E-Commerce Letters 电子商务评论, 2024, 13(3), 8710-8718
Published Online August 2024 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2024.1331066

Hans汉斯

基于LDA主题模型的商品在线评论文本挖掘分析

窦欣怡

贵州大学管理学院，贵州 贵阳

收稿日期：2024年6月28日；录用日期：2024年7月8日；发布日期：2024年8月23日

摘　要

互联网的快速发展给各大电商平台和生产厂家带来机遇的同时也带来了挑战。用户在互联网上购物的同时，产生了海量的评论数据，而在这些评论文本中包含着许多有价值的潜在信息，因此通过对商品评论信息的分析，不仅能让企业掌握更多自身产品和服务中的具体细节信息，同时能够进一步分析用户的消费行为，从本质上发现用户的需求偏好，推进企业实施科学经营决策。本文的研究对象是笔记本电脑，使用爬虫技术获取联想拯救者Y9000P的用户评论，对数据进行预处理、分词与词性标注，采用余弦相似度的方法进行主题数寻优，确定主题数后建立隐藏式狄利克雷模型(Latent Dirichlet Allocation)，挖掘用户高频关注的产品属性，用词典匹配的方法匹配情感词，进行情感倾向分析，得到用户对产品的意见、态度、购买偏好、购买习惯以及购买动机。

关键词

文本挖掘，隐藏式狄利克雷模型，行为分析，在线评论，情感分析

Text Mining Analysis of Online Reviews of Commodities Based on LDA Topic Model

Xinyi Dou

School of Management, Guizhou University, Guiyang Guizhou

Received: Jun. 28th, 2024; accepted: Jul. 8th, 2024; published: Aug. 23rd, 2024

Abstract

The rapid development of the Internet has brought opportunities as well as challenges to major e-commerce platforms and manufacturers. While users are shopping on the Internet, they gener-

文章引用: 窦欣怡. 基于 LDA 主题模型的商品在线评论文本挖掘分析[J]. 电子商务评论, 2024, 13(3): 8710-8718.
DOI: 10.12677/ecl.2024.1331066

ate a large amount of comment data, and these comment texts contain many valuable potential information. Therefore, through the analysis of commodity comment information, enterprises can not only grasp more specific details of their own products and services, but also further analyze users' consumption behavior, discover users' demand preferences in essence, and promote enterprises to implement scientific management decisions. The research object of this paper is notebook computer. It uses crawler technology to obtain user comments of Lenovo savior Y9000P, preprocesses the data, segment and label the part of speech, uses cosine similarity method to optimize the number of topics, establishes LDA model after determining the number of topics, excavates the product attributes that users pay high attention to, uses dictionary matching method to match emotional words, carries out emotional tendency analysis, and obtains users' opinions, attitudes, purchase preferences, habits and motivations.

Keywords

Text Mining, Latent Dirichlet Allocation Model, Behavior Analysis, Online Comments, Sentiment Analysis

1. 引言

随着电子商务的不断发展，以及人们消费观念和习惯的转变，商品的传统销售渠道正逐渐向现代销售渠道改变，网上购物已经成为主流。与此同时，网购平台发展逐渐成熟，线上消费产生了大量评论数据，针对这些数据中隐藏的信息进行挖掘分析，获取用户在购买产品时的心理、需求和满意度等一系列行为信息，可帮助企业发现用户的偏好习惯、需求关注以及产品购买行为的影响因素等，进而推进企业实施科学经营决策。

然而，对企业来说，想要从海量的评论数据中获取有价值的信息，就必须对数据进行进一步加工，并借助恰当的工具进行分析，寻找用户评论数据中所隐藏的关联信息，进而更好地对竞争产品进行对比分析、规划产品的未来营销策略。

文本挖掘是数据挖掘的分支，属于一项新兴的领域，经过几年的发展，目前文本挖掘已经成为一项热门研究。通过对用户评论进行文本数据挖掘，进而分析用户行为，从大量的评论数据中挖掘到有价值的信息，这将对企业做出有针对性的营销策略具有重要意义。本文选用 LDA 模型对处理后的数据进行建模，挖掘用户高频关注的产品属性，获取用户对这些特征的意见和态度，了解消费者的购买偏好、购买习惯以及购买动机，结合分析结果与行为科学理论，分析用户购买行为。

2. 相关研究

2.1. LDA 主题模型相关研究

主题模型(Topic Model)作为一种有效的文本数据挖掘方法，已经受到了学术界的广泛重视。在文本挖掘过程中，由于大量的数据是非结构化的，所以要从这些数据中得到相应的、需要的信息是较为困难的。主题模型可以识别文档中的主题，挖掘隐藏在主题中的信息，在主题聚合、提取非结构化文本中信息、特征选择等方面有着广阔的应用前景。Blei [1]等人提出的 Latent Dirichlet Allocation (LDA)是其中最具代表

性的模型，通过引入了主题和词汇的潜在狄利克雷分布，有效的解决了传统主题模型中的过拟合问题。

国内，许多在线评论的研究中都有主题模型的应用，例如评论满意度分析、评论推荐、无效评论发现和评论情感分析等[2]。关鹏[3]等为了有效确定科技情报分析中 LDA 主题模型的最优主题数目，利用主题相似度度量潜在主题之间的差异，结合困惑度，给出了一种既考虑主题抽取效果，也考虑模型对新文档的泛化能力的 LDA 最佳主题数量的算法。实验结果显示，与单纯运用困惑度相比较，所提出的最佳 LDA 主题数目的确定方法的主题抽取查准率较高。孙红[4]等人将二项分布应用到 LDA 的基本模型中，进而提出一种具有判别学习能力的 LDA 模型，该模型增加了词语的识别能力。通过理论分析与对比试验，结果显示，改进后的 LDA 模型在处理聚类问题方面要比 LDA 和 VSM 好得多。能够发现，在用户评论分析领域，LDA 模型已经得到广泛运用，相关研究也较为成熟，这为本文提供了一定的参考。

综上所述，国外关于文本挖掘的研究早在 20 世纪 50 年代就已开始，经过半个多世纪的发展，英文文本挖掘在数据挖掘算法研究的基础之上已经形成了比较完善的理论与技术体系，并且随着文本挖掘技术的不断发展进步，国内也已有了比较完善的文本挖掘技术[5]，在文本挖掘算法和情感倾向分析领域的研究都取得了不错的成绩，并且开始热衷于对一些非逻辑性、开放度较高的文本进行研究。

2.2. 消费者行为分析相关研究

网购平台的消费者行为主要为消费者消费之后到平台进行评论评分反馈、评论评分、评论的文本和评论情感等特征。许多学者研究表明在线评论对市场和销售存在重要的影响，并且重点研究消费者的评论评分特征，包括评分长度和评分等特征，表明评论长度和产品质量相关，体现消费者的评论意愿。并且越来越多的学者转向对评论的情感内容进行分析。

2017 年，彭丽徽[6]从用户的评论长度，图片数量，情感强度和产品属性特征等对在线评论的有用性进行研究。构建多元线性回归模型，以亚马逊电商平台的手机在线评论数据进行实证，对不同品牌声誉感知的手机进行研究和验证。说明在线评论特征有用性需要考虑品牌声誉的影响。2019 年，罗汉洋[7]研究网络口碑的影响机制和作用路径，将消费者信任作为中介变量，将用户性别、用户需求和用户感知作为调节变量，构建模型。得出结论:评论数量对男性消费者感知评论可信度的影响更大，并且感知可信度对男性情感信任的影响也更大;对于女性消费者，感知评论理性强度对其感知评论可信度的作用更强。

2.3. 情感分析相关研究

用户的评论中，很可能隐藏着用户的某些情感倾向，情感分析与分类则是对具有感情色彩的文本进行分析、处理、归纳和推理的一种过程。Nogueira [8]等人利用观点挖掘技术和社会网络分析，从消费者的角度提出了一个品牌资产分析模型。该模型的应用结果表明，品牌资产可以从虚拟社交网络中的数据中进行分析，揭示消费者在这种环境下是如何感知品牌的。随着互联网的发展，Yang Cheng [9]等人提出了一种从电子商务平台获取有用在线评论的新方法，构建了产品评价指标体系，并提出了基于在线评论的意见挖掘和情感分析的产品改进策略。该方法可应用于产品评价和改进，尤其适用于需要迭代设计的产品，并可在线浏览大量用户评论。Shigang Hu [10]等人利用评论者可信性分析和细粒度特征情感分析来扩充启发式驱动的用户兴趣分析，以设计一个稳健的推荐方法。该模型包括候选特征提取、评价者可信度分析、用户兴趣挖掘、候选特征情感分配和推荐五个模块。其推荐方法不但利用了数字评分，而且利用了与特征、顾客偏好和评价者可信度相关的情感表达，对各种替代产品进行定量分析。

综上，本文尝试借助 LDA 文本挖掘模型对在线评论信息展开挖掘研究，首先采集京东商城中的相关评论为数据源，接着对爬取的数据做预处理，最后利用 LDA 模型进行建模，最后根据建模结果对消费者行为特征进行深入研究。

3. 研究方法

LDA 模型是一种表示文档层与主题层，主题层与词汇层之间的联系的三层贝叶斯模型。LDA 模型通常用一种概率分布来确定一篇文档的主题，并以概率分布的形式表示各个主题。其中，一篇文档可以对应一个或者多个主题，一个主题对应一个词语表达[11]。

LDA 模型将每个文档视为单词频率向量，然后在数学上将文本注释数据信息矢量化，以便于建模。然后建立一个单词袋模型，LDA 模型见图 1 所示。

图中个符号含义见表 1 所示。

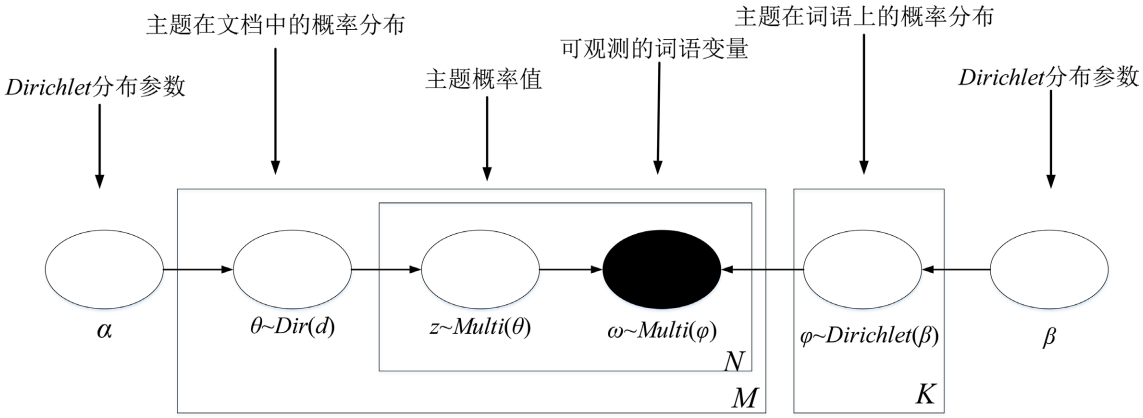

Figure 1. Diagram of LDA topic model
图 1. LDA 主题模型图

Table 1. Meaning of symbols in the LDA topic model diagrams
表 1. LDA 主题模型图中符号含义

符号	含义
N	词语的数量
M	文档的数量
K	主题的数量
α	每篇文档中潜在主题的多项式分布的狄利克雷先验参数，是 K 维向量(记作 $\vec{\alpha}$)
β	每个主题下词语的多项式分布的狄利克雷先验参数，是 N 维向量(记作 $\vec{\beta}$)
θ~Dir(d)	文档 d 的主题概率分布，$P(\theta)$~Dirichlet(α)，是 K 维向量(记作 $\vec{\theta}$)，表示文档中各个主体所占的比重；所有文档的主题概率分布 $\{\vec{\theta}_d\}_{d=1}^{M}$，为 M × N 阶矩阵(记作 $\underline{\theta}$)
φ~Dirichlet(β)	主题词在词语表上的概率分布 $P(\varphi)$~Dirichlet(β)，是 N 维向量(记作 $\vec{\varphi}$)；所有主题词在词语表达式上的概率分布 $\left\{\{\vec{\varphi}_{ji}\}_{i=1}^{K}\right\}_{j=1}^{N}$，为 K×N 阶矩阵(记作 $\underline{\varphi}$)，其中元素 φ_{ij} 表示第 j 个词语归属于第 i 个主题
z~Multi(θ)	由概率分布 $P(\theta)$产生的离散随机变量，表示文档 d 中词语 n 属于主题 z 的概率
ω~Multi(φ)	可观测的离散随机变量(即文档中的词语)，且由 $P(\omega\|z,\varphi)$条件概率生成

由图知，LDA 的联合概率为：$P(\theta,z,\omega\,|\,\alpha,\beta) = P(\theta\,|\,\alpha)\prod_{n=1}^{N}P(z_n\,|\,\theta)P(\omega_n\,|\,z_n,\beta)$。在建模过程中，将 ω 作为观察变量，θ 和 z 作为隐藏变量，再使用 EM 算法求解出 α 和 β 的值。

建立模型首先需要确定 LDA 模型下主题的个数，确定选择多少个主题才是最合适的。选择合适的主题数一般有三种方法，第一种是根据经验直接确定主题数。第二种方法是基于困惑度，它通常用来度量某个概率分布或者概率模型预测样本的好快程度。第三种是利用各主题间的余弦相似度来度量主题间的相似程度。从词频入手，计算它们的相似度，用词越相似，则内容越相近[12]。

本文选用的是第三种方法，首先用余弦相似度函数将每两个主题组合，计算它们的余弦相似度，余弦相似度值越小，则表示它们越相似，进而对应的模型也是最优的。具体步骤如下：

1) 取初始主题数 k 值，得到初始模型，计算各主题之间的相似度(平均余弦距离)。

2) 增加或减少 k 值，重新训练模型，再次计算各主题之间的相似度。

3) 重复步骤(2)直到得到最优 k 值。

假定 A 和 B 是两个 n 维向量，A 是$(A_1, A_2, \ldots, A_n)$，B 是$(B_1, B_2, \ldots, B_n)$，则 A 与 B 的夹角θ的余弦值通过公式(1)计算：

$$\cos\theta = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2 \sum_{i=1}^{n} (B_i)^2}} \tag{1}$$

4. 数据处理及建模分析

4.1. 数据采集选取

通过对比发现京东商城中的电子产品用户评论量较多、评论相关度较高、评论内容较为客观，因此选取销量最高的联想拯救者 Y9000P 笔记本电脑进行分析，采集的内容包括评论内容、评论时间以及评分，总共爬取到用户评论 5769 条。

为了保证后续分析工作的准确性，需要对采集到的评论进行初步处理，首先去重，将重复采集的评论以及无意义的评论进行删除，以提高文本的准确性[13]。之后，由于许多用户评论习惯用表情、颜文字等非规范性符号，例如“ ╰(‘□’)╯ ”、“enmm”等，需要对这些内容进行删除。最后剩余数据 2600 条。

4.2. 分词、去停用词以及词性标注

文本挖掘中数据处理最基本的步骤便是分词。分词是将一个单词序列切分成单个单词的过程，是保证后续任务顺利进行的首要工作。本文选用最常用的 jieba 分词包，它是一个专门的中文分词包，是 Python 内的一个分词开源库。

部分分词结果见表 2：

Table 2. Example of segmentation results
表 2. 分词结果示例

非常 棒 配置 一流 全 金属 外壳 手感 很 棒
有 小键盘 和 全 尺寸 方向 键 接口 很 丰富 功能 也 很 不错。
不过 配备 的 全 功率 充电器 相当 大 重量 也 不轻 出门 的 话 还是 比较 重 的

通过上表可以看出，“的”“也”“了”等词语出现频率很高，但这些词语并没有实际的意义以及分析价值，不仅会影响结果的准确新，还会加大工作量，所以需要设置停用词。

经过去除停用词后，原本冗长的评论就会变得精简许多，大大减少了词语数量，节省了空间，为后

续工作准确性奠定了基础。进行完分词和去除停用词处理后，需要对词语进行词性标注，以便后续的研究分析。Jieba 分词包中，词典词性标记采用的是 ICTCLAS 的标记方法。

4.3. 基于 LDA 主题模型的文本语义挖掘

数据预处理完成后，进行词云图的绘制，它能够把用户评论的内容简单直接的呈现出来。在词云图中，出现频率越高的词语会更大以及更靠近图的中心，大小不一、颜色各异的字体非常有利于区分重要主题和次要主题。

本文选取词频最高的前 100 个词语，使用 WordCloud 模块绘制词云图，见图 2 所示：

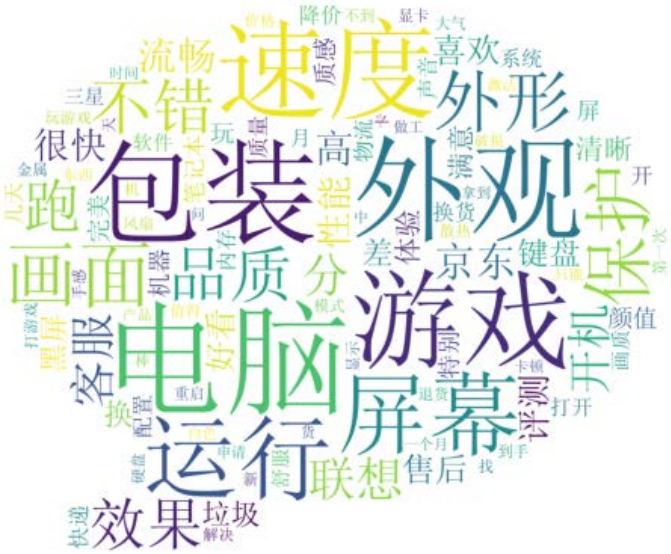

Figure 2. Clouds of the top 100 words in terms of word frequency
图 2. 词频前 100 词云图

通过词云图可以看出，“速度”“包装”“外观”“画面”“游戏”这些词语在图中的字体更大，颜色更明显，说明这些词语出现的频率较高，同时这也反映出用户在购买笔记本时的关注重点。可以看出图中词汇除中性词外，大多是褒义词，如“喜欢”“清晰”“不错”“满意”等，说明大部分用户对这款笔记本呈好评态度，产品符合大多数用户的喜好。通过词云图，能够对用户关于此款笔记本的关注点有了一个简单的了解，但词云图仅仅只是根据词频统计进行绘制，只展示了词语出现的频率，丢失了很多关联信息，因此需要进一步的建模分析。

由于在商品评论中存在用户的情感倾向，可以将情感倾向理解为用户对该商品的态度是积极、消极还是中立的，也就是通常所说的正面情感、负面情感、中性情感。

对评论情感倾向进行分析，进而分别确定正面评论和负面评论的主题数寻优，首先要匹配情感词，主要采用词典匹配的方法，本文使用的是知网发布的“情感分析用词语集(beta 版)”进行情感词匹配。将其中的“中文正面评价”“中文正面情感”两个词表合并，并给每个词语赋予初始权重 1，作为正面评论情感词表。将“中文负面评价”“中文负面情感”两个词表合并，并给每个词语赋予初始权重−1，作为负面评论情感词表。

用余弦相似度函数将每两个主题组合，计算它们的余弦相似度[14]，以下分别是正面评论和负面评论的余弦相似度曲线，如图 3 所示：

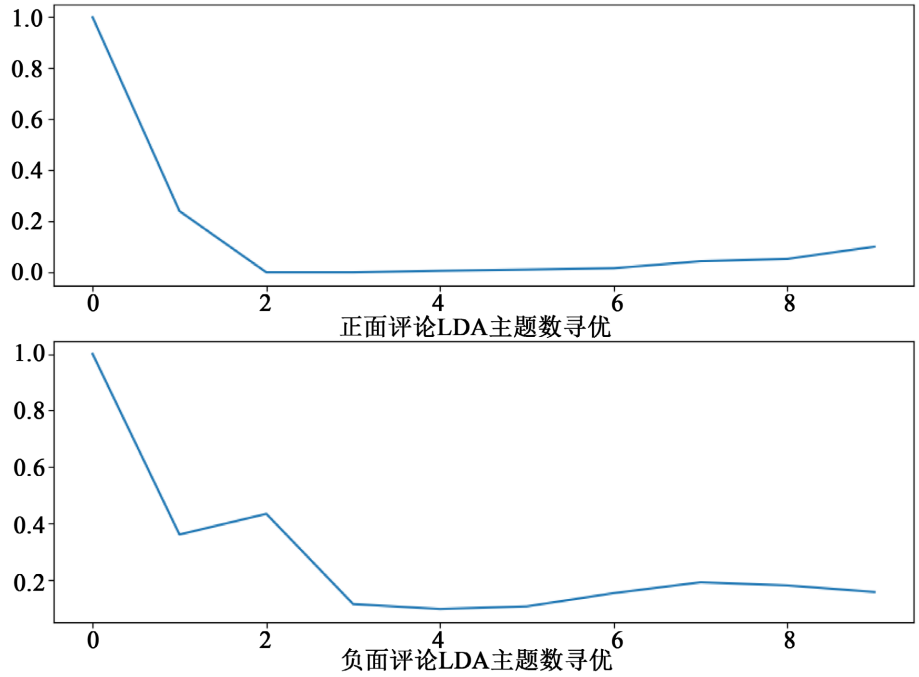

Figure 3. Positive and negative review topic count search for optimization
图 3. 正、负面评论主题数寻优

由图 3 的余弦趋势图可以看出，正面情感词的主题个数 K 均取 2 时，余弦相似度曲线达到最低点，负面情感词的主题数 K 取 3 时，余弦相似度曲线达到最低点，因此正、负面情感词的主题数分别取 2 和 3 较为合适。

根据主题数寻优结果，使用 Python 的 Gensim 模块对正、负面评论数据分别构建 LDA 主题模型，设置主题数 K 分别为 2 和 3，经过 LDA 主题分析后，每个主题下生成最可能出现的前 10 个词语以及对应的概率。每个主题频率前 10 的词语如表 3 所示：

Table 3. Positive comment subject words
表 3. 正面评论主题词

主题 1		主题 2	
主题词	词频	主题词	词频
外观	0.027	不错	0.044
速度	0.019	喜欢	0.028
很快	0.018	电脑	0.020
游戏	0.018	画面	0.018
运行	0.015	效果	0.017
外形	0.014	品质	0.017
包装	0.014	跑	0.013
联想	0.013	屏幕	0.011
性能	0.013	质感	0.011
满意	0.009	高	0.010

根据表 3 能够看出，我们可以发现关于主题 1 的显著趋势。主题 1 主要集中在该电脑的外观设计和实际性能上。从“速度”和“运行”这两个词汇中，我们可以清晰地感受到用户对于电脑处理速度和整体运行流畅性的赞同，这无疑体现了用户对于电脑性能方面较高的满意程度。在购买笔记本电脑时，性能始终是消费者最为关注的因素之一，因为它直接关系到用户在使用过程中的实际体验。

另一方面，“包装”“外形”和“满意”等词汇频繁出现，则反映出用户对这款电脑外观设计的高度认可与重视。这不仅包括产品的外观造型和材质选择，更涵盖了产品的包装设计和细节处理。此外，从用户反馈中还可以看出，物流服务的效率和准确性也是用户所关注的重点，因为良好的物流体验可以确保产品安全、快速地送达用户手中。

在主题 2 中，我们可以看到用户的关注点转向了这款电脑的屏幕效果。通过“画面”“品质”和“质感”等词汇，我们可以推断出这款电脑的屏幕质量较好。而“喜欢”和“不错”等正面评价则进一步印证了用户对这款电脑屏幕的高度认可。这种对于屏幕效果的重视，也体现了现代消费者对于电子产品视觉体验的追求和关注。

Table 4. Negative review subject words
表 4. 负面评论主题词

主题 1		主题 2		主题 3	
主题词	词频	主题词	词频	主题词	词频
系统	0.045	差	0.035	电脑	0.033
游戏	0.029	垃圾	0.030	售后	0.030
黑屏	0.017	屏幕	0.027	客服	0.025
声音	0.017	高	0.024	换	0.017
质量	0.016	不好	0.021	京东	0.017
机	0.016	包装	0.019	开机	0.016
卡	0.014	解决	0.017	体验	0.016
机器	0.012	退货	0.016	跑	0.015
慢	0.011	外观	0.015	软件	0.015
跑	0.010	激活	0.012	联想	0.014

根据表 4 可以看出，对于主题 1，主要是用户对于电脑性能等方面的态度，能够看到部分用户对于屏幕、声卡、系统等方面格外关注。主题 2 则反映出用户对于电脑外观、包装的不满，因此即使是电子产品，包装以及外观同样是用户关注的重点。主题 3 中，“京东”“客服”“售后”等词语出现频率较高，这说明用户在购买电脑后出现的售后问题并没有得到很好的解决。

根据表 4 我们可以进一步解读用户对电脑产品的多元态度。首先，对于主题 1，用户的焦点主要集中在电脑的核心性能上。从数据中可以看到，部分用户对于屏幕、声卡以及操作系统等方面有着特别的关注。他们希望这些核心组件能够达到自己的期望标准，以确保电脑在使用过程中能够流畅、高效地运行。屏幕清晰度、声卡音质以及系统稳定性等因素，都是用户在选择电脑时的重要考虑因素。

其次，主题 2 则揭示了用户对于电脑外观和包装的不满。即使是电子产品，其外观设计和包装细节同样是用户非常关注的方面。用户期望电脑不仅性能出色，同时在外观上也要具有吸引力和质感。包装的完好程度、设计的精致程度等都会影响到用户的购买体验和满意度。因此，制造商在提升产品性能的

同时，也不应忽视外观和包装的设计。

最后，主题 3 中“京东”“客服”和“售后”等词汇的高频出现，暗示了用户在购买电脑后遇到的售后问题并未得到妥善解决。这表明在电子产品的购买过程中，售后服务的质量同样重要。用户期望在遇到问题时能够得到及时、有效的解决方案，而不仅仅是产品本身的质量保证。因此，对于电商平台和制造商来说，提升售后服务质量，增强用户满意度，将是未来需要重点关注和改进的方面。

5. 结论

本文借助 LDA 主题模型对在线产品评论展开研究，进而分析用户购买行为以及用户关注度，用户的购买行为受多方面因素影响，以联想拯救者 Y9000P 在京东的用户评论为例进行实例分析，并从中获得了用户的购买行为信息：首先用户对于该电脑的性能尤为关注，其次电脑屏幕以及外观也是用户决定是否购买以及是否满意的关键因素，最后，用户对于客服和售后的满意度较低。

通过分析，企业首先应发展新的宣传方式，让人们通过口碑了解产品，树立消费者的价值感知、价格感知等。其次，企业可采用体验营销的模式，令用户在不激活系统的情况下更好地体验产品，进而收集用户反馈，了解用户需要，更好地改进产品。最后，为提高产品口碑，企业应注重对消费者服务需求的感知，优化服务过程，培训更为专业的客服，为用户提供满意的售后服务。

参考文献

[1] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.

[2] 王庆福, 王兴国. 基于 LDA 的网络评论主题发现研究[J]. 无线互联科技, 2016(11): 103-104.

[3] 关鹏, 王曰芬. 科技情报分析中 LDA 主题模型最优主题数确定方法研究[J]. 现代图书情报技术, 2016(9): 42-50.

[4] 孙红, 俞卫国. 改进 LDA 模型的短文本聚类方法[J]. 软件导刊, 2021, 20(9): 1-6.

[5] 张尧政, 邓少灵. 基于文本情感分析的企业网络舆情应对策略比较研究[J]. 电子商务, 2019(5): 32-35.

[6] 彭丽徽, 李贺, 张艳丰, 陈远方. 基于品牌声誉感知差异的在线评论有用性影响因素实证研究[J]. 情报科学, 2017, 35(9): 159-164.

[7] 罗汉洋, 李智妮, 林旭东, 于素敏. 网络口碑影响机制: 信任的中介和性别及涉入度的调节[J]. 系统管理学报, 2019, 28(3): 401-418.

[8] Nogueira, E. and Tsunoda, D.F. (2018) A Proposed Model for Consumer-Based Brand Equity Analysis on Social Media Using Data Mining and Social Network Analysis. *Journal of Relationship Marketing*, **17**, 95-117. https://doi.org/10.1080/15332667.2018.1440141

[9] Yang, C., Wu, L., Tan, K., Yu, C., Zhou, Y., Tao, Y., *et al*. (2021) Online User Review Analysis for Product Evaluation and Improvement. *Journal of Theoretical and Applied Electronic Commerce Research*, **16**, 1598-1611. https://doi.org/10.3390/jtaer16050090

[10] Hu, S., Kumar, A., Al-Turjman, F., Gupta, S., Seth, S. and Shubham (2020) Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation. *IEEE Access*, **8**, 26172-26189. https://doi.org/10.1109/access.2020.2971087

[11] 秦春秀, 祝婷, 赵捧未, 张毅. 自然语言语义分析研究进展[J]. 图书情报工作, 2014, 58(22): 130-137.

[12] 张良均. R 语言数据分析与挖掘实战[M]. 北京: 机械工业出版社, 2015.

[13] 李春晓, 李辉, 刘艳筝, 等. 多彩华夏: 大数据视角的入境游客体验感知差异深描[J]. 南开管理评论, 2020, 23(1): 28-39.

[14] 刘兵, 郑承利. 基于 EMD 特征提取的高频面板数据自适应聚类方法[J]. 统计与决策, 2022, 38(10): 16-20.