

现代情报
Journal of Modern Information
ISSN 1008-0821, CN 22-1182/G3

《现代情报》网络首发论文

题目：消费者感知价值下企业产品在线评论本体构建研究
作者：李萍，何有世
收稿日期：2024-08-26
网络首发日期：2025-03-17
引用格式：李萍，何有世. 消费者感知价值下企业产品在线评论本体构建研究[J/OL]. 现代情报. <https://link.cnki.net/urlid/22.1182.G3.20250317.1114.002>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

消费者感知价值下企业产品在线评论本体构建研究

Research on the Construction of Enterprise Product Online Review Ontology under Consumer Perceived Value

李萍^{1,2} 何有世¹
Li Ping^{1,2} He Youshi¹

(1. 江苏大学管理学院, 江苏 镇江, 212013; 2. 盐城工学院经济管理学院, 江苏 盐城 224051)
(1. School of Management, Jiangsu University, Zhenjiang 212013, China; 2. School of Economics and Management, Yancheng Institute of Technology, Yancheng 224051, China)

摘 要: [目的/意义] 有效构建产品在线评论本体, 有助于企业深入挖掘在线评论中的产品属性及其相关的消费者情感倾向, 从而进一步用于挖掘消费者需求以辅助商业决策。[方法/过程] 为了使企业能够精准地获取消费者反馈信息, 本文以消费者感知价值为视角, 采用 LDA 模型从在线评论文本语料中挖掘出感知价值主题因素作为本体知识分布的主题面; 在此基础上, 采用句法依存分析、Word2Vec 模型以及 SO-PMI 算法等方法进行本体知识内容的填充。[结果/结论] 以家用智能扫地机器人在线评论文本为语料, 对本文提出的本体构建方法进行了实践与评估, 为产品在线评论本体知识模型的搭建提供了一种新的视角。

关键词: 消费者感知价值; 产品在线评论; 本体构建; 评论挖掘; LDA 模型

分类号: G254

Abstract: [Purpose/Significance] Effectively constructing an online product review ontology helps enterprises to deeply explore the product attributes and related consumer emotional tendencies in online reviews, and further use them to explore consumer needs to assist business decision-making. [Method/Process] In order to enable enterprises to accurately obtain consumer feedback information, the paper took the perspective of consumer perceived value and used LDA model to mine perceived value topic factors from online review corpus as the topic surface of ontology knowledge distribution; On this basis, the paper used methods such as syntactic dependency analysis, Word2Vec model, and SO-PMI algorithm to fill in the ontology knowledge content. [Result/Conclusion] The paper takes the online reviews of household intelligent sweeping robots as the corpus, practices and evaluates the ontology construction method proposed in the paper, providing a new perspective for the construction of the product online review ontology knowledge model.

Key words: consumer perceived value; product online review; ontology construction; review mining; LDA model

近年来, 随着电子商务市场的不断扩大和网络购物的持续发展, 在线评论数据呈现爆发式的增长。在线评论数据包含大量影响消费者购买决策的重要信息, 如产品价格、功能与服务等属性及其相关的消费者情感倾向, 可以被企业用于挖掘消费者需求以辅助商业决策。然

收稿日期: 2024-08-26

基金项目: 国家社会科学基金项目“基于领域本体的企业产品网络口碑危机预警研究”(项目编号: 16BGL088)。

作者简介: 李萍(1982-), 女, 博士研究生, 研究方向: 电子商务与商业智能。何有世(1964-), 男, 博士, 教授, 博士生导师, 研究方向: 电子商务与商业智能。

而，在线评论作为消费者生成内容，具有表达不规范、结构复杂等特点，这无疑会影响企业对在线评论的有效使用，从而降低辅助决策的效率。本体是特定领域中概念及其相互关系的形式化表达，能较好地解决领域内知识描述、共享与抽取等问题^[1]。在线评论中涉及的产品概念领域性较强，构建产品在线评论本体不仅可以规范产品概念表述，还可以揭示产品概念之间的语义关系，如产品特征词间、特征词与情感词之间的关联，这对于企业深入挖掘评论中的产品属性及其情感具有重要价值。

与此同时，海量的在线评论文本数据中包含大量的产品属性信息与消费者情感信息，将其作为本体构建的语料，可以弥补依赖领域专家知识构建的本体对消费者口语词汇理解的不足^[2]。因此，在线评论中的词汇分析成了产品在线评论本体构建的关键。当前，大部分学者主要采用文本统计和文本语义分析等方法从在线评论中抽取产品属性类、情感类词汇以及它们的组合^[3-4]。这类本体构建方法的优点在于产品领域概念及其关系的提取较为准确，但较少关注产品领域概念分类与组织的依据。为产品领域概念的分类与组织提供依据，有助于自上而下地描述产品属性类以及类间的层次关系，进而突出产品在线评论分析的重点属性及内容。面对大量复杂的在线评论数据，如何将产品领域概念进行合理分类与组织以更为准确地描述消费者需求，使得企业能够从在线评论中获取真正有用的消费者反馈信息，是产品在线评论本体构建亟须解决的重要问题。

在市场营销领域，消费者感知价值是消费者对于产品多个方面的认知和评价，是影响消费者购买决策和行为的重要因素之一。在“消费者至上”的市场环境中，消费者感知价值的有效评估可以帮助企业更好地了解消费者的需求，从而提升消费者的满意度。在线评论作为一种重要的产品口碑信息披露形式，包含消费者对于产品各个方面的感知评价，将消费者感知价值作为产品领域概念分类和组织的依据无疑更贴近企业进行产品在线评论挖掘的应用需求。

因此，本文从消费者感知价值视角出发，以产品在线评论文本为语料，采用主题模型挖掘评论中的感知价值主题因素作为领域知识分布的主题面，试图建立一个可以准确描述消费者感知的本体概念框架。最后，采用自然语言处理技术从在线评论文本中提取产品概念及其关系以完成本体概念框架的知识内容填充，从而构建出一个更符合企业评论挖掘实际需求的产品在线评论本体知识模型。

1 相关研究

1.1 产品在线评论本体构建

近年来，随着自然语言处理技术的发展，本体与在线评论情感分析技术的结合正逐渐成

为研究热点，而构建和完善本体就成了研究的核心和关键。海量的在线评论数据包含大量的产品属性和情感信息，学者们已经逐渐意识到产品在线评论本体对于有效挖掘产品在线评论的价值，开始致力于产品在线评论本体构建方法的研究。

类（或概念）描述了产品领域涉及的对象和知识，准确定义类与类层次以形成本体概念结构，是有效构建产品在线评论本体的关键。在已有产品在线评论本体构建研究中，学者们定义类与类层次的思路主要有自底向上和自顶向下两种。第一种自底向上是先定义具体类，然后再把这些类泛化成综合性的类。郭冲等^[3]从抽取在线评论的细粒度意见要素着手，按照评价搭配抽取、搭配倾向预测和特征聚合的步骤实现用于细粒度意见挖掘的领域情感本体构建。张晓勇等^[4]将电商网站的产品评论作为本体构建的数据，首先利用条件随机抽取产品候选术语，然后运用深度学习知识优化聚类方法，进而生成本体的概念层次体系。耿骞等^[5]提出了一种基于领域新词发现的本体演化方法，采用深度学习的算法模型从在线评论文本语料中自动识别新特征，并利用 K-means 算法进行聚类得到最终领域新词，由此实现对本体结构及内容的调整。第二种自顶向下是先定义领域中综合的类，然后再对这些类进行细化。邓斯予等^[6]在构建本体中，首先结合智能手机商品信息以及用户评论中的情感分类情况搭建本体框架，然后插入自然语言处理后的特征——情感词对。Ali M M 等^[7]将产品生命周期管理思想引入本体概念框架中，在此基础上采用自然语言处理技术从在线评论中抽取概念进一步完善本体概念框架。翟夏普等^[8]则通过抽取在线评论中不同粒度的评价对象来初始化本体概念框架。

1.2 消费者感知价值

消费者感知价值的思想最早由管理大师彼得·德鲁克于 1954 年提出，他认为消费者购买和消费的不是产品，而是价值^[9]。虽然其研究没有直接给出感知价值的概念，但消费者层面感知价值思想的提出却为感知价值维度研究奠定了基础。感知价值是一个多维度的变量，这种感知是由消费者的多重感知属性决定，而非单一因素可决定的^[10]。Zeithaml V A^[11]率先从产品的质量和价格两个维度提出了感知价值二维权衡观，认为感知价值是“消费者在权衡利得与利失的基础上形成的对产品或服务效用的总体评价”。之后，Sheth J N 等^[12]从消费者角度构建了一个更为详细的感知价值多维度研究框架，确定了社会价值、情感价值、功能价值、认知价值和情境价值 5 个感知价值维度。在此基础上，Sweeney J C 等^[13]针对耐用消费品，设计了感知价值 PERVAL 量表，将感知价值归纳为功能价值、经济价值、情感价值与社会价值 4 个维度。这些均为后续消费者感知价值多维度研究奠定了良好的基础。

然而，从体验式营销来看，消费者的感知价值本质上是情境性的，消费者的感知价值结

构在不同领域具有不同的维度^[14]。传统商务环境中构建的消费者感知价值维度模型并不一定完全适用于电子商务模式。孟庆良等^[15]结合电子商务平台的特殊性，在感知价值维度中加入网购情境，构建了电子商务模式下的消费者感知价值度量模型，从产品质量、成本、获取产品的时间、方便性、购买享受、情感联系等多种因素衡量消费者感知价值。此外，在感知价值维度划分研究越来越精细的同时，研究方法也日益多样化，部分学者开始采用大数据分析技术开展感知价值多维度研究。Liu C T 等^[16]以电子商务市场为研究背景，通过对大量智能手机产品在线评论数据进行内容分析和特征提取，构建消费者感知价值多维度模型，最终确定了产品质量、过程感知（销售服务及物流等因素）、情感价值和风险认知 4 个感知价值维度。龚雨璐^[17]以图书馆文化创意产品评论文本为研究对象，通过高频词统计和主题分析将图书馆文创产品感知价值划分为 4 个维度：情感价值、社会价值、功能价格和功能质量。

1.3 文献述评

综上所述，如何有效构建产品在线评论本体以提升在线评论信息挖掘的价值已经引起了学术界的广泛关注。当前的研究大多侧重于本体构建自动化程度的提升，主要表现为产品领域概念或概念间关系自动抽取算法的改进，其优点是产品领域概念及其关系的提取较为准确。将这类本体应用于在线评论情感分析之中，虽然能够有效地实现在线评论中不同评价对象的抽取以及评价对象层级的情感分析^[18]，但忽略了评论文本包含的语义信息，即每个评价对象未必都是消费者重点关注的。对于企业而言，若能根据语义信息先从评论文本中抽取出消费者重点关注的评价对象，然后再对该评价对象进行情感分析，则评论挖掘更具有商业应用价值。

在已有产品在线评论本体构建研究中，自底向上的本体概念框架确立方式，虽然能保证从在线评论中获取的产品领域概念较为全面，但由于产品领域概念的分类和组织多依赖于算法模型，往往可能会导致领域知识描述与企业挖掘在线评论的信息需求不符。而自顶向下的本体概念框架确立方式，可将企业挖掘在线评论的信息需求预先纳入领域知识范围的确定和类的定义中，抽取的产品领域概念也就更贴近企业评论挖掘的实际应用需要。然而，这种自顶向下的本体概念框架确立方式在当前产品在线评论本体构建研究中还比较匮乏。在为数不多的研究中，学者们主要围绕产品本身的属性特征来确定领域知识范围以定义综合性的类，尚未考虑企业挖掘在线评论的信息需求。消费者感知价值是消费者对产品的认知，厘清消费者感知价值的维度结构，可以帮助企业了解消费者在不同维度上的价值感知，从而更好地获取消费者对产品的关注点和需求。因此，本文从消费者感知价值视角出发，依据感知价值多维度思想建立产品在线评论本体的语义描述和表达，从而有助于企业从在线评论中挖掘出真

正有用的消费者反馈信息。

2 感知价值主题分析与本体构建

2.1 在线评论的感知价值主题分析

确立在线评论的感知价值主题,即依据感知价值多维度思想对在线评论传递的产品感知评价信息进行分类,是本文产品在线评论本体构建研究的关键。

主题模型是以非监督学习的方式发现大量文本数据中潜在主题的文本挖掘技术^[19]。将主题模型引入在线评论文本分析中,不仅可以识别在线评论文本中存在的主题,还可以获取主题相关的信息,如概念、词汇、情感等^[20]。在众多主题模型算法中,隐含狄利克雷分布(LDA)在在线评论文本挖掘中有着较为广泛的应用,且成功被应用于消费者满意度分析研究之中。例如, Lucini F R 等^[21]采用 LDA 模型从在线评论中识别客户满意维度和形容词,并将其用来预测客户对航空公司的推荐。林伟振等^[22]使用 LDA 模型从在线评论中识别顾客满意维度,并结合机器学习算法构建顾客满意度模型。尤天慧等^[23]利用 LDA 模型从在线评论中提取顾客满意度评估属性,并通过构建向量自回归模型分析顾客满意度与评估属性之间的动态关系。消费者对产品的感知价值是消费者满意度的核心驱动因素,消费者满意维度反映了自身所感知的产品属性。因此,将 LDA 主题模型引入在线评论的感知价值主题分析之中具有一定的可行性。

然而, LDA 模型仅能呈现算法得出的词汇聚类,至于每个类(即主题)的解释性还需要依赖于主题标签的选择^[24]。感知价值多维度模型作为一种用于理解消费者对产品或服务价值的认知框架,为从 LDA 模型分析结果中有效识别感知价值主题提供了重要的依据。鉴于此,在线评论的感知价值主题分析主要包括 3 个步骤: □采用 Python 语言配置网络爬虫采集产品在线评论数据,通过对评论文本进行数据清洗、分词及去停用词等预处理,形成规范的待处理评论语料。□采用 LDA 模型对预处理后的评论语料进行主题抽取,主要包括主题数量确定、主题发现及主题关键词获取等。□基于已有消费者感知价值研究,依据感知价值多维度模型对主题抽取结果进行语义分析,将部分没有实际意义以及与感知价值无关的主题进行剔除,从而确定在线评论的感知价值主题,并以其主题为类别获取感知价值主题关键词。

2.2 感知价值主题与本体框架映射

在本体中,类描述了领域中的对象和知识,是本体的核心和基础。前文 2.1 节中确立的感知价值主题是依据感知价值多维度思想形成的消费者评价主题分类体系。该体系中包含的感知价值维度信息及其对应的感知价值主题类别,可分别对应领域本体中的类,并形成上下

类关系。此外,针对产品隐含特征抽取以及情感词领域依赖性问题,产品在线评论本体中还应添加意见类,以方便建立特征词与情感词的关联。同时,为了消除情感词的领域依赖性,可以将意见类分为静态极性情感词类和动态极性情感词类。其中,静态极性情感词类的情感极性稳定,即情感词本身就承担了正向(或负向)情感,如“好”“差”,故又将其分为正面情感词类和负面情感词类。而动态极性情感词的情感极性会因修饰的产品特征不同而不同,如“高”与“价格”在一起是负面情感,而与“配置”在一起是正面情感,故在此不进行分类。上述类形成产品在线评论本体框架雏形,如图 1 所示。其中,评论对象类的层次展开将依据前文 2.1 节的研究结果进行。

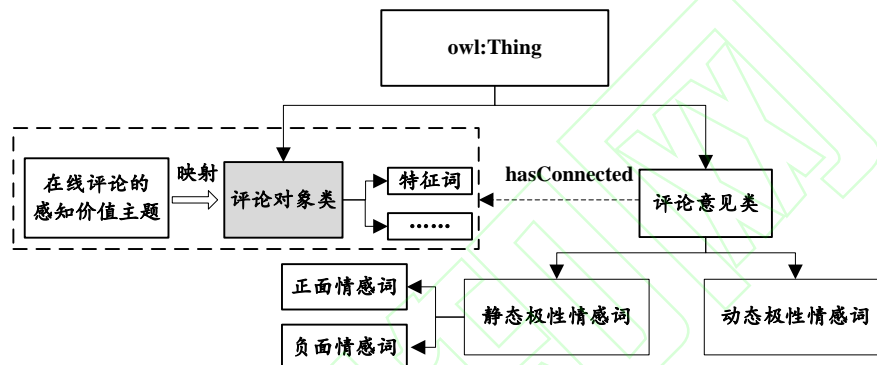


图 1 产品在线评论本体框架雏形

Fig. 1 Prototype of Product Online Review Ontology Framework

2.3 本体知识模型搭建

确定好产品在线评论本体框架,在此基础上进一步进行本体知识内容的填充,从而获取完整的本体知识模型。由图 1 可知,本体知识模型可以看成是一个图结构,节点是评论对象类特征词和评论意见类情感词等本体概念,边则是这些概念之间的语义关系。因此,评论对象类特征词和评论意见类情感词等概念的获取以及概念之间语义关系的确定是构建产品在线评论本体知识模型的核心内容。

2.3.1 评论对象类特征词抽取

在线评论中包含大量消费者关注的与产品特征相关的概念,是获取评论对象类特征词的主要来源。为了便于评论对象类特征词按主题归类,本文将前文 2.1 节中获取的感知价值主题关键词作为候选评论对象类特征词。在文本中,有效的评论对象类特征词一般是以名词(n)、动词(v)、名词短语和动名词(nv)等词组形式存在的^[25]。然而,通过 LDA 模型提取的主题关键词对词性是没有限制的,只要与主题相关的词语都能被 LDA 模型提取出来。因此,本文采用 Python 的 jieba 中文分词库对各感知价值主题的关键词进行词性标注和词频统计。在此基础上,根据评论对象类特征词的词性及词频阈值的设置,从各感知价值主题的关键词

中筛选评价对象类特征词，从而构建按感知价值主题归类的评论对象类特征词集。

2.3.2 情感词抽取

在获取评论对象类特征词的基础上，引入依存句法分析抽取特征词对应的情感词。特征词与其对应的情感词在评论文本语义层面一般存在直接或间接的修饰关系。依存句法分析以评论对象类特征词为中心，通过分析句子成分间的语义联系实现特征词对应情感词的抽取。本文设计抽取规则，使用哈工大语言技术平台 LTP 的汉语言处理模块对产品评论语句进行依存句法分析^[26-27]。若评论语句中存在主谓关系（SBV）、动宾分析（VOB）、动补结构（CMP）、定中关系（ATT）或状中结构（ADV）等依存关系，与前文 2.3.1 节获取的评论对象类特征词集匹配，且其依赖的语法元素词性为形容词（a）、其他名词修饰符（b）、成语（i）和描述用语（z），则该语法元素对应的词为情感词。

2.3.3 情感词情感倾向识别

考虑到静态极性情感词和动态极性情感词的差异，为了确保在线评论情感分析的准确性，本文将这两类情感词分开来进行情感词情感倾向识别研究。具体步骤为：□在前文 2.3.2 节情感词抽取的基础上，根据动态极性情感词的特征，采用人工筛选的方式将动态极性情感词从评论意见类情感词集中过滤掉，保留情感倾向相对稳定的静态极性情感词，如“漂亮”“垃圾”“好看”等作为候选情感词。□采用情感倾向点互信息（SO-PMI）算法对候选情感词进行情感倾向划分，以实现评论意见类情感词中静态极性情感词情感倾向的识别。

2.3.4 概念相似词识别

在产品在线评论本体中，概念相似词识别是对产品在线评论本体概念进行整合的重要途径。通常，评论对象类特征词的相似词识别主要用于揭示特征词之间的等同关系与层级关系，比如，“外表”和“外形”是同义词，“颜色”是“外观”的下位特征词。而评论意见类情感词的相似词识别主要用于揭示情感词之间的等同关系，比如，“好看”和“漂亮”是同义词。目前，相似词识别方法主要有 3 种：借助已有知识库、上下文相关性与文本相似度。其中，借助上下文相关性识别相似词的方法较简单，对于词频较高的词效果较好，适用于在线评论文本语料^[28]。因此，本文采用 Word2Vec 模型来识别概念相似词。相对于其他词向量模型，该模型在语义聚类 and 同义词发现中具有较强的适用性^[29-30]。

2.3.5 本体知识模型生成

在经过上述步骤的处理后，可以获取按感知价值主题归类的本体概念及其关系。产品在线评论本体概念主要包括评论对象类特征词和评论意见类情感词，其语义关系涉及部分关系、属性关系以及关联关系 3 种。其中，部分关系和属性关系是特征词节点之间的关系；关联关

系为特征词节点和情感词节点之间的关系，需要对其中的情感极性进行标注。依据上述处理结果，将特征词和情感词等概念及概念间关系添加至本体框架，以实现产品在线评论本体知识模型的构建。

3 实证研究

3.1 研究对象

近年来，国务院、科技部、工信部等政府主管部门积极推进制造业向高端跨越，鼓励企业加快产品研发，形成竞争优势，家用智能扫地机器人（简称“扫地机器人”）作为一种智能化软硬件结合的产品，是体现当下产品升级发展的典型代表。当前扫地机器人行业处于快速发展期，竞争日趋激烈，企业有效获取产品口碑信息以洞察消费者需求对提升产品竞争优势具有重要的商业价值。京东在线购物平台作为国内的老牌家电网络零售渠道，凭借其规模、物流及服务优势，在扫地机器人的线上销售渠道中处于领先地位，其提供的评论数据较为丰富，足以支持本文的研究分析。

3.2 感知价值主题确定

3.2.1 文本语料获取及预处理

选取京东在线购物平台上的 20 款扫地机器人产品，利用 Python 语言编程抓取在线评论文本总计 37 399 条，去除过短数据、无意义符号以及系统默认评论，最终得到 29 865 条有效产品在线评论文本作为本体构建语料。所有有效的在线评论文本数据采用分词、去停用词等预处理。在分词方面，本文使用 Python 的 jieba 中文分词工具^[31]。为了保证分词的准确性，本文参考产品说明书以及电商网站产品页面信息构建扫地机器人产品自定义词典，主要涉及产品部件及功能方面的词汇，如部件类词汇有“电控水箱”“集尘袋”等，功能类词汇有“扫拖一体”“自清洁”等。在去停用词方面，停用词表的构建以哈工大停用词、百度停用词以及四川大学机器词典智能实验室的停用词为基础。同时，为了保证研究的普遍性，将扫地机器人品牌名称（如“追觅”“科沃斯”等）也加入了停用词表里面。

3.2.2 感知价值主题分析结果

根据前文 2.1 节的处理方法，本文首先采用 Python 语言及 sklearn 库对预处理后的扫地机器人在线评论文本语料进行 LDA 主题分析。LDA 模型无法根据一般认知设定预期结果，可采用困惑度计算法确定主题数量^[32]，如图 2 所示。当主题数量为 9 时，LDA 主题模型困惑度产生局部最小值，之后困惑度值逐渐增加，取其拐点作为主题数量即可，无需过度增加主题数目，故将 LDA 模型的最优主题数量 K 确定为 9。同时，设置 α 为 50/K， β 为 0.01，迭代次数为 300 次，应用 LDA 模型抽取扫地机器人在线评论主题及其关键词如表 1 所示。

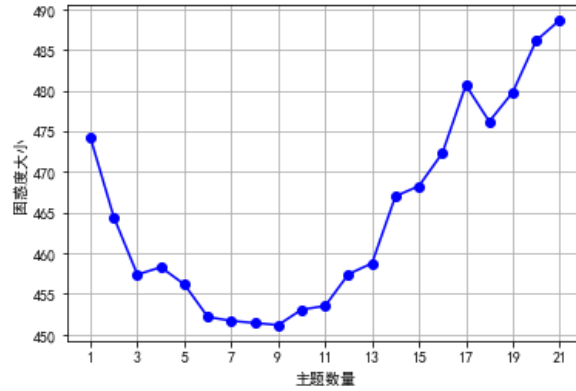


图 2 主题—困惑度变化图

Fig. 2 Topic- Confusion Change Chart

表 1 扫地机器人在线评论主题及其关键词

Tab.1 Themes and Keywords of Reviews on Sweeping Robots

在线评论主题	关键词（前 20 个）	主题描述
Topic1	充电 小时 清扫 分钟 左右 完成 房子 大概 面积 打扫 电量 不到 基站 回去 日常 充满 回充 时间 电池 续航时间	续航性能
Topic2	拖布 自动 清洗 烘干 拖地 功能 路线 自动集尘 基站 集尘 智能 打扫 扫拖一体 避障 干净 自动洗拖布 静音 一体 智能化 地面	清洁性能
Topic3	清扫 地图 设置 区域 房间 自动 建图 智能 全屋 打扫 定时 模式 扫描 禁区 工作 障碍物 强劲 足够 门槛 障碍 App	导航性能
Topic4	操作 简单 App 语音 手机 控制 视频 智能 功能 远程 遥控 父母 便捷 摄像头 识别 监控 随时 对话 在家 清晰	交互性能
Topic5	外观 好看 大气 漂亮 白色 设计 时尚 美观 简洁 上档次 颜色 外 观设计 高级 质感 简约 风格 造型 材质 占地 精致	外观设计
Topic6	客服 问题 售后 解决 服务 耐心 态度 机器 产品 回复 退货 东西 服务态度 联系 售后服务 专业 咨询 解答 换货 处理	客户服务
Topic7	很快 物流 包装 满意 速度 收到 发货 快递 京东 客服 宝贝 送货 服务 服务态度 体验 小哥 配送 到货 下午 使用	物流服务
Topic8	价格 活动 618 赠品 双十一 配件 优惠 便宜 划算 预售 京东 下单 力度 报价 首发 降价 没用 出现 保价 宣传	经济性
Topic9	干净 解放双手 家里 拖地 扫地 打扫 真的 非常 每天 不同 终于 特别 干干净净 神器 方便 时间 不错 再也 光脚 智能	使用体验

通过对上述在线评论主题的关键词及其表达的语义信息进行分析,再结合扫地机器人产品的特性,本文借鉴 Sheth J N 等^[12]的五维度模型、Sweeney J C 等^[13]设计的耐用品感知价值 PERVAL 量表以及电子商务模式下消费者感知价值维度等相关研究^[15-16],实现扫地机器人在线评论主题到感知价值维度的映射。Topic 1、Topic 2、Topic3 与 Topic4 关注扫地机器人续航、清洁、导航以及交互等性能的效用, Topic5 关注扫地机器人外观设计的效用,因此将这 5 个主题归为功能价值维度; Topic6 与 Topic7 关注扫地机器人客户服务及物流服务的效用,因此将这两个主题归为服务价值维度; Topic8 关注扫地机器人经济性(如价格、优惠

活动等)的效用,因此将该主题归为经济价值维度。Topic9 关注扫地机器人使用体验的效用,鉴于使用体验是来自产品具体的特点,如性能、服务等,因此不考虑该主题。基于此,最终获取扫地机器人评论的感知价值主题如表 2 所示。

表 2 扫地机器人评论的感知价值主题分析结果

Tab.2 Perceived Value Theme Analysis Results of Reviews on Sweeping Robots

感知价值维度	在线评论主题	感知价值主题
功能价值	Topic 1	续航性能
	Topic 2	清洁性能
	Topic 3	导航性能
	Topic 4	交互性能
	Topic 5	外观设计
服务价值	Topic 6	客户服务
	Topic7	物流服务
经济价值	Topic 8	经济性

3.3 本体框架搭建

根据前文 2.2 节的处理方法,本文依据阮冈纳赞分面理论思想^[33]搭建扫地机器人在线评论本体框架,如图 3 所示。评论对象类结构是由表 2 中的感知价值维度与其对应的感知价值主题映射得到的,主要包括功能类、服务类和经济性类。其中,功能类与续航类、清洁类、导航类、交互类、外观类构成上下位类关系,服务类与物流服务类和客户服务类构成上下位类关系。

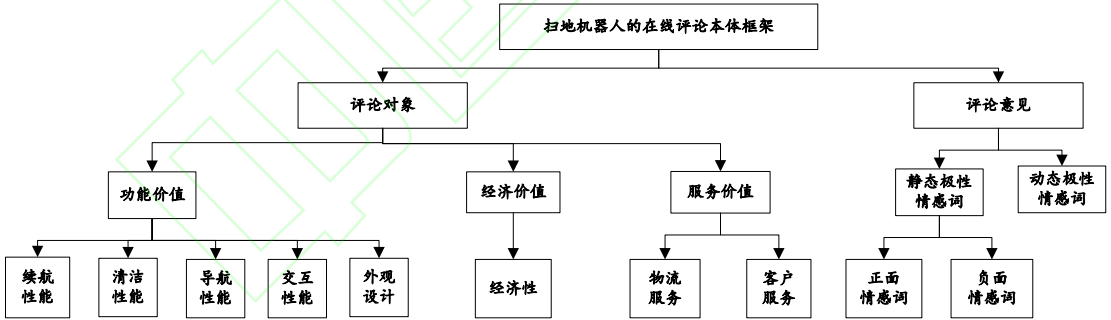


图 3 扫地机器人在线评论本体框架

Fig. 3 Online Review Framework for Sweeping Robots

3.4 本体知识内容填充

3.4.1 特征词与情感词收集

根据前文 2.3.1 节的处理方法,首先以感知价值主题为类别构建扫地机器人评论对象类特征词集,如表 3 所示。

表 3 扫地机器人评论对象类特征词

Tab.3 Object Class Feature Words for Reviews on Sweeping Robots

感知价值主题	评论对象特征词（部分）
外观设计	外观 外形 外观设计 造型 做工 质感 颜值 包装 颜色
清洁性能	清扫 噪声控制 吸尘 擦洗 拖地 除尘 扫拖一体 集尘 扫拖一体 清洁
导航性能	地图 建图 绘制地图 路线规划 避障 越障 模式扫描 禁区 算法
续航性能	耗电量 电池 充电时间 电池容量 电力 续航时间 电池消耗 充电
交互性能	交互 语音 对话 巡视 摄像 实时视频 远程 遥控 语音指令 手机操作
物流服务	物流配送 快递 送货速度 配送 运送 发货 包装 配货
客户服务	解释 解决 客服 客户服务 回复 售前 服务态度 回应 售后服务
经济性	价格 价钱 活动价 赠品 性价比 优惠价 价位 活动 促销

采用 LTP 的 python 接口 pyltp 库对前文预处理之后的扫地机器人评论进行句法依存分析。在此基础上，以上述评论对象类特征词为中心，根据前文 2.3.2 节的方法实现情感词的抽取。在此基础上，以感知价值主题为类别整理扫地机器人评价对象类特征词与情感词对，如表 4 所示。

表 4 评价对象类特征词与情感词对（部分）

Tab.4 Evaluation Object Class Feature Words and Sentiment Words Pairs (Partial)

感知价值主题	特征词	情感词
外观设计	外形	漂亮 大气 耐看 帅气 时尚 好看 高端 精致
	外表	漂亮 秀气 大气 优雅 耐看 简约 好看 高端
	颜色	漂亮 耐看 洋气 美 大方 优雅
	体积	合适 占地 正合适 小巧玲珑 不小
	扫地	干净利落 清爽 一尘不染 无噪声 卖力 利索
清洁性能	拖地	湿 水印 令人满意 干净 赞 印记
	扫拖一体	强大 全能 一步到位 全面 实力
	集尘	噪声 自动化 强大 省心
续航性能	续航时间	长 持久 超长 挺长 满意
	充电时间	快 长
导航性能	建图	准 准确 快且准 快速 慢
交互性能	App 操作	实用 方便 复杂 摆设 简单
经济性	价格	合理 优惠 高 贵 小贵 便宜 公道
客户服务	服务态度	热情周到 耐心 有问必答
物流服务	送货	又快又好 给力 完好无损 快 快速

3.4.2 情感词情感倾向识别

根据前文 2.3.3 节的方法识别情感词情感倾向，首先从情感词中筛选动态极性情感词，并手工建立动态极性情感词存储集合，如表 5 所示。

表 5 动态极性情感词存储集合（部分）

Tab.5 Partial Dynamic Polar Sentiment Word Storage Sets (Partial)

情感词	评价对象	情感倾向
快	电量消耗	负向
快	配送	正向
高	性价比	正向
高	价格	负向
低	工作噪声	正向
低	价格	正向

采用 SO-PMI 算法对上述筛选过的情感词，即静态极性情感词进行情感倾向识别。情感基准词的选取是采用 SO-PMI 算法准确识别情感词情感倾向的关键。本文采用学术界常用的方法^[34-35]，挑选词频高的褒义词和贬义词构建情感基准词表，如表 6 所示。

表 6 情感基准词表

Tab.6 Sentiment Benchmark Vocabulary List

褒义词				贬义词			
不错	干净	方便	时尚	不足	烦恼	严重	复杂
完美	很棒	轻松	强大	遗憾	很脏	虚假	糟心
聪明	容易	很好	准确	落差	最贵	尴尬	混乱
耐心	好用	漂亮	合适	伤害	苦恼	懒惰	烦琐
自由	便宜	优秀	完整	很傻	不爽	烦心	不满

在基准词的基础上进行 SO-PMI 计算，根据计算结果来确定静态情感词的情感倾向。最终通过人工整理，得到了 763 个正面情感词和 232 个负面情感词，如表 7 所示。

表 7 部分静态极性情感词（部分）

Tab.7 Static Polar Sentiment Words(Partial)

正面情感词				负面情感词			
香香的	勤勤恳恳	锃亮	高档	犹豫	乱撞	顽固	美中不足
上档次	便捷	物超所值	整洁	忽悠	莫名其妙	心疼	头疼
耐用	先进	神奇	快乐	发霉	一塌糊涂	郁闷	一塌糊涂
勤劳	爱不释手	完好无损	敏捷	不通	超时	脏兮兮	失效
强劲	灵活	舒服	厉害	太笨	无效	恶劣	一言难尽

3.4.3 概念相似词识别

根据前文 2.3.4 节的处理方法，本文采用 python 的第三方库 gensim 中自带的 Word2Vec 函数来训练词向量。结合各个参数设置适用场景，经过反复实践调整，将 vector_size 设置为 100，window 为 5，min_count 为 3，epochs 为 7，negative 为 10，sg 为 1。经过 Word2Vec 工具训练后，得到了一个包含所有词的词向量模型，然后调用 Word2Vec 提供的方法进行相似度计算。表 8 列示了部分利用词向量得到的相似词。

表 8 相似词列表（部分）

Tab.8 Similar Word List(Partial)

特征词	相似词及其相似度	情感词	相似词及其相似度
外观	(外观设计, 0.8457) (外形, 0.8257) (样式, 0.7993) (外表, 0.7940) (外形, 0.7917) (工艺, 0.7768) (款式, 0.7123) (颜色, 0.7382) (做工, 0.6923) (质感, 0.6298)	好看	(漂亮, 0.9580) (耐看, 0.9268) (大气, 0.9255) (美观, 0.9229) (精美, 0.9195) (精致, 0.9153) (美观大方, 0.8838) (简约, 0.8698)
价格	(价, 0.8392) (价钱, 0.8839) (价位, 0.7721) (促销, 0.6073) (性价比, 0.6160) (打折, 0.6055) (优惠价格, 0.5885)	划算	(实惠, 0.9421) (值, 0.8713) (给力, 0.8584) (经济, 0.8531) (优惠, 0.8468) (便宜, 0.8086) (亲民, 0.7856)
扫地	(吸尘, 0.8238) (洗地, 0.8199) (扫, 0.8159) (扫吸, 0.8155) (连扫, 0.7926) (拖扫, 0.7794) (清洁, 0.6054)	干净	(舒心, 0.9026) (清爽, 0.8767) (整洁, 0.8613) (干净利落, 0.8604) (发亮, 0.8547) (发光, 0.8544)
配送	(物流配送, 0.7479) (送货, 0.7374) (配货, 0.7373) (快递, 0.7368)	快	(给力, 0.9071) (飞快, 0.8788) (迅速, 0.8518) (高效, 0.7526) (快捷, 0.7408)

在使用 Word2Vec 模型时，当两个词的相似度大于 0.5 时，通常可将其视为相似词。由表 8 可以看出，通过词向量计算得到的相似度较为可靠。根据词向量相似度计算值，对扫地机器人特征词与情感词进行整理，主要包括：

- 评价对象类概念归并。对评价对象类特征词的同义词进行归并，相似度值越高，其成为同义词的可能性就越大。比如，“外观”与“外观设计”“外形”“样式”“外表”“外形”相似度超过或接近 0.8，可将其抽取出来归并为同一概念。为了便于概念归并，本文选择词频最高的特征词作为节点共享概念。
- 评价对象类概念层级划分。基于相似度值结合扫地机器人产品实际情况识别概念的上/下位关系词，由此建立评论对象类产品特征概念的层级关系。比如，“颜色”“做工”“质感”等属于“外观”的下属层级概念。
- 意见类情感词的同义词识别及其扩充。与评价对象类特征词归并类似，基于相似度值获取意见类情感词同义词。同时，为了进一步提高情感分析的准确性，本文选用知网 HowNet 情感词典中的正面情感词和负面情感词，对意见类情感词进行同义词扩充。

3.4.4 本体知识模型形成

基于上述步骤的处理结果，将特征词和情感词等概念及概念间关系填充至扫地机器人在线评论本体框架，并在本体中标注情感词，从而确立扫地机器人在线评论本体知识模型，如图 4 所示。

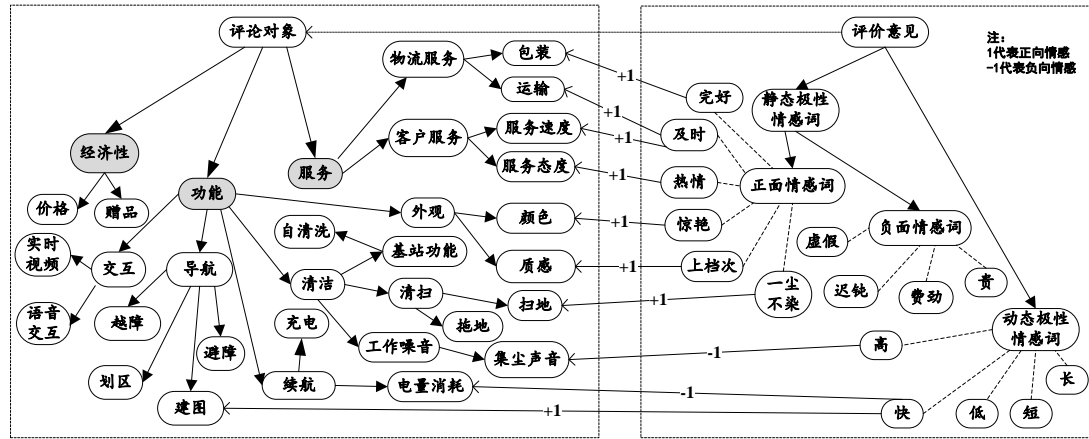


图 4 扫地机器人在线评论本体知识模型（部分）

Fig. 4 Knowledge Model of Online Review Ontology for Sweeping Robots (Partial)

值得注意的是，扫地机器人作为高科技智能电子产品，其产品概念领域性强。在扫地机器人评论对象类概念体系设计中，需要依据产品说明书及电商网站产品参数说明对部分概念及其关系进行修正。比如，对于清洁类的子类设置，按照产品说明书中部件的介绍，可以将清洁类分为基站功能类和清扫功能类，其中基站功能类又可以分为烘干、自清洁和集尘等子类。再如，工作噪声是衡量扫地机器人清洁性能的重要指标，故将工作噪声设为清洁类的子类。

3.5 本体信息评估分析

为了验证采用上述本体构建方法确立的扫地机器人在线评论本体的有效性和准确性，本文从本体概念层与结构层两个方面将其与传统方法构建的本体进行对比评估分析。传统的本体构建方法采用已有文献^[3-4]的研究方法，运用本文构建扫地机器人在线评论本体的评论文本语料构建对比本体。

3.5.1 本体概念层对比分析

图 5 对比了本文方法与传统方法构建的扫地机器人在线评论本体概念体系。传统本体构建方法是采用自底向上的构建思路，从评论语料中抽取产品实体和关系来识别本体底层概念，在此基础上，归纳和抽象出中层概念和顶层概念，如底层概念“型号”和“属性”归纳为中层概念“品牌”，进而再抽象为顶层概念“产品”。而本文方法采用自顶向下的构建思路，从消费者感知价值视角确定领域范围和类，在此基础上依次明确中层概念和底层概念，如顶层概念“功能”细化为中层概念“续航”，再具体为底层概念“充电”和“电池续航”。显然，本文方法获得的本体概念虽没有传统本体构建方法全面，但都是消费者高度关注的，以“功能”“服务”“经济性”为评论对象类名更符合消费者需求，本体概念体系也就更贴近企业产品评论挖掘需求。

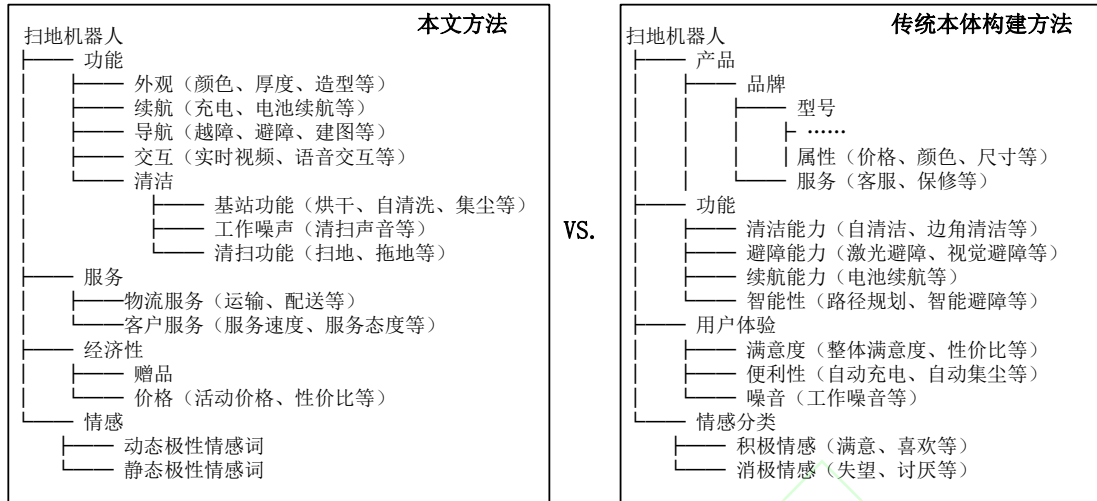


图 5 本体概念层体系片段对比

Fig.5 Comparison of Fragments of Ontology Concept Layer System

3.5.2 本体结构层对比分析

根据 OntoQA 提供的评价指标^[36], 本文选取较为常用的继承丰富度 (IRs)、关系丰富度 (RR) 和属性丰富度 (AR) 3 个指标进行本体结构对比分析, 具体计算公式如公式 (1) ~ (3) 所示:

$$IRs = \frac{\text{所有类的子类总数}}{\text{类的数量}} \quad (1)$$

$$RR = \frac{\text{其他关系的数量}}{\text{继承类的关系数量} + \text{其他关系的数量}} \quad (2)$$

$$AR = \frac{\text{所有类的属性总数}}{\text{类的数量}} \quad (3)$$

对上述本体结构的评估指标进行量化计算, 如表 9 所示。

表 9 本体结构对比评估结果

Tab.9 Comparison and Evaluation Results of Ontology Structure

方法对比	IRs	RR	AR
本文方法	0.703	0.478	0.873
传统本体构建方法	0.931	0.594	0.739

依据表 9 中不同指标的评估结果可知, 在继承丰富性方面, 本文方法构建的扫地机器人在线评论本体 IRs 值为 0.703, 低于传统本体 IRs 值, 表明相对于传统本体, 本文构建的本体更偏向垂直结构, 本体知识描述更为细致。在属性丰富性方面, 本文方法构建的扫地机器人在线评论本体 AR 值为 0.873, 高于传统本体 AR 值, 表明相对于传统本体, 本文构建的本体包含属性更为丰富, 本体具有较强的知识表达能力。然而, 在关系丰富性方面, 本文方法构建的扫地机器人在线评论本体 RR 值为 0.478, 低于传统本体 RR 值, 表明相对于传统本体, 本文方法构建的本体概念关系丰富度较低, 主要原因是本文构建的本体从消费者感知

价值视角明确了本体的知识范畴，故本体实体之间的关联相对于传统本体较少。由此可见，本文构建的本体质量较高，但在后期更新本体时，应进一步扩充语料库以丰富本体概念间关系。

4 结语

面对大量复杂的产品在线评论数据，如何将产品领域概念进行合理分类与组织以更为准确地描述用户需求，使得企业能够从在线评论中获取真正有用的消费者反馈信息，是产品在线评论本体构建亟须解决的问题。针对该问题，本文提出了一种基于消费者感知价值的产品在线评论本体构建方法。不同于以往研究大都采用自底向上的本体构建思路，本文以消费者感知价值多维度思想为依据界定本体知识范围和核心概念，自顶向下确定产品在线评论本体的类及层次，为产品在线评论本体研究提供了新的视角。与已有学者构建的本体类及层次相比，本文重点考虑了企业关注的产品感知概念类，更能贴近企业进行在线评论挖掘的信息需求。然而，消费者感知价值是一种主观构造，其往往会因为时间、环境以及人等因素的不同而产生差异。在未来的研究中，可以引入时间序列、环境变量、人的特征等因素，从更多的视角对消费者感知价值维度进行更深入的探讨，以提高本体知识描述的全面性和准确性。另外，扩充在线评论文本语料以完善和丰富本体语义关系，在后续的研究中，尝试引入泛化能力表现更强的深度学习和大语言模型技术，以提升本体关系动态更新的准确性。

参考文献

- [1] 李爱华, 徐以则, 迟钰雪. 本体构建及应用综述[J]. 情报理论与实践, 2023, 46(11): 189-195.
- [2] 翟羽佳, 王芳. 基于文本挖掘的中文领域本体构建方法研究[J]. 情报科学, 2015, 33(6): 3-10.
- [3] 郭冲, 王振宇. 面向细粒度意见挖掘的情感本体树及自动构建[J]. 中文信息学报, 2013, 27(5): 75-83, 92.
- [4] 张晓勇, 章成志, 周清清. 基于电商产品评论的产品概念层次体系自动构建研究[J]. 情报理论与实践, 2016, 39(6): 120-125.
- [5] 耿骞, 邓斯予, 靳健. 融合词语义表示和新词发现的领域本体演化——以产品评论数据为例[J]. 图书情报工作, 2021, 65(8): 85-96.
- [6] 邓斯予, 耿骞, 靳健, 等. 基于产品评论分析的领域知识库构建与应用[J]. 情报理论与实践, 2019, 42(11): 115-122, 127.
- [7] Ali M M, Doumbouya M B, Louge T, et al. Ontology-Based Approach to Extract Product's Design Features from Online Customers' Reviews[J]. Computers in Industry, 2020, 116: 103175.
- [8] 翟夏普, 安源, 龙艺璇. 本体和深度学习融合的在线评论细粒度情感分析[J]. 北京邮电大学学报, 2023, 46(5): 125-131.
- [9] Drucker P F. The Practice of Management[M]. New York: Harper and Row Publishers, 1954.
- [10] Al-Sabbahy H Z, Ekinci Y, Riley M. An Investigation of Perceived Value Dimensions: Implications for Hospitality Research[J]. Journal of Travel Research, 2004, 42(3): 226-234.
- [11] Zeithaml V A. Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of

- Evidence[J]. Journal of Marketing, 1988, 52(3): 2-22.
- [12] Sheth J N, Newman B I, Gross B L. Why We Buy What We Buy: A Theory of Consumption Values[J]. Journal of Business Research, 1991, 22(2): 159-170.
- [13] Sweeney J C, Soutar G N. Consumer Perceived Value: The Development of a Multiple Item Scale[J]. Journal of Retailing, 2001, 77(2): 203-220.
- [14] El-Adly M I, Eid R. An Empirical Study of the Relationship between Shopping Environment, Customer Perceived Value, Satisfaction, and Loyalty in the UAE Malls Context[J]. Journal of Retailing and Consumer Services, 2016, 31: 217-227.
- [15] 孟庆良, 韩玉启, 吴正刚. 电子商务模式下顾客价值度量模型的构建及应用[J]. 科技进步与对策, 2006(6): 148-150.
- [16] Liu C T, Wang S S, Jia G Z. Exploring E-Commerce Big Data and Customer-Perceived Value: An Empirical Study on Chinese Online Customers[J]. Sustainability, 2020, 12(20): 8649.
- [17] 龚雨璐. 基于感知价值的图书馆文化创意产品评论文本分析[J]. 图书馆理论与实践, 2024(2): 130-136.
- [18] 郑丽娟, 王洪伟. 基于情感本体的在线评论情感极性强度分析: 以手机为例[J]. 管理工程学报, 2017, 31(2): 47-54.
- [19] Vayansky I, Kumar S A P. A Review of Topic Modeling Methods[J]. Information Systems, 2020, 94: 101582.
- [20] Omar M, On B W, Lee I, et al. LDA Topics: Representation and Evaluation[J]. Journal of Information Science, 2015, 41(5): 662-675.
- [21] Lucini F R, Tonetto L M, Fogliatto F S, et al. Text Mining Approach to Explore Dimensions of Airline Customer Satisfaction Using Online Customer Reviews[J]. Journal of Air Transport Management, 2020, 83: 101760.
- [22] 林伟振, 刘洪伟, 陈燕君, 等. 基于在线评论的顾客满意度研究——以健康监测穿戴产品为例[J]. 数据分析与知识发现, 2023, 7(5): 145-154.
- [23] 尤天慧, 陶玲玲, 袁媛. 基于在线评论的顾客满意度评估方法[J]. 运筹与管理, 2023, 32(12): 144-150.
- [24] Onan A, Korukoglu S, Bulut H. LDA-Based Topic Modelling in Text Sentiment Classification: An Empirical Analysis[J]. International Journal of Computational Linguistics and Applications, 2016, 7(1): 101-119.
- [25] 张璞, 李逍, 刘畅. 基于规则的评价搭配抽取方法[J]. 计算机工程, 2019, 45(8): 217-223.
- [26] 周知, 方正东. 融合依存句法与产品特征库的用户观点识别研究[J]. 情报理论与实践, 2021, 44(7): 111-117.
- [27] 王克勤, 刘朝明. 基于在线评论的重要度绩效竞争对手分析的产品设计改进方法[J]. 计算机集成制造系统, 2022, 28(5): 1496-1506.
- [28] 陆勇, 侯汉清. 基于模式匹配的汉语同义词自动识别[J]. 情报学报, 2006, 25(6): 720-724.
- [29] 陈丹华, 王艳娜, 周子力, 等. 基于Word2Vec的WordNet词语相似度计算研究[J]. 计算机工程与应用, 2022, 58(3): 222-229.
- [30] Zhang D W, Xu H, Su Z C, et al. Chinese Comments Sentiment Classification Based on Word2Vec and SVM Per[J]. Expert Systems with Applications, 2015, 42(4): 1857-1863.
- [31] 孙宝生, 敖长林, 王菁霞, 等. 基于网络文本挖掘的生态旅游满意度评价研究[J]. 运筹与管理, 2022, 31(12): 165-172.
- [32] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3(4/5): 993-1022.
- [33] Fernández-López M, Gómez-Pérez A. Overview and Analysis of Methodologies for Building Ontologies[J]. The Knowledge Engineering Review, 2002, 17(2): 129-156.
- [34] 张志飞, 苗夺谦, 岳晓冬, 等. 强语义模糊性词语的情感分析[J]. 中文信息学报, 2015, 29(2): 68-78.
- [35] 齐梦娜, 朱丽平, 李宁. 基于卷积神经网络与情感倾向点互信息算法的农产品情感词典构建[J]. 计算

机应用, 2022, 42(S2): 10-13.

[36] Tartir S, Arpinar I B, Moore M, et al. OntoQA: Metric-Based Ontology Quality Analysis[C]// Proceedings of IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, Houston, USA, 2005.

