

10th International Conference on Information Technology and Quantitative Management

Analyzing Public Perception of Educational Books via Text Mining of Online Reviews

Liu Ye¹, Yin Yimeng², Cheng Wei³

*School of Management Science and Engineering,
Hebei University of Economics and Business
Shijiazhuang, Hebei 050061, China*

Abstract

With the development of the era of big data, online reviews have become an important basis for influencing the decisions of consumers, sellers and producers. The research on the publishing of educational books from the perspective of public demand side can provide new ideas for the development of book publishing. This paper explores the public's perception of educational books by text mining reviews of educational books on Douban Books, using a combination of keyword analysis, topic analysis and aspect-level sentiment analysis using tools such as Python. The results indicate that the content quality of educational books strongly influences users' perception. This study reveals the public's perspectives on different aspects of educational books, aiding publishers and authors in gaining deeper insights into areas that require attention and improvement. Based on the analysis results, scientific recommendations are proposed to enhance the quality and publishing standards of educational books, assisting publishers in making publishing decisions that meet the evolving needs of the public. This research contributes to the development of educational book publishing by providing valuable insights and recommendations.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the Tenth International Conference on Information Technology and Quantitative Management

Keywords: book publishing research; Aspect level sentiment analysis; Text mining; Douban Books

1. Introduction

In recent years, with the continuous development of the digital economy, the "big data" environment has gradually taken shape and massive data have emerged. Text mining technology has become the focus of people's attention. Using the text mining method to mine massive online data can quickly discover the important information hidden in the text, so as to achieve better decision-making. Research has been done in various fields, Xu Lijuan (2022)[1] used text data mining techniques to process and analyze the collected questionnaire data. The needs of higher vocational students for teaching materials were explored from several aspects. Similarly, Chen Dong rui (2022) [2] used the method of text

mining to analyze the campus network text to identify the influencing factors of college students' mental health. Goldberg David M (2022) [3] proposed a method to obtain intelligence on product innovation by analyzing online reviews with text mining tools. Text mining technology is not only widely used in various fields, but also commonly used as a tool for academic research.

As we all know, educational books are not only an important way for people to acquire knowledge and improve themselves, but also an important part of the education industry, which has an important impact on people's learning and growth. Online reviews of books are the real feedback of readers on books, expressing users' views on books, and are an important channel for publishers to understand readers' needs and evaluations. Therefore, this paper will explore the publication of educational books by mining online reviews of books at multiple levels of granularity and identifying users' attitudes and evaluations of the whole book and their individual attributes, with the aim of providing new ideas and methods for educational book publishing, which in turn will provide useful references for book publishing.

2. Related research

2.1. Aspect-level sentiment analysis

Sentiment analysis (SA) is the study that discover the feelings, opinions or emotions expressed of people about something, products or services from texts. The result of the Sentiment analysis is the subjective attitude, which are generally classified as positive or negative. Aspect-level sentiment analysis has certain advantages over traditional sentiment analysis tasks and has become one of the hot spots in the field of natural language processing in recent years. Aspect Sentiment Triplet Extraction (ASTE) is the latest subtask in the area of aspect-level sentiment analysis, which consists of three major tasks, including identifying aspect terms, extracting opinion words, and judging sentiment polarity. By identifying the terms of aspect and viewpoint, and then predicting sentiment tendencies for different aspects, the triplet of aspect level <aspect, viewpoint, emotion >[4] is obtained, so as to determine the polarity of opinions, summary of results and multi-dimensional comparison.

2.2. User online review mining

2.2.1 The impact and usefulness of online reviews

Online user reviews mainly consist of reviews on social networks and review sites. Users' online comments are mainly composed of text, expressing their opinions on social events or related products. Since consumers can speak freely when writing comments online, they are more likely to convey their true thoughts and reflect more truthful and comprehensive information.[5] The emotions and opinions expressed by users in online reviews can not only influence the opinions of other users, thus providing value to the consumer decision-making process, but also enhance the understanding of the enterprise about the views and opinions of the users, thus guiding corporate decisions. In addition, Xia Yingxue (2022) [6] demonstrated that customer satisfaction has no effect on customers' intention to write comments in the subsequent consumption stages. Other related studies have shown that the effect of online reviews on customer satisfaction, trust and consumer intention to write reviews diminishes over time.

2.2.2 Using online comments for text analysis

With the popularity of online comments, more and more researchers have started to analyze it. These researchers use text analytics techniques, such as sentiment analysis and topic analysis, to study online reviews under a certain topic or theme in order to gain a deeper understanding of people's attitudes towards a certain topic or product. For example, Mu Di (2022) [7] used LDA model to conduct qualitative research on the negative comments of UAV products on e-commerce platforms. Another study conducted sentiment analysis and topic analysis on user reviews of an app and assessed the user experience of the app accordingly. Online reviews can reflect user needs. In particular, complaints often indicate deficiencies in a product or service. Li Qiaoxing (2023) [8] proposes a method for user demand mining that can combine online reviews with complaint information. However, these studies lacked the analysis of different product attributes. Zhang Jing (2021) [9] uses a fine-grained sentiment analysis method to extract

consumer needs for product attributes, and analyses online reviews from different aspects to obtain more detailed information.

2.2.2 Book review studies

The growth of online reviews has not been limited to physical goods such as computers and mobile phones; online reviews of experiential goods such as books, movies and restaurants have also received widespread attention. Ali Muhammad Asghar (2021) [10] conducted a study on restaurant reviews to explore the influence of online reviews on consumers' restaurant choices. Book review is text that users use to reflect their own perceptions, preferences and other feedback information after reading books. At present, there are a lot of studies on book reviews, such as studying the influence of book reviews on consumers' purchasing behavior and using book reviews to recommend books [11][12].

However, there are more articles exploring the overall sentiment tendency in current research. Analyzing the overall sentiment tendency can provide sentiment trends, but overall sentiment tendency analysis may not accurately capture details. Aspect-level sentiment analysis can provide more specific information to help people understand different aspects of sentiment tendencies and better understand consumer feedback. Therefore, combined with the above background and literature review, this paper attempts to use aspect level sentiment analysis method to analyze book reviews.

3. Research design

This paper adopts the method of "reverse text mining", which has the advantage of improving the accuracy and efficiency of topic analysis, because aspect-level sentiment analysis first can narrow the scope of topic analysis and reduce the error and computational volume of topic analysis. The specific steps include: (1) Obtain user online comment text and pre-processing the comment text; (2) conducting aspect-level sentiment analysis on the pre-processed comment text to complete the task of aspect category word extraction and sentiment tendency determination; (3) conducting keyword analysis on the aspect category words and phrases extracted from the comment text; (4) inputting the extracted aspect category words into the LDA topic model; and (5) performing a topic analysis on the aspect category words and phrases extracted from the comment text. to achieve the topic analysis task; (5) construction of a network graph by calculating the contribution value of each feature word extracted by LDA for the topic to which it belongs. (6) Analysis of the sentiment propensity profiles of secondary aspects to achieve a multidimensional fine-grained assessment measure of educational books.

3.1. Data collection and pre-processing

Public online reviews of educational books are widely distributed across social media platforms and third-party review platforms. The official website of Douban Books contains a large number of educational books and has a lot of rich review information, therefore, this paper chooses Douban Books as the information source for public perception evaluation exploration of educational books. Using the Octopus data acquisition tool, the reviews of 28 Dou ban education books were automatically collected, and a total of 8050 reviews were finally collected. After data collection, pre-processing work such as de-duplication and deactivation of words was carried out on the review data, and the review content was used as the main textual research data.

3.2. Text analysis methods

3.2.1 Aspect-level sentiment analysis

Aspect sentiment analysis is a popular branch of research in text sentiment analysis, focusing on mining fine-grained sentiment information with the aim of extracting the sentiment tendencies of each aspect for several aspects of an entity. In this study, we used a deep learning model rather than a sentiment dictionary approach (Cynthia Whissell, 2009)[13] or a support vector machine approach (Sharma, 2013) [14]. The aspect-level sentiment analysis of the

Aspect sentiment analysis was carried out on the processed texts ,and 6384 aspect category words and phrases were extracted. Synonym substitution was carried out for aspect words to eliminate ambiguity and ensure the accuracy of the word frequency statistics. The keyword analysis results were interpreted through visualization to identify the degree of attention given to different aspects in online reviews. As shown in Figure 2, word frequency determines the size of each word. The analysis shows that when people reading educational books ,they are most concerned about are story and content, which are the most important factors for readers. In addition, words such as "childhood", "child" and "memory" also appear more frequently, indicating that people have a greater interest and concern for childhood times, family relationships and memories. In addition, the language, writing and other aspects of attention, indicating that readers also pay great attention to the language expression and art of books when reading. In addition to this, some practical content, such as finance, logic and case studies, also received attention from readers. In general, people read not only the story itself, but also the language expression, practicality, and its relevance to their own experiences.

4.2. Topic analysis in book reviews

By training the LDA topic model using educational book review data as input, binary combinations of the form <aspect, review entity> can be obtained, and the results are obtained through the review text of the LDA topic model (see Table 1), with the aspect categories arranged in descending order of "article-topic" relevance. Users' concerns about educational books are mainly focused on the thematic aspect categories of content quality, language use, reader satisfaction social influence and other themes.

Table 1. LDA Topic Model Review Text Mining Results.

Primary aspect	Secondary aspects	review entity
Content quality	Article topics	culture, history, science, art, religion, philosophy, humanities, nature, society, life
	Article structure	rhythm, bridge, plot twist, logic, context, conception, thought, beginning, end, structure
	Storyline	scene, drama, complex, direction, story, climax, clues, plot, twist, dilemma
	Characters	character, personality, traits, image, destiny, experiences, values, behavior, language, motivation
Language use	Description Techniques	landscape, character, psychology, action, frontal, profile, detail, sketch, white space, detail
	Language expression	literary talent, language, writing style, phrasing, diction, profundity, vividness, flexibility, richness
	Rhetorical devices	metaphor, simile, symbolism, personification, hyperbole, repetition, prose, irony, couplet, questioning
Reader satisfaction	Article Appeal	recommended, attractive, effective, good read, good review, novel, interesting, informative, like, want to read
	Reader suitability	beneficial, practical, easy, readable, introductory, popular science book, enlightening, inspiring, practical, suitable
	Publication quality	paper quality, print quality, circulation, pagination, quotation, finish, layout
Social Impact	Cultural influences	customs, history, geography, way of life, literature and art, norms of behavior, ways of thinking, values, traditional culture, Buddhism
	Social Impact	ethos, resources, policies, systems, politics, equity, interests, power, democracy
	Educational Insights	parenting, teaching, guiding, inspiring, enlightening, educational value, educational significance, educational popularization, educational propaganda

4.3. Topic-featured word network graph construction and analysis

In order to better understand the relationship between topics and keywords in text data, this paper also conducts topic-feature word network graph construction, by calculating the contribution value of each feature word extracted by LDA to the topic to which it belongs. The contribution value reflects the importance and representativeness of the feature word to the topic. The topic-feature word network diagram is shown in Figure 3, where Different colors

feelings, ratings and other aspects. Thus the content quality aspect triggers a stronger perception by users. The above analysis shows that the statistics in the graph correspond to the actual situation. According to the data, the analysis of the sentiment tendency for each secondary theme, except for the publication quality and education revelation secondary themes, which have a higher negative share, all other secondary themes have a positive share of over 80%, and overall the readers' sentiment tendency for book reviews is relatively positive. Among the various secondary themes, storyline and article theme are the two areas that readers are most concerned about, with a high number of reviews, and the positive percentages for both are around 85%, indicating that storyline and article theme are factors that readers value more and rate them more highly. The lower number of reviews for publication quality and reader applicability, but the higher negative share, suggests that these two areas are not as important to readers, but that the book's publication quality and educational revelation may not be ideal for readers who are concerned about these two areas. The relatively high number of positive reviews in the area of language expression and descriptive techniques suggests that readers are satisfied with the language expression and descriptive techniques of the book. The low number of reviews on social impact and the high percentage of positive comments suggest that the book may have had a positive social impact on some groups of people.

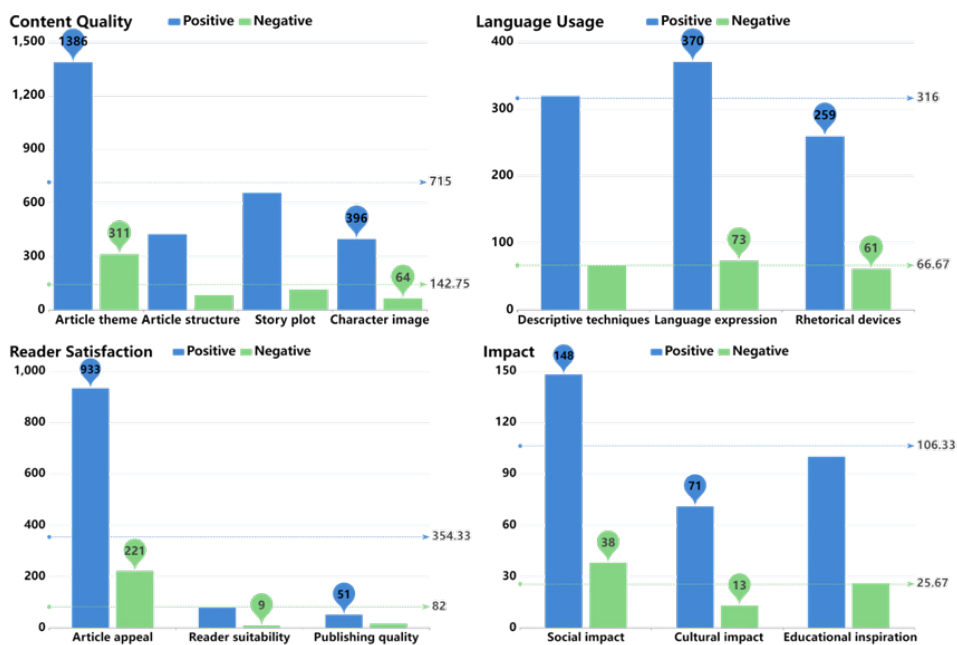


Fig. 3 .Thematic sentiment analysis results

5. Suggestions for book publishing

First, The content of the book should be "practical": Content is the heart and soul of a publication. Readers expect accurate, authoritative content that is well supported by factual evidence. Book content should be meaningful and cover as much relevant information as possible, especially in educational books, where it is recommended to add more detail or expand the breadth. The content should be practical and useful: readers are usually interested in whether the content is practical and whether it will help them solve problems or make progress. For different age groups, publishers should present books that are appropriate for their cognitive level.

Second, The use of language should be 'sensible': books should use language that readers can easily understand and avoid overly complex or vocabulary and sentence patterns. At the same time, avoid mistakes as much as possible, such as misspellings and grammatical errors. These errors can affect the reader's reading experience and may influence the evaluation of the book. Literary rhetoric is a linguistic art, and its appropriate use can enhance the linguistic beauty

and artistry of a text, but it should be controlled according to the characteristics of the readership and content, and excessive or inappropriate use should be avoided.

Third, The expression of ideas should have "meaning": content education publishers should pay attention to social development and change, focus on the impact on society, provide more socially meaningful publications, and promote social development and progress. They should also focus on the promotion of cultural excellence and the creation of culturally influential publications. At the same time publishers should focus on delivering the right educational values when publishing educational books.

6. Conclusion and outlook

Based on the text mining technology, this paper extracts the main aspects of users' attention to educational books, and identifies users' emotional tendency towards each aspect. The main contributions of this paper include: extending the application of aspect level sentiment analysis to the field of public cultural services. Through the analysis of user comments on books, it is found that the aspect of the content quality of educational books has aroused strong perception among users, among which the storyline and article theme are the secondary aspects that that readers value more and rate them more highly. This paper extracts and summarizes the aspects that arouse different emotions of users, and based on the analysis results, puts forward scientific suggestions for improving the quality and publishing of educational books, which can be referred to in the future.

There are still some limitations in the process of this study. First, the data source is limited: this paper only selects the official website of Douban Books as the data source, and many book review platforms only attract specific readers, which may lead to the limitations of the research results. Secondly, although this study considers different aspects of users' experience of books from the perspective of books, future studies need to analyse the changes of user experience over time from the perspective of users, taking into account the life cycle effect of online reviews.

References

- [1] Xu L, Zhao Y, Xi Z, et al. Analysis of Needs of Higher Vocational College Students for Textbooks Based on Text Mining Techniques[C]//2022 3rd International Conference on Modern Education and Information Management (ICMEIM 2022). Atlantis Press, 2022: 693-705.
- [2] Chen Dongrui et al. Online Evaluation Method of College Students' Mental Health Based on Campus Network Text Mining[J]. Mathematical Problems in Engineering, 2022, 2022.
- [3] Goldberg D M, Abrahams A S. Sourcing product innovation intelligence from online reviews[J]. Decision Support Systems, 2022, 157: 113751.
- [4] Peng H, Xu L, Bing L, et al. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(05): 8600-8607.
- [5] Alzate M, Arce-Urriza M, Cebollada J. Mining the text of online consumer reviews to analyze brand image and brand positioning[J]. Journal of Retailing and Consumer Services, 2022, 67: 102989.
- [6] Xia Y, Ha H Y. Do Online Reviews Encourage Customers to Write Online Reviews? A Longitudinal Study[J]. Sustainability, 2022, 14(8): 4612.
- [7] Mu D, Yue C, Chen A. Are we working on the safety of UAVs? An LDA-based study of UAV safety technology trends[J]. Safety science, 2022, 152: 105767.
- [8] Li Q, Yang Y, Li C, et al. Energy vehicle user demand mining method based on fusion of online reviews and complaint information[J]. Energy Reports, 2023, 9: 3120-3130.
- [9] Zhang J, Zhang A, Liu D, et al. Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews[J]. Knowledge-Based Systems, 2021, 228: 107259.
- [10] Ali M A, Ting D H, Ahmad-ur-Rahman M, et al. Effect of online reviews and crowd cues on restaurant choice of customer: moderating role of gender and perceived crowding[J]. Frontiers in Psychology, 2021: 5183.
- [11] Saraswat M, Srishti. Leveraging genre classification with RNN for Book recommendation[J]. International Journal of Information Technology, 2022, 14(7): 3751-3756.
- [12] Khatun F, Chowdhury S M M H, Tumpa Z N, et al. Sentiment analysis of amazon book review data using lexicon based analysis[C]//Computational Vision and Bio-Inspired Computing: ICCVBIC 2019. Springer International Publishing, 2020: 1303-1309.
- [13] Whissell C. Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language[J]. Psychological reports, 2009, 105(2): 509-521.

- [14] Sharma A, Dey S. A boosted svm based ensemble classifier for sentiment analysis of online reviews[J]. *ACM SIGAPP Applied Computing Review*, 2013, 13(4): 43-52.
- [15] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.
- [16] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. *Journal of machine Learning research*, 2003, 3(Jan): 993-1022.