Journal of Global Information Management
Volume 32 • Issue 1

Text Mining-Based Study on Consumer Satisfaction in the Mobile Phone Market

Qun Zhou, School of Information Resource Management, Renmin University of China, China*
https://orcid.org/0000-0002-0792-9221

Meihua Chen, School of Emergency Management, Shandong Academy of Governance, China

Junying Chen, School of Information Management, Wuhan University, China

Keren Chen, Seoul School of Integrated Sciences and Technologies, South Korea

Sang-Bing Tsai, International Engineering and Technology Institute, Hong Kong
https://orcid.org/0000-0001-6988-5829

ABSTRACT

In the current context of rapid technological advancement, smartphones have become an indispensable part of people's daily lives. This has led to an increasing focus on the satisfaction of consumers with smartphone products, as understanding consumer emotions and satisfaction has become a key factor for manufacturers and retailers to enhance the quality of products and services. This study delves into the satisfaction of consumers with smartphones in the market through an in-depth application of text mining techniques, leveraging advanced technologies such as natural language processing, sentiment analysis, and topic modeling. Our research methodology encompasses the process of collecting and preprocessing a substantial volume of consumer reviews from online shopping platforms. Subsequently, we apply Latent Dirichlet Allocation (LDA) for topic modeling and Extreme Learning Machine (ELM) for sentiment analysis.

KEYWORDS

Consumer Satisfaction, ELM, LDA, Sentiment Analysis, Text Mining

INTRODUCTION

With changes in people's lifestyles and the widespread use of the internet, online shopping has become increasingly accepted and ingrained in daily routines (Lissitsa and Kol, 2016). The scale of users engaging in online shopping has gradually expanded. It is evident that as relevant technologies and supporting services mature and improve, the scale of e-commerce has grown significantly. As the popularity of online shopping continues to rise, the volume of product reviews also experiences continuous growth. For consumers, referencing these reviews can assist them in making informed decisions (Khan et al., 2023b). Browsing through these reviews allows users to gain a more comprehensive understanding of products and even learn about real user experiences in advance. This, in turn, enables more effective decision-making, leading to the purchase of ideal products.

DOI: 10.4018/JGIM.344835
*Corresponding Author

For businesses, analyzing these reviews and associated data provides a deeper insight into user needs and the expected user experience (Lewis, 1983; Grönroos, 1982). This understanding helps businesses better grasp their product positioning and areas for improvement, allowing for targeted product enhancements and ensuring the success of their products. However, automating the process of crawling consumer reviews and presenting consumer focal points in a meaningful way remains a significant challenge (Zhou et al., 2007).

Text mining is a branch of data mining, a concept formally introduced by Feldman in 1995 (Feldman and Hirsh, 1996). It leverages computer technology to extract implicit and high-value information from semi-structured or unstructured text that is of interest to users (Hotho et al., 2005). By extracting structured information from text and conducting research, text mining achieves highly automated analysis of textual data, making it applicable in various scenarios (Zhao and Chen, 2022). Hei-Chia Wang et al. constructed an integrated summarization system based on text mining algorithms (Wang et al., 2020; Khan et al., 2022). Lei C et al. utilized stacked variational autoencoder technology to extract features from system text, proposing an effective SVAE text feature extraction model (Che et al., 2020). Sandra Maria Correia Loureiro et al. presented a new text mining approach from the consumer's perspective, using aggregated dictionaries based on consumer brand authenticity and brand involvement, offering new insights for brand development (Rosado-Pinto et al., 2020). These methods reflect the development of text mining technology, which is highly significant for machine learning of text semantics. However, these methods have not been further connected with consumer sentiment analysis.

With the development of artificial intelligence (Khan et al., 2023a; Ning et al.,2024; Tian et al., 2024), Natural Language Processing (NLP) technology has been rapidly developed. Sentiment analysis, also known as opinion mining, is a common application in NLP methods and is a crucial information analysis technique (Medhat et al., 2014). It is one of the most commonly used methods for evaluating textual data. With the rise of the Internet, sentiment analysis has gained widespread attention, particularly for its success in various business domains (Wankhade et al., 2022). It is now extensively applied in fields such as financial stock market analysis, marketing, social sentiment analysis, and consumer satisfaction analysis (Hussein, 2018). For instance, Hiroki Sakaji et al. introduced a Gradient-Interpretable Neural Network text visualization method from a sentiment analysis perspective. This method visualizes market sentiment scores and sentiment gradient scores within words and conceptual units across entire financial documents, facilitating easy understanding for non-experts regarding financial reports (Ito et al., 2020). Min L et al. proposed a new method called Emotion Cause Span Extractio, based on Emotion Cause Identification, which enables more precise extraction of emotional factors in textual data (Li et al., 2021). Additionally, they presented a sentiment classification joint learning framework based on emotion cause span extraction, aiming to better extract the underlying causes behind certain emotions in the text. Chauhan D S et al. introduced a multi-modal sentiment analysis approach based on recurrent neural networks. Their proposed model, incorporating an autoencoder mechanism, learns the interactions between patterns involved in various existing state-of-the-art systems, demonstrating effectiveness in sentiment and emotion analysis (Chauhan et al., 2019). However, after obtaining sentiment analysis results in this study, further exploration is needed on how to integrate it with text mining techniques.

Currently, research on text mining and sentiment analysis has been gaining momentum, and there has been considerable research on the relationship between consumer evaluations and behavior in the smartphone domain. However, the application of text mining technology in the smartphone field remains relatively limited. Traditional research methods for consumer satisfaction in the smartphone domain still heavily rely on offline approaches such as consumer phone interviews or household surveys, which are time-consuming and yield limited effective information. Moreover, they may face challenges such as unanswered calls or difficulty in conducting household surveys due to consumers being unavailable or unwilling to participate. Therefore, integrating text mining of online shopping reviews with consumer satisfaction in the smartphone domain not only enriches the methods for studying consumer satisfaction but also addresses the gap in text mining in the smartphone field. This comprehensive approach is expected to positively contribute to the development of smartphone manufacturers and the enhancement of brand competitiveness.

Based on the above analysis, this study focuses on consumer reviews of mobile phones on a certain online shopping platform, proposing a method that integrates text data mining and sentiment analysis. This approach aims to extract key information from the reviews, gaining a deeper understanding of consumer sentiments, and encouraging improvements for mobile manufacturers and online shopping platforms. The main research steps include:

(1) Firstly, consumer reviews of several mobile phones were scraped from a specific online shopping platform. Subsequently, the obtained consumer online shopping review texts were cleaned using the Jieba word segmentation library and stop-word library, including sentence segmentation and stop-word removal, to facilitate subsequent analysis of consumer satisfaction with solid wood beds.
(2) The Latent Dirichlet Allocation (LDA) model was employed to extract keywords from the reviews and reveal topics related to consumer satisfaction. Additionally, third-party libraries such as SnowNLP in Python were utilized to analyze the features and sentiment tendencies of consumer satisfaction texts. This analysis involved frequency statistics of feature words and feature association analysis in terms of text features. Regarding sentiment tendencies, a corpus of positive and negative sentiment comments in the smartphone consumer domain was constructed.
(3) Sentiment analysis was conducted on the text by training an Extreme Learning Machine (ELM) model. The results were aligned with the output from the LDA model to identify features that consumers are more inclined to focus on. This comprehensive approach is expected to provide valuable insights and directions for improvement in the industry.

This paper is structured as follows: Section 1 introduces the significance of the research and outlines the research objectives. Section 2 provides an overview of the relevant technical background. Section 3 elaborates on the specific process of the proposed methodology. Section 4 presents the experimental results and analysis. Finally, Section 5 offers a comprehensive summary of the entire document.

RELATED BACKGROUND

Consumer Behavior Analysis

In traditional retail, businesses typically rely on surveys or expert market insights to understand consumer behaviors during the purchasing process (Luomala, 2002). However, the challenge in the era of big data e-commerce is the constantly evolving preferences, characteristics, motivations, and emotions of online consumers (Kotler, 2001), making traditional consumer behavior analysis theories less effective.

Currently, analysis of online product reviews primarily focuses on 'word of mouth' and references, studying how existing product evaluations impact potential consumers. However, there's limited research on the usefulness of information in these reviews, with most studies concentrating on the relationship between reviews and user behavior, also known as 'word-of-mouth marketing.' Major research platforms include Amazon, Tmall, JD, and WeChat, where research centers on positive and negative reviews and comment quantity. The quantity of reviews reflects product acceptance and can influence consumer interest and sales (Vinodhini, 2012). More reviews suggest higher sales potential, as seen when consumers prefer products with more reviews on platforms like Amazon. Additionally, increased online movie reviews correlate positively with box office revenue. The quality of reviews directly affects consumer satisfaction, with negative reviews holding more weight in purchasing decisions. Businesses benefit from analyzing negative reviews to understand buyer dissatisfaction and improve services and experiences. However, focusing solely on positive and negative reviews provides limited insights into consumer preferences and emotions.

To enhance e-commerce sales activities, more precise analysis and data mining techniques are necessary for understanding consumer preferences and emotions. This approach provides more accurate guidance for businesses in the current e-commerce landscape (Son and Hua, 2022).

Fundamentals of Text Mining Methods

Sentence Segmentation and Stopword Handling

Sentence segmentation, refers to the process of using tokenization techniques to break down a text into individual words or sequences of words, with separation achieved through spaces. The purpose of sentence segmentation is to break down a block of text into individual sentences, making the text more manageable and facilitating further linguistic analysis or processing. Sentence segmentation serves as the foundation and prerequisite for in-depth text mining. Text data, after undergoing tokenization, can be transformed into a mathematical vector format, facilitating subsequent analysis (Srinivasan and Dyer, 2021).

Currently, the mainstream Chinese word segmentation tools include JieBa, THULAC, pyltp, and others. Due to its high efficiency and accuracy in word segmentation, as well as the ability to import custom dictionaries for specific terms such as "product information" and "price", JieBa is selected for Chinese word segmentation in this article. The JieBa library offers key functionalities such as word segmentation, custom dictionary usage, keyword extraction, and part-of-speech tagging. It supports three segmentation modes: precise mode, full mode, and search engine mode.

Stopwords can introduce significant interference in the text processing pipeline. They inherently carry limited useful information and can exert a certain inhibitory effect on other words, thereby significantly impacting the efficiency and accuracy of text processing. Generally, stopwords can be broadly categorized into two types: general stopwords and domain-specific stopwords. General stopwords are the most widespread in the text, appearing in nearly all documents. Examples include meaningless degree adverbs, modal particles, prepositions, and other words that lack substantive information, such as "although", "if","but" etc. These words have high frequencies but do not contribute meaningful information, so they need to be removed in text mining to avoid influencing research outcomes.

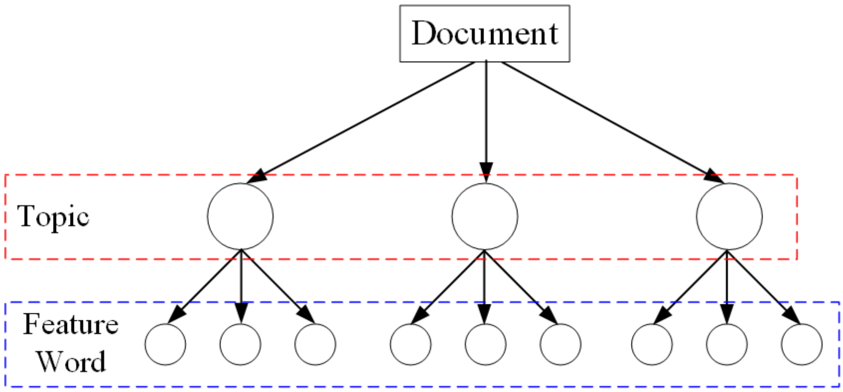
The other category is domain-specific stopwords, which are applicable to specific domains. For instance, in the domain of product reviews, terms like "thing" and "feeling" may appear frequently but contain limited useful information and should be excluded in the processing. Therefore, this study aims to construct a new stopword dictionary based on Chinese word segmentation to filter out irrelevant terms.

LDA Topic Model

LDA topic model is a probabilistic generative model extensively employed in the field of NLP. Its primary objective is to unveil latent structures or themes within a corpus of textual data (Cao et al., 2009; Jelodar et al., 2019). LDA is a popular method among topic extraction models, assuming that each document is a mixture of topics and each topic is a distribution of words. Compared to other models like Latent Semantic Analysis (LSA), Non-negative Matrix Factorization (NMF), and Hierarchical Dirichlet Process (HDP), LDA stands out for its simplicity and interpretability. With clear topic-word distributions provided by LDA, it is easier to understand and interpret the generated topics. Additionally, LDA is widely used and supported by various implementations and libraries, making it more accessible and deployable in practice.

At its core, LDA operates as an unsupervised learning algorithm, tasked with categorizing a set of documents into distinct topics. Each topic is characterized by a distribution of words, and every

Figure 1. The structure of the LDA topic model



document is treated as a blend of multiple topics. The underlying assumption is that the distribution of topics in documents and the distribution of words in topics are drawn from a Dirichlet distribution.

The model construction process involves the generation of each word in a document through a multi-step procedure. Initially, a topic distribution is selected for the document, followed by the assignment of a topic to each word. Finally, a word is randomly chosen from the vocabulary distribution associated with the selected topic. This entire process is effectively captured by a probabilistic graphical model.

As shown in Fig.1, LDA relies on two crucial parameters – the document-topic distribution and the topic-word distribution. These parameters are estimated through training on the available data. The versatility of LDA extends to various applications, including text mining, information retrieval, and recommendation systems. Its ability to uncover hidden thematic structures within a document collection proves invaluable for extracting latent insights from textual data. In summary, the LDA topic model stands as a robust and adaptable tool, offering profound insights into the latent structures inherent in textual data.

Methods for Sentiment Analysis

Text sentiment analysis refers to the process of extracting, processing, and analyzing text data with emotional nuances using NLP techniques. Currently, the most common methods for sentiment classification are based on dictionary-based sentiment analysis and machine learning-based sentiment analysis. Dictionary-based methods involve constructing sentiment dictionaries, intensity lexicons, and negation lexicons. These methods identify sentiment words in sentences and perform sentiment analysis by summing up the weighted values of these words. While dictionary-based sentiment analysis yields result with good readability, it relies on continually refining dictionaries to enhance domain adaptability.

On the other hand, machine learning-based sentiment analysis methods require substantial amounts of manually labeled datasets for training. These methods involve extracting text features, training models, and then classifying the sentiment of the text to be analyzed. This approach tends to offer higher reliability in sentiment classification. However, it necessitates a significant effort in data annotation for training purposes.

Given the widespread application and high reliability of machine learning-based sentiment analysis methods, this paper opts to utilize this approach for sentiment analysis. This method involves the use of machine learning techniques to achieve sentiment analysis, owing to its proven effectiveness.

Table 1. Mobile phone online comment text data sources

Number	Name	Comments	Words
1	Mobile phone A	528	3254
2	Mobile phone B	1254	6523
3	Mobile phone C	2208	12053
4	Mobile phone D	1326	6132
5	Mobile phone E	687	4658
6	Mobile phone F	903	6203
7	Mobile phone G	840	5214
8	Mobile phone H	775	5325
9	Mobile phone I	2127	9872
10	Mobile phone J	1876	9657

METHOD

The consumer behavior analysis scheme proposed in this article mainly involves data acquisition and preprocessing, keyword extraction, and sentiment analysis.

Data Acquisition and Preprocessing

Data Acquisition

Due to restrictions on a certain shopping website, the web client can only display the first 250 pages of content for each product. Consequently, each mobile phone can only obtain information on the first 250 pages of online shopping comments. This study utilized Python and Web Scraper as web scraping tools to collect comment text, and the data collection took place in January 2022. The collected content includes specific comments, follow-up comments, comment times, and evaluation types. A total of 12,524 comments were crawled for 10 mobile phone products, amounting to 68,891 words. The time span of these comments covers from January 2022 to January 2023.

Considering the need for machine learning methods in the subsequent analysis of satisfaction sentiment tendencies, the comments data for the aforementioned 10 mobile phone products were divided into training data and test data at a ratio of 7:3. Below is some detailed information on the online shopping comment text data for mobile phones, as shown in Table 1.

Data Preprocessing

Due to the substantial workload of the data collection task in this study, frequent operations during the collection process may result in some missing values and duplicate data. Additionally, due to the diverse nature of online shopping review text information, there may be a portion of data with no value. To ensure the accuracy of subsequent analysis results, it is necessary to clean the above invalid data, transforming it into high-quality data that can be efficiently utilized, and avoiding any impact on the analysis results. The specific data cleaning rules and steps are as follows:

(1) Remove Null Values: Exclude data with other information such as comment time but with empty comment content or entirely empty entries. Such comments are excluded because frequent visits and data crawling triggered the anti-crawling mechanism of the shopping website, causing the relevant information not to be displayed.

(2) Eliminate Duplicate Data: Exclude multiple comments with the same content when the time interval is short or when the comments have the same timestamp. The reasons for such duplicate comments include improper operations during repeated crawling, consumers copying others' comments for positive reviews, cashback, and rebate discounts, as well as duplicates generated by merchants engaging in order brushing.
(3) Exclude Valueless Data: This type of data refers to information in comments that is unrelated to the product. The main reason for such comments is the automatic comment feature on major e-commerce shopping websites. When consumers fail to provide comments within a specified period after purchasing a product, the system defaults to positive feedback and automatically generates comments like "The reviewer did not make a timely evaluation, the system defaults to positive!" or "This user did not write a review." These comments hold no meaningful value for subsequent research and need to be removed.

Subsequently, to obtain more accurate and non-redundant segmentation results, this study chose the Jieba segmentation precise mode based on the characteristics of the mobile phone online shopping review text under investigation. In addition, to save storage space, reduce text feature dimensions, and avoid interference with the model's understanding of the text during the segmentation process, it is necessary to remove stop words to increase the density of meaningful key feature words. To achieve this, the study integrated the stop word libraries from the Machine Intelligence Laboratory of Sichuan University, the stop word list from Harbin Institute of Technology, and Baidu's stop word list. Additionally, adjustments were made based on the actual situation of mobile consumer review text data, and some stop words related to the electronics product domain were added. This resulted in a stop word list suitable for this study.

LDA Model for Keyword Extraction

The core issue of this paper is how to scientifically and reasonably extract useful information from a massive volume of user reviews and construct a suitable evaluation indicator system for consumer satisfaction with mobile phones. The specific process is as follows:

(1) Determining the Optimal Number of Topics

The determination of the optimal number of topics is crucial for identifying deficiencies in express service quality in this study, as the results of topic classification may vary with different numbers of topics. The construction process of the LDA topic model requires manually specifying the value of K, the number of topics. Therefore, before topic modeling, it is essential to determine the optimal number of topics.

Currently, perplexity and topic coherence are two methods for determining the number of topics. Perplexity and topic coherence are two methods used to determine the optimal number of topics in a topic model. Perplexity assesses the model's predictive performance on unseen data, with lower values indicating better performance. On the other hand, topic coherence measures the degree of association between words within topics and the distinctiveness between topics, with higher values indicating more interpretability and coherence. Combining these two metrics provides a comprehensive evaluation of the quality and effectiveness of the topic model. Perplexity can be understood as the uncertainty of a document d belonging to a topic K. Generally, a lower perplexity value indicates higher model accuracy, and the K value corresponding to the lowest point or inflection point of perplexity is often considered the optimal number of topics. The formula for calculating topic perplexity is as follows:

$$perplexity\left(D\right)=\exp\left(-\frac{\sum_{i=1}^{M}\ln P\left(d_{i}\right)}{\sum_{i=1}^{M}N_{i}}\right) \tag{1}$$

where ***D*** represents the test set, ***M*** is the number of documents, ***di*** refers to the word sequence in document *d*, and the probability of each word occurring collectively. ***N***i indicates the total number of words in document *d*.

A higher coherence score indicates a better model; therefore, the peak coherence score may represent the optimal number of topics. The formula for calculating topic coherence is as follows:

$$coherence\left(T\right)=\sum_{\left(\nu_{i},\nu_{j}\right)\in T}score\left(\nu_{i},\nu_{j}\right) \tag{2}$$

where ***T*** represents a specific topic; *υi* and *υj* are words within that topic; *score*() is a scoring function measuring the semantic closeness of word pairs within the topic. In practical applications, the *score*() function typically employs the UMass algorithm, which assesses similarity based on the co-occurrence frequency between words. The formula is as follows:

$$score\left(\nu_{i},\nu_{j}\right)=\log\frac{D\left(\nu_{i},\nu_{j}\right)+\varepsilon}{D\left(\nu_{j}\right)} \tag{3}$$

where ***D***(*x*,*y*) represents the number of documents in which both *x* and *y* co-occur, ***D***(*x*) is the number of documents containing the word *x*, and ε is a smoothing coefficient used to avoid a numerator of zero, typically taking a small value such as 10^{-12}.

(2) LDA Topic Extraction

After determining the optimal number of topics, the process involves extracting consumer satisfaction evaluation indicators for mobile phones through the induction of the LDA topic model. The procedure is as follows: firstly, conduct Chinese word segmentation and stop word processing on consumer reviews after purchase, and then import the LDA topic model for modeling; secondly, after model processing, K sets of topic words representing K topics can be obtained, with each set of topic words distributed with certain probabilities across the topics; finally, by summarizing each set of topic words, the corresponding satisfaction evaluation for mobile phone consumers can be derived. Figure 2 illustrates how the corresponding consumer evaluations for mobile phones are summarized based on the topics extracted through LDA.

(3) Text Topic Prediction

Text topic prediction is the foundation of text classification. In this study, the LDA topic model is utilized to predict and classify topics for all texts, assigning each text a specific theme. The process of applying the LDA topic model to predict topics in texts is illustrated in Figure 3. After determining the optimal number of topics ***K***, a new LDA model is constructed by setting the number of topics to the optimal value. This optimal model is then employed to predict topics for all texts. The process of topic prediction is as follows: firstly, the LDA model processes each text by conducting Chinese word segmentation, removing stop words, and filtering by part-of-speech, ensuring that each text retains only

Figure 2. LDA model topic extraction diagram

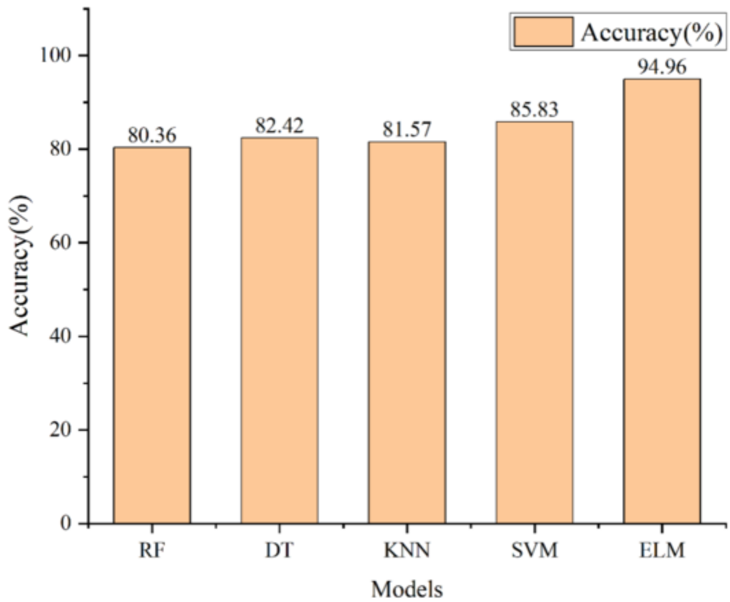
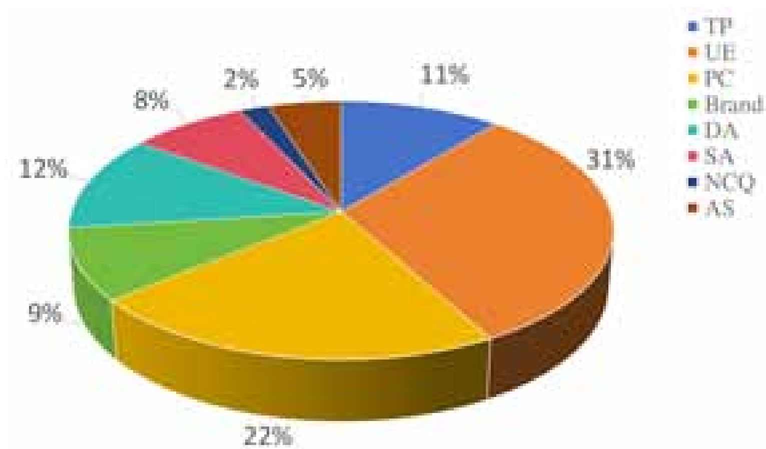


Figure 3. LDA model predicts text topic diagram



key terms sufficient to summarize the entire content. Secondly, based on the keywords and extracted topic results, the LDA model predicts topics for each text. The predicted texts are distributed over each topic with certain probabilities, indicating that each text is probabilistically associated with K topics. Finally, the LDA model selects the topic with the highest probability distribution for each text as the predicted topic result.

Sentiment Analysis

Text sentiment analysis is a technique that quantifies and qualifies the sentiment values of subjective text with emotional tones through NLP. It enables the extraction of information about the sentiment inclination and weights of the text. Depending on the granularity of the analysis, text sentiment analysis can be categorized into coarse-grained and fine-grained analyses. Coarse-grained sentiment analysis assesses the overall sentiment inclination of the text and makes sentiment polarity judgments for large paragraphs of text. Fine-grained sentiment analysis, on the other hand, goes beyond sentiment polarity assessment to determine the intensity of emotions, involving the computation of sentiment numerical values.

In this study, we have opted for fine-grained sentiment analysis. Initially, we trained sentiment analysis models specifically for the mobile consumer domain. Subsequently, the trained models were utilized for sentiment inclination analysis. The analysis process is illustrated in Figure 4.

The training of the sentiment analysis model in the mobile phone consumer domain will employ machine learning methods. To mitigate the impact of data scale and reduce manual labeling operations, this study will utilize the SnowNLP library for sentiment score calculation analysis, assessing the sentiment value on a per-comment basis for online purchase reviews of solid wood beds. SnowNLP is a Python library that provides NLP functionalities, particularly tailored for processing Chinese text. It offers various tools for tasks such as part-of-speech tagging, sentiment analysis, text summarization, keyword extraction, and more. SnowNLP utilizes machine learning algorithms and statistical models to analyze and understand Chinese text, making it a valuable resource for developers and researchers working with Chinese language data in their projects. The library demonstrates high accuracy in analyzing shopping-related comments, as it is trained on e-commerce review texts. The underlying principle of sentiment calculation in the SnowNLP library involves using the Naive Bayes classification algorithm. The sentiment score ranges from 0 to 1, with values closer to 0 indicating a more negative consumer sentiment and values closer to 1 indicating a more positive consumer sentiment.

ELM, depicted in Figure 5, is a distinctive machine learning model. In contrast to conventional algorithms, ELM randomly selects input weights and hidden layer biases during training, dynamically adjusting the number of neurons in the hidden layer. This approach significantly reduces training time, making ELM suitable for large-scale datasets and real-time applications. In addition, ELM often exhibits competitive generalization performance compared to more complex algorithms. It can effectively learn nonlinear relationships in data, making it suitable for a wide range of regression and classification tasks. Furthermore, optimal output weights are determined without the need for multiple iterations, reducing the risk of getting stuck in local optima. This streamlined approach significantly shortens the training time of the ELM model, leading to a more efficient convergence towards the global optimum (Han et al., 2014). These features make ELM a valuable tool for various machine learning applications, particularly in real-time scenarios and with large-scale datasets. Therefore, ELM is adopted to realize emotion classification in this paper.

Figure 4. Flow chart of emotional tendency analysis

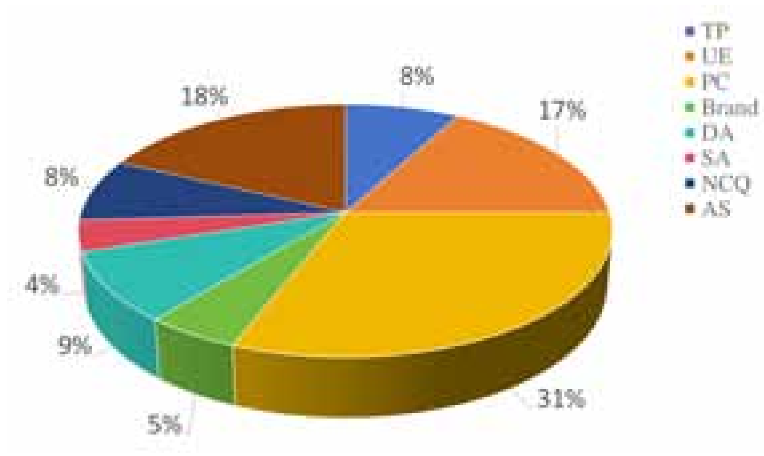


Figure 5. ELM Model

After completing the model training, it is essential to validate its accuracy. In this phase, 200 accuracy calculation data points are initially labeled through manual annotation. Positive sentiments expressed by consumers in smartphone reviews are labeled as "1," while negative sentiments are marked as "-1." Subsequently, the model, retrained using SnowNLP, scores these 200 accuracy calculation data points, predicting the probability of positive sentiment. If the probability of positivity is greater than or equal to 0.5, it is classified as positive sentiment; otherwise, it is considered negative sentiment. Since SnowNLP returns values between 0 and 1 instead of discrete labels (1 or -1), a conversion is applied—scores equal to or higher than 0.5 are recorded as satisfaction (labeled as "1"), while scores below 0.5 are marked as dissatisfaction (labeled as "-1"). This facilitates comparison with the actual labels assigned through manual annotation.

For the validity of the ELM model, in the code implementation, the process involves encoding the data into characters, updating the model path to incorporate the newly trained model, iterating through each accuracy calculation data for predictions, adding a new column for predicted labels alongside the manually annotated actual labels, comparing these labels to determine correctness, and ultimately outputting the accuracy of the predictions. These steps are essential for evaluating the performance of the model and assessing its accuracy in predicting sentiments based on the provided accuracy calculation data.

RESULTS AND DISCUSSION

Experiment Environment

In this experiment, the computer setup is as follows: an Intel Core i9-10900CPU serves as the central processing unit, complemented by a Quadro P2200 GPU. The server is equipped with 32 GB of memory to facilitate high-performance computing required for deep learning tasks. The experiment employs the ELM model, with the following model parameters: the input layer utilizes default randomly initialized weights, and the number of hidden layer neurons is dynamically adjusted based on the specific requirements of the problem. Notably, the ELM model's training process does not involve the use of backpropagation, eliminating the need for setting a learning rate. This experimental environment is designed to provide ample computational power for the effective training and evaluation of the deep learning model.

Evaluation Indicators

In the context of the binary classification problem, samples are categorized into positive and negative outcomes. The classification of a sample produces four possible results, as outlined in Table 2.

Table 2. The classification results

Sample	Results	Symbol
Positive	Positive	True Positive (TP)
Positive	Negative	False Negative (FN)
Negative	Positive	True Negative (TN)
Negative	Negative	False Positive (FP)

In addition, this paper evaluates classifier performance using four key metrics: accuracy, precision, recall, and F-measure. Accuracy signifies the proportion of correctly classified samples in the total dataset, precision focuses on the accuracy of predicting positive instances, recall emphasizes the model's capability to identify actual positive instances, and F-measure is a harmonic mean of precision and recall, suitable for a comprehensive assessment across diverse data distributions. Widely applied in deep learning, machine learning, and statistics, these metrics offer multiple perspectives for a thorough evaluation of classifier performance. They are expressed as

$$acc = \frac{TP + TN}{TP + FN + FP + TN} \tag{4}$$

$$pre = \frac{TP}{TP + FP} \tag{5}$$

$$rec = \frac{TP}{TP + FN} \tag{6}$$

$$F_1 = \frac{2 \times rec \times pre}{rec + pre} \tag{7}$$

where acc, pre, rec, and F_1 are the accuracy, precision, recall, and F-measure, respectively.

Experiments Results

Determine the Optimal Number of Topics

The programming code in Python 3.7 utilizing the Gensim library was employed to calculate the topic perplexity and topic coherence scores for various numbers of topics in the LDA model. The results are presented in Figure 6 and Figure 7. Comparing the two methods for determining the optimal number of topics, the topic perplexity graph exhibits an upward trend, making it challenging to identify the optimal number of topics. Consequently, this study opts for the topic coherence method to determine the optimal number of topics.

As depicted in Figure 7, the topic coherence scores peak at 8 and 12, suggesting that these values may represent the optimal number of topics. To further ascertain the optimal number of topics, the LDAvis visualization tool is introduced to visualize the topic model. By applying LDAvis, topic

Figure 6. Topic confusion degree trend

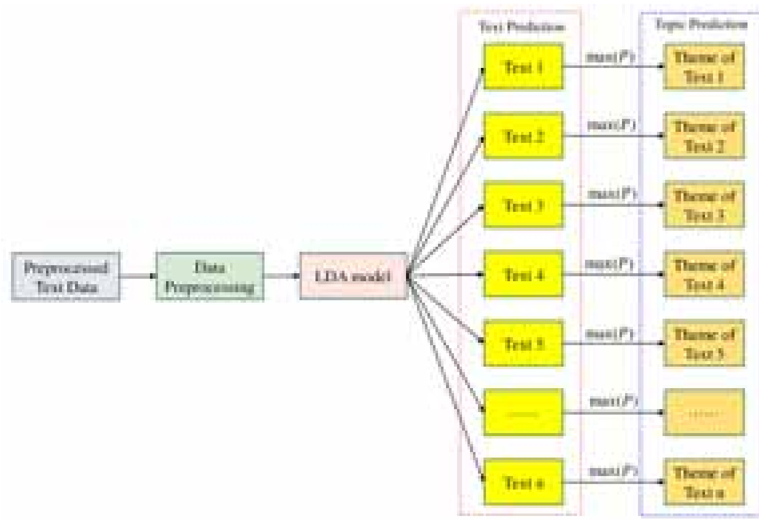
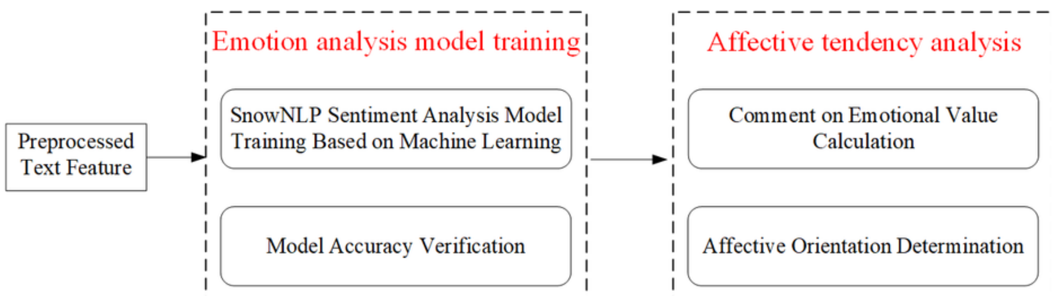


Figure 7. Theme consistency score trend

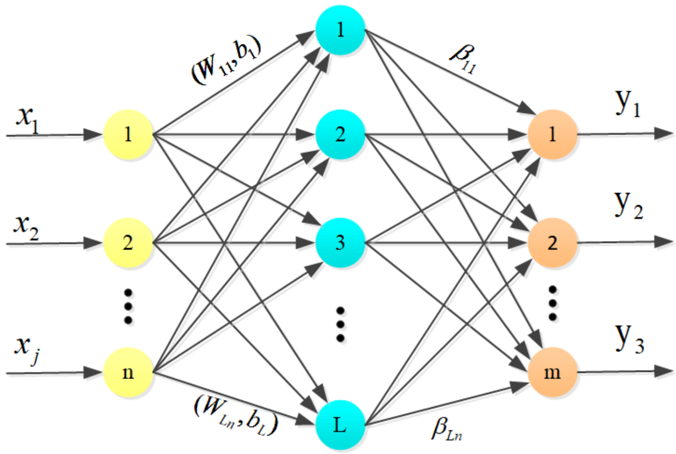


distribution graphs for 7, 8, 9, 10, 11, and 12 topics were generated for comparison. Notably, at 7 and 12 topics, multiple topics exhibit overlapping, indicating that some topic words are assigned to multiple topics, resulting in poorer model classification performance. Conversely, the topic distribution graph for 8 topics reveals a more dispersed arrangement between topics, suggesting that topic words can be well-distinguished, leading to improved model classification performance. Hence, this study selects the LDA model with K=8 topics for extracting themes from consumer satisfaction comments on mobile phones.

Analysis of LDA Topic Extraction Results

Drawing on relevant literature and empirical patterns, this study sets two hyperparameters in the LDA topic model as $\alpha=50/k$ and $\beta=0.01$, with an iteration count of 100 times. The number of topics is set to K=8 for extracting topics from mobile phone consumer reviews. The output includes the main topic words for each topic, as shown in Figure 8. Since LDA cannot automatically generate names for each topic, manual summarization is required. Figure 8 displays the main topic words for each topic. By combining the meaning of each set of topic words, consulting relevant literature, and seeking advice from experts in the field, we summarized twelve defect topics affecting mobile phone consumer satisfaction: Technical Performance (TP), User Experience (UE), Price and Cost-

Figure 8. LDA model theme extraction results



effectiveness (PC), Brand, Design and Appearance (DA), Software and Applications (SA), Network Communication Quality (NCQ), After-sales Service (AS).

After extracting eight topics representing quality defects in mobile phone consumer satisfaction, the LDA topic model is employed to predict topics for the entire text. The topic with the highest prediction probability is selected as the assigned topic for each text. Through data compilation, the statistical results presented in Figure 8 are obtained. From the analysis of the results, it is evident that among the collected mobile phone consumer reviews, a significant 35.6% of comments emphasize the user experience, 22.3% focus on the brand of the phone, and 18.4% prioritize technical performance. Therefore, the conclusion can be drawn that consumers in the mobile phone market are more concerned about the user experience and the brand effect of the phone.

Sentiment Recognition Analysis

In this paper, the sigmoid activation function is chosen to determine the number of hidden nodes in ELM. The classification performance is evaluated based on the accuracy of sentiment recognition in mobile phone consumption. The results, as shown in Figure 9, indicate that the classification accuracy is highest when the number of hidden nodes is set to 25. Therefore, the hidden node count is set to 25.

To further highlight the classification performance of ELM, this study compared it with common classification models such as Random Forest (RF) (Li et al., 2021), Decision Tree (DT) (Song and Ying, 2015), K-Nearest Neighbors (KNN) (Peterson, 1883), and Support Vector Machine (SVM) (Wu et al., 2022). Evaluation metrics included accuracy, precision, recall, and F1 score. The specific experimental results are presented in Table 3. From the table, it is evident that ELM achieved an accuracy of 95.87% and precision of 94.12%, which are the highest among these models. Overall, ELM demonstrates excellent performance across these evaluation metrics, potentially making it an effective tool for addressing this problem. Additionally, due to the characteristics of ELM, it also exhibits certain advantages in terms of runtime compared to these models.

To further validate the classification performance of these models, we utilized a manually annotated sentiment classification dataset as the test set to observe the classification accuracy of these models. The experimental results are presented in Figure 10. From Figure 10, it can be observed that, compared to other models, ELM exhibited the best classification performance on the test set, achieving an accuracy of 94.96%. This indicates that the ELM model trained in this study possesses excellent generalization performance.

Figure 9. The performance comparison of ELM with different hidden nodes

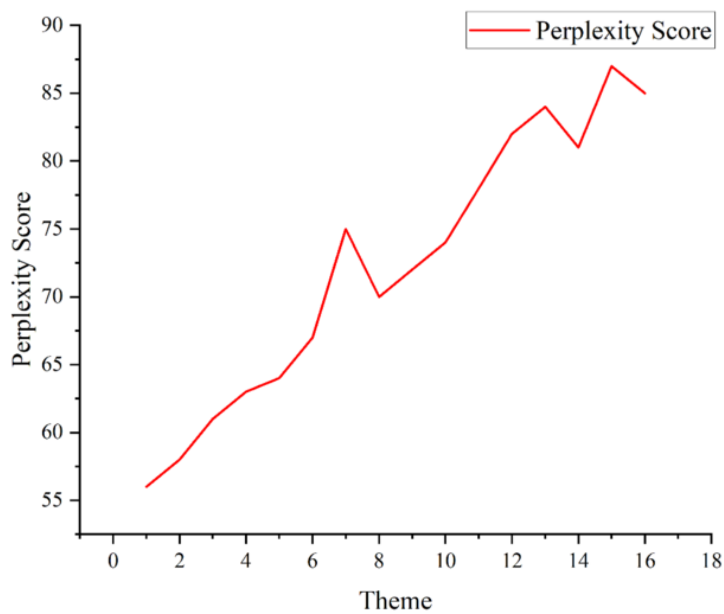


Table 3. Empirical comparison of different models on training set

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 value (%)
RF	82.58	81.52	85.8	85.3
DT	83.76	82.43	86.7	86.2
KNN	85.03	84.38	88.95	89.72
SVM	87.89	86.57	89.93	89.81
ELM	**95.87**	**94.12**	**93.74**	**94.64**

The Relationship Between Emotion and Theme

In order to identify the direction for product improvement and enhance online shopping experience based on consumer reviews, establishing the connection between smartphone consumer satisfaction and topics is crucial. In this section, we categorize smartphone consumer satisfaction based on their reviews and associate these reviews with the corresponding topics from the LDA model. The specific results are depicted in Figures 11 and 12. From Figure 11, it can be observed that the majority of dissatisfied consumers are unhappy due to poor user experience with the phone. Following this is the dislike for the brand. On the other hand, as seen from Figure 12, satisfied consumers are mostly fond of specific brands and also value user experience and after-sales service. Overall, user experience and cost-effectiveness are key indicators influencing consumer satisfaction. Additionally, brand impact and hardware performance play crucial roles in satisfying customers. On the other hand, after-sales service and the external design of the phone are identified as significant factors contributing to customer dissatisfaction.

Figure 10. Accuracy of different models on test set

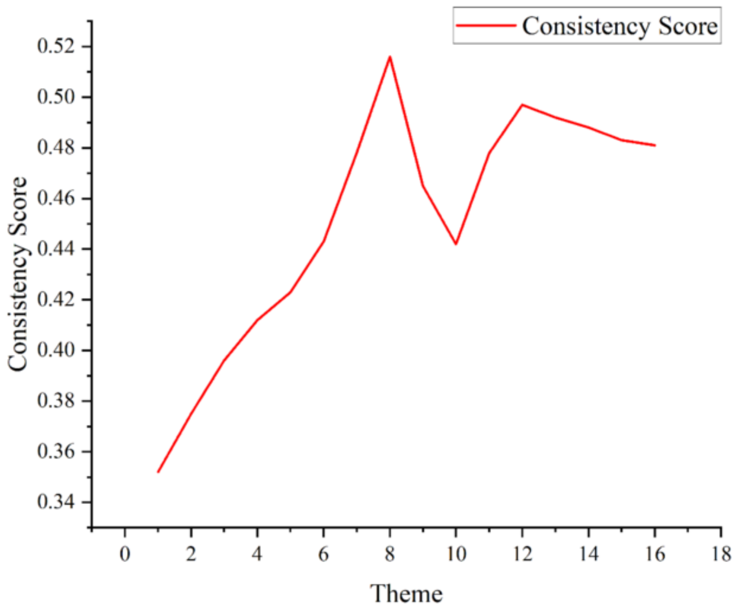


Figure 11. The relationship between consumer satisfaction and theme

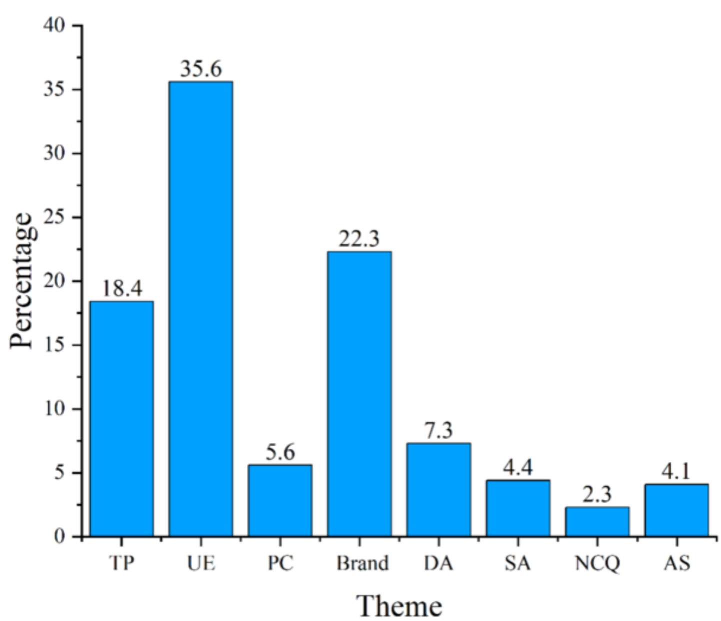
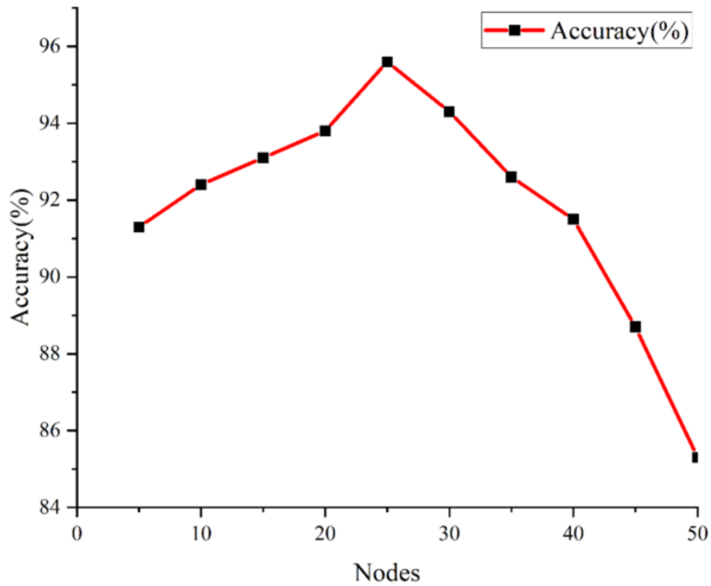


Figure 12. The relationship between consumer dissatisfaction and theme



CONCLUSION

This study conducted an in-depth analysis of consumer reviews on mobile shopping platforms to uncover key factors influencing consumer satisfaction and dissatisfaction with mobile products. Initially, we collected and preprocessed consumer reviews from a specific online shopping platform. Subsequently, leveraging the LDA topic model, we successfully extracted key topics related to mobile consumption. We then applied the ELM model for sentiment analysis, obtaining a more accurate classification model to assess sentiment tendencies in the comments. Finally, by establishing the relationship between satisfaction and topics, we further identified themes influencing user satisfaction and dissatisfaction, including TP, UE, PC, brand, DA, SA, NCQ, and AS. This offers targeted improvement directions for mobile manufacturers and online shopping platforms to meet consumer needs, enhance product quality, and optimize overall service experience. In the process of constructing an emotional corpus in the field, this article did not consider other cases besides positive and negative emotions, such as neutral emotions. Additionally, expressions that are more colloquial or ironic still cannot be accurately identified. Therefore, future research could consider exploring the risks associated with the evolution of positive and negative emotions, and introducing more advanced deep learning algorithms to analyze text, thereby obtaining more accurate sentiment analysis results.

CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

PROCESS DATES

Received: January 21, 2024, Revision: February 25, 2024, Accepted: March 23, 2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Qun Zhou, zhouq1105@ruc.edu.cn

FUNDING AGENCY

This research is funded by the Youth Fund of the Natural Science (ZR2023QG166).

REFERENCES

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7-9), 1775–1781. doi:10.1016/j.neucom.2008.06.011

Chauhan, D. S., Akhtar, M. S., Ekbal, A., & Bhattacharyya, P. (2019, November). Context-aware interactive attention for multi-modal sentiment and emotion analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 5647-5657). doi:10.18653/v1/D19-1566

Che, L., Yang, X., & Wang, L. (2020). Text feature extraction based on stacked variational autoencoder. *Microprocessors and Microsystems*, *76*, 103063. doi:10.1016/j.micpro.2020.103063

Feldman, R., & Hirsh, H. (1996, August). Mining Associations in Text in the Presence of Background Knowledge. In KDD (pp. 343-346).

Grönroos, C. (1982). An applied service marketing theory. *European Journal of Marketing*, *16*(7), 30–41. doi:10.1108/EUM0000000004859

Han, K., Yu, D., & Tashev, I. (2014, September). Speech emotion recognition using deep neural network and extreme learning machine. In Interspeech 2014. doi:10.21437/Interspeech.2014-57

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. *Journal for Language Technology and Computational Linguistics*, *20*(1), 19–62. doi:10.21248/jlcl.20.2005.68

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University. Engineering Sciences*, *30*(4), 330–338. doi:10.1016/j.jksues.2016.04.002

Ito, T., Sakaji, H., Izumi, K., Tsubouchi, K., & Yamashita, T. (2020). GINN: Gradient interpretable neural networks for visualizing financial texts. *International Journal of Data Science and Analytics*, *9*(4), 431–445. doi:10.1007/s41060-018-0160-8

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, *78*(11), 15169–15211. doi:10.1007/s11042-018-6894-4

Khan, A. A., Laghari, A. A., Li, P., Dootio, M. A., & Karim, S. (2023a). The collaborative role of blockchain, artificial intelligence, and industrial internet of things in digitalization of small and medium-size enterprises. *Scientific Reports*, *13*(1), 1656. doi:10.1038/s41598-023-28707-9 PMID:36717702

Khan, A. A., Laghari, A. A., Shaikh, Z. A., Dacko-Pikiewicz, Z., & Kot, S. (2022). Internet of Things (IoT) security with blockchain technology: A state-of-the-art review. *IEEE Access : Practical Innovations, Open Solutions*.

Khan, A. A., Shaikh, A. A., & Laghari, A. A. (2023b). IoT with multimedia investigation: A secure process of digital forensics chain-of-custody using blockchain hyperledger sawtooth. *Arabian Journal for Science and Engineering*, *48*(8), 10173–10188. doi:10.1007/s13369-022-07555-1

Kotler, P. (2001). *Marketing management, millenium edition*. Prentice-Hall, Inc.

Lewis, R. (1983). The marketing aspects of service quality. Emerging perspectives on service marketing, 99.

Li, M., Zhao, H., Su, H., Qian, Y., & Li, P. (2021). Emotion-cause span extraction: A new task to emotion cause identification in texts. *Applied Intelligence*, *51*(10), 1–13. doi:10.1007/s10489-021-02188-7

Li, Y., Hao, W. A. N. G., & Xizi, J. I. N. (2021). Self-organizing Neural Network Algorithm Based on Random Forest Optimization [J]. *Journal of Jilin University Science Edition*, *59*(2), 351–358.

Lissitsa, S., & Kol, O. (2016). Generation X vs. Generation Y–A decade of online shopping. *Journal of Retailing and Consumer Services*, *31*, 304–312. doi:10.1016/j.jretconser.2016.04.015

Luomala, H. (2002). Why We Buy-The Science of Shopping. *Psychology and Marketing*, *19*(1), 111–120. doi:10.1002/mar.1004

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. doi:10.1016/j.asej.2014.04.011

Ning, E., Wang, C., Zhang, H., Ning, X., & Tiwari, P. (2024). Occluded person re-identification with deep learning: A survey and perspectives. *Expert Systems with Applications*, *239*, 122419. doi:10.1016/j.eswa.2023.122419

Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia Journal*, *4*(2), 1883. doi:10.4249/scholarpedia.1883

Rosado-Pinto, F., Loureiro, S. M. C., & Bilro, R. G. (2020). How brand authenticity and consumer brand engagement can be expressed in reviews: A text mining approach. *Journal of Promotion Management*, *26*(4), 457–480. doi:10.1080/10496491.2020.1719955

Song, Y., & Hua, X. (2022). Implementation of Data Mining Technology in Bonded Warehouse Inbound and Outbound Goods Trade. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(3), 1–18. doi:10.4018/JOEUC.291511

Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Jingshen Yixue*, *27*(2), 130. PMID:26120265

Srinivasan, S., & Dyer, C. (2021, August). Better Chinese Sentence Segmentation with Reinforcement Learning. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 (pp. 293-302). doi:10.18653/v1/2021.findings-acl.25

Tian, S., Li, L., Li, W., Ran, H., Ning, X., & Tiwari, P. (2024). A survey on few-shot class-incremental learning. *Neural Networks*, *169*, 307–324. doi:10.1016/j.neunet.2023.10.039 PMID:37922714

Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: A survey. *International Journal (Toronto, Ont.)*, *2*(6), 282–292.

Wang, H. C., Chen, W. F., & Lin, C. Y. (2020). Note Sum: An integrated note summarization system by using text mining algorithms. *Information Sciences*, *513*, 536–552. doi:10.1016/j.ins.2019.11.011

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780. doi:10.1007/s10462-022-10144-1

Wu, Z., Xiaodong, Z. H. U., Yuanning, L. I. U., Chaoqun, W. A. N. G., & Zhiyong, Z. H. O. U. (2022). Iris Quality Evaluation Method Based on GA-SVM Model [J]. *Journal of Jilin University Science Edition*, *60*(1), 89–0098.

Zhao, S., & Chen, J. (2022). The Influence of E-Book Teaching on the Motivation and Effectiveness of Learning Law by Using Data Mining Analysis. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(3), 1–17. doi:10.4018/JOEUC.295092

Zhou, L., Dai, L., & Zhang, D. (2007). Online shopping acceptance model-A critical survey of consumer factors in online shopping. *Journal of Electronic Commerce Research*, *8*(1).

Qun Zhou is a PhD candidate major in Library and information science in School of Information Resource Management, Renmin University of China. Her research interests include information analysis and emergency intelligence research, social media mining and analysis.

Meihua Chen is a lecturer at Shandong Academy of Governance. Her research focuses primarily on intelligence theory and emergency intelligence studies.

Junying Chen is a Phd candidate major in Library and information science in School of Information Management, Wuhan university. His research interests include informatrics, text mining and visualization.

Keren Chen is a postdoctoral fellow in the Seoul School of Integrated Sciences & Technologies. His recent research interests in Sustainable Science & Technology and Environmental Studies.

Sang-Bing Tsai is currently the fellow at International Engineering and Technology Institute (IETI). Dr. Tsai is the Editor-in-Chief for the Journal of Organizational and End User Computing and International Journal of Information Technologies and Systems Approach. His recent research interests in Computer Science, Big Data, Green Operation and Applied Mathematics. He has well over 200 published peer-reviewed journal articles.