WEP
Warwick
Evans
Publishing

Sentiment Analysis of Chinese R&B Music Album Short Reviews Using SKEP Topic Model

Sa Xiao[*]

College of Computer Science and Technology, Jilin University, Changchun, 130012, China

*Corresponding author: xiaosa2122@mails.jlu.edu.cn

Abstract. In the current online environment, music reviews can be a good reflection of the popularity of the music, which can attract others to listen and pay attention to the music. A sentimental analysis of these reviews and the identification of emotional attitudes are essential for the dissemination of music. Based on the classic Chinese R&B album "David Tao" and its related reviews, this paper aims to analyze sentiment tendencies, analyze word frequency, and extract comment topics based on the Baidu AI Cloud interface of the SKEP algorithm. Research has found that listeners have a relatively positive emotional attitude towards this album, with positive reviews accounting for about 85%, and the reviews cover a variety of hot topics. The album's popularity, word-of-mouth evaluation, influence and other indicators all show a high level, highlighting that this album indeed provides a certain direction and strategy for the development of Chinese R&B. The charm of the album's musical style has been fully reflected and attracted widespread attention and recognition. It not only injects new vitality into the development of Chinese R&B music, but also provides valuable reference and inspiration for those interested in this music style. The research in this article emphasizes the important role of music in cultural inheritance and innovation, and also provides useful reference and guidance for the sustainable development of the music industry.

Keywords: SKEP; Chinese R&B; Sentiment analysis.

1. Introduction

With the continuous development of the times, more and more users will choose to express their views, opinions, emotions, and attitudes on social networks and public platforms. After listening to a song or an album, music lovers can leave reviews on the relevant websites, which are valuable references for others. However, as the number of reviews increases, judging whether a song or album is good or not based on just a handful of reviews may seem too one-sided. Therefore, to have a more comprehensive understanding of users' views and emotions, more emotional features need to be mined from a large number of reviews so that albums can be analyzed more in-depth and comprehensively. Such analysis can provide more valuable information for music lovers and help music practitioners better understand the needs and preferences of their audiences, thus improving the quality and popularity of their music works.

Regarding related sentiment analysis studies, Monsalve Pulido Julian et al. combined text, images, and numbers to propose a multimodal sentiment analysis model for Spanish-language data in the field of tourism to perform sentiment analysis on tourism data [1].Bowen Xiao et al. performed sentiment analysis on textual data collected from four Reddit communities to understand social media platforms Reddit to understand public perceptions and attitudes towards NIPT [2]. Dohun Kim et al. used LDA and SERVQUAL models to compare changes in airline passengers' perceptions of airline services before and during COVID-19, analyzed which keywords influenced passengers' satisfaction with airline services, and compared negative keyword changes [3]. Mahammad Khalid Shaik Vadla et al. used the Bert model and T5 for sentiment analysis to understand customer emotions and preferences in a dynamic product design environment, providing data-driven insights for product designers and research developers [4]. J P Chart Pascual et al. conducted an analysis of all references published between 2008 and 2022 to SRD and the names of the main drugs used in BD and found that English-language tweets about economic and legal aspects showed negative sentiment, whereas Spanish-

language tweets seeking advice showed surprise [5]. Victoria Vought et al. used demographic data recorded on a publicly available website to conduct a sentiment analysis of pediatric ophthalmologists' ratings using VADER and word-frequency analysis to assess pediatric ophthalmology patient satisfaction and identify trends in patient sentiment that can be used by pediatric ophthalmologists seeking to improve healthcare delivery [6].

Therefore, this experiment primarily extracts the short review texts of the "David Tao" album from the Douban website and the comment content of the "David Tao" album analysis videos on the Bilibili website. It utilizes the Baidu AI Cloud interface based on the SKEP algorithm to mine and analyze the comment texts of this album. This is to obtain the commenters' emotional attitudes towards the album, and thereby provide a more comprehensive review of the album's impact on Chinese R&B. It also aims to offer some suggestions for those who want to understand and get in touch with Chinese R&B.

2. Data and Methodology

2.1. Data sources

R&B, Rhythm and blues is a music genre that has been playing an important role in the world's popular music scene since its birth in the 1940s. Chinese R&B, on the other hand, has been developing since the last thirty years. Among them, David Tao, known as the godfather of Chinese R&B, whose first album "David Tao" is regarded as the pioneer of Chinese R&B. Through this album, David Tao turned R&B into a kind of pop, and made the Chinese music scene enter a golden age with his unique R&B style, which made R&B style spread in the Taiwanese music scene in a real sense and influenced several generations. Part of the data for this study comes from the music reviews and ratings of the album "David Tao" on the Douban website, while the other part of the data comes from "This album made him the godfather of Chinese R&B? Full analysis of David Tao 's album of the same name" on the bilibili website Comments on the video.

Douban, as a community website, provides information about books, movies, music, and other works, and a large number of ratings and reviews related to books and videos exist on the website. Bilibili, is an iconic brand and leading video community for China's younger generation, with a diverse and large number of videos within the site, building an ecosystem that constantly generates high-quality content and a multicultural community that covers more than 7,000 circles of interest. Therefore, this experiment is based on Python crawler technology and crawled the relevant short reviews of the album "David Tao" on the Douban website and the video titled "This album made him the godfather of Chinese R&B? Full analysis of David Tao's album of the same name" on the bilibili website's comments, a total of 10,553 data related to short reviews were obtained. Then, check the Douban rating; 33,490 people rated it. After obtaining the data, data preprocessing was performed to eliminate duplicate and irrelevant reviews. Finally, 8216 pieces of relevant data were obtained, including comments, number of likes, whether it was useful, commenter ID, date, etc., of which 669 were valid comment data.

2.2. SKEP

SKEP is a sentiment pre-training algorithm based on Sentiment Knowledge Enhancement proposed by Baidu's research team.This algorithm uses an unsupervised approach to mine sentiment knowledge automatically, and then utilizes the sentiment knowledge to construct pre-training goals, so that the machine learns to understand the semantics of the sentiment. SKEP provides a unified and powerful representation of the semantics of sentiment for various types of sentiment analysis tasks [7] .

The input to the SKEP model has two parts, one is the object (Aspect) to be evaluated and the other aspect is the corresponding comment text. After splicing the two, they can be passed into the SKEP model, which semantically encodes the text string and then takes the token output vector at the CLS

location as the final semantic encoding vector. The next step is to perform sentiment classification based on this semantic coding vector.

First, some a priori knowledge is obtained from large-scale textual data in an unsupervised manner. The following are obtained using PMI's assessment: sentiment Word, Word Polarity, Aspect-sentiment Pair respectively.

$$WP(w) = \sum_{WP(s)=+} PMI(w,s) - \sum_{WP(s)=-} PMI(w,s) \tag{1}$$

Next, a Mask is required in three steps:

① Mask Aspect-sentiment Pair, but at most two pairs are randomly masked in a sentence.

② Mask Sentiment Word: In a sentence, the number of tokens to be masked cannot exceed 10% of the total number of tokens in the current sentence.

③ Mask Generic Word. However, there are certain prerequisites: the proportion of tokens accounted for by sentiment words in the second step does not exceed 10%.

Next, for emotional pre-training, this algorithm needs to use the objective optimization function L, which consists of three objective optimization functions, namely $L = L_{sw} + L_{wp} + L_{ap}$, where Sentiment Word Objectives (L_{sw}), Word Polarity Objectives (L_{wp}) and Aspect-sentiment Pair target(L_{ap})

① Sentiment Word Objectives L_{sw}

$$\breve{y}_i = \text{softmax}(\breve{x}_i W + b) \tag{2}$$

$$L_{sw} = \sum_{i=1}^{i=n} m_i \times y_i \log \breve{y}_i \tag{3}$$

② Aspect-sentiment Pair Objectives L_{ap}

The methods ① and ② are similar, except that the former calculates the loss of the token, while the latter calculates the loss of the polarity, which can be interpreted as another type of token, with the difference that there are only two total classes: positive and negative.

③ Aspect-sentiment Pair Objectives L_{ap}

$$\overline{y_a} = \text{sigmoid}(\breve{x}_i W_{ap} + b_{ap}) \tag{4}$$

$$L_{ap} = \sum_{i=1}^{a=A} y_a \log \overline{y_a} \tag{5}$$

3. Analysis of Results

In order to better discover the listeners' emotional attitudes towards this album, this section analyzes the relevant data from the three aspects of evaluation distribution, word frequency analysis, and sentiment analysis, respectively, as follows.

3.1. Distribution of evaluations

The album's evaluation distribution is analyzed to understand the audience's attitude towards the album. The evaluation is divided into five levels, increasing from 1 to 5. The proportion is shown in Fig. 1. Among them, 1-star proportion accounts for 0.10%, 2-star proportion accounts for 0.10%, 3-star proportion accounts for 2.10%, 4-star proportion accounts for 14.00%, and 5-star proportion

accounts for 83.70%. In short, the overwhelming majority of the five-star rating stations, and the 1-3 stars ratio is even less than 3%. This shows that the audience deeply loves the album and is worth listening to.
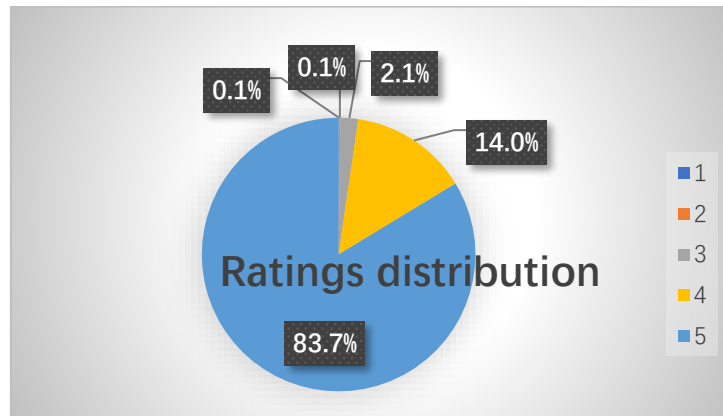

Fig. 1 Ratings distribution

3.2. Word frequency analysis

The frequency of words appearing in the short reviews can be analyzed to determine the public's emotional attitude towards the album preliminarily. First, use the Jieba library and nltk library to segment the Chinese and English words in the comments, respectively. Then, the Chinese and English stop words in the text are processed respectively based on the Baidu stopwords list and the English stopwords list in the nltk library, and the stopwords are filtered. Finally, the collections database was used to complete word frequency statistics, and a word cloud diagram was drawn. After translating the Chinese vocabulary into English, Fig. 2 was obtained.

As shown in Fig. 2, it can be seen that the frequently mentioned words include "DavidTao", "r&b", "album", etc. Among them, words related to emotional attitudes are "love", "like", "good", "best", etc. Basically, emotional attitudes are biased towards the positive. This shows that this album has a high reputation in the eyes of the public and is an excellent work.
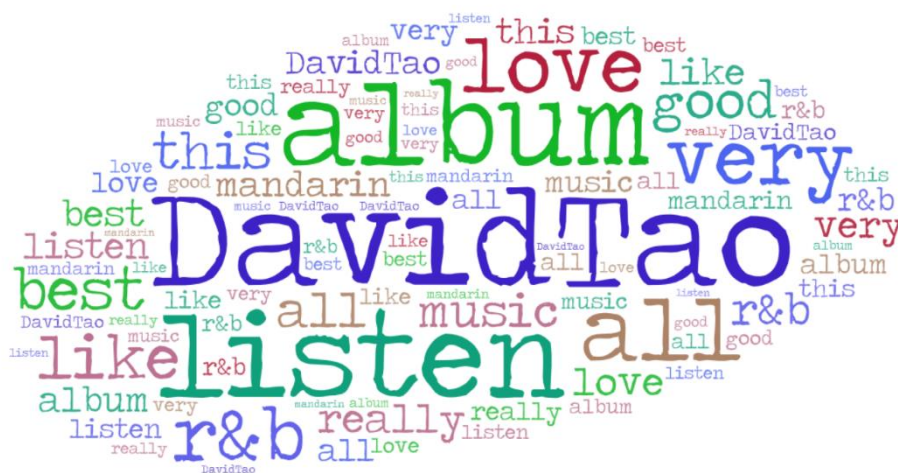

Fig. 2 Word cloud diagram

3.3. Sentiment analysis

The obtained short review text is subjected to sentiment analysis, in which the sentiment polarity is divided into three levels: 0: negative, 1: neutral, 2: positive. The five most popular comments (excluding controversial comments) were extracted:

(1) His first time defined his status as the father of R&B in the Chinese music scene.

(2) He is the first person to do R&B in China.

(3) He did not surpass this album later. The melancholy on the cover is Tao Zhe's background.

(4) Tao Zhe's first album is a grand debut of R&B music in the Chinese music scene. It is also the warmest and calmest work in Tao Zhe's music career. Compared with subsequent works, it has a very strong integrity. The song "Beach" is indeed a classic.

(5) The idea is awesome, the arrangement is awesome, and the singing is awesome.

It can be seen from Table 1 that the emotional polarity of the five most popular comments is all 2, which further indicates that the public loves the album. And it can be initially determined that the hot topics discussed by the audience are music types, creative concepts, impact on the music scene, etc.

Table 1. Information related to the five most popular comments

Comment number	Number of useful comments	Star rating	Emotional polarity
(1)	1152	5	2
(2)	693	5	2
(3)	661	5	2
(4)	242	5	2
(5)	168	5	2

Based on the SKEP algorithm, the sentiment analysis results are derived from statistical analysis and plotted as shown in Fig. 3, in which the number of reviews with sentiment polarity of 2 is 571, with a proportion of 85%, accounting for the vast majority; the number of comments with sentiment polarity of 1 is less, with a total of 12 comments, with a proportion of 2%; and the number of comments with sentiment polarity of 0 has a total of 86 comments, with a proportion of 13%. It can be judged that most people have a favorable view of this album, but there is still a small percentage of people who have a neutral or negative view. Overall, this album is still popular and worth listening to.
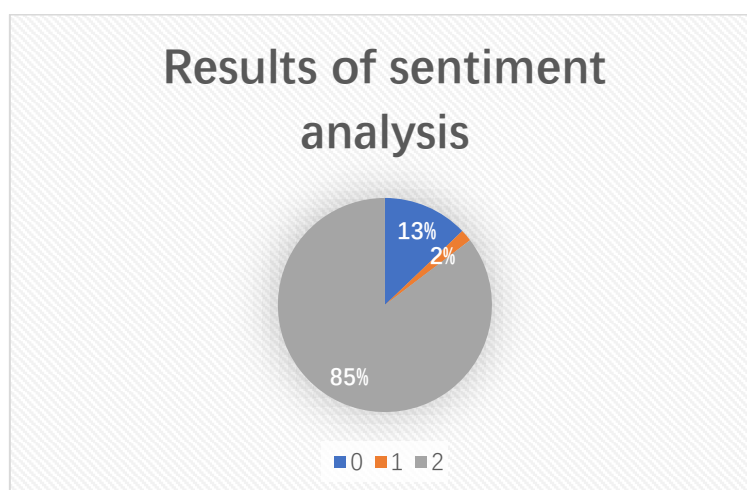

Fig. 3 Sentiment analysis results

Then summarize the topics involved and the topics of comments: music type, album influence, album style, listening experience, etc. The positive and negative review texts related to these four topics were extracted, and the key evaluation information obtained is shown in Table 2. Among them, the positive words in the "music type" topic are: novel, injecting vitality, and freshness; the positive words in the "album influence" topic are: the beginning of R&B style, classic textbook, leading the times; the positive words in the "album style" topic are: warm, calm, amazing, and strong in integrity;

the positive words in the "listening experience" topic include: awesome, very good, perfect production, very nice, looped, never tire of listening; and the negative words corresponding to these related topics are: too much English is strange, can't get used to listening to it, The negative words associated with these topics are: too much English is weird, can't stand it, don't like it, can't listen to it, greasy, non-mainstream, juvenile.

Table 2. Related topics Positive and negative evaluation keywords

Related Topics	Positive	Negative
Music type	novel, energizing, fresh	too much English is weird, can't stand it, don't like it, can't listen to it, greasy, non-mainstream, juvenile
Album influence	the beginning of R&B style, classic textbook, leading the times	
Album style	warm, calm, stunning, and comprehensive	
Listening experience	awesome, very good, perfect production, very nice to listen to, can be played repeatedly, and you will never tire of listening to it	

4. Discussion

This experiment analyzes the results of relevant data from three aspects: distribution of evaluation, word frequency analysis, and sentiment analysis. The results obtained are also progressive layer by layer. First, in the distribution of evaluation part, the audience's attitude towards this album is generally understood, and then in the word frequency analysis, the public's emotional attitude towards the album was initially judged. Finally, sentiment analysis was performed to obtain the experimental results, which provided an in-depth understanding of the album's popularity among the public.

However, the jieba library used in this experiment for Chinese word segmentation and the Baidu stopwords list do not have a very good effect on the word exclusion of the text data used in this experiment, and further processing is still needed. And SKEP still has some errors in analyzing the sentiment of sentences, for example, the review "The whole album is breathtakingly incredible" is analyzed as having a sentiment polarity of 0, i.e., negative, which is clearly the opposite of what it actually is. In addition, this study only conducted an emotion-related analysis of the review text, summarized and extracted conclusions from an emotional perspective, and commented on and interpreted the album without conducting more in-depth and detailed research. There is also a lack of further analysis and prediction on the future development of Chinese R&B and a lack of more in-depth quantitative research on issues such as the types of songs preferred by contemporary netizens.

5. Conclusion

This paper utilizes the Baidu AI Cloud interface based on the SKEP algorithm to perform sentiment analysis on the reviews. By refining the topics of the reviews and summarizing the frequency of vocabulary, this study finds that most of the words with a high frequency of occurrence are positive words, and that the majority of the emotional attitudes towards the comments of this album are positive, with only a small portion being negative. Generally speaking, most people like this album very much, but there are also a few controversies. Overall, it has played a certain positive role in the development of the Chinese music scene. This album has a certain influence on the Chinese music scene and can highlight the charm of related music types and increase the popularity of Chinese music.

At the same time, it can understand the audience's emotional attitude towards this album, so that the public has a clear understanding of this album, which is conducive to the overall understanding of this album.

This study can make certain references for music lovers' listening direction to avoid wasting a lot of time; at the same time, it will analyze the influence of the album on the popularity of Chinese R&B, and then show the characteristics of this music style of Chinese R&B, which can make more people support Chinese music and enhance the popularity of Chinese music.

However, this paper has only conducted a sentiment analysis, summarizing the conclusions and extracting them from an emotional perspective to comment on and interpret the album. Further analysis and predictions development of Chinese R&B are needed, as well as more in-depth quantitative research on issues such as the types of songs preferred by contemporary netizens. Future research in this area may integrate multiple social or rating platforms to more comprehensively analyze music lovers' preferences, personalize song recommendations, and provide valuable insights into the development of music and popular aesthetics.

References

[1]   Monsalve Pulido Julian, Parra Carlos Alberto, Aguilar Jose. Multimodal model for the Spanish sentiment analysis in a tourism domain. Social Network Analysis and Mining, 2024, 14 (1).

[2]   Xiao B ,Yan J ,Hayeems Z R. Public Opinions and Attitudes toward Non-invasive Prenatal Testing on Reddit: Content and Sentiment Analysis. Public health genomics, 2024.

[3]   Kim D ,Lim C ,Ha K H . Comparative analysis of changes in passenger's perception for airline companies' service quality before and during COVID-19 using topic modeling. Journal of Air Transport Management, 2024, 115 102542-.

[4]   Vadla S K M ,Suresh A M ,Viswanathan K V . Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT. Algorithms, 2024, 17 (2).

[5]   Pascual C P J ,Torres M M ,Ortega A M , et al. Areas of interest and sentiment analysis towards second generation antipsychotics, lithium and mood stabilizing anticonvulsants: Unsupervised analysis using twitter. Journal of affective disorders, 2024, 351:649-660.

[6]   Vought V ,Vought R ,Sharma R , et al. Evaluating Pediatric Ophthalmic Care Using Sentiment Analysis of Physician Review Sites. Journal of pediatric ophthalmology and strabismus, 2024, 1-8.

[7]   Tian H , Gao C , Xiao X ,et al. SKEP: Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis. ACL, 2020.