味, 薄, 味道, 瑕疵, 起球, 腥味, 难闻, 老气, 漏毛, 露毛, 粘毛, 弄脏, 速度慢,

本文以准确率(Accuracy)、F1 值作为指标衡量方法有效性。

Accuracy 表示准确率, 其计算公式为 $Accuracy = (TP + TN) / (TP + TN + FP + FN)$ 。

本文使用评论数据中 2000 条数据, 其中含有负面评论 199 条, 对 2000 条评论数据进行人工情感倾向判别, 并进行人工标注。将 2.3 中所述实验方法所得结果中, 情感值小于 0 的评论归为负向评论, 大于等于 0 的评论归为正向评论。评价结果混淆矩阵如表 5.11 所示。实验结果的各项指标如表 5.12 所示。

表 5.11 评价结果展示

Table 5.11 Evaluation results

真实情况	预测结果	
	正例	反例
正例	171	28
反例	9	1792

表 5.12 情感分类结果

Table 5.12 Sentiment classification results

Accuracy	Precision	Recall	F1
98.15%	95%	85.93%	90.24%

其中准确率为 98.15%，F1 值为 90.24%，效果较好。通过在实际数据集上的实验证明了本文构造的领域情感词典、扩充的情感词汇本体以及情感计算规则的有效性。

此外，为了对比该方法在本数据集上的优势，本文使用随机森林(Random Forest)结合不同的特征和参数设置在相同的数据集上进行对比实验。

1) 特征构建

本文使用两种文本特征构建的方法分别进行实验，其一是使用 Doc2Vec 构造评论文本特征，其二是使用情感词、否定词以及程度词的 Word2Vec 训练所得加和向量作为情感向量构造评论文本特征。

a) Word2Vec

本文使用 5.2.1 节中商品特征提取时训练所得词向量模型作为情感词、否定词、程度词向量化的基础，与此同时，由于情感词、否定词和程度词属于一个整体，因此，本文采用加和方式作为以情感词为中心的情感表达词向量，使得向量表达更能表现评论文本的情感。

b) Doc2vec

Doc2vec 又称为 Paragraph2vec、sentence embedding，是一种非监督式算法，该模型的思想基础是 Word2vec。与 Word2vec 类似，Doc2vec 也具有两种模型结构，即 DBOW(distributed bag of words)和 DM (distributed memory)。二者区别在于预测向量时是否考虑上下文。DM 模型根据给定上下文中的模型来预测目标词，DBOW 根据目标词来预测上下文概率。相较于 Word2vec，Doc2vec 增加了一个与词向量长度相等的段落向量 paragraph id，该向量具有固定长度，不仅加入了文本语义信息，同时具有更好地泛化能力。

2) Random Forest 算法

随机森林(Random Forest)算法是有监督学习算法，该分类方法是在决策树基础上发展而来，可以看作是以决策树作为基分类器并结合而成，且决策树之间没有关联，当一个新的样本进入随机森林，让森林中的每棵决策树分别判断这个样本应该属于哪一类，选取投票数最多的一类作为最终的分类结果。随机森林的主要步骤为：

- 采用 Bootstrap 方法进行重复采样，随机生成 N 个训练集 T_1, T_2, \dots, T_N ;
- 使用每个训练集，生成对应的决策树 D_1, D_2, \dots, D_N ，并选择分裂特征集，采用这些特征中最好的分裂方式对该节点进行分裂；
- 每棵决策树逐层生长，不采用剪枝策略；
- 对于任意一个需要进行预测的样本 P ，利用每个决策树进行分类，得到类别集合为 $D_1(P), D_2(P), \dots, D_N(P)$;
- 采用投票的方式确定 P 最终所属类别。

3) 数据集准备

为了和本文提出的情感分析方法进行有效对比,对比实验中采用和本文情感分析方法预测时的 2000 条评论数据进行实验效果的对比。由于部分对比算法是有监督学习,因此,人工打标 10999 条评论文本(人工打标文本不包括预测时使用的 2000 条数据),原则为:正向情感打标为 0,负向情感打标为 1。打标文本中,正向情感为 10543 条,负向情感为 456 条。

4) 实例分析

a) Random Forest + Doc2vec

随机森林由于采用了未剪枝策略,因此,随机森林依靠树深度(max-depth)和最小分裂样本数(min-samples-split)缓解过拟合问题。为了选取最优的算法参数以达到最优的实验效果,本文计算了不同参数对应的准确率(Accuracy)等指标进行衡量。参数设置及其相关指标如附录 B 所示。指标在各参数表现情况如图 5.2 所示。

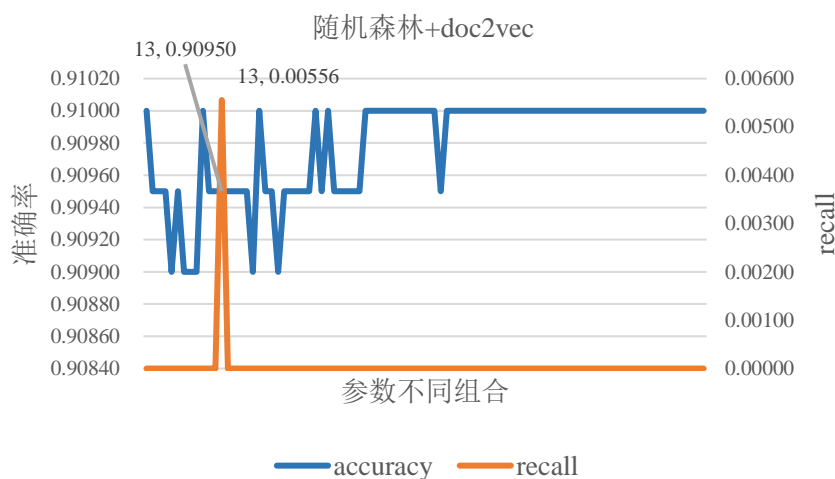


图 5.2 Doc2vec 结合随机森林在不同参数下指标变化情况

Figure 5.2 Doc2vec combines random forest performance under different parameters

从图 5.2 中可以看出,在参数的第 13 个组合状态(min_samples_split=12, max_depth=40)时,准确率(Accuracy)和召回率(Recall)最高,此时,Accuracy 为 90.95%,但是 Recall 仅为 0.56%。其原因可能从两个方面导致:1、训练集中的数据极不平衡,导致算法学习负面情感特征不充分;2、特征构造不充分,特征向量无法充分表达评论中所包含的情感。

b) 情感向量 + Random Forest

为了解决上述问题中特征向量不能很好的表达评论情感问题,本文使用情感词、否定词、程度词在 Word2vec 模型中训练所得的词向量的加和作为特征向量,使用该方法的一个原因在于以情感词为核心、否定词和程度词为补充修正的向量表达相较于 Doc2vec 具有更丰富的情感信息,更有利于情感表达。

参数组合可能值和 a) 中相同，即与附录 B 中的参数组合相同，在相同参数、相同数据集的情况下，具体参数及其对应指标表现如附录 C 所示，各指标的情况如图 5.3 所示。

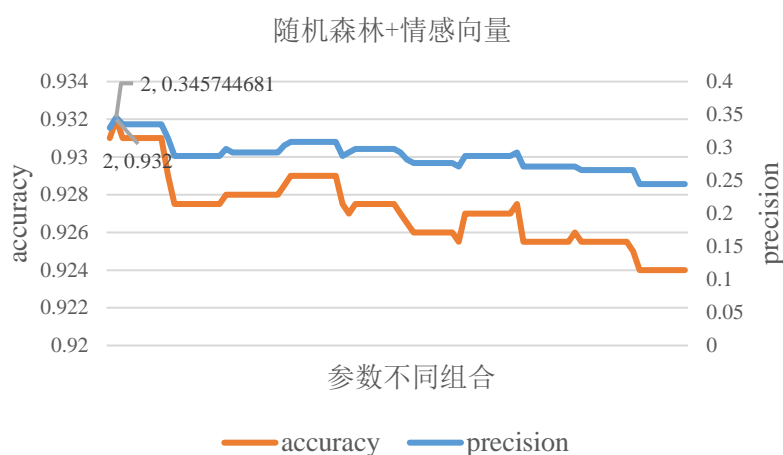


图 5.3 情感向量结合随机森林在不同参数下指标变化情况

Figure 5.3 Sentiment vector combined with random forest performance under different parameters

从图 5.3 中可以看出，在参数的第 2 个组合状态 ($\text{min_samples_split}=2$, $\text{max_depth}=20$) 时，准确率 (Accuracy) 和召回率 (Recall) 最高，此时，Accuracy 为 93.2%，Recall 为 34.57%。相较于 Doc2vec 作为特征向量在随机森林中的表现有了明显的提升，这也说明在情感分析的任务中，情感词汇较单纯的使用语义向量进行情感分类更有利。

c) 情感向量 + Random Forest + 均衡训练集

为了解决在 a) 中数据不均衡的问题，我们将训练数据集中的 456 条人工打标为 1 的文本与同等数量的人工打标为 0 的文本混合后作为训练数据集，结合情感向量和随机森林算法进行对比实验，参数组合仍和附录 B 中相同，在相同参数、均衡数据集下，各个参数下的指标表现情况见附录 D 所示。各指标的情况如图 5.4 所示。

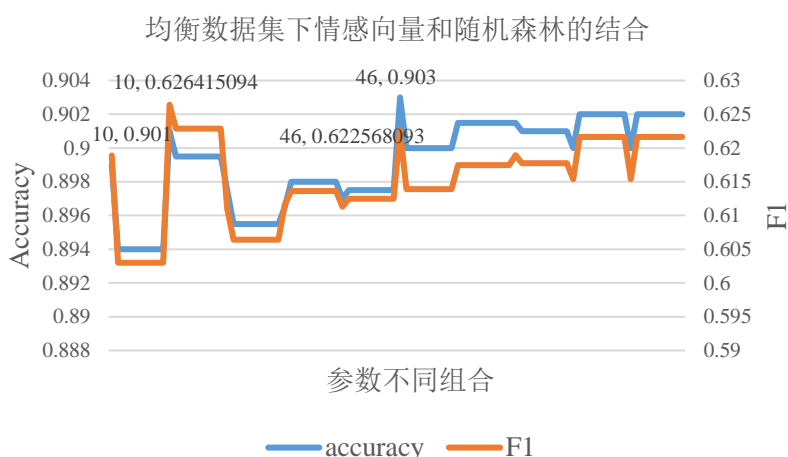


图 5.4 均衡数据集下情感向量结合随机森林在不同参数下指标变化情况

Figure 5.4 Sentiment vector under equilibrium data set combined with random forest performance under different parameters

图 5.4 中引入了 F1 值指标，该指标综合考虑了 Precision 和 Recall 两个指标。在这里考虑 F1 指标的原因在于：从附录 D 中可以看出 Recall 指标有了非常大的改善，因此，我们需要对预测效果的其他指标进行进一步的衡量，以便于和本文所提出的情感分析的方法对比，这就需要引入 Precision 指标，为了综合考量 Precision 和 Recall 两个指标，故需要引入 F1 值指标对该模型进行评价。

从图 5.4 中可以看出，在参数的第 10 个组合状态($\text{min_samples_split}=12$, $\text{max_depth}=10$)和第 46 个组合状态($\text{min_samples_split}=52$, $\text{max_depth}=10$)时，准确率 (Accuracy) 和 F1 值均较高，此时，Accuracy 分别为 90.1% 和 90.3%，F1 值分别为 62.64% 和 62.26%。相较于 b) 在随机森林中的表现有了明显的提升，这也说明在机器学习进行情感分析的任务中，构造情感表达较为突出的向量以及均衡的数据集有利于情感分类。

综合上述的实例分析讨论，对随机森林在不同参数水平下，结合不同的特征向量和改进训练数据集的方法，使得模型在准确率、召回率、精准率上均有了非常突出的提升，但是从实验 c) 中可以看出，经过优化训练所得的模型依旧较本文所使用的模型在衡量指标表现上有差距。

回溯本文使用的评论文本情感分析的方法，通过定位特征词语、情感词语，并结合详细的计算规则、否定词、程度词以及特征权重对评论文本精细计算了评论文本的情感强度。相比于机器学习模型而言，比较容易控制整个计算过程，并且计算流程清晰，较易理解。

(3) 其他指标量化

其他指标的量化按照 3.2.3 节中所述方法逐一进行量化。对于在线评论长度计算，

将去除停止词和标点符号等无效字符后剩余评论文本的有效长度；

对于用户打分的计算，直接使用用户的已有打分即可；

对于评论时间量化，本文的评论数据爬取结束时间为 2019 年 1 月 21 号，因此，本文按照 2019 年 1 月 22 号 00:00:00 为截止时间，计算时间差后，转化为天数作为评论时间跨度的量化值。

5.2.2 指标权重计算

本文按照 4.1.1 节所述的层次分析法步骤，逐步确定在线评论有用性指标权重。层次分析法的流程为如图 5.5 所示。

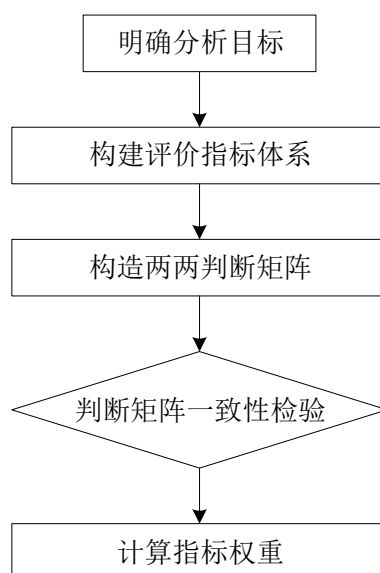


图 5.5 层次分析法流程图

Figure 5.5 Analytic hierarchy process flow chart

本文采用问卷调查的方式构建层次分析法的判断矩阵。根据在线评论有用性的各个指标的相对重要性，使用问卷星设计问卷，并发放问卷。问卷发放时间为 2019 年 6 月 29 日，回收时间为 2019 年 7 月 1 日，共收集 100 份问卷，其中有效问卷 70 份。问卷题目设置如附录 E 所示。

本文使用 python 实现 4.1.1 节层次分析法的各个步骤后，最终通过一致性检验的判断矩阵如表 5.13 所示：

表 5.13 判断矩阵

Table 5.13 Judgment matrix

指标	文本长度	情感强度	星级打分	评论的全面性 与相关性	评论发布距 今时间跨度
评论文本长度	1.000	0.239	0.272	0.195	0.306
评论情感强度	4.184	1.000	3.344	3.534	2.976
用户的星级打分	3.676	0.299	1.000	0.272	0.35
评论的全面性与相关性	5.128	0.283	3.676	1.000	1.801
评论距今的时间跨度	3.268	0.336	2.857	0.555	1.000

将该矩阵存储值 Excel 中，并按照 4.1.1 节所示的层次分析法算法流程，将该矩阵以 python 的 dataframe 格式读入，层次分析法的实现代码实现如表 5.14 所示。算法最后输出各指标的权重如表 5.15 所示。

表 5.14 层次分析法代码

Table 5.14 Analytic hierarchy process code

IP 池构建代码

```
import numpy as np
import pandas as pd
weight = pd.read_excel("weight.xlsx")
RI_dict = {1: 0, 2: 0, 3: 0.58, 4: 0.90, 5: 1.12, 6: 1.24, 7: 1.32, 8: 1.41, 9: 1.45}
def get_w(array):
    row = array.shape[0]
    a_axis_0_sum = array.sum(axis=0)
    b = array / a_axis_0_sum
    b_axis_1_sum = b.sum(axis=1)
    w = b_axis_1_sum / row
    AW = (w * array).sum(axis=1)
    max_max = sum(AW / (row * w))
    CI = (max_max - row) / (row - 1)
    CR = CI / RI_dict[row]
    if CR < 0.1:
        print("特征向量:%s' % w)"
        return w
```

表 5.14 (续)

Table 5.14 (Continuation)

IP 池构建代码
<pre>def main(array): if type(array) is np.ndarray: return get_w(array) if __name__ == '__main__': e = weight.values e = main(e)</pre>

表 5.15 指标权重

Table 5.15 Index Weight

评论文本长度	评论情感强度	用户星级打分	评论信息的全面性与相关性	评论发布距今的时间跨度
0.055	0.116	0.170	0.396	0.263

5.2.3 评论有用性量化结果展示

为了说明在线评论有用性量化结果,本节以商品 id 为 31373205887 为例,通过将原有排序 TOP5 的评论与按有用性所得的新排序进行比较,排序前后比较结果如表 5.16 所示:

表 5.16 本文模型排序前 5 条各属性指标值

Table 5.16 The top five attributes of this model are sorted

评论内容	id	C-R	情感强度	评论时间	用户打分	评论长度	原文排序
衣服款式好看,做工精致,布颜色好看,手感柔软,真的好喜欢,推荐给大家,物流相当给力	C_1	3.088926	7.360729	2018-11-15 09:22:14	5	30	1 2
衣服款式不错.做工精细.面料舒适。最重要的是我很喜欢,推荐大家购买	C_2	1.823825	4.362174	2018-11-22 07:32:40	4	20	2 5

表 5.16 (续)

Table 5.16 (Continuation)

评论内容	id	C-R	情感强度	评论时间	用 户 打 分	评 论 长 度	原 本 排 序
很不错的羽绒服，做工细致，毛领不错，拉链很好，希望是鸭鸭正品	C_3	1.460281	7.785765	2018-12-12 19:42:45	5	22	3 3
老板服务态度不错，回复也很快，推荐的尺码也很标准，衣服的质量也很好，摸着也舒服，推荐购买	C_4	1.54776	4.784954	2018-12-28 17:12:33	5	30	4 4
宝贝已经收到了，物流很快，颜色非常好看，没有色差，版型也非常好看，上身效果好，显气质。面料也特别舒服。	C_5	3.813977	10.905915	2018-12-17 09:42:48	5	32	5 1

以原有排序为 1 的评论和本文排序为 1 的评论进行结果解释。本文认为 C_5 优于 C_1 的原因如下：首先，C_1 所提及的商品特征包括：款式、做工、颜色、手感、物流；C_5 所提及的商品特征包括物流、颜色、色差、版型、上身效果、气质、面料，即 C_5 描述的商品特征比 C_1 更丰富；其次，C_5 含有非常、很、特别等程度词被多次提及，因此情感强度值较大，说明情感表达相对明确；且 C_5 用户评论时间距今较近，参考价值相对较大；最后，C_5 的评论长度较 C_1 长，描述更详尽。综合上述指标，本文认为 C_5 优于 C_1。

5.3 商品模糊综合评价

本部分商品评价过程中共进行了两次修正，其底层部分是按照商品特征进行分类的，通过计算每个商品特征的情感值，用于了解评论者对商品特征的观点和情感表现。第一次修正是使用评论有用性量化值进行修正的。其原因在于评论数据的相对有用性会影响其作为数据源用于商品评价时的参考价值。通过有用性系数的修正，使得对高有用性评论的情感表达得到重点关注而弱化了低有用性评论的参考价值。第二次修正是使用商品特征权重实现的，其原因在于对商品进行评价过程中，应当侧重用户关注

最多的因素进行评价，这样可以更符合大众预期。

5.3.1 模糊评价矩阵构建

模糊评价矩阵是关于商品特征各个指标以及评论有用性指标属于各个评价等级的隶属度矩阵。具体步骤如下：

(1) 商品特征部分指标量化

记每条评论中提及商品特征情感值为 s_{ij} ，表示第 i 次提及第 j 个商品特征时的情感值， $i = 1, 2, \dots, n$ ， n 表示某条评论提及某特征的总次数， $j = 1, 2, \dots, m$ ， m 表示商品特征的总个数，故一条评论中某个特征的情感值为 $\sum_{i=1}^n s_{ij}$ 。由此构成指标向量 $S = (\sum_{i=1}^n s_{i1}, \sum_{i=1}^n s_{i2}, \dots, \sum_{i=1}^n s_{i7})$ ，向量维度为 1×7 。

这里是按照商品特征分别求和计算情感值，并用于填充模糊评价矩阵。因此，需要将每条评论中所提及的情感按照商品特征分类计算。

(2) 有用性部分指标量化

将每条评论的有用性得分 ρ 修正指标向量 S ，得到 $S' = (\sum_{i=1}^n s'_{i1}, \sum_{i=1}^n s'_{i2}, \dots, \sum_{i=1}^n s'_{i7})$ ，简记为 $S' = (s_1, s_2, \dots, s_7)$ ，向量维度为 1×7 ，其中， $s_1 = \sum_{i=1}^n s_{i1} \times \rho$ ， $s_2 = \sum_{i=1}^n s_{i2} \times \rho$ ， $\dots s_7 = \sum_{i=1}^n s_{i7} \times \rho$ 。对于每个商品id对应的评论数据集而言， S' 是矩阵形式，此时的 s_k 维度为 $p_t \times 1$ ， p_t 表示评论集大小。

(3) 定义评价等级

本文定义五个评价等级 $L = \{l_1, l_2, \dots, l_5\}$ 。根据每个商品特征的指标值分布确定评价等级。这里，对每个商品特征确定等级阈值 $F = \{(0, \Gamma(s_1)), (0, \Gamma(s_2)), \dots, (0, \Gamma(s_7))\} = \{(0, \gamma_{11}, \dots, \gamma_{13}), (0, \gamma_{21}, \dots, \gamma_{23}), \dots, (0, \gamma_{71}, \dots, \gamma_{73})\}$ ，其中 Γ 是求解 s_k 上、中、下分位数的函数， γ_{ij} 表示 s_i 的第 j 个分位数。这里阈值的计算，是以全体数据集为基础的。由于商品的评价是相对的，因此，关于某一个指标的好坏，应该是当前指标值相对于全体该指标对应的值中的优劣。所以，本文根据各个商品特征的情感值分布情况确定阈值，并进行相对的比较。

(4) 模糊评价矩阵构建

构建模糊评价矩阵 Λ 为：

$$\Lambda = \begin{bmatrix} \frac{\sum_{i=1}^{t_k} f_0(s_1)}{t_k} & \dots & \frac{\sum_{i=1}^{t_k} f_4(s_1)}{t_k} \\ \vdots & \ddots & \vdots \\ \frac{\sum_{i=1}^{t_k} f_0(s_7)}{t_k} & \dots & \frac{\sum_{i=1}^{t_k} f_4(s_7)}{t_k} \end{bmatrix}$$

其中 $f_d = \begin{cases} 1 & \text{if } s_{1j} > (0, \Gamma(s_1))[d], \\ 0 & \text{else} \end{cases}$, $d = 0, 1, 2, 3, 4$, s_{1j} 表示 s_1 的第 j 个元素, $s_{1j} > (0, \Gamma(s_1))[d]$ 表示 s_{1j} 大于第 d 个阈值, t_k 表示第 k 个商品 id 的数据集大小。

这里需要注意阈值与构建模糊评价矩阵时使用数据维度的区别。在划分阈值时, 是按照全部数据集中对于每一类商品特征情感值的分布进行等级划分的, 而在构建模糊评价矩阵时, 其内部元素是以商品 id 维度进行计算的。这并不矛盾。标准的产生, 是以全体数值的数据分布为基础的, 这是因为商品优劣是所有商品之间相对的判断; 模糊评价矩阵是则是对每个商品 id 计算得分的, 在计算商品 id 的某个特征在某个等级上的比例时, 如果仍然除以全体数据集大小, 这对于评论数据少的商品 id 计算时不公平的, 因此, 在构建模糊评价矩阵的时候, 应该以商品 id 为维度进行内部计算。

模糊评价矩阵是模糊综合评价法的核心内容。模糊评价矩阵是根据商品特征的情感表达构建的。由于对于不同的商品、不同的商品特征, 用户对其的评价是不同的, 本文使用商品特征的各个类别进行情感值归纳, 具体来说, 将每条评论的情感值归结到各个商品特征上, 由于每条评论可能不止一次提及同一商品特征, 因此, 本文将对应描述同一个商品特征的情感值相加, 得到每条评论中关于商品特征的情感得分。总的来说, 指标量化主要是对商品进行得分和等级的划分与计算。由于商品特征的维度对于每个商品来说是一致的, 因此, 可以通过比较每一项商品特征的情感值的分布情况, 判断出以商品 id 为维度的商品对应各个特征的情况。比如, 对于“上身效果”这一商品特征而言, 每条评论对其表达的情感会有所不同, 假设“上身效果”特征中情感值的分布集中于偏向于下四分位数区间, 那么可以认为用户对于该商品特征的反响普遍较差, 但是若此时某件商品 id 的“上身效果”情感值较高, 那么本文会认为对于这个商品 id 的“上身效果”特征而言, 是表现较优的。

5.3.2 商品评价得分

记 α 为商品特征指标权重, ω 为评价等级 $L = \{l_1, l_2, \dots, l_n\}$ 的分值。故商品评价得分 $score$ 计算如式(5.1)所示:

$$score = \alpha \times \Lambda \times \omega \quad (5.1)$$

本文使用商品特征重要性作为商品特征指标权重。此外, 由于本文评价等级划分

时，第一级别阈值为 0，对于商品特征情感值小于 0 的商品应该予以惩罚，因此，对于情感值小于 0 的等级予以负数得分表示，即 $\omega = (-5, 1, 3, 5, 7)$ 。

5.3.3 商品模糊综合评价结果展示

本文使用试验数据集，按照 5.2~5.3 节所述的步骤，首先进行在线评论有用性指标的量化计算，其次使用层次分析法和 TOPSIS 排序算法进行在线评论有用性量化，最后构建模糊矩阵，并代入模糊综合评价法进行商品评价，部分结果如表 5.17 所示。

表 5.17 商品评价部分结果展示

Table 5.17 Commodity evaluation part results show

商品 id	商品得分
19452729708	1.280969
19234032309	2.665214
10665592348	3.070967
32032960418	4.265852
31728550238	5.411344

由表 5.17 可知，id 为 19452729708 的商品评价等级介于二等和三等之间，id 为 31728550238 的商品评价等级介于四等和五等之间，即 id 为 31728550238 的商品优于 id 为 19452729708 的商品。商品进行评价后，用户可以根据页面展示的商品排序和得分对商品本身有一定的感知，进而辅助用户快速做出决策。

5.3.4 商品评价模型验证与分析

为了验证本文提出的商品评价模型的有效性，本文从评论数据集中选取部分好评和差评的评论集合构造为新的验证数据集。验证数据集选取 10 条商品评论，为了更好的区分两个商品，本文将商品 id 为 1 的数据集中好评数较商品 id 为 2 的多，通过拉开二者之间的差距，以便于人工更好的评判本模型的效果。构造的验证数据集如表 5.18 所示。

表 5.18 验证数据集示例

Table 5.18 Verify data set example

productid	comment_id	comment	creationTime	score	有用性
1	C_1	衣服到货了，物流很快回来就试穿了，很合身	2018-11-15 20:18:40	5	0.299890
1	C_2	发货速度快，衣服大小合适，颜色和展示一样。很好，满意！	2019-01-06 01:04:46	5	0.307038
1	C_3	衣服挺不错的，价格也很喜欢，卖家发货速度挺快的。	2018-10-08 13:49:48	5	0.307311
1	C_4	衣服收到了，非常不错，尺码标准，颜色跟图片一样，没有色差，非常满意。	2018-12-30 18:44:57	5	0.309750
1	C_5	比较薄不是纯黑，衣服和袖子的颜色都不一样，其它还好！	2019-01-20 05:24:20	5	0.356257
2	C_6	衣服不咋滴，不值这个价。太薄了，发货又慢，包装超级简单，就一个袋子装着，买亏了，适合夏天穿，	2019-01-16 18:15:39	5	0.300867
2	C_7	真薄啊，适合南风天穿！还有一股子鸭子的骚臭味，造孽啊！	2018-12-20 22:42:54	3	0.281599
2	C_8	跟图片简直就是天壤之别，一打开衣服里面就毛毛直飞，感觉就像几十块钱的衣服一样。拉链都是坏的，毛领完全看不得。客服推荐 XL，买的 L 码都大了，承诺当天发货冒发，第二天下午再发的货，一点点信誉都没有	2019-01-12 12:39:03	1	0.306907
2	C_9	这个羽绒服太薄了，想一层透明的纸一样，完全没有保暖效果，太差劲了，大家不要再上当了。	2018-12-13 12:38:10	2	0.296606
2	C_10	快递已经收到，包装很好，无异味，衣服面料柔软舒适，做工精细，无多余线头，大小合身，穿着很有精神，这个价位性价比高，比实体店便宜，五星好评	2019-01-20 10:28:17	5	0.457104

(1) 首先判断商品评价结果的合理性。

首先，从表 5.18 中可以看出，id 为 1 的商品用户评论表述普遍较好，而 id 为 2 的商品用户评论表述普遍较差，因此可以得出 id 为 1 的商品明显较 id 为 2 的商品优质的观点。

其次，通过构建本文所述的模糊评价矩阵代入模糊评价法，计算得出 id 为 1 的商品得分为 2.657760，id 为 2 的商品得分为 1.189443。因此，从商品模糊综合评价上来看，id 为 1 的商品较 id 为 2 的商品优。这也和之前的人工预判是一致的，初步验证了模型的可行性。

(2) 其次判断有用性修正系数的正确性。

为了验证本文引入在线评论有用性这一系数是有意义的，本文做了如下的对比实验，对照实验中模糊评价矩阵不考虑在线评论有用性系数。

具体来说，对照实验为：将商品特征对应的情感值，在不引入在线评论有用性的前提下，只按照等级进行划分以及计算各个等级的占比，并构建模糊评价矩阵，通过模糊综合评价法得出 id 为 1 的商品得分为 4.154546，而 id 为 2 的商品得分为 1.173424。

首先，从对照实验中，我们可以看出，对照实验仍然可以得到 id 为 1 的商品较 id 为 2 的商品优质的观点。

其次，对比本文模型的商品评分与对照实验的商品评分。商品 id 为 1 与商品 id 为 2 的对比得分如表 5.19 所示。

表 5.19 不同商品在不同模型下评分比较

Table 5.19 Comparison of different commodities under different models

商品 id	本文模型评分	原始评分(对照实验)
1	2.657760	4.154546
2	1.189443	1.173424

综合表 5.19 可进行如下分析：

首先，对于 id 为 1 的商品得分，对照实验的得分高于本文的模型评分。而对于商品 id 为 2 的得分，对照实验的得分低于本文的模型评分。

其次，我们对分别对各个商品结合上述结果进行探讨：

➤ 对于 id 为 1 的商品而言，本文模型得分低于对照实验得分。从理论上讲，通过观察构成 id 为 1 商品的在线评论可知，C_5 的在线评论由于其评论时间距今时间跨度小，其有用性较高，具有较高的可参考性，且其评论情感得分为-0.377472，属于负面评论，因此，该商品的评分被 C_5 拉低。同时，从时间上看，C_1、C_2、C_3、C_4 评论距今时间跨度较大，因此，可参考性相较于 C_5 而言略低，用户更愿意相信时间跨度较小的评论。故最终的商品得分较对比实验而言，略倾向于 C_5。

➤ 对于 id 为 2 的商品而言, 本文模型得分高于对照实验得分。从理论上, 通过观察构成 id 为 2 商品的在线评论可知, C_10 的在线评论距今时间跨度较近, 且商品特征描述丰富, 情感强度为 11.345884, 属于正向评论, 其有用性量化值较大, 这使得该条评论具有较大的参考价值, 因此, 该商品的评分被 C_10 提高。但是由于 C_6、C_7、C_8、C_9 评论中, 各个商品特征情感贬损倾向强烈, 因此, C_10 提高分数并不明显。

综合上述两点考虑, 我们可以发现, 在线评论有用性对于商品评价具有调节功能, 其调节的强度可大可小, 这取决于评论数据集中, 在线评论针对商品特征的情感强度大小及其情感值的分布。这是合理的。在线评论有用性表达的是一种对评论的可参考程度, 其并不能掩盖历史评论数据对商品评价所带来的价值和贡献。最后所得的商品评价得分只是从有用性的角度, 突出强调了距今时间跨度较短等优质特征的在线评论对商品所持有的观点和态度, 进而达到了修正、调节的作用。

5.4 研究启示

本章对前述的理论方法进行了实证研究, 通过对商品评价模型的研究与实现, 为相关研究和实践带来了如下思考与启示。

首先, 从理论研究的角度来看, 目前关于在线评论应用文本挖掘与处理, 获取深层次信息的相关研究非常广泛: 这些研究的研究角度不同, 使用数据源的领域、语言类型不同, 研究目标不同等等, 这在一定程度上为在线评论挖掘营造了良好的研究氛围并打下了良好的研究基础。然而, 以在线评论为数据源的研究中, 对商品进行评价的相关研究就显得较为单薄, 研究方法较为单一, 这有可能是该主题所具有的商品特征、个性色彩较重且用户对于商品的评价具有“因人而异”的个性, 因此, 学者对于商品评价的研究较少。但是, 商品评价的研究目标或者其初衷, 并不是对商品做出符合于每位消费者特性的评价, 其更多注重的是, 基于现有的资源(在线评论)为消费者展示真实的评论集合对商品的评价倾向, 这也是本文研究的出发点和意义。也就是说, 本文为用户揭示了一个事实, 这个事实是用户难以通过阅读海量评论进行总结的, 并且本文的研究思路和逻辑具有普适性。除此以外, 在商品评价研究中应该加入新的元素以改进原有方法中存在的不足。本文在商品评价过程中加入了对评论元素的区分, 也就是说, 不同的评论元素对于商品评价所具有的影响程度是不同的。区分不同的评论元素依靠的是其具有的有效有用性大小。这里的有效有用性为用户感知的有用性。引入在线评论有用性的原因在于, 商品评价过程中, 不应该平等的参考所有的在线评论, 例如, 对于日期较远的评论, 在商品评价时, 其参考价值会较新评论较低, 因此, 有必要区别对待评论数据集, 这样处理后的商品评

价更合理。这对于商品评价的研究具有启示意义，在进行商品评价过程中，可以从数据源处理的角度、评价模型构建角度等多方面进行优化，进而不断丰富、不断深入商品评价的相关研究。

在线评论的有用性是用户较为主观的感受，在该主题下，研究者进行了大量的研究，尽管没有对影响在线评论有用性影响因素达成一致的指标集，但是本文将其研究结论较为一致的指标用于有用性评价。根据研究者对部分指标设置目的表述，我们发现对于这部分指标量化过程显得稍有粗略。因此，本文根据这部分指标设置的目的，在指标量化时保证准确率前提下进行细化。这对于在线评论有用性研究具有启示性，我们不必拘泥于计数等方法，可以通过文本挖掘的多种方法相结合完成对指标的细化，使指标值更具有实际意义。

除了上述的理论研究方面的启示外，本文认为，对于商品评价的研究是具有跨领域研究可能性的，无论是将评论有用性与商品评价相结合还是其他领域的知识、思路等与商品评价的研究相结合，对于商品评价的研究而言，是一种优化、更是理论上的丰富，并且非常有意义。通过商品评价在学术研究领域的关注，也会促进将研究成果应用于实践，通过真正地将研究应用于实践，才能让用户受益。

此外，对于本文相关联的研究领域，如推荐系统的研究中，研究的热点是如何提高推荐效率、如何提高推荐准确率、如何通过社交网络更好地把握用户的兴趣等问题，但是忽略了对推荐列表的整理，满足用户的兴趣固然重要，但是考虑了商品质量的推荐更有利于提高转化率；又如商品排序研究，商品评价研究的延伸即可实现商品排序。因此，对于商品评价的研究，其可扩展的研究范围很广泛，应当在学术研究中电子商务领域以评论作为数据源的相关研究中占有相对重要的一席之地。

其次，从实践角度看，本文主要是将文本挖掘方法应用于电商平台的在线评论数据集，以从中获取知识和探索新的发现，优化用户体验和提高平台商品转化率。实践角度不妨分为本文研究的应用角度和商品评价的实践意义两个方面进行介绍。

从本文的应用角度来看，正如前面所说，本文的研究是嵌入式的，也就是说，本文的思路和逻辑是可以进行多重应用的，例如，本文的研究是可以在实践中进行用户交互的，当用户在选择本方法进行商品排序时，用户可以自行定义对各个商品的重要性排序，并且在评论有用性评价中可以对影响因素重要性进行自定义，甚至用户可以不进行任何操作，仅按照默认方式即统计方法得出权重，这不仅可以实现用户的个性化选择且简单易行，更重要的是，这可以使得商品评价结果更符合用户的习惯；又如，商品排序可以嵌入到平台推荐系统中，根据用户以往的习惯进行特征权重的确定等多种个性化操作，这会更有利于用户体验的提升。

从本文研究的实践意义来看，可以分为平台、用户、商家三个角度进行介绍。

商品评价研究对于平台而言是非常重要的，它不仅是平台人性化、个性化、科技化的体现，更有利于提升用户体验、增强用户粘性。平台将其拥有的信息充分利用，更好地回馈给用户。通过自动化的手段，平台不仅为用户提供了优质的商品，还提高了商品的点击率与转化率。与此同时，在平台治理方面，平台可以重点关注商品得分较低的商家，重点改进这些商家的商品质量、服务质量等，有效改善平台治理效果。

商品评价研究对于用户而言，是极其便利的。我们可以从用户购物的多个阶段进行分析。首先，用户浏览商品阶段，通过商品评价模型计算按照其关键词得到的搜索页商品得分，并进行商品的排序，有利于缩短用户的浏览、翻看的时间，使其从入口阶段即可高效快捷。随后，当用户点开一个商品时，用户无需逐条查看该商品下上万条评论，由于在计算时可以得到各个商品特征的得分，因此，有利于用户根据其关注的商品特征判断是否购买，省去了用户浏览海量评论的困扰。最后，当用户选择购买商品之后，用户选择了商品评分高的商品，并且按照其关注的角度确认了商品，用户收到商品后的满意度便是可以进行口碑传播的开端。由此正向良性循环，在增强用户体验、提高购买效率的同时，也使得平台得到了更多高粘性用户。

商品评价研究对于商家而言，可以分为长尾商品商家、优质商品商家、劣质商品商家进行介绍。首先，本文的商品评价研究对长尾商品是非常有利的，如果该商品十分小众，但是购买好评率很高，那么在本文计算方式中依旧有机会优先展示在用户面前，这对于长尾商品而言，该方法为其提供了公平的竞争机会，并鼓励其更好的运营，增加了长尾商品的商家对平台的信任。其次，对于优质商品的商家而言，这里的优质主要是指好评数量较高的商家，其本身的曝光率和流量可能已经比较高，用户通过该方法进入该商品时信任度较高，更有利于商品的转化率，与此同时，由于有了长尾商品的竞争，也有利于优质商家继续提升自己的服务等多方面的质量，有利于通过商家自身来减少平台治理的难度。最后，对于劣质商品的商家，由于本文的商品评价可以进行商品特征对应得分，因此，商家可以针对自己的薄弱环节进行重点提升，并且由于商品得分较低的商品会被向后搁置，因此这也有利于激励商家尽快做出改进。

5.5 本章小结

本章通过对商品评价指标进行量化，并结合模糊综合评价法构建了模糊评价矩阵，该矩阵是模糊综合评价法的重点，等级划分和隶属度计算是构建该矩阵的核心。本章以实证研究的方法，验证了本文所介绍方法的可行性。通过本章的计算，

可以得出商品 id 维度的评价得分，以期能够为用户推送相对优质的商品。最后通过实验示例证明本文模型的可行性。

本章的创新点在于，首先对评论信息的全面性与相关性加入了特征重要性的考量。其次，本文的情感分析规则定义详尽，其 F1 值达到 90.24%，取得了较好的分析效果，并且与随机森林算法进行比较，进一步证明了本文情感分析方法的优势。最后在线评论有用性的计算引入商品评价模型，并通过示例计算，证明了本文商品评价模型的可行性。

结 论

1. 研究总结

随着电子商务的快速发展，用户可以在线上浏览更多的商品，相较于传统的线下购物，线上用户有了更多选择心仪商品的机会。但是，当用户看到搜索页展示的大量搜索结果，以及用户被动接受的推荐列表时，仍会对这些商品有着零感知的心理。用户需要阅读商品介绍以认识商品、浏览用户评论以熟悉商品等繁杂的流程后，才会做出购买决策。过程的繁杂不仅会降低用户的购物体验，也会降低商品的转化率。如何让用户减少评论的阅读量，快速建立对商品的初识，以辅助用户迅速做出决策，是电商平台从增强用户体验的角度应当思考的问题。

本文以电子商务平台为背景，从已经购买过商品的用户发表的在线评论中挖掘信息，用于评价商品。这种对商品的评价不会依赖于商家的美化宣传，而更具有参考性。然而，并不是所有的在线评论都具有参考意义。通过文献阅读，选取评论信息的全面性与相关性、情感强度、评论文本长度、用户打分、评论发布距今的时间跨度等五个指标，并对指标进行量化，通过层次分析法为指标赋权值，代入 TOPSIS 排序算法，用于量化评论有用性。该指标作为修正系数，调整各个商品特征维度的情感值构建的模糊评价矩阵，代入模糊综合评价法，实现基于在线评论有用性的对商品的评价。最终，通过示例计算，验证本文商品评价模型的可行性。

2. 研究局限

本文通过选取指标、量化指标，完成对在线评论有用性的计算，并将其作为修正系数加入模糊评价矩阵中，通过模糊综合评价法得出商品综合得分。但是本研究仍具有局限性：

第一，本文对于评论有用性研究所使用的指标，基本上是大多数学者认可的指标集合。但是仍然有一些针对不同商品类型的有用性指标，尽管这些指标仍存有争议，但是随着今后研究的不断深入，指标逐渐明朗化后，可以有选择性地丰富评价指标。

第二，本文通过问卷调查构建层次分析法的判断矩阵，尽管判断矩阵通过了一致性检验，但是仍存在问卷填写人员的主观性。

3. 研究展望

在电子商务研究领域，在线评论的研究是一项非常重要的课题。由于在线评论数据量大、信息丰富等特点，学者可以从多个角度对其进行挖掘。随着用户体验逐渐被重视，用户自主发表的评论信息更值得研究。

第一，研究所使用的数据源可以尝试选取多个商品类别的评论数据，将研究不同商品类型的评论有用性作为下一步研究的方向。

第二，本文从用户对商品特征的情感反馈以及评论自身的特征去评价商品，未来可以从更多的维度、结合更多的信息更完善的评价商品。

第三，本文确定在线评论有用性权重时，采用了 AHP 方法，但是该方法带有一定的主观色彩，是否有更好的方法确定指标权重，使得其结果更为客观使得未来进行尝试和探索。

参考文献

- [1] 李亚琴. 电子商务平台用户在线评论比较研究[J]. 现代情报, 2017(07):81-85.
- [2] Bickart B, Schindler R M. Internet forums as influential sources of consumer information[J]. Journal of Interactive Marketing, 2001, 15(3): 31-40.
- [3] Duan W, Gu B, Whinston A B. The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry[J]. Journal of Retailing, 2008, 84(2): 233-242.
- [4] Senecal S, Nantel J. The influence of online product recommendations on consumers' online choices[J]. Journal of Retailing, 2004, 80(2): 159-169.
- [5] 林钦和,刘钢,陈荣华. 基于情感计算的商品评论分析系统[J]. 计算机应用与软件, 2014, 31(12): 39-44.
- [6] 张梦莹, 邓三鸿, 王昊, et al. 基于有用性排序的在线评论与销量的关系研究[J]. 现代情报, 2019, 39(02):154-162.
- [7] Li X, Zhu S, Yu J, et al. A commodity evaluation algorithm based on commodity review abstracts[C]//2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2018: 1-5.
- [8] 陈颖, 付晓东, 岳昆, 等. 一种考虑用户评分标准不一致性的在线商品评价方法[J]. 小型微型计算机系统, 2017, 38(6): 1317-1322.
- [9] 王乾, 傅魁. 基于 LSTM-AE 神经网络的商品评价综合评分计算方法研究[J]. 北京邮电大学学报 (社会科学版), 2018 (4): 4.
- [10] 李梵蓓. 商品质量的模糊综合评价[J]. 内蒙古财经学院学报, 1994(4):92-95.
- [11] 王涛. 电子商务企业产品评价指标体系的建立及模糊评价[J]. 现代情报, 2007 (1): 67.
- [12] 杨静. 电子商务网站商品模糊评价模型的构建[J]. 现代情报, 2013, 33(4): 31-33.
- [13] 徐勇, 张慧, 陈亮. 一种基于情感分析的 UGC 模糊综合评价方法——以淘宝商品文本评论 UGC 为例[J]. 情报理论与实践, 2016, 39(6): 64-69.
- [14] Chatterjee P. Online Reviews: Do consumers use them?[J]. Advances in Consumer Research, 2001, 28.
- [15] Mudambi S M, Schuff D. What makes a helpful review? A study of customer reviews on Amazon. com[J]. MIS quarterly, 2010, 34(1): 185-200.
- [16] 刘伟, 徐鹏涛. O2O 电商平台在线点评有用性影响因素的识别研究——以餐饮行业 O2O 模式为例[J]. 中国管理科学, 2016, 24(5) 168-176.

- [17] 李中梅, 张向先, 郭顺利. 移动商务环境下 O2O 用户在线评论有用性影响因素研究[J]. 情报科学, 2017, 35(2): 130-137.
- [18] Quaschnig S, Pandelaere M, Vermeir I. When consistency matters: The effect of valence consistency on review helpfulness[J]. Journal of Computer-Mediated Communication, 2014, 20(2): 136-152.
- [19] 吴江, 刘弯弯. 什么样的评论更容易获得有用性投票*——以亚马逊网站研究为例[J]. 数据分析与知识发现, 2017, 1(9): 16-27.
- [20] 王晗, 李存林, 杨世瀚. 基于 word2vec 的评论有用性研究[J]. 广西民族大学学报 (自然科学版), 2018 (1): 14.
- [21] 田依林, 黎盈盈, 滕广青. 基于商品类型的在线负面评论有用性影响因素比较研究[J]. 现代情报, 2019, 39(08): 111-119+131.
- [22] 郝媛媛, 叶强, 李一军. 基于影评数据的在线评论有用性影响因素研究[J]. 管理科学学报, 2010, 13(8): 78-88.
- [23] Han A, Ji-Fan R. Do market differences influence online consumer's behavior? An investigation of online movie reviews usefulness across markets[C]//2017 International Conference on Management Science and Engineering (ICMSE). IEEE, 2017: 575-580.
- [24] Zhou Y, Yang S. Roles of review numerical and textual characteristics on review helpfulness across three different types of reviews[J]. IEEE Access, 2019, 7: 27769-27780.
- [25] Singh J P, Irani S, Rana N P, et al. Predicting the “helpfulness” of online consumer reviews[J]. Journal of Business Research, 2016, 70.
- [26] 艾时钟, 曾鑫. 基于 Ebay 评论数据中的情感总量与信息熵对评论有用性的影响[J]. 软科学, 2019, 33(07): 129-132+144.
- [27] 陈市, 李峥, 吴晓松. 基于模糊层次分析法的购物评价效用排序模型研究[J]. 企业经济, 2017, 36(9): 29-34.
- [28] 王建文. 基于信息采纳视角的在线评论有用性排序研究[J]. 现代计算机(专业版), 2019(11): 67-71.
- [29] 蔡晓珍, 徐健, 吴思竹. 面向情感分析的用户评论过滤模型研究[J]. 数据分析与知识发现, 2014, 30(4): 58-64.
- [30] 田依林, 滕广青, 黄微. 社会化在线评论情感因素测量与有用性分析[J]. 现代情报, 2018, v.38; No.324(06): 21-29.
- [31] Xiang C, Sun Y. Research on the evaluation index system of online reviews helpfulness[C]//International Conference on Service Systems & Service Management. IEEE, 2016.
- [32] Huang A H, Chen K, Yen D C, et al. A study of factors that contribute to online review helpfulness[J]. Computers in Human Behavior, 2015, 48: 17-27.
- [33] 李昂, 赵志杰. 在线评论有用性研究综述[J]. 商业经济研究, 2019(06): 77-80.

- [34] Hong H, Xu D. Research of online review helpfulness based on negative binary regress model the mediator role of reader participation[C]//2015 12th International Conference on Service Systems and Service Management (ICSSSM). IEEE, 2015: 1-5.
- [35] 孙佳圣.公开商品评价对销售的影响研究——产品评价的策略性展示[J].中国软科学,2017(08):184-192.
- [36] 郭顺利, 张向先, 李中梅. 面向用户信息需求的移动 O2O 在线评论有用性排序模型研究——以美团为例[J]. 图书情报工作, 2015, 59(23): 85-93.
- [37] 李志宇. 在线商品评论效用排序模型研究[J]. 数据分析与知识发现, 2013 (4): 62-68.
- [38] 闵庆飞, 覃亮, 张克亮. 影响在线评论有用性的因素研究[J]. 管理评论, 2017 (2017 年 10): 95-107.
- [39] 王倩倩. 一种在线商品评论信息可信度的排序方法[J]. 情报杂志, 2015, 34(3): 181-185.
- [40] 毛郁欣,朱旭东.面向 B2C 电商网站的消费者评论有用性评价模型研究[J].现代情报,2019,39(08):120-131.
- [41] Brody S, Elhadad N. An unsupervised aspect-sentiment model for online reviews[C]//Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010: 804-812.
- [42] Santosh D T, Vardhan B V, Ramesh D. Extracting product features from reviews using Feature Ontology Tree applied on LDA topic clusters[C]//2016 IEEE 6th International Conference on Advanced Computing (IACC). IEEE, 2016: 163-168.
- [43] 彭云, 万常选, 江腾蛟, 等. 一种词聚类 LDA 的商品特征提取算法[J]. 小型微型计算机系统, 2015, 36(7): 1458-1463.
- [44] Liu L, Du W, Wang H, et al. A bipartite graph model for implicit feature extraction in Chinese reviews[C]//2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). IEEE, 2017: 915-920.
- [45] Wu H, Liu T, Xie J. Fine-grained product feature extraction in Chinese reviews[C]//2017 International Conference on Computing Intelligence and Information System (CIIS). IEEE, 2017: 327-331.
- [46] Devasia N, Sheik R. Feature extracted sentiment analysis of customer product reviews[C]//2016 International Conference on Emerging Technological Trends (ICETT). IEEE, 2016: 1-6.
- [47] Wai M S, Aung S S. Simultaneous opinion lexicon expansion and product feature extraction[C]//2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). IEEE, 2017: 107-112.
- [48] Rodrigues A P, Chiplunkar N N. Mining online product reviews and extracting product features

- using unsupervised method[C]//2016 IEEE Annual India Conference (INDICON). IEEE, 2016: 1-6.
- [49] 邱云飞,陈艺方,王伟,邵良杉.基于词性特征与句法分析的商品评价对象提取[J].计算机工程,2016,42(07):173-180.
- [50] 郭博,李守光,王昊,等. 电商评论综合分析系统的设计与实现——情感分析与观点挖掘的研究与应用[J]. 数据分析与知识发现, 2017, 1(12): 1-9.
- [51] 黄仁, 张卫. 基于 word2vec 的互联网商品评论情感倾向研究[J]. 计算机科学, 2016, 43(S1): 387-389.
- [52] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [53] Le Q, Mikolov T. Distributed representations of sentences and documents[C]//International conference on machine learning. 2014: 1188-1196.
- [54] Brin S, Page L. The anatomy of a large-scale hypertextual web search engine[J]. Computer networks and ISDN systems, 1998, 30(1-7): 107-117
- [55] 徐立.基于加权 TextRank 的文本关键词提取方法[J].计算机科学,2019,46(S1):142-145.
- [56] 颜端武,杨雄飞,李铁军.基于产品特征树和 LSTM 模型的产品评论情感分析[J/OL].情报理论与实践:1-7[2019-12-04].<http://kns.cnki.net/kcms/detail/11.1762.G3.20190816.1008.004.html>.
- [57] Aleebrahim N , Fathian M , Gholamian M R , et al. Sentiment classification of online product reviews using product features[C]// International Conference on Data Mining & Intelligent Information Technology Applications. IEEE, 2011.
- [58] Sharaff A, Soni A. Analyzing sentiments of product reviews based on features[C]//2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2018: 710-713.
- [59] Bijari K, Zare H, Veisi H, et al. Deep Sentiment Analysis using a Graph-based Text Representation[J]. arXiv preprint arXiv:1902.10247, 2019.
- [60] 彭浩,徐健,肖卓.基于比较句的网络用户评论情感分析[J].现代图书情报技术,2015(12):48-56.
- [61] 杜嘉忠,徐健,刘颖.网络商品评论的特征 - 情感词本体构建与情感分析方法研究[J].现代图书情报技术,2014(05):74-82.
- [62] 樊振,过弋,张振豪,等. 基于词典和弱标注信息的电影评论情感分析[J]. 计算机应用, 2018, 38(11): 3084-3088.
- [63] Bross J, Ehrig H. Automatic construction of domain and aspect specific sentiment lexicons for customer review mining[C]//Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, 2013: 1077-1086.
- [64] 张艳辉, 李宗伟. 在线评论有用性的影响因素研究:基于产品类型的调节效应[J]. 管理评论, 2016, 28(10):123-132.
- [65] 赵志滨,刘欢,姚兰,等. 中文产品评论的维度挖掘及情感分析技术研究[J]. 计算机科学与探索,

- 2018, 12(3): 341-349.
- [66] 李然,林政,林海伦,王伟平,孟丹.文本情绪分析综述[J].计算机研究与发展,2018,55(01):30-52.
- [67] 聂卉,刘梦圆.基于细粒度情感的评论感知效用预测研究[J].情报理论与实践,2019,42(09):117-122+153.
- [68] 吴良庆,张栋,李寿山,陈瑛.基于多任务学习的多模态情绪识别方法[J/OL].计算机科学:1-10[2019-12-04].<http://kns.cnki.net/kcms/detail/50.1075.TP.20190814.1712.070.html>.
- [69] 王伟,王洪伟. 特征观点对购买意愿的影响:在线评论的情感分析方法[J]. 系统工程理论与实践, 2016, 36(1): 63-76.
- [70] 张艳丰, 李贺, 彭丽徽, 等. 基于情感语义特征抽取的在线评论有用性分类算法与应用[J]. 数据分析与知识发现, 2017, 1(12): 74-83.
- [71] 吴江, 刘弯弯. 基于信息采纳理论的在线商品评论有用性影响因素研究[J]. 信息资源管理学报, 2017, 7(1): 47-55.
- [72] Ahluwalia R, Burnkrant R E, Unnava H R. Consumer response to negative publicity: The moderating role of commitment[J]. Journal of marketing research, 2000, 37(2): 203-214.
- [73] 游浚,张晓瑜,杨丰瑞.在线评论有用性的影响因素研究——基于商品类型的调节效应[J].软科学,2019,33(05):140-144.
- [74] 李启庚, 赵晓虹, 余明阳. 服务型产品在线评论信息特征对评论感知有用性与购买意愿的影响[J]. 工业工程与管理, 2017 (6): 21.
- [75] 张艳丰, 李贺, 翟倩,等. 基于模糊 TOPSIS 分析的在线评论有用性排序过滤模型研究——以亚马逊手机评论为例[J]. 图书情报工作, 2016(13):109-117.
- [76] 苗蕊,徐健.评分不一致性对在线评论有用性的影响——归因理论的视角[J].中国管理科学,2018,26(05):178-186.
- [77] 李涵昱,钱力,周鹏飞. 面向商品评论文本的情感分析与挖掘[J]. 情报科学, 2017, 35(1): 51-55, 61.
- [78] 徐琳宏, 林鸿飞, 潘宇, et al. 情感词汇本体的构造[J]. 情报学报, 2008(2):180-185.
- [79] 王萝娜, 李端明. 运用加权 TOPSIS 法的学术图书影响力评价方法研究[J]. 图书情报工作, 2018(23).
- [80] Yang Z, Jun M. Consumer perception of e-service quality: from internet purchaser and non-purchaser perspectives[J]. Journal of Business strategies, 2002, 19(1): 19.

附录 A 数据源部分展示

本文使用的实验数据集如表 A1 所示。

表 A1 数据源部分展示

Table A1 Data source part display

brand	comment	creationTime	nickname	Color	productSize	productid	score
骆驼 (CAMEL)	给老公买的!说是穿着很方便!很喜欢! 薄薄的很轻松!喜欢!我也买了一件很 喜欢!	2019-01-14 18:14:22	j***n	黑色男	XL	3313215	5
骆驼 (CAMEL)	东西挺好的,就是太薄了	2018-12-29 19:08:05	j***i	胭脂红女	M	3313215	5
骆驼 (CAMEL)	双十二搞活动买得,很划算,一下买了 四件,质量不错,很轻很薄,穿起来还 算暖和吧。	2019-01-01 22:52:24	j***g	藏蓝男	L	3313215	5
骆驼 (CAMEL)	衣服很轻很薄,颜色适合老人家,穿码 数刚好,含绒量有点少了	2019-01-15 21:31:55	岭***园	藏蓝男	XL	3313215	5
骆驼 (CAMEL)	终于在 13 号凌晨 12 点整抢到三件羽 绒服,价钱十分划算,颜色好看	2019-01-13 22:04:36	岭***园	胭脂红女	XL	3313215	5

表 A1 (续)

Table A1 (Continuation)

brand	comment	creationTime	nickname	Color	productSize	productid	score
骆驼 (CAMEL)	此用户未填写评价内容	2019-01-16 21:50:58	岭***园	银灰女	L	3313215	5
骆驼 (CAMEL)	以前有一件抽烟烧破了, 又来一件一模一样的还便宜很多	2019-01-18 22:35:30	1***p	藏蓝男	L	3313215	5
骆驼 (CAMEL)	挺轻薄的羽绒服, 比较修身, 颜色也不错。自营就是好	2019-01-02 13:05:39	w***5	藏蓝男	L	3313215	5
骆驼 (CAMEL)	超棒的衣服, 便宜送货快!	2019-01-13 16:57:00	G***T	藏蓝男	L	3313215	5
骆驼 (CAMEL)	双十二搞活动买得, 很划算, 一下买了四件, 质量不错, 很轻很薄, 穿起来还算暖和吧。	2019-01-01 22:50:02	j***g	黑色女	M	3313215	5
骆驼 (CAMEL)	还是很不错的, 下次我还会买在的。价格实惠, 得确比较轻薄	2019-01-09 11:17:51	克***8	藏蓝男	XL	3313215	5
骆驼 (CAMEL)	京东产品值得拥有京东产品值得信赖五星好评五星好评五星好评五星红旗迎风飘扬五星红旗迎风飘扬!	2019-01-01 10:30:33	j***v	藏蓝男	L	3313215	5
骆驼 (CAMEL)	很不错。确实比较薄, 保暖性能却还行。如果需要更好保暖的, 建议买厚一些的。	2019-01-09 08:27:39	z***m	藏蓝男	L	3313215	5
骆驼 (CAMEL)	羽绒服质量不错, 也很保暖, 大小合适, 值得推荐购买	2019-01-14 21:34:00	j***0	藏蓝男	L	3313215	5
骆驼 (CAMEL)	没想到这么便宜可以买到羽绒服?, 质量虽无法与鹅绒比, 但适合在南方穿	2018-12-27 23:00:16	1***5	藏蓝男	L	3313215	5

表 A1 (续)

Table A1 (Continuation)

brand	comment	creationTime	nickname	Color	productSize	productid	score
骆驼 (CAMEL)	衣服轻薄, 上身效果也不错挺收身的, 是短装的刚开始看图片还以为是中长装的, 大小合适刚刚好, 整体不错	2019-01-15 19:02:13	l***2	黑色女	M	3313215	5
骆驼 (CAMEL)	轻薄, 却够暖和。是羽绒的, 很不错。	2019-01-02 23:00:54	y***9	藏蓝男	L	3313215	5
骆驼 (CAMEL)	京东自营, 买了好多东西, 非常好	2019-01-10 08:36:59	j***i	藏蓝男	L	3313215	5
骆驼 (CAMEL)	这是第二件了, 今天通知了我, 但没时间! 明天去拿!	2019-01-14 20:50:22	j***f	藏蓝男	L	3313215	5
骆驼 (CAMEL)	商品不错, 服务热情周到, 点赞	2019-01-08 00:04:08	c***r	藏蓝男	L	3313215	5
骆驼 (CAMEL)	不错不错, 很保暖, 下次还会再来, 快递小哥态度很好	2019-01-02 18:13:50	转***谁	藏蓝男	L	3313215	5
骆驼 (CAMEL)	买小了, 换的货, 舒适。	2019-01-06 17:25:38	j***4	藏蓝男	L	3313215	5
骆驼 (CAMEL)	一直在京东买, 省时省力更省钱啊?	2019-01-18 13:09:57	j***z	藏蓝男	L	3313215	5
骆驼 (CAMEL)	衣服很好, 是正品羽绒服。不错!	2019-01-15 18:22:36	Q***3	藏蓝男	L	3313215	5
骆驼 (CAMEL)	衣服非常轻, 初冬穿上正合适, 非常保暖, 一次买了三件	2018-12-30 18:06:32	烨***张	藏蓝男	L	3313215	5
骆驼 (CAMEL)	轻薄好看, 穿得舒服, 价格便宜。	2019-01-07 13:13:33	j***7	藏蓝男	L	3313215	5
骆驼 (CAMEL)	衣服版型好看, 而且很轻薄。	2019-01-18 11:27:24	叶***佳	藏蓝男	L	3313215	5
骆驼 (CAMEL)	给公公买的, 合身	2019-01-19 09:05:31	m***g	藏蓝男	L	3313215	5

附录 B Doc2vec 指标表现

随机森林结合 Doc2vec 向量在不同参数水平下，指标表现如表 B1 所示：

表 B1 随机森林结合 Doc2vec 向量在不同参数水平指标表现

Table B1 Random forest combined with Doc2vec vector performance at different parameter levels

id	min_samples_split	max_depth	recall	accuracy
1	2	10	0.00000	0.91000
2	2	20	0.00000	0.90950
3	2	30	0.00000	0.90950
4	2	40	0.00000	0.90950
5	2	50	0.00000	0.90900
6	2	60	0.00000	0.90950
7	2	70	0.00000	0.90900
8	2	80	0.00000	0.90900
9	2	90	0.00000	0.90900
10	12	10	0.00000	0.91000
11	12	20	0.00000	0.90950
12	12	30	0.00000	0.90950
13	12	40	0.00556	0.90950
14	12	50	0.00000	0.90950
15	12	60	0.00000	0.90950
16	12	70	0.00000	0.90950
17	12	80	0.00000	0.90950
18	12	90	0.00000	0.90900
19	22	10	0.00000	0.91000
20	22	20	0.00000	0.90950
21	22	30	0.00000	0.90950
22	22	40	0.00000	0.90900

表 B1 (续)

Table B1 (Continuation)

id	min_samples_split	max_depth	recall	accuracy
23	22	50	0.00000	0.90950
24	22	60	0.00000	0.90950
25	22	70	0.00000	0.90950
26	22	80	0.00000	0.90950
27	22	90	0.00000	0.90950
28	32	10	0.00000	0.91000
29	32	20	0.00000	0.90950
30	32	30	0.00000	0.91000
31	32	40	0.00000	0.90950
32	32	50	0.00000	0.90950
33	32	60	0.00000	0.90950
34	32	70	0.00000	0.90950
35	32	80	0.00000	0.90950
36	32	90	0.00000	0.91000
37	42	10	0.00000	0.91000
38	42	20	0.00000	0.91000
39	42	30	0.00000	0.91000
40	42	40	0.00000	0.91000
41	42	50	0.00000	0.91000
42	42	60	0.00000	0.91000
43	42	70	0.00000	0.91000
44	42	80	0.00000	0.91000
45	42	90	0.00000	0.91000
46	52	10	0.00000	0.91000
47	52	20	0.00000	0.91000
48	52	30	0.00000	0.90950
49	52	40	0.00000	0.91000
50	52	50	0.00000	0.91000

表 B1 (续)

Table B1 (Continuation)

id	min_samples_split	max_depth	recall	accuracy
51	52	60	0.00000	0.91000
52	52	70	0.00000	0.91000
53	52	80	0.00000	0.91000
54	52	90	0.00000	0.91000
55	62	10	0.00000	0.91000
56	62	20	0.00000	0.91000
57	62	30	0.00000	0.91000
58	62	40	0.00000	0.91000
59	62	50	0.00000	0.91000
60	62	60	0.00000	0.91000
61	62	70	0.00000	0.91000
62	62	80	0.00000	0.91000
63	62	90	0.00000	0.91000
64	72	10	0.00000	0.91000
65	72	20	0.00000	0.91000
66	72	30	0.00000	0.91000
67	72	40	0.00000	0.91000
68	72	50	0.00000	0.91000
69	72	60	0.00000	0.91000
70	72	70	0.00000	0.91000
71	72	80	0.00000	0.91000
72	72	90	0.00000	0.91000
73	82	10	0.00000	0.91000
74	82	20	0.00000	0.91000
75	82	30	0.00000	0.91000
76	82	40	0.00000	0.91000
77	82	50	0.00000	0.91000
78	82	60	0.00000	0.91000

表 B1 (续)

Table B1 (Continuation)

id	min_samples_split	max_depth	recall	accuracy
79	82	70	0.00000	0.91000
80	82	80	0.00000	0.91000
81	82	90	0.00000	0.91000
82	92	10	0.00000	0.91000
83	92	20	0.00000	0.91000
84	92	30	0.00000	0.91000
85	92	40	0.00000	0.91000
86	92	50	0.00000	0.91000
87	92	60	0.00000	0.91000
88	92	70	0.00000	0.91000
89	92	80	0.00000	0.91000
90	92	90	0.00000	0.91000

附录 C 情感向量指标表现

随机森林结合情感向量在不同参数水平下，指标表现如表 C1 所示：

表 C1 随机森林结合情感向量在不同参数水平指标表现

Table C1 Random forest combined with sentiment vector performance at different parameter levels

min_samples_split	max_depth	recall	accuracy
2	10	0.32979	0.931
2	20	0.34574	0.932
2	30	0.33511	0.931
2	40	0.33511	0.931
2	50	0.33511	0.931
2	60	0.33511	0.931
2	70	0.33511	0.931
2	80	0.33511	0.931
2	90	0.33511	0.931
12	10	0.31383	0.929
12	20	0.28723	0.9275
12	30	0.28723	0.9275
12	40	0.28723	0.9275
12	50	0.28723	0.9275
12	60	0.28723	0.9275
12	70	0.28723	0.9275
12	80	0.28723	0.9275
12	90	0.28723	0.9275
22	10	0.29787	0.928
22	20	0.29255	0.928
22	30	0.29255	0.928
22	40	0.29255	0.928

表 C1 (续)

Table C1 (Continuation)

min_samples_split	max_depth	recall	accuracy
22	50	0.29255	0.928
22	60	0.29255	0.928
22	70	0.29255	0.928
22	80	0.29255	0.928
22	90	0.29255	0.928
32	10	0.30319	0.9285
32	20	0.30851	0.929
32	30	0.30851	0.929
32	40	0.30851	0.929
32	50	0.30851	0.929
32	60	0.30851	0.929
32	70	0.30851	0.929
32	80	0.30851	0.929
32	90	0.30851	0.929
42	10	0.28723	0.9275
42	20	0.29255	0.927
42	30	0.29787	0.9275
42	40	0.29787	0.9275
42	50	0.29787	0.9275
42	60	0.29787	0.9275
42	70	0.29787	0.9275
42	80	0.29787	0.9275
42	90	0.29787	0.9275
52	10	0.29255	0.927
52	20	0.28191	0.9265
52	30	0.2766	0.926
52	40	0.2766	0.926
52	50	0.2766	0.926
52	60	0.2766	0.926

表 C1 (续)

Table C1 (Continuation)

min_samples_split	max_depth	recall	accuracy
52	70	0.2766	0.926
52	80	0.2766	0.926
52	90	0.2766	0.926
62	10	0.27128	0.9255
62	20	0.28723	0.927
62	30	0.28723	0.927
62	40	0.28723	0.927
62	50	0.28723	0.927
62	60	0.28723	0.927
62	70	0.28723	0.927
62	80	0.28723	0.927
62	90	0.28723	0.927
72	10	0.29255	0.9275
72	20	0.27128	0.9255
72	30	0.27128	0.9255
72	40	0.27128	0.9255
72	50	0.27128	0.9255
72	60	0.27128	0.9255
72	70	0.27128	0.9255
72	80	0.27128	0.9255
72	90	0.27128	0.9255
82	10	0.27128	0.926
82	20	0.26596	0.9255
82	30	0.26596	0.9255
82	40	0.26596	0.9255
82	50	0.26596	0.9255
82	60	0.26596	0.9255
82	70	0.26596	0.9255
82	80	0.26596	0.9255

表 C1 (续)

Table C1 (Continuation)

min_samples_split	max_depth	recall	accuracy
82	90	0.26596	0.9255
92	10	0.26596	0.925
92	20	0.24468	0.924
92	30	0.24468	0.924
92	40	0.24468	0.924
92	50	0.24468	0.924
92	60	0.24468	0.924
92	70	0.24468	0.924
92	80	0.24468	0.924
92	90	0.24468	0.924

附录 D 均衡数据集指标表现

通过构建均衡数据集，随机森林结合 Doc2vec 向量在不同参数水平下，指标表现如表 D1 所示：

表 D1 均衡数据集下不同参数水平指标表现

Table D1 Performance of different parameter level indicators under equilibrium data set					
min_samples_split	max_depth	accuracy	recall	precision	F1
2	10	0.899	0.87234	0.47953	0.618868
2	20	0.894	0.85638	0.46532	0.602996
2	30	0.894	0.85638	0.46532	0.602996
2	40	0.894	0.85638	0.46532	0.602996
2	50	0.894	0.85638	0.46532	0.602996
2	60	0.894	0.85638	0.46532	0.602996
2	70	0.894	0.85638	0.46532	0.602996
2	80	0.894	0.85638	0.46532	0.602996
2	90	0.894	0.85638	0.46532	0.602996
12	10	0.901	0.88298	0.48538	0.626415
12	20	0.8995	0.88298	0.48116	0.622889
12	30	0.8995	0.88298	0.48116	0.622889
12	40	0.8995	0.88298	0.48116	0.622889
12	50	0.8995	0.88298	0.48116	0.622889
12	60	0.8995	0.88298	0.48116	0.622889
12	70	0.8995	0.88298	0.48116	0.622889
12	80	0.8995	0.88298	0.48116	0.622889
12	90	0.8995	0.88298	0.48116	0.622889
22	10	0.8975	0.85638	0.47493	0.611006
22	20	0.8955	0.85638	0.46939	0.606403
22	30	0.8955	0.85638	0.46939	0.606403
22	40	0.8955	0.85638	0.46939	0.606403

表 D1 (续)

Table D1 (Continuation)

min_samples_split	max_depth	accuracy	recall	precision	F1
22	50	0.8955	0.85638	0.46939	0.606403
22	60	0.8955	0.85638	0.46939	0.606403
22	70	0.8955	0.85638	0.46939	0.606403
22	80	0.8955	0.85638	0.46939	0.606403
22	90	0.8955	0.85638	0.46939	0.606403
32	10	0.8965	0.86702	0.47246	0.611632
32	20	0.898	0.8617	0.47647	0.613636
32	30	0.898	0.8617	0.47647	0.613636
32	40	0.898	0.8617	0.47647	0.613636
32	50	0.898	0.8617	0.47647	0.613636
32	60	0.898	0.8617	0.47647	0.613636
32	70	0.898	0.8617	0.47647	0.613636
32	80	0.898	0.8617	0.47647	0.613636
32	90	0.898	0.8617	0.47647	0.613636
42	10	0.897	0.8617	0.47368	0.611321
42	20	0.8975	0.8617	0.47507	0.612476
42	30	0.8975	0.8617	0.47507	0.612476
42	40	0.8975	0.8617	0.47507	0.612476
42	50	0.8975	0.8617	0.47507	0.612476
42	60	0.8975	0.8617	0.47507	0.612476
42	70	0.8975	0.8617	0.47507	0.612476
42	80	0.8975	0.8617	0.47507	0.612476
42	90	0.8975	0.8617	0.47507	0.612476
52	10	0.903	0.85106	0.4908	0.622568
52	20	0.9	0.84574	0.48182	0.6139
52	30	0.9	0.84574	0.48182	0.6139
52	40	0.9	0.84574	0.48182	0.6139
52	50	0.9	0.84574	0.48182	0.6139
52	60	0.9	0.84574	0.48182	0.6139

表 D1 (续)

Table D1 (Continuation)

min_samples_split	max_depth	accuracy	recall	precision	F1
52	70	0.9	0.84574	0.48182	0.6139
52	80	0.9	0.84574	0.48182	0.6139
52	90	0.9	0.84574	0.48182	0.6139
62	10	0.9015	0.84574	0.48624	0.617476
62	20	0.9015	0.84574	0.48624	0.617476
62	30	0.9015	0.84574	0.48624	0.617476
62	40	0.9015	0.84574	0.48624	0.617476
62	50	0.9015	0.84574	0.48624	0.617476
62	60	0.9015	0.84574	0.48624	0.617476
62	70	0.9015	0.84574	0.48624	0.617476
62	80	0.9015	0.84574	0.48624	0.617476
62	90	0.9015	0.84574	0.48624	0.617476
72	10	0.9015	0.85106	0.48632	0.618956
72	20	0.901	0.85106	0.48485	0.617761
72	30	0.901	0.85106	0.48485	0.617761
72	40	0.901	0.85106	0.48485	0.617761
72	50	0.901	0.85106	0.48485	0.617761
72	60	0.901	0.85106	0.48485	0.617761
72	70	0.901	0.85106	0.48485	0.617761
72	80	0.901	0.85106	0.48485	0.617761
72	90	0.901	0.85106	0.48485	0.617761
82	10	0.9	0.85106	0.48193	0.615385
82	20	0.902	0.85638	0.48788	0.621622
82	30	0.902	0.85638	0.48788	0.621622
82	40	0.902	0.85638	0.48788	0.621622
82	50	0.902	0.85638	0.48788	0.621622
82	60	0.902	0.85638	0.48788	0.621622
82	70	0.902	0.85638	0.48788	0.621622
82	80	0.902	0.85638	0.48788	0.621622

表 D1 (续)

Table D1 (Continuation)

min_samples_split	max_depth	accuracy	recall	precision	F1
82	90	0.902	0.85638	0.48788	0.621622
92	10	0.9	0.85106	0.48193	0.615385
92	20	0.902	0.85638	0.48788	0.621622
92	30	0.902	0.85638	0.48788	0.621622
92	40	0.902	0.85638	0.48788	0.621622
92	50	0.902	0.85638	0.48788	0.621622
92	60	0.902	0.85638	0.48788	0.621622
92	70	0.902	0.85638	0.48788	0.621622
92	80	0.902	0.85638	0.48788	0.621622
92	90	0.902	0.85638	0.48788	0.621622

附录 E 调查问卷

电商商品在线评论有用性影响因素重要性调查问卷

尊敬的专家：

您好！非常感谢您能在百忙之中抽出时间来回答我们的问题。

本问卷是电商商品在线评论有用性研究的重要组成部分。您的经验和意见将对本次研究工作提供非常重要的帮助。影响在线评论有用性的因素有很多，我们调研的目的是对部分因素的重要性进行分析。请回忆在网购过程中，评论的哪些要素会让您认为这条评论是有用的，结合您的知识和经验填写我们的问卷。

请您针对问卷提到的两个指标的相对重要性加以比较。问卷采用 1-9 标度法，请在相应的数字下打“√”，数字标度的含义及说明如表 E1 所示，问卷结构设计如表 E2 所示：

表 E1 标度法赋值含义

Table E1 Scale method assignment meaning

标度赋值	定义	含义
1	同等重要	两个指标，同等重要
3	稍微重要	两个指标，前者比后者，稍重要
5	相当重要	两个指标，前者比后者，明显重要
7	明显重要	两个指标，前者比后者，强烈重要
9	绝对重要	两个指标，前者比后者，极端重要
2, 4, 6, 8		两个指标，重要度介于上述标度两者之间
倒数 1/n		两个指标，比之上述标度重要性相反

表 E2 问卷结构设计

Table E2 Questionnaire structure design

	Y 比 X 重要	X 比 Y 重要	1	2	3	4	5	6	7	8	9
评论文本长度(Y)											
评论内容的情感强度											
(X)											
评论文本长度(Y)											
用户的星级打分(X)											
评论文本长度(Y)											
评论描述产品信息的相关性与全面性(X)											
评论文本长度(Y)											
评论的时间跨度(X)											
评论内容的情感强度											
(Y) 用户的星级打分											
(X)											
评论内容的情感强度											
(Y) 评论描述产品信息的相关性与全面性											
(X)											
评论内容的情感强度											
(Y) 评论的时间跨度											
(X)											
用户的星级打分(Y)											
评论描述产品信息的相关性与全面性(X)											
用户的星级打分(Y)											
评论的时间跨度(X)											
评论描述产品信息的相关性与全面性(Y)											
评论的时间跨度(X)											

攻读硕士学位期间承担的科研任务与主要成果

1. 学术论文

- [1] 耿立校, 王然, 赵丽桥. 基于评论有用性的在线商品评价模型研究[J]. 图书馆理论与实践. (在审中)
- [2] 耿立校, 赵丽桥, 王然. 基于传播者特性的微博舆情扩散模型与仿真研究[J]. 信息资源管理学报. (在审中)

2. 软件著作权

- [1] 耿立校, 王然, 赵丽桥, 等. 基于 web 的图书管理系统 V1.0[Z]. 北京: 中华人民共和国国家版权局, 2019.

致 谢

两年半的研究生时光转瞬即逝，在即将结束之际，感慨万千。两年半的时光中，接触了以前从未学习过的新知识，得到了老师的谆谆教诲和精心的指导，收获了珍贵的师门和同学的情谊，特此向所有帮助过我的良师益友们表达我衷心的感谢！

首先，衷心感谢我的导师——耿立校老师。在科研的道路上，感谢老师细心、耐心的教导、指点，为我拨开迷雾，指点迷津。在学习探索的道路上，感谢老师将我带入了这个全新的世界，给予了正确的指引，才让我有机会去尝试、探索以前从未接触过的领域，并有所收获。耿老师严谨、踏实的工作作风使我敬佩，宽以待人的处世之道令我尊重，积极乐观的心态使我钦佩。老师的言传身教，我会铭记在心，老师的教诲，我会时刻谨记、遵守。在此，向耿立校老师致以最衷心的感谢和诚挚的敬意！

其次，衷心感谢我的同门师兄师姐、师弟师妹以及我的同窗。感谢你们在论文撰写过程中和生活中对我毫无保留的出谋划策，给了我非常多的帮助，谢谢你们陪我度过了美好的研究生生涯。

最后，衷心感谢我的父母、家人。感谢你们在这条道路上的陪伴，你们的鼓励是我前进下去的动力。感谢你们无条件的支持和默默的奉献，帮助我渡过了无数的难关，迎来了收获成果的喜悦。感谢你们为我披荆斩棘，在曲折的道路上始终亮起一束温暖的灯光。

衷心感谢所有帮助过我的老师、同学、家人，是你们的支持和鼓励，才让我义无反顾、无所顾虑的走完了研究生生涯。研究生的学习时光结束了，但是情谊不会结束，感激之心不会熄灭，不断学习的品质不会迷失，希望在未来的我们仍能不忘初心、砥砺前行。