# Text Analysis of Mobile Phones Online Reviews based on LDA Model

Min Jiang[a]
*School of Management, Shanghai University, Shanghai, China*

Keywords: Online Reviews, Topic Mining, LDA Model.

Abstract: With online shopping is becoming more and more frequent in our daily life, it will generate a large number of user online reviews. This paper used the LDA (Latent Dirichlet Allocation) model to extract the topic of three brands of mobile phone online reviews. We found that the focus of users' attention is mainly on battery life, running speed, mobile phone appearance, and photograph effect. This work can increase data support for businesses to understand the needs of consumers, to provide targeted products and services, and also provide some suggestions for customers to purchase mobile phones.

## 1 INTRODUCTION

According to the data from *The 47th China Statistical Report on Internet Development* released by China Internet Network Information Center (CNNIC) in 2021 (China Internet Network Information Center, 2021), by December 2020, the number of Internet users in China reached 989 million, and the Internet penetration rate reached 70.4%. Among them, the number of mobile phone users is 986 million, accounting for 99.7% of the total number of Internet users. Moreover, by December 2020, the scale of online shopping users in China reached 782 million, an increase of 72.15 million compared with March 2020, accounting for 79.1% of the total Internet users. It can be seen that online shopping is becoming more and more frequent in our daily life.

With the maturity and large-scale popularization of Internet technology, people's life has also undergone great changes, online shopping has gradually become a way of life for people, and somehow affecting store shopping (Xi, Zhen, Cao, and Xu, 2020). Online shopping not only expands the choice of consumers but also provides consumers with a large amount of product information, including the display of the product's attributes and other consumers' feedback on the use of the product. A lot of users' reviews are affecting the purchase decisions of potential consumers (Hu, 2019), but also data to help merchants better understand the needs of consumers (Kim and Na, 2021).

However, with the massive comment data generated by online shopping, it is more difficult for users and businesses to extract the information they are interested in quickly and efficiently. Therefore, it is very necessary to process these comments with the help of Internet technology.

In this paper, we used the data collector to obtain the comment data of a certain mobile phone in the official flagship stores of Apple, Huawei, and Xiaomi, three popular brands. Taking this as the research data, the LDA (Latent Dirichlet Allocation) model is used to analyze the reviews, the advantages of each brand mobile phone are obtained, and the main concerns of consumers when buying mobile phones are comprehensively obtained. Finally, we put forward some suggestions for consumers and businesses according to the analysis results.

## 2 MATERIALS AND METHODS

### 2.1 Research Method and Data Collection

At present, there are two main ways to crawl web data, one is to develop data grab procedures, the other is to use developed mature software. Fig.1 shows the data collection and cleaning process of this paper. In this paper, we use mature software to achieve the purpose of obtaining data efficiently, the bottom layer

[a] https://orcid.org/0000-0003-2044-6766

Jiang, M.
Text Analysis of Mobile Phones Online Reviews based on LDA Model.
DOI: 10.5220/0011344600003437
In *Proceedings of the 1st International Conference on Public Management and Big Data Analysis (PMBDA 2021)*, pages 342-347
ISBN: 978-989-758-589-0

of this kind of software is also to execute the data capture program, but the front end provides the user with a friendly and simple operation interface. A total of 1980 reviews of a certain mobile phone in Huawei official flagship Store, 1980 reviews of a certain mobile phone in Xiaomi official flagship Store, and 1980 reviews of a certain mobile phone in Apple Store from January 6th, 2021 to May 12th, 2021 are collected by using the mature software. A total of 5940 comments were collected, then we used the Excel and Python languages to clean data.

Latent Dirichlet Allocation (LDA) model is a three-layer Bayesian classical topic model based on a probability graph proposed by David Blei in 2003 (Blei, Ng, & Jordan, 2003), it is an unsupervised text mining technique that extracts topics from initial documents. The LDA model treats a document as a collection of words with no order between them. It assumes that a document has multiple topics and that each topic corresponds to a different word. In the construction process of a document, first, select a topic with a certain probability, and then select a word under this topic with a certain probability, to generate the first word of the document, repeat this process, and then generate the whole article. The use of LDA is the reverse process of the above document generation process, that is, according to an obtained document, to find out the topic of the document and the words corresponding to these topics. LDA model is widely used in the field of text mining, such as text topic recognition, text classification, and text similarity calculation (Wang, Zou, & Liu, 2018).

## 2.2 Data Analysis

As we can see in Fig.1, data cleaning is divided into five steps:

Step 1: deal with irrelevant data, mainly with all kinds of emojis in sentences, by deleting or replacing them with near-sense text;

Step 2: delete blank data, mainly those without comments, the system automatically gives “this user did not fill in the comments!” data like this;

Step 3: data deduplication, delete completely duplicate data;

Step 4: sentence filtering, remove sentences with a length less than 4. Those sentences with a length less than 4 are not of practical significance, so this part of the sentence is removed;

Step 5: delete the repeated words in the sentence, such as “I like, like, like, like this phone”, eliminate the repeated “like” in the sentence.

After the data cleaning, Apple, Xiaomi, and Huawei had 1856, 1979, and 1965 reviews respectively.

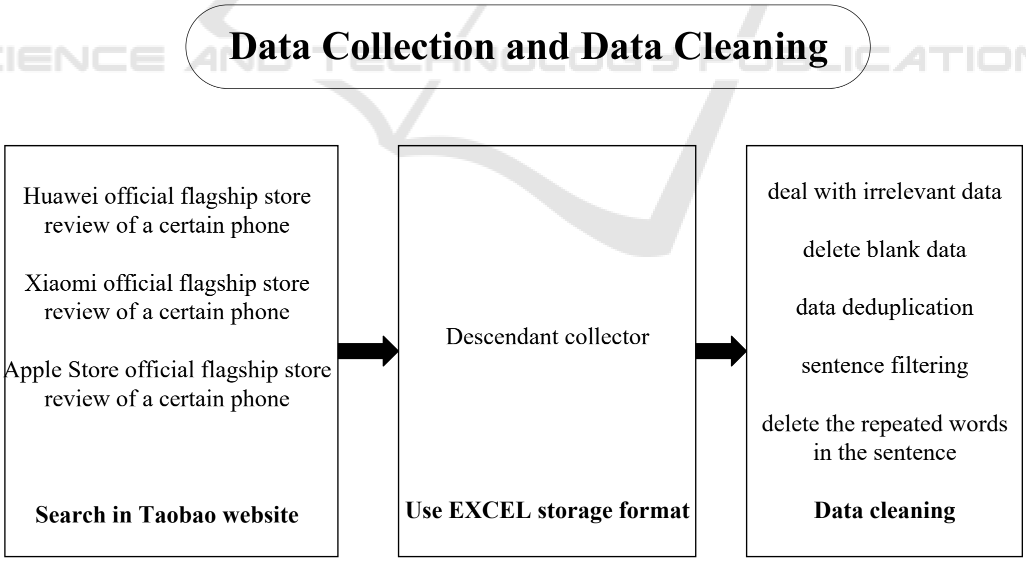

Data Collection and Data Cleaning
Huawei official flagship store review of a certain phone
Xiaomi official flagship store review of a certain phone
Apple Store official flagship store review of a certain phone
Search in Taobao website
Descendant collector
Use EXCEL storage format
deal with irrelevant data
delete blank data
data deduplication
sentence filtering
delete the repeated words in the sentence
Data cleaning

Figure 1: Flow chart of data collection and data cleaning.

Fig. 2 shows the whole experimental design flow chart of this paper.

Firstly, reviews data sets are collected and processed according to mobile phone reviews of various brands.

Secondly, Chinese word segmentation is the most critical step in Chinese text processing, and the quality of word segmentation directly affects the results of text mining. In this paper, Jieba is used for Chinese word segmentation. At the same time, a customized dictionary is set. The language used by Chinese consumers in comments is more casual and diverse than that of foreign comments, with more colloquial expressions. Some popular Internet terms used in Chinese comments do not even exist in Chinese dictionaries. For this, we set up a special custom dictionary, mainly manually inputting those words that the segmentation cannot recognize.

Then remove the stop words. The stop words used in this paper are reconstructed by integrating the four most commonly used Chinese stop words, including 2311 in total, and some customized stop words are added according to the analysis purpose and data characteristics.

Finally, the topics are extracted through the LDA model. We use Python 3.0 to build the analysis model. First of all, visualize topics by using the LDAvis tool, and then extract the top 10 features for each topic, summarize each topic through topic feature words.

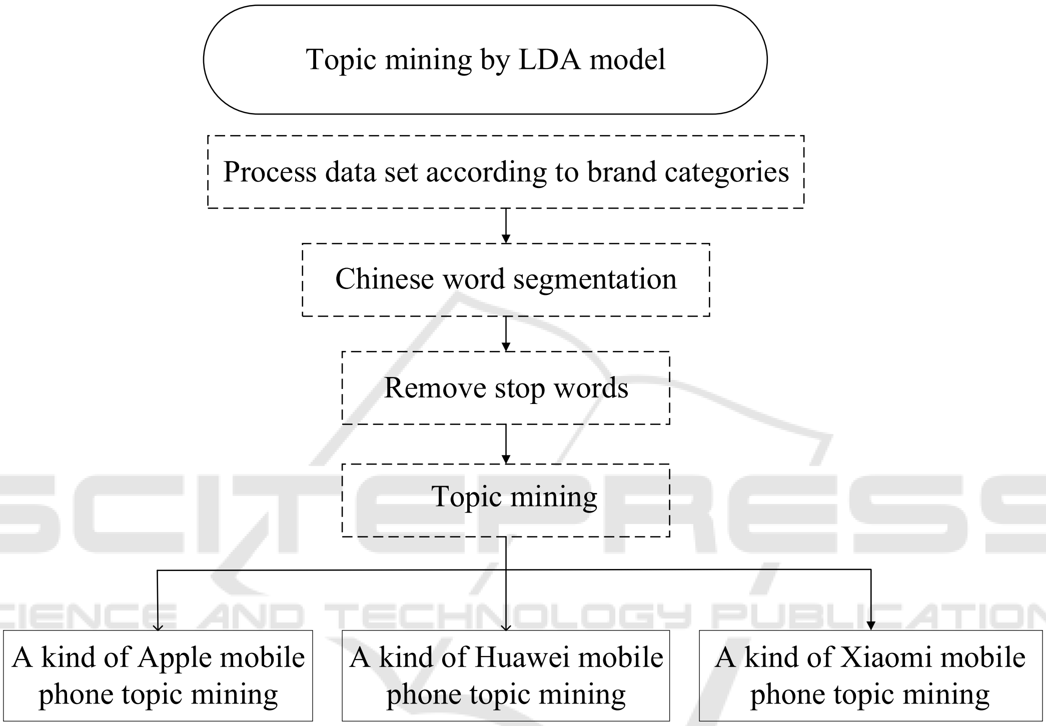

Topic mining by LDA model
Process data set according to brand categories
Chinese word segmentation
Remove stop words
Topic mining
A kind of Apple mobile phone topic mining
A kind of Huawei mobile phone topic mining
A kind of Xiaomi mobile phone topic mining

Figure 2: Experimental design flow chart.

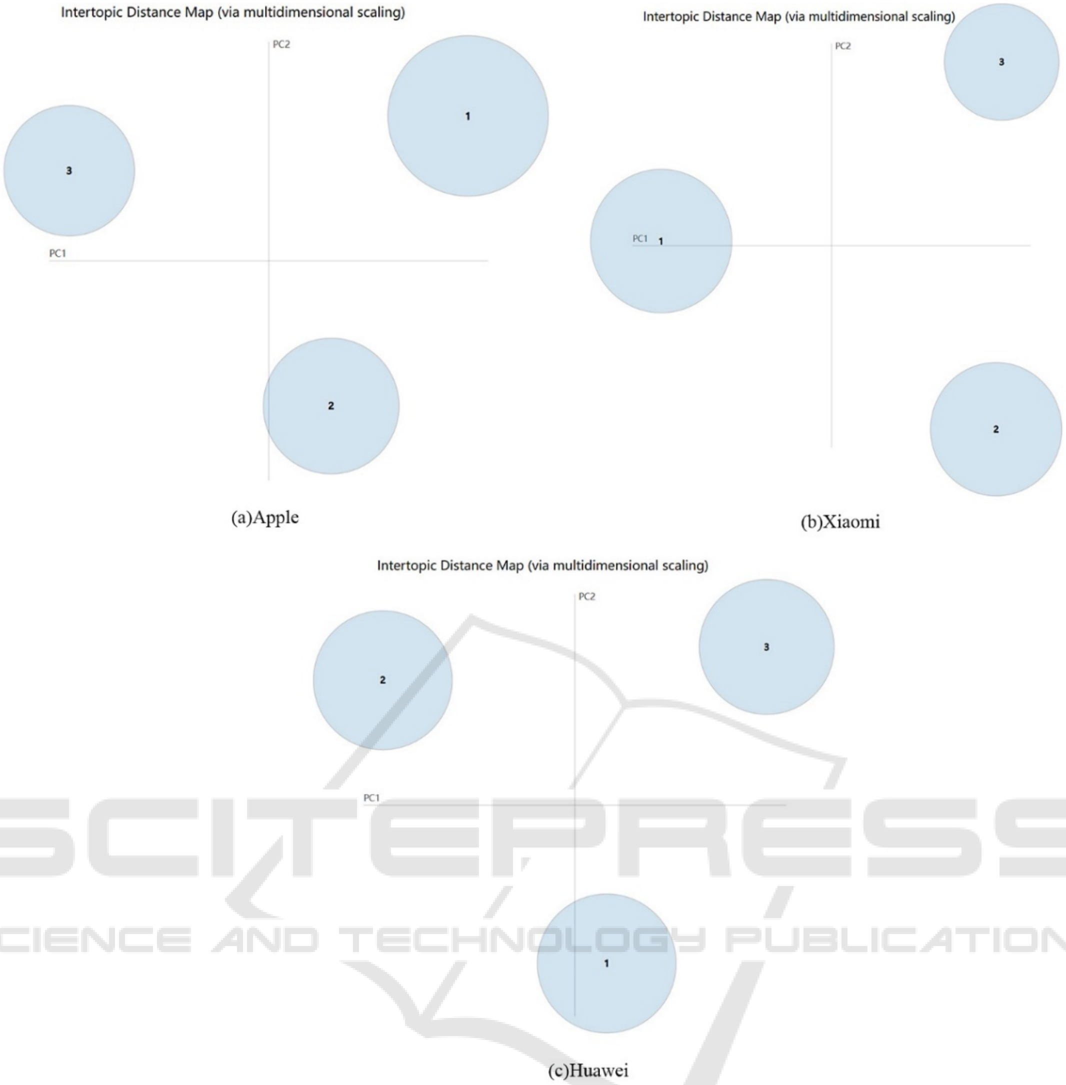
## 3 RESULTS & DISCUSSION

The topic analysis of this paper is to analyze the mobile phone reviews of the three brands respectively, summarize the previous studies of scholars, we set the number of topics as 3 (Wang, Li, Liu, & Zhang, 2021).

The LDAvis tool visually shows the distance between different topics in a two-dimensional vector space, with the size of different circles representing the amount of text the topic contains.

After repeated visual adjustment, as can be seen in Fig. 3, the LDAvis tool intuitively displays the distance of different topics in two-dimensional vector space, and the size of different circles represents the amount of text contained in the topic. In the two-dimensional space vector, there are obvious differences between the three themes, and the distance between the centers of topics 1, 2, and 3 is evenly dispersed without overlap. This result shows that the specified three subject categories are acceptable, and then we confirm that the final number of topics is 3.

In this paper, we use the gensim package in Python 3.0 to analyze the LDA topic model, due to space limitations, the top 10 keywords are listed in Table 1 to Table 3, and then use the keywords to estimate topic meaning.



(a)Apple

(b)Xiaomi

(c)Huawei

Figure 3: Visual display of topic analysis of mobile phone reviews.

Table 1: Topic and keyword of Apple mobile phone reviews.

| Topic | Keywords |
|---|---|
| Topic 1 | ‘certified products’, ‘fast’, ‘feel’, ‘satisfied’, ‘delivery’, ‘battery’, ‘logistics’, ‘official website’, ‘battery life’, ‘green screen’ |
| Topic 2 | ‘speed’, ‘photograph’, ‘battery’, ‘effect’, ‘run’, ‘battery life’, ‘display’, ‘tone quality’, ‘good-looking’, ‘communication’ |
| Topic 3 | ‘photograph’, ‘battery life’, ‘smoothly’, ‘battery’, ‘fast’, ‘run’, ‘speed’, ‘very fast’, ‘effect’, ‘display’ |

From Table 1, by combining the keywords information of the three topics, the characteristics of Apple mobile phones can be summarized as follows:

(a)Good commodity quality and good service;

(b)Good photo effect and tone quality, and it looks beautiful;

(c)The phones run smoothly, but the battery life is poor.

Table 2: Topic and keyword of Xiaomi mobile phone reviews.

| Topic | Keywords |
|---|---|
| Topic 1 | ‘photograph’, ‘charging’, ‘fast’, ‘screen’, ‘speed’, ‘smoothly’, ‘clear’, ‘function’, ‘system’, ‘effect’ |
| Topic 2 | ‘photograph’, ‘effect’, ‘speed’, ‘run’, ‘battery’, ‘battery life’, ‘display’, ‘tone quality’, ‘communication’, ‘features’ |
| Topic 3 | ‘photograph’, ‘effect’, ‘fast’, ‘feel’, ‘battery life’, ‘run’, ‘high’, ‘speed’, ‘smoothly’, ‘screen’ |

As can be seen in Table 2, the characteristics of Xiaomi mobile phones can be summarized as follows:

(a)Good stability and durable battery;

(b)Good service, contained customer service processing speed, good after-sales;

(c)Cost-effective, value for money.

Table 3: Topic and keyword of Huawei mobile phone reviews.

| Topic | Keywords |
|---|---|
| Topic 1 | ‘photograph’, ‘fast’, ‘run’, ‘effect’, ‘smoothly’, ‘speed’, ‘battery life’, ‘support’, ‘battery’, ‘charging’ |
| Topic 2 | ‘speed’, ‘photograph’, ‘fast’, ‘support’, ‘effect’, ‘run’, ‘good-looking’, ‘satisfied’, ‘beautiful’, ‘charging’ |
| Topic 3 | ‘photograph’, ‘battery’, ‘effect’, ‘battery life’, ‘speed’, ‘run’, ‘color’, ‘good-looking’, ‘display’, ‘fast’ |

In the same way, from Table 3, the characteristics of Huawei mobile phones can be summarized as follows:

(a)For personal emotion, customer support domestic products;

(b)Fast running performance and good-looking;

(c)Take pictures ultra high definition and ultra-wide Angle.

## 4 CONCLUSIONS

This paper uses the text of mobile phone users' reviews in the official flagship store to extract the topic. According to our analysis:

(a)Users mainly focus on camera effect, running speed, appearance level, battery life, etc.

(b)The main difference between the three brands of mobile phones is that Apple mobile phones' advantage lies in the running speed of the phones; Xiaomi's advantage lies in its cost performance; The advantage of Huawei phones is that their camera performance is good by comparison.

(c)As a business, it is necessary to spend money, material, and human resources on the focus of users, whether in promoting new products or displaying product information.

For individual users, each of the three brands has its advantages and disadvantages. By comparison, Apple's advantage lies in the running speed of the phone; Xiaomi's advantage lies in its cost performance; the advantage of Huawei mobile phones is that their camera performance is good by comparison. Individual users can choose mobile phone brands according to their own needs.

For businesses, it is necessary to spend money, material, and human resources on the focus of users, whether in promoting new products or displaying product information. According to our analysis, users' several focuses are battery life, photo effect, operation speed, communication signals, tone quality, the stability of the system, and mobile phone appearance level. As a result, businesses should constantly pay attention to these qualities of mobile phones when designing products, will be limited energy and financial resources in these aspects, targeted to meet customer demand. When merchants in the product description are also available in these respects more ink, due to the bounded rationality of consumers, does not have the time and energy to pursue the “best”, but instead “satisfied”, so bluntly show that consumers care about commodity information can help consumers make purchase decisions quickly.

In this paper, users’ online reviews of Apple, Xiaomi, and Huawei mobile phones are selected as the research data to explore users’ concerns about smartphones and the advantages and disadvantages of each brand of mobile phone. In future research, more interesting information can be mined, such as factors influencing user satisfaction, to provide effective management and marketing strategies for businesses. To carry on the comprehensive analysis of the whole online mobile phone market, it will also be an interesting direction to include the relevant information of mobile phone brands and sales volume into the analysis.

## REFERENCES

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. the Journal of machine Learning research, 3, 993-1022.

China Internet Network Information Center. (2021). The 47th China Statistical Report on Internet Development. http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm.

Hu, F. (2019). The relationship analysis between online reviews and online shopping based on B2C platform technology. Cluster Computing, 22 (2), 3365-3373.

Kim, C., & Na, Y. (2021). Consumer reviews analysis on cycling pants in online shopping malls using text mining. Fashion and Textiles, 8 (1), 1-21.

Xi, G., Zhen, F., Cao, X., & Xu, F. (2020). The interaction between e-shopping and store shopping: Empirical evidence from Nanjing, China. Transportation Letters, 12 (3), 157-165.

Wang, L., Zou, L., & Liu, X. (2018). Visualizing Document Correlation Based on LDA Model. Data Analysis and Knowledge Discovery, 2 (03), 98-106.

Wang, X., Li, Y., Liu, T., & Zhang, L. (2021). Research on the Collaborative Model of Sentiment Analysis and Topic Mining of Micro-blogging Users in the Context of COVID-19. Journal of the China Society for Scientific and Technical Information, 40 (03), 223-233.