J Vis
https://doi.org/10.1007/s12650-024-00966-2

REGULAR PAPER

Longfei Chen · Qianyu Liu · Chenyang Zhang · Yangkun Huang · Zhenhui Peng · Haipeng Zeng · Zhida Sun · Xiaojuan Ma · Quan Li

Amplifying the music listening experience through song comments on music streaming platforms

Received: 27 November 2023 / Revised: 7 January 2024 / Accepted: 26 January 2024

Abstract Music streaming services are increasingly popular among younger generations who seek social experiences through personal expression and sharing of subjective feelings in comments. However, such emotional aspects are often ignored by current platforms, which affect the listeners' ability to find music that triggers specific personal feelings. To address this gap, this study proposes a novel approach that leverages deep learning methods to capture contextual keywords, sentiments, and induced mechanisms from song

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12650-024-00966-2.

L. Chen · Q. Liu · Q. Li
School of Information Science and Technology, ShanghaiTech University, Shanghai, China
E-mail: chenlf@shanghaitech.edu.cn

Q. Liu
E-mail: liuqy2022@shanghaitech.edu.cn

L. Chen · Q. Liu · Q. Li (✉)
Shanghai Engineering Research Center of Intelligent Vision and Imaging China, Shanghai, China
E-mail: liquan@shanghaitech.edu.cn

C. Zhang
Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, USA
E-mail: zhang414@illinois.edu

Y. Huang
Tandon School of Engineering, New York University, New York, USA
E-mail: yh1875@nyu.edu

Z. Peng
School of Artificial Intelligence, Sun Yat-Sen University, Guangzhou, China
E-mail: pengzhh29@mail.sysu.edu.cn

H. Zeng
School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, China
E-mail: zenghp5@mail.sysu.edu.cn

Z. Sun
College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
E-mail: zhida.sun@outlook.com

X. Ma
Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
E-mail: mxj@cse.ust.hk

Published online: 10 March 2024

comments. The study augments a current music app with two features, including the presentation of tags that best represent song comments and a novel map metaphor that reorganizes song comments based on chronological order, content, and sentiment. The effectiveness of the proposed approach is validated through a usage scenario and a user study that demonstrate its capability to improve the user experience of exploring songs and browsing comments of interest. This study contributes to the advancement of music streaming services by providing a more personalized and emotionally rich music experience for younger generations.
Keywords Music comments · Music streaming services · Visualization

1 Introduction

The integration of music into individuals' daily lives has become increasingly prevalent. In particular, the emergence of music streaming platforms, including *YouTube Music* and *NetEase Cloud Music*, has provided users with not only fundamental music listening capabilities but also a forum-like comment section that enables listeners to share their sentiments, preferences, and personal narratives. These social functionalities have significantly altered the mode of music consumption and have facilitated connections between individuals with similar musical preferences, overcoming geographical constraints (Wang and Fu 2020; Chen 2018).

The act of leaving comments on music has become a widespread means for listeners to express their agreement, sympathy, or disagreement with others by sharing their experiences in embedded forums on music streaming platforms or by responding or "liking" other users' comments. In this study, we extract listeners' interests, preferences, and emotions toward songs, as well as related personal experiences from a vast number of music comments. The results obtained have the potential to assist users in exploring songs with specific emotions (e.g., happiness or homesickness) or in particular scenarios (e.g., work and parties). In such instances, users can peruse a set of song lists labeled with various keywords provided by the platform or other users that correspond to a specific mood or situation. Subsequently, the user may select a specific song from the list to listen to, either randomly or based on prior experience. While listening to the song, the user can also browse comments from other listeners, some of which may be of interest to them. We refer to this process as the "Library → Song List → Song → Comment" exploration process. However, exploring music comments to help users find songs and comments of interest is not a trivial task for several reasons. First, existing song lists on most platforms only show basic music metadata, such as the song's title, singer, and album name, while comments, which are the primary means of conveying personal feelings, are not readily accessible to users in the song list interface. As a result, it is challenging for users to identify songs in a given list that may elicit the appropriate nuances of their emotions. Second, the manner in which comment sections are organized on current music streaming platforms makes it difficult for listeners to locate posts that are relevant to their interests. Most platforms only offer two types of comment filtering: "sort by time" and "sort by popularity." "Sort by time" tends to neglect early comments, while "sort by popularity" tends to overlook minority comments. A considerable number of comments are arranged in chronological order or sorted by popularity, requiring users to sift through them one by one to find the ones that interest them. Third, music comments can be complex and convoluted. Comments over time can be intertwined to address several topics, some of which are not entirely related to the song itself. For instance, ardent fans may boast about their group identity, which may be unrelated to the music (Ren et al. 2012; Sugiana and Hafiar 2018). Specific events, such as Michael Jackson's passing, may also spark discussions in comments (Siddiqi and Sharan 2015).

In order to assess the viability of the concept of enhancing songs with information derived from music comments, aimed at aiding users in the discovery of songs and comments that pique their interest, we carried out a questionnaire involving a cohort of 104 music streaming service users who were actively engaged. The results revealed that 75.49% of the participants expressed a keen interest in the integration of a music comment feature within existing platforms, with 61.04% of them having actively contributed comments themselves. Moreover, an additional 58.82% of the respondents held the belief that prevalent mainstream platforms inadequately acknowledge the significance of comments in enhancing the overall user experience. In this study, we introduce two features that enhance an existing music streaming application. These features are the comment-related preview tags for songs and map metaphor visualization of comments. In order to generate the preview tags for each song in the song list, language models (LM) are used to analyze all the comments. This analysis includes the extraction of high-frequency keywords, topic detection, sentiment analysis, and identification of the induced mechanism. The eight preview tags for each song are then displayed on the song list page. The aim of this feature is to optimize user experience in the "Song List ⟶ Song" subprocess and provide a personalized music exploration for song seekers. The comment map feature on the comment details page structures and summarizes the comments, which facilitates the identification of personally relevant interests in the "Song → Comment" subprocess. The proposed approach undergoes evaluation via a usage scenario, followed by a user study assessing the usefulness, effectiveness, user interactions, and impact of our approach. The primary contributions can be summarized as follows.

- We conducted a formative study aimed at comprehending the requirements of users when utilizing music streaming platforms and capturing their browsing experiences with music comments.
- We fine-tuned pre-trained LMs to extract keywords of high frequency, identify topics, assess sentiment, and determine induced mechanisms from chronologically organized music comments. Subsequently, we enhanced an existing music streaming platform with innovative visual designs to empower end-users with a comprehensive overview of song comments and their dynamics.
- We conducted a usage scenario and a user study involving end-users to validate the efficacy of our proposed approach. Through these activities, we derived valuable insights and presented evidence demonstrating the effectiveness of our approach.

2 Related work

2.1 Sentiments and induced mechanisms behind music comments

Over the past decades, there has been significant attention given to empirical studies on musical sentiments (Marin and Bhattacharya 2010; Zeng et al. 2020, 2019). These studies have consistently reported that music can effectively communicate sentiments to the listener and may also induce the listener's own sentiments. For example, Marin et al. Marin and Bhattacharya (2010) suggest that "*automatic affective responses to stimuli are essentially relevant for subsequent cognitive, emotional, and behavioral reactions: emotions induced by music crucially influence the (emotional) processing of pictures, facial expressions, films, and words.*" As a result, listeners' comments on music reflect the sentiments they experience toward the song, which may be a result of the music itself or the induced effects of the music.

To comprehend what triggers musical sentiments, the field of psychology has developed two complex explanatory studies based on human cognitive theory. The first study by Schere et al. Scherer and Zentner (2001) proposes a broad interaction profile based on the Component Process Model (CPM), which assumes that there are three emotions associated with music and that experts can calculate the contribution of four factors (structural, performance, listener, and contextual features) in inducing sentiments. However, the weights of the parameters involved in the above factors require experimental data to be calculated and are challenging to quantify directly from the comment text. Alternatively, Juslin et al. Juslin (2013) proposed the BRECVEMA framework, which summarizes eight induced mechanisms, at least one of which generates musical sentiments. In this work, we incorporate the underlying sentiment and the induced mechanism as additional properties of the comment text to help users better explore the comments of a song.

2.2 Comment data processing

The predominant focus of comment data processing studies has been sentiment analysis (Pang et al. 2002). However, binary classification of sentiments as positive or negative is deemed inadequate for offering comprehensive insights for human understanding. Following the sentiment classification approach proposed by Chuang et al. Chuang and Wu (2004); Eckman (1972), we categorize sentiments into six types: *angry*, *neutral*, *sad*, *fear*, *surprise*, and *happy*. Subsequently, we employ fine-tuned deep-learning models to predict the sentiment category of each comment. Similarly, we utilize another fine-tuned model to predict the category of the induced mechanism. To categorize the induced mechanisms, we adopt the *BRECVEMA* framework (Juslin 2013), as introduced by Juslin. Our approach involves referencing and aligning with this framework for efficient classification.

Moreover, comment analysis research has also investigated methods for detecting keywords (Bharti and Babu 2017; Hasan and Ng 2014). Statistical-based techniques utilize frequency measures to choose the top n candidates based on the linguistic corpus. Graph-based methods (Beliga et al. 2015) use bag-of-words (BOW) with co-occurrence metrics, which generates an n-dimensional vector for each document. Linguistic approaches utilize the linguistic features of the words for keyword detection (Barzilay and Elhadad 1999; Hulth 2003). However, the aforementioned techniques are limited by their reliance on surface-level word frequency, statistics, or rule-based approaches, constraining their ability to capture intricate semantic information. Various machine learning methods, such as Naïve Bayes (Uzun 2005), SVM (Cortes and Vapnik 1995), HMM (Baum et al. 1970), and CRF (Lafferty et al. 2001), have been explored for keyword extraction. However, these methodologies often treat words as independent features and struggle to capture the intricate relationships and contextual dependencies between them. In contrast, deep learning methods (Graves et al. 2013; Graves and Schmidhuber 2005; Zhang et al. 2016) exhibit enhanced capability in extracting keywords by effectively harnessing contextual semantic information. These models can automatically learn linguistic patterns and contextual relationships within the text, thereby providing a more comprehensive understanding of the underlying data. By incorporating attention mechanisms, contextual encoding, and other techniques, these deep learning models further demonstrate improved accuracy in inferring and predicting the significance of individual words. To annotate keywords based on their contextual relevance to the given sentence, our study employs a self-supervised deep learning approach (Grootendorst 2020).

2.3 Visualization of sentiments and topics in text

Over the past decade, a diverse range of visualization techniques has emerged for the analysis of sentiment and thematic patterns in text, encompassing both simplistic infographics and sophisticated visual analytics systems (Kucher et al. 2018). For example, one such approach known as *Review Spotlight* (Yatani et al. 2011) utilizes a tag cloud consisting of prominent adjective/noun pairs to provide a summary of customer reviews. Another technique, referred to as *OpinionBlocks* (Alper et al. 2011), employs multiple coordination bars and text tags to explore polarities and noteworthy keywords associated with specific product features. However, these existing methods fail to adequately capture the temporal evolution of comments. Additionally, Zhao et al. (2014) introduced a technique called *PEARL* to analyze the sentiment changes of individual users on social media over time. However, this approach falls short in the context of our study, as the analysis of a single user is insufficient to address our requirements.

The visualization of topics has been a widely explored area in the literature, typically using stacked graphs (Byron and Wattenberg 2008; Dörk et al. 2010; Havre et al. 2002; Leskovec et al. 2009). Interactive visualizations have been proposed by *TIARA* (Liu et al. 2009; Wei et al. 2010), which combines text summarization with interactive visualization, and *Textflow* (Cui et al. 2011), which employs a river-based visualization to display merging and splitting relationships between topics. Word clouds with sentiment analysis have also been utilized for comment visualization (Kucher et al. 2018; Wang et al. 2018). Chen et al. (2016, 2017, 2019) developed a structured semantic space for exploring the social media model of ego-centric information diffusion and event evolution. In this approach, users involved in retweeting central posts are mapped to a hexagonal grid based on the similarity and temporal order of their behavior, with the edges of each hexagonal cell constructed to follow the retweets and replies between them. However, in the case of music comments, most of them are original with no forwarding relationships, making it challenging to construct relationships and track the evolution. Motivated by the aforementioned studies, we propose a novel visualization approach that utilizes the map metaphor to restructure music comments.

3 Formative study

In order to better understand the actual needs of users when using music streaming platforms and the process by which users connect through music comments, we conducted a formative study, wherein participants were invited to respond to a questionnaire that aimed to elicit their experiences in browsing music comments.

3.1 Procedure

We designed a questionnaire to discover users' usage scenarios and preferences when viewing, receiving, analyzing, and writing music comments. First, we designed questions to gather information on users' usage scenarios and preferences regarding viewing, receiving, analyzing, and writing music comments. Initially, the questionnaire probed participants on their usage habits of the commenting function and preferences for

Table 1 A total of 102 valid responses were obtained in the questionnaire survey

Have a habit of browsing comments while using music streaming platforms?							
Yes (77/102)						No (25/102)	
Score different types of comments based on level of interest							
Comment Type	AVG	SD	Comment Type	AVG	SD		
Contextual Background	4.25	0.93	Latest Updates	2.56	1.03	Why you don't like browsing comments?	
Expert Analysis	4.16	0.97	Personal Experiences	2.49	1.23		
Shared Feeling	3.92	0.93	Real-time Commentary	2.27	1.14	No habit of browsing	(18/25)
Trending Highlights	3.56	0.97	Social Media Trends	2.08	1.09	No content of interest	(10/25)
Creative Team Insights	3.47	1.21	Fan Sentiments	1.90	0.97	Affect experience of listening	(6/25)
Literary Assessment	2.90	1.28	Concise Remarks	1.57	0.85	Massive data with no classification	(5/25)
Creative Excellence	2.86	1.23	\	\	\	Unable to empathize	(5/25)
When do you usually browse comments?			Have comments influenced your attitude to a song?				
Before listening	While listening	After listening	Yes	No	Unknown		
(6/77)	(59/77)	(12/77)	(20/77)	(37/77)	(20/77)		

Participants were divided into a positive group (left side) and a negative group (right side) based on their engagement with the comments feature. Those in the positive group regularly browsed comments, and their interest in various types of comments was assessed using a 5-point Likert scale question. The questionnaire also inquired about other usage habits. Conversely, participants in the negative group rarely used the comments feature, and we explored the reasons for their lack of interest in browsing comments

Table 2 The rankings provided by participants indicate the distribution of the influence of song-related information on their listening behavior. Rankings range from 1 (most influential) to 8 (least influential), offering insights into participant preferences and priorities in the context of song selection

Ranking	Information Term								
	1	2	3	4	5	6	7	8	Average Rank
Song' Name / Lyrics	47	26	13	10	1	4	0	1	2.11
Genre / Style	18	32	12	18	7	10	5	0	3.14
Creator	18	23	19	13	13	10	5	1	3.34
Comments	6	5	17	21	25	16	7	5	4.52
Visual Information: Song / Album / Video Cover	2	8	17	10	13	21	24	7	5.13
Playlists	3	1	11	14	20	12	25	16	5.58
Recommendations from Others	6	6	6	10	8	13	15	38	5.91
Release Date	2	1	7	6	15	16	21	34	6.17

various types of music comments. The comment data were subdivided into 13 sub-items (their definitions refer to Appendix A), each of which was assessed using a Likert scale. To determine the positivity of browsing music comments, participants rated a variety of comment types on a scale of 1 to 5, ranging from "don't want to see it at all" to "very much want to see it." The questionnaire also investigated the interaction between commenting and listening to songs. Additionally, we identified the information carried by songs on multiple music platforms and asked users to rank them from highest to lowest based on their level of interest. Finally, the questionnaire investigated the user interface design that would impact the user's listening experience.

3.2 Participants

Our study recruited a total of 104 participants through social media posts and received 102 responses. Of these respondents, 64 were males, 37 were females, and one preferred not to specify their gender. The majority of respondents fell within the 16–28 age group, with 78 participants, while 21 participants fell within the 29–45 age group, and only two were above the age of 45. Educational levels varied, with one participant having a junior high school education, three having a high school education, 65 having an undergraduate degree, 29 having a postgraduate degree, and four holding a Ph.D. The demographic distribution of our sample suggests that our results are valid. Specifically, younger and middle-aged individuals tend to use music streaming services more frequently and seek positive social experiences. To further investigate listeners' challenges and requirements in using music streaming services, we conducted semi-structured interviews with four participants, referred to as P1 to P4. These participants were selected based on their frequent use of music streaming services and varied commenting habits. P1, P2, and P3 were music enthusiasts who were knowledgeable about at least one genre of music and ranged in age from 20 to 23. P4 was an amateur songwriter registered on NetEase Cloud Music and had released about 10 songs in three years.

We conducted the interviews using a teleconferencing tool and audio-recorded them with consent. The interviews lasted between 30 and 50 minutes and were divided into three parts, similar to the questionnaire. We followed an iterative coding process (Hruschka et al. 2004) to analyze the usage habits and needs of streaming music users. Two researchers coded the four interview transcripts separately, using a codebook for text-based questions. We then calculated Cohen's kappa to assess inter-rater reliability, with the average score across all codes being 0.87 (SD=0.10, ranging from 0.74 to 1.00) and an average of 93.7% agreement. After independent coding, the researchers discussed interpretations and discrepancies until they reached a consensus on the codebook and then adjusted their code data.

3.3 Results

The questionnaire findings are presented in Tables 1 and 2. The results revealed a number of challenges that the participants encountered while utilizing the commenting feature on music streaming platforms.

C.1 Users' interests are variable and need-dependent Table 1 depicts the general preference of individuals regarding comments. Participants within the positive group (77 out of 102) exhibited a strong inclination toward neutral content, specifically favoring *Contextual Background* and *Expert Analysis* slightly more than *Shared Feeling*. However, this trend is not absolute, particularly when specific scenarios are simulated, such as loss of love or separation, resulting in significant differences in participants' choices. Furthermore, P4 noted a similar situation, stating that "*as a music creator, I would prefer to see comments that point out problems with my songs than unjustified praise. But when I am blocked from creating and depressed, negative comments tend to make me self-doubt and I would feel that reading some positive comments would be a better choice.*"

C.2 Massive comments pile up, making it difficult for users to get to the content of interest As indicated in Table 1, approximately 25% (25 out of 102) of the survey participants expressed minimal interest in exploring music comments. Within the negative group, nearly 72% (18 out of 25) attributed their dislike to the overwhelming volume of uncategorized content within the comment section. Specifically, the arduous task of finding relevant comments of interest, coupled with the high prevalence of extraneous and superfluous information, which includes fan comments, represents a significant obstacle to the enjoyment of music. In agreement with these sentiments, P1 articulated that the current comment section can be a source of frustration, stating that "*everyone says whatever they want on the Internet,*" leading to futile discussions.

C.3 Existing platforms ignore the impact of comments on users' decisions Based on the findings, users who actively follow comments tend to engage with them while listening to music. Moreover, owing to the platform's interaction logic, users often check comments after entering the playback page. As depicted in Table 1, 25% (20 out of 77) of the positive group expressed that their perception of a song could be influenced by browsing comments. When combined with the ranking results in Table reftab:selection-ranking, it becomes apparent that the positive impact of comments on influencing listening behavior ranks second only to the fundamental information about the song, including the title, genre, lyrics, creator, etc. Participant P3 highlighted the untapped potential for enhancing the existing commenting functionality, stating, "*Existing music platforms seem to have overlooked the power of comments in engaging the listener.*"

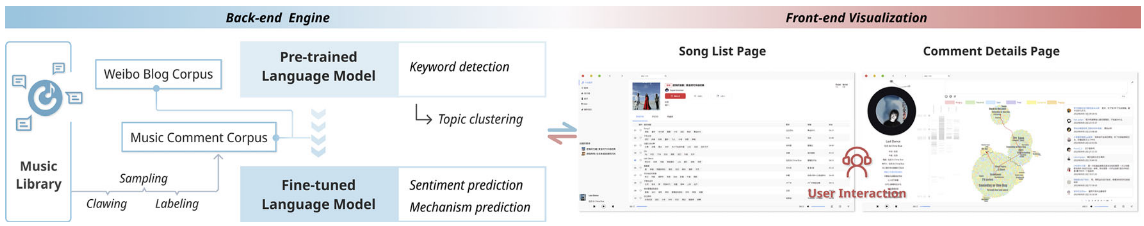

Fig. 1 The pipeline of the architecture consists of a data processing module, a back-end engine, and a front-end visualization interface

3.4 Design goals

Drawing upon the findings we have identified and user expectations, we have formulated the following design objectives.

G.1 Use comments to help users make decisions In our formative study, we discovered that users take into account the comments of others when choosing songs. However, current music platforms provide inadequate support for this aspect of the user experience (*C.3*). To enhance this aspect, we propose the incorporation of a visual preview of the comments. By providing users with the ability to preview comments while browsing the song list, they can gain a general idea of the song's characteristics, which can aid them in selecting songs that align with their interests.

G.2 Provide a browser-friendly way of organizing comments A significant number of users refrain from perusing comments on mainstream music platforms not because they are entirely indifferent to them, but due to the inadequate organization of the comments section. The current approach, which involves presenting users with a high volume of comments simultaneously, is not conducive to the exploration of comment trends (*C.2*). As such, there is an immediate need for a more effective organizational strategy that promotes efficient browsing of comments.

G.3 Help users efficiently find comments of interest Users may have varying preferences for the categories of comments they wish to view, depending on their particular scenario or emotional state. However, accommodating these diverse emotional states can be challenging due to the inherently variable nature of human emotions (*C.1*). In light of this challenge, we propose the development of a system that leverages the sentiment conveyed within comments to facilitate a more seamless identification of comments that are better suited to the user's current emotional state.

G.4 Reduce the cognitive burden on users The questionnaire conducted among participants revealed certain limitations in the functionality of existing music streaming platforms (*C.1*). To provide an improved user experience, it is necessary to incorporate visualization techniques wisely to address these shortcomings. In addition, since the target users are the general public, it is crucial to reduce the cognitive burden that the system imposes on users.

4 System design

4.1 Approach overview

In accordance with our design objectives, we propose a pipeline that facilitates users in effectively exploring music comments and leveraging them to make informed decisions. The system pipeline, as depicted in Fig. 1, comprises a *Back-end Engine* and a *Front-end Visualization Interface*. The Back-end Engine preprocesses the raw music comment data to construct a corpus of comments, which is then utilized to fine-tune the pre-trained language model. Both the pre-trained and fine-tuned language models are capable of extracting sentiment, induced mechanisms, and keywords. These keywords are subsequently employed to perform topic clustering using the LDA topic model. The outputs obtained from these processes are then transmitted to the Front-end Visualization Interface. This interface features two pages that are modeled on the NetEase Cloud Music player: the first is a song list page, which presents eight keywords per song as a visual preview, and the second is a comment details page that furnishes an overview of the comments, with a chronologically segmented structure that enables users to delve deeper into the details as per their requirements.

4.2 Back-end engine

In order to facilitate the organization and exploration of music comments, we employ language models to process the raw comment data obtained through web crawling on the *NetEase Cloud Music* platform. Specifically, the language models are utilized to derive topics from the keywords extracted from each individual comment, as well as to predict the sentiment and induced mechanisms associated with the comments.

4.2.1 Data processing

Our dataset comprises a total of 206, 134, 406 music-related comments, which were collected from 3, 000 distinct songs, resulting in a cumulative data size of 280 GB. The collected data are subjected to a three-step processing methodology. First, the relevant features of each comment message, such as the comment ID, content, and time of posting, are extracted. Second, statistical information pertaining to the comments for each song is obtained, which includes data on the number of comments received during different time periods, listener demographics, and geographic locations. Third, we manually labeled 8000 comments for the purpose of fine-tuning the language model in the prediction task. A summary of the key statistics of our dataset is presented in Table 3.

4.2.2 Contextual keywords and topic detection

In this study, we utilize a self-supervised method (Grootendorst 2020) for the purpose of labeling keywords based on their contextual relevance to the sentence. This is achieved by leveraging the pre-trained Language model and the feature vector of the sentence (Vaswani et al. 2017), which can be used to learn the context of the sentence. The embedding representations of the keywords are obtained, and the cosine similarity between the keyword and the sentence is measured. If the similarity value exceeds a certain threshold, the word is marked as a keyword. In our approach, we select the top 5 words with the highest similarity as the keywords of the sentences. To further enhance the analysis of the collected music comment dataset, we apply the eLDA (Ensemble Latent Dirichlet Allocation) topic model (Brigl 2018). We train a collection of 8 topic models, each of which extracts a cluster of 20 topics. The topic clusters are subsequently grouped by DBSCAN (Ester et al. 1996).

4.2.3 Sentiment and induced mechanism prediction

In this study, we have chosen the pre-trained language model *RoBERTa-wwm-ext-large, Chinese* (Cui et al. 2020) as it demonstrates favorable suitability for Chinese language semantics and exhibits commendable performance across various NLP tasks. Our research encompasses two classification tasks, specifically geared toward predicting the sentiment and induced mechanism of music comments. To accomplish these tasks, we develop two separate language models that undergo fine-tuning. Regrettably, no substantial publicly available dataset of labeled Chinese musical comments, pertaining to sentiment or induced mechanism, exists. Consequently, we manually label the collected data, consisting of approximately 10, 000 music comments, for the sentiment prediction task. Subsequently, the labeled data are partitioned into training, validation, and test sets in a ratio of 6:2:2. These subsets play a crucial role in fine-tuning the pre-trained model. Throughout the fine-tuning process, we initialize the weights of *RoBERTa-wwm-ext-large, Chinese* using the *HuggingFace Transformers* framework (Wolf et al. 2020). Employing an Adam optimizer with a learning rate of 5×10^{-5} and utilizing sparse categorical entropy as the loss function, we fine-tune the model until it attains optimal performance on the validation set. The same fine-tuning and evaluation procedures are applied to other commonly used models, allowing us to compare their performance against our approach based on EM [1] and F1 metrics. A detailed presentation of the results obtained is available in Table 4.

[Par33] EM (Exact Match) calculates the percentage of ground truth answers that match the predicted outcomes.

Table 3 Statistical information of the corpus

Max Length	280
Avg Length	21.41
Vocabulary Size	419,201
Token #	5,819,300,217
Sentence #	29,367,459
Sentence with emoticon #	19,382,522

Table 4 Experimental results in the sentiment prediction task

Models	Train		Test	
	EM	F1	EM	F1
SVM	59.8	61.4	59.2	60.3
BiLSTM	65.2	73.1	69.8	72.4
Chinese-BERT-wwm	66.3	85.6	70.5	87.7
RoBERTa-wwm-ext	67.4	87.2	72.6	89.4
Ours	68.5	88.4	74.2	90.6

In the induced mechanism prediction task, we meticulously label a dataset comprising 8, 000 music comments. Notably, the distribution of labels across the four induced mechanism categories, namely *Music Evaluation*, *Personal Memory and Experience*, *Background Information*, and *Others*, is approximately equal, with a ratio of 1 : 1 : 1 : 1. These categories have been formulated and condensed based on Juslin's framework (Juslin 2013). They carry the following meanings: (1) *Music Evaluation*: Comments in this category focus on aspects such as lyrics, melody, rhythm, performance style, recording mode, and production team (excluding the singer). They mirror the immediate attitudes of listeners, evoke visual imagery, and involve aesthetic judgments regarding the quality of the music clip. (2) *Personal Memories and Experiences*: Within this category, comments are triggered by a song and evoke personal memories or experiences associated with certain events in the listeners' lives. (3) *Contextual Information*: Comments falling into this category go beyond the song itself, encompassing broader social events or human knowledge that may be relevant to a wider audience. They provide contextual information alongside the listener's perspective on the song. (4) *Others*: This category encompasses music comments that are either incomprehensible and philosophical or involve listeners simply copying and pasting lyrics they enjoy. Such comments cannot be effectively classified into any of the aforementioned three categories. Through this labeling process, we establish a comprehensive dataset that accounts for the diverse induced mechanisms present in music comments.

4.3 Front-end visualization

The primary principle of our design is to leverage and enhance familiar visual metaphors. We adhere to the guiding principle of "overview first, zoom and filter, then details-on-demand" (Shneiderman 1996) for users to actively explore music comments, thereby enhancing their music listening experience. In this section, we will present the front-end interface of our system, comprising two distinct pages. Subsequently, we will introduce the visual encodings and the successive stages involved in the reconstruction of comments using a map metaphor-based approach.

4.3.1 Song list page

The Song List Page comprises three primary views, which are illustrated in Fig. 2. First, the Song List Library View presents multiple alternative song lists for users to browse. Second, the Selected Song List View displays comprehensive details about the songs in the selected song list, such as the song title, singer's name, album name, and eight high-frequency keywords extracted from music comments, presented as visual preview tags. Third, the Music Player View enables users to pause, resume, switch the currently playing song, and click the album cover image to navigate to the Comment Details Page. This study aims to investigate the impact of incorporating visual preview tags on users' song selection behavior, and consequently, the entire layout of the interface closely adheres to *NetEase Cloud Music*.

Fig. 2 Song List Page. **A** Song List Library View, **B** Selected Song List View, and **C** Music Player View. Each song in the Song List view has eight keywords hints extracted from music comments

4.3.2 Comment details page

As depicted in Fig. 3, the Comment Details Page comprises two primary views. First, the Lyric View (Fig. 3A), situated on the left, serves as a music player that displays lyrics. Second, the Comment View, located on the right, integrates the comment browsing (Fig. 3B.1–B.3) and comment posting modules (Fig. 3B.4), enabling the user to selectively view the text of the original comment in both abstract and sequential formats. Inspired by the visualization of social media information (Chen et al. 2016, 2017, 2019), we propose a map-metaphor-based comment overview design (Fig. 3B.1) to graphically organize comment texts as countries, counties, and railways (Fig. 4). The features of music comments are mapped onto various visual elements, which we present in a sequence from the entire to the partial view.

Time Period. Song comments often reflect diverse influences. For instance, when the song *Immigrant Song* was featured in the movie *Thor: Ragnarok*, it experienced a surge in comments. Similarly, when singer Ed Sheeran faced allegations of plagiarism, critical comments emerged under many of his songs. The internet's advanced nature allows any social event to thrust a song into the spotlight, leading to a substantial increase in comments within a specific timeframe. To effectively capture this temporal clustering phenomenon, we organize the comment map's structure based on chronological order. For a song, we first determine a time series based on the earliest and latest posting times of all its comments to ensure that every comment falls within this range. Then, using a time segmentation algorithm, we divide this time series into distinct periods by considering the growth rate of comments. Specifically, our approach follows a top-down methodology inspired by Keogh and Pazzani Keogh et al. (2004), recursively segmenting the time series until a predefined stopping condition is met. The algorithm assesses all possible divisions, pinpointing the most suitable positions for splitting. The resulting subsequences are tested to ensure their approximation error falls below a specified threshold. If the error exceeds the threshold, the algorithm continues to recursively segment the subsequences until all segments achieve approximation errors below the threshold.

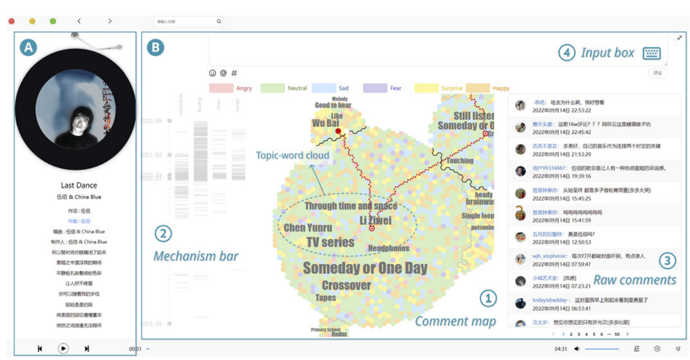

Fig. 3 Comment Details Page. **A** Lyric View, and **B** Comment View. Inside the Comment View, there is **B.1** a comment map in the middle, the topic-word cloud above the comment map summarizes all the topics, **B.2** four induced mechanism bars with a timeline on the left, **B.3** raw comment content on the right, and **B.4** an input box on the top

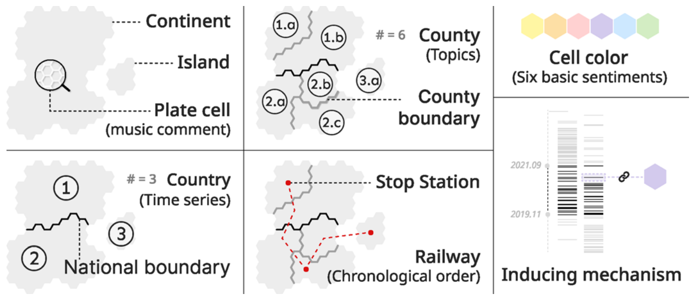

Fig. 4 A legend explaining our map metaphors. **A** Country corresponds to the group of comments in the same period. **B** The center of each country has a stop station marked by a red circle. And the railway crosses all the countries' stop stations. **C** Each topic within a specific time period is associated with a county

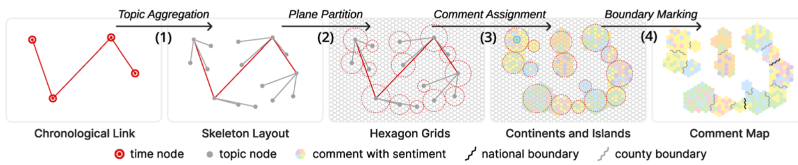

Fig. 5 Comment map construction. (1) Extract topics in different time periods to generate a skeleton layout. (2) Split the skeleton layout into hexagonal grids using Voronoi diagrams. (3) Comments are assigned plate cells to generate a map layout, and (4) complementary boundaries are added to separate different time periods and topic clusters

As depicted in Fig. 4A, we define a "country" as a group of comments within the same time period. Furthermore, in Fig. 4B, we illustrate the center of each "country" denoted by a stop station represented by a double concentric circle, except for the station corresponding to the earliest time period, which is marked with a solid circle. To convey the temporal order of these time periods associated with the countries, we employ a "railway" metaphor represented by a red stroke traversing the centers of all countries. Additionally, if there are bordering relationships between countries, we emphasize the boundaries between them using black strokes.

Topic. Through unsupervised methods, several topics can be derived from comments within a specific time period. In our visualization, each topic is associated with a "county" (Fig. 4C), which serves as a subunit of the country. The layout algorithm determines whether counties should be adjacent or not. If a county shares borders with other counties, we employ dark gray strokes to demarcate their boundaries, allowing users to distinguish between topics. Moreover, word clouds are utilized to represent topics through keywords, aiding users in gaining an intuitive understanding of their content. The size of the keywords reflects their prevalence in the topic; the larger the size, the more frequently it occurs in the comments under that topic.

Sentiment. Each comment is represented as a puzzle piece, and these puzzle cells combine to form the territory of the country. The corresponding puzzle pieces are assigned one of six colors based on the sentiment class labels predicted by our back-end model, following the color-sentiment mapping relationship presented in Hanada (2018). Specifically, the mapping is as follows: *Happy – Orange*, *Angry – Red*, *Sad – Blue*, *Fear – Violet*, *Surprise – Yellow*, and *Neutral – Green*. Additionally, users can obtain the raw contextual information related to a comment by clicking on its cell, enabling them to better understand the sentiment conveyed.

Induced Mechanism. The underlying mechanism that motivates users to write music comments is largely influenced by the semantic content of the comments. Therefore, we leverage this aspect to provide users with insight into a particular topic. The four mechanisms are visually represented as light gray stripes (Fig. 3B.2) that run parallel to the vertical timeline on the left. If a user selects a particular county associated with a topic, the corresponding mechanism is highlighted with dark gray, providing a cue to the user.

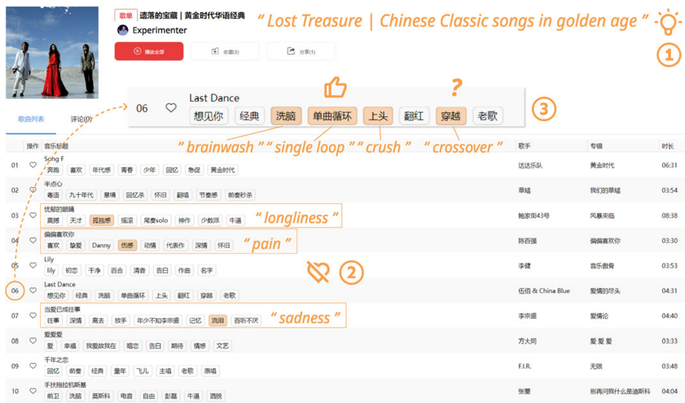

Fig. 6 Usage scenario: Music preview tags bring surprises. (1) The user enters a song list consisting of Chinese classic songs, (2) negative emotional tags make the user not want to listen to certain songs, and (3) some interesting tags attract the user

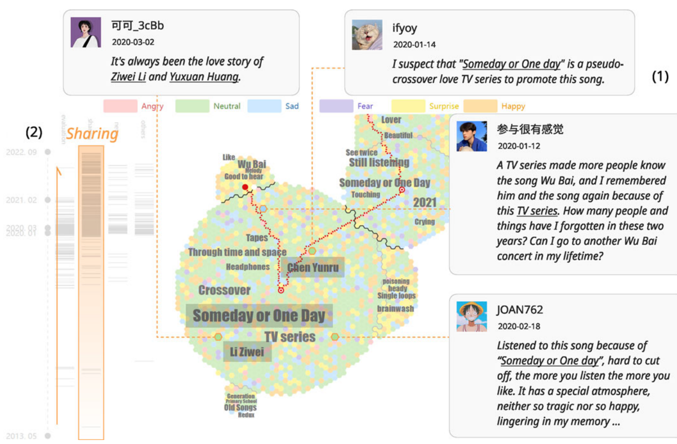

Fig. 7 Usage Scenario: Impact of external events on song comment topics. (1) Comments help the user understand the keywords in word clouds representing topics, and (2) the comments induced by *Sharing* mechanism become more and more common over time

4.3.3 Map construction

The generation of a comment map entails the initial identification of a chronological link comprising a sequence of time nodes. These time nodes are defined as the cut points derived from the application of a time segmentation algorithm to the comments. Once the chronological link has been established, a subsequent four-step process, as depicted in Fig. 5, is required to produce the final comment map.

Topic Aggregation An unsupervised learning approach is employed to extract topics from the comments associated with each time node. Subsequently, each extracted topic is added to the corresponding time node as a child node, referred to as the "topic node," whose radius is determined by the square root of the number of comments it contains. Once all topic nodes have been added, we replace each time node with its child node possessing the largest radius. This yields a topic tree, upon which the bubble tree drawing algorithm (Grivet et al. 2006) is applied to produce the skeleton layout.

Plane Partition After the generation of the skeleton layout, a set of hexagons is overlaid within its bounding box. The Voronoi diagram is used to divide these hexagons. The seed points of the Voronoi diagram are projected as the center points of the hexagons into the bounding box, where the number of seed points, denoted as N_s, is determined as $N_s := \frac{W \times H}{\pi}$, where W and H are the width and height of the bounding box, respectively. This design is intended to make the number of hexagons per unit area approximately equal to the number of comments for the following operations. Alternatively, other types of shapes embedded in the plane, such as squares and triangles, have been considered. However, using squares may cause a distinct jaggedness that leads to unaesthetic effects, while using triangles requires dividing them into

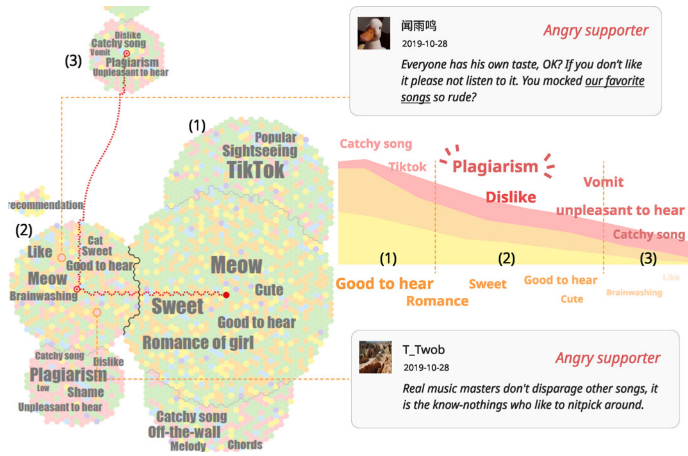

Fig. 8 Usage scenario: Changes in the induced sentiment direction of song comments over time

upper and lower triangles, and encoding the same information using two basic shapes may lead to confusion in understanding. In contrast, the hexagonal layout is a better choice.

Comment Assignment In this stage, we proceed to assign a distinct hexagonal grid to each comment. We adopt a sequential approach, starting from the center point of each topic node and assigning comments in chronological order. A comment will seek out an unassigned hexagonal grid closest to the center point. Once all comments have been assigned, the skeleton layout transforms into a map layout.

Boundary Marking Upon generating the map layout, it is necessary to establish boundaries to delimit the visual metaphors depicted in the map. To this end, we employ black strokes, referred to as "national boundaries," to separate topic plates in adjacent time series. Additionally, gray strokes, or "county boundaries," are employed to separate nearby topic plates in the same time series. The national boundaries and county lines serve to distinguish the different plates visually. In instances where closed plates in different time series or different topics are not adjacent to one another, the national boundaries and county lines are not distinctly marked. The resulting gaps between these boundaries can aid users in distinguishing between them.

5 Usage scenario

We demonstrate the effectiveness of our approach in exploring song comments by presenting a usage scenario based on actual music comments extracted from *NetEase CloudMusic*.

Alice, an oldies fan, selected a song list of 13 Chinese classical songs. As shown in Fig. 6, after observing the tags of several songs on the song list page, she chose "Last Dance" because it had tags such as "brainwash," "crush," and "single loop." Alice found these tags to be indicative of a song that people would want to listen to repeatedly. She also had questions about some tags, such as "crossover" and "someday or one day." She double-clicked on the song and navigated to the comment details page. Alice examined the timeline in the comment view, which is divided into four stages, and then moved to the comment map to get an overview of the comments corresponding to these stages. She quickly identified the earliest comment area and discovered that early comments were about songwriters and melodies, supported by the mechanical bars on the left that are mainly concentrated in "evaluation." By examining the color of the cells in this area, Alice determined that the early comments were overwhelmingly positive.

Subsequently, Alice's attention was captured by the second area with the red "railway," which had the highest number of comments during this period, following her exploration of the first area. As the second time period on the left timeline spanned from January to March 2020, she was surprised to find that a considerable number of comments were generated in only three months, far more than in the previous seven years. This timeframe contained three topics, and she decided to view the largest one first. However, she noticed the presence of incomprehensible keywords (e.g., "Someday or One Day," "Li Ziwei," "Chen Yunru") and clicked on some comments to see the original text. As depicted in Fig. 7(1), some comments helped her grasp the meaning of these words: "Someday or One Day" is a TV series about time travel and love, while "Li Ziwei" and "Chen Yunru" are the characters in the TV series, which features the song. This resolved her previous confusion about the reasons for the surge in comments—a popular TV series drew attention to the song. Alice also observed that the cells during this period exhibited a considerable amount of green, with some blue added, indicating that the sentiment behind the comments was neutral to sad. Upon clicking on these comments, she realized that they were mostly discussing the TV show rather than the song itself. Moreover, during the same period, comments expressing *surprise* and *happiness* were primarily related to the songs rather than external events. Alice then investigated the third and fourth time periods and found that the dominant sentiment remained neutral. Additionally, as depicted in Fig. 7(2), comments triggered by the *Sharing* feature became increasingly prevalent over time. After concluding her analysis, she remarked that "*the song itself can elicit positive sentiments, but external events can influence the sentiment trend, causing the discussion to shift away from the song and towards neutral or sad sentiments.*" Furthermore, she added that "*it is challenging to capture this change in the original comment section of existing music streaming platforms.*"

Alice's curiosity piqued as she ventured to explore another song named "Say Meow Meow," a controversial and popular tune on the internet in recent years, with divergent opinions among its listeners. She intended to employ our system to investigate how the comments on this song have evolved over time. As depicted in Fig. 8, in the first period, the listeners of the first topic generally expressed favorable opinions toward the song, describing it as nice, sweet, and romantic, evoking emotions of *surprise* and *happy*. The second topic pertained to the short video platform *TikTok*, where several listeners stated hearing the song on this platform before coming to the music streaming platform, with the sentiment being *neutral*. However, some *angry* comments expressed a dislike for *TikTok*. The third topic centered on *angry* sentiments, and the keywords of the topic suggested that these comments criticized the song as an off-the-wall catchy tune. Alice observed that the topic parallels for different time periods were the same, signifying the two groups of listeners who liked and disliked the song. The color of the cells in this area indicated that even among those who liked the song, there were also *anger*, which were mainly a response to negative comments. Conversely, the group that disliked the song deemed it a catchy tune but used plagiarism as a potent weapon to criticize it. Most of the negative comments concerning the song were about its alleged similarity to another tune. In the third period, the sentiment trend was overwhelmingly negative where red and green cells dominated, with few positive sentiments, and most negative comments were still related to plagiarism. Alice deduced that the comments on a popular song's sentiment changed over time. After its release, two major groups, namely supporters and opponents, emerged rapidly. Initially, the supporters outnumbered the opponents, with the induced sentiment showing a positive trend. However, over time, the opponents accused the song of being plagiarized, which gradually reversed the listeners' perception of the song, and the induced emotions turned negative.

6 User study

In order to evaluate our proposed augmented music application, we formulated two research questions as follows: ***RQ1: To what extent is the augmented music application usable and effective in an actual listening process?*** and ***RQ2: How do users interact with and respond to this application during real-world listening sessions?*** To address these questions, we conducted a user study under controlled laboratory conditions where participants utilized our enhanced music app and the baseline music application, i.e., *NetEase Cloud Music*.

6.1 Participants

A total of 20 participants (designated as P1–P20) were enlisted for our study. Of these participants, 14 were males and 6 were females, with a mean age of 22.6 years (SD = 1.47). Participants were assigned to one of four groups, with five participants randomly assigned to each group, following the formative study.

6.2 Procedure

We evaluated our improved music application against the baseline by conducting a comparison. To reduce the impact of ordering effects, participants were asked to explore two song lists in a counterbalanced order using two music apps. The song lists were designed to resemble existing song lists on music streaming platforms, with songs in the same list sharing a common theme. For instance, one song list contained classic songs, while the other featured songs that evoke happiness. We ensured that all songs had a minimum of hundreds of comments, but none were excessively popular, to avoid potential biases. Prior to the tasks, we provided participants with instructions on how to use both music apps and allowed them five minutes to familiarize themselves with each app. Participants completed two tasks for each app. The first task was to **identify three songs in a specific list that matched three detailed descriptions**, while the second task was to **identify the key features of comments of the designated songs**, and participants could listen to any music in the list while completing the task. After each exploration round, participants were required to complete a 5-point Likert scale in-task questionnaire to gather feedback on the *usability*, *effectiveness*, and *user experience* of the music app. The questionnaire included options ranging from strongly disagree (1) to strongly agree (5). Finally, we conducted semi-structured interviews with participants, consisting of several open-ended questions.

6.3 Results

Our study employs both quantitative and qualitative methods to analyze the data. The quantitative analysis involves the use of the Friedman test with post hoc Wilcoxon signed rank test, and descriptive statistics are employed to summarize the responses to the in-task questionnaire.

6.3.1 RQ1: How usable and effective is the augmented music application in a real listening process?

We investigated the usability of our proposed music app experienced by participants compared to the baseline app. The Friedman tests showed that there are significant differences in terms of *Helpfulness in music exploration* ($\chi^2(2) = 18.968, p < 0.0001$), *Helpfulness in comment browsing* ($\chi^2(2) = 33.641, p < 0.0001$), *Satisfaction with the visual representation* ($\chi^2(2) = 30.349, p < 0.0001$), *Likelihood for future use* ($\chi^2(2) = 21.076, p < 0.0001$). Apparently, our music app had a better visual representation of comments and helped users enjoy more comments of interest ($\mathrm{AVG} = 4.500, \mathrm{SD} = 0.761; \mathrm{AVG} = 4.400, \mathrm{SD} = 0.821$) than the baseline ($\mathrm{AVG} = 2.200, \mathrm{SD} = 1.005; \mathrm{AVG} = 1.800, \mathrm{SD} = 0.616$). We also investigated how our music apps can enhance the user experience in different aspects. Both apps demonstrate significant differences in terms of *Improving pre-impression* ($\chi^2(2) = 24.479, p < 0.0001$), *Making it easier to have an overview of comments* ($\chi^2(2) = 30.857, p < 0.0001$), *Establishing connections with other listeners* ($\chi^2(2) = 14.862, p = 0.0007 < 0.001$), *Making sentiments of comments easier to understand* ($\chi^2(2) = 36.308, p < 0.0001$), *Clarifying the chronological order of the comments* ($\chi^2(2) = 26.078, p < 0.0001$), and *Enhancing engagement of comments* ($\chi^2(2) = 15.741, p < 0.0001$). Our app was more effective in obtaining an overview of comments, understanding sentiment, and clarifying the temporal order of comments ($\mathrm{AVG} = 4.600, \mathrm{SD} = 0.754; \mathrm{AVG} = 4.400, \mathrm{SD} = 0.940; \mathrm{AVG} = 4.500, \mathrm{SD} = 0.761$) than the baseline ($\mathrm{AVG} = 1.600, \mathrm{SD} = 0.940; \mathrm{AVG} = 1.500, \mathrm{SD} = 0.688; \mathrm{AVG} = 1.750, \mathrm{SD} = 1.118$). These results are consistent with qualitative feedback from participants in the interviews. Our app was considered more usable and helpful than the baseline because our app "*reduces the time wasted on aimless browsing*" (P2), "*increases the exposure of comments of multiple contents*" (P5), "*filters uninteresting information*" (P9), "*provides an attractive visual design for boring text messages*" (P17).

6.3.2 RQ2: How will users interact with and be influenced by this application during real listening sessions?

Users performed fewer "click to listen" operations and complete searching tasks more quickly. Music search can be time-consuming, especially when users are presented with numerous similar songs. Tags can streamline the search process and help users find relevant songs more quickly by allowing them to narrow down their search based on their interests and preferences. This can significantly reduce the number of songs that need to be considered. Our app's tags were reported to have allowed a user (M6) to exclude songs belonging to genres that did not align with their tastes.

Users got more information from the comment map, even though they actually viewed fewer comments. In Task 2, participants had to identify and summarize two main topics from the comments section using the baseline app, but had difficulty due to the large number of comments and had to spend more time manually searching for relevant information. Our app improved efficiency by providing the entire comment timeline and semantically similar groups of comments, resulting in improved accuracy and ease of summarization.

Users felt more positive about their listening and browsing experience with our app. The two music applications were found to significantly differ in terms of users' emotional experiences, including happiness ($\chi^2(2) = 21.283, p = 0.000 < 0.01$), exhaustion ($\chi^2(2) = 13.406, p = 0.009 < 0.01$), hopefulness ($\chi^2(2) = 22.085, p = 0.000 < 0.01$), relaxation ($\chi^2(2) = 18.269, p = 0.001 < 0.01$), and satisfaction ($\chi^2(2) = 23.677, p = 0.000 < 0.01$). Our application elicited higher levels of happiness ($t = -5.586, p = 0.000 < 0.01$), hopefulness ($t = -6.749, p = 0.000 < 0.01$), satisfaction ($t = -6.631, p = 0.000 < 0.01$), and lower levels of exhaustion ($t = 4.196, p = 0.000 < 0.01$). Furthermore, users could easily empathize with other listeners and were more attracted to the comments ($t = -4.700, p = 0.000 < 0.01; t = -4.342, p = 0.000 < 0.01$). However, no significant differences were found in terms of feeling dysphoric or overwhelmed.

The results of the Friedman tests indicated significant differences between the two music apps with regards to users' workload on selecting music ($\chi^2(2) = 26.619, p = 0.000 < 0.01$), workload on understanding comments ($\chi^2(2) = 26.286, p = 0.000 < 0.01$), and attention load ($\chi^2(2) = 9.740, p = 0.045 < 0.05$). Specifically, our music app demonstrated a noteworthy reduction in workload in both selecting music and understanding comments ($\mathrm{AVG} = 2.050, \mathrm{SD} = 0.826; \mathrm{AVG} = 2.100, \mathrm{SD} = 0.968$) when compared to the baseline app ($\mathrm{AVG} = 4.050, \mathrm{SD} = 0.826; \mathrm{AVG} = 4.300, \mathrm{SD} = 0.923$). However, no significant differences in cognitive load were observed.

Categorizing comments logically can reduce the workload of users. Traditional comment organization poses difficulties in identifying and organizing relevant comments on specific topics as commenters may participate in multiple threads, leading to a shortage of comments related to particular topics. Some participants in our study found it challenging to comprehend the significance of comments related to Ziwei Li due to this approach. However, users of our comment map did not experience such confusion.

The attentional and cognitive load of users is negligible. Most participants (18 out of 20) found the cognitive load associated with the keyword cues to be low since users only need to attend to eight words per song, which is faster than listening to song clips. The new visualization also reduces cognitive load. A few participants (P2, P4, P14, and P20) usually browse only the most recent comments, but they showed interest in exploring more due to the appealing nature of the comment map, according to P14.

The additional information forced users to think about what they really wanted, which introduced a new workload. Participants showed interest in the tags and expressed concern for their preferences during exploration. P6 found that tags helped them identify the music they want to listen to, while P8 commented on the effort the tags required. P12 mentioned the workload was heavy but worth it to discover more songs they liked.

7 Discussion and limitation

Comment Specificity in Streaming Music Platforms. Comments on streaming music platforms like *YouTube Music* and *NetEase Cloud Music* focus more on music rather than overall user experience. The informality of comments and lack of text make data mining difficult (Song et al. 2014). We used self-tagged data to adapt a pre-trained language model to achieve some accuracy, but accurate capture of sentiment and induced mechanisms in comment texts still requires training on a larger corpus. Music comments on social media or online forums may be more centered on user experience. This study aims to improve browsing efficiency on streaming music platforms using a map metaphor.

Visual Design and Learning Curve. The design of our system followed a user-centered process. We selected a popular streaming music platform, *NetEase Cloud Music*, as the baseline for our study. To identify the limitations of the platform and the challenges users faced when exploring song comments, we conducted a formative study with 104 participants. During the design process, we prioritized ease of use and aesthetics, employing intuitive visual metaphors such as countries, counties, and railroads to represent comments. Since our system was built on top of an existing music platform, we aimed to maintain interface consistency with the baseline platform as much as possible. Once introduced and explained, users were able to quickly adapt to our visualizations and explore the system with ease.

Generalizability and Scalability. Regarding generalizability, our visual design has potential applications in various scenarios involving the viewing of comment text, such as video comments, social media comments, and product comments. The back-end model is also capable of processing different types of text inputs; however, it requires tuning to better suit the characteristics of the new corpus for improved performance. In terms of scalability, the front-end visualization is not capable of handling vast amounts of comment data. The back-end engine used for detecting sentiments, induced mechanisms, and keywords utilizes a typical machine learning process that can be adapted to other data types with only retraining. However, as our approach uses supervised machine learning, it may face challenges in obtaining sufficient labeled training data when targeting other specific tasks.

Limitations. This work has several limitations. First, although our approach facilitates a seamless transition for users from a comprehensive overview to a more detailed examination of comments within the comment map, the sheer volume of comments can pose challenges in terms of navigation. To mitigate this issue, one possible solution is to employ comment sampling or aggregate similar comments within a single hexagonal grid. Second, it is important to highlight that our approach intentionally omits the visual depiction of intricate connections between comments. This deliberate omission aims to prevent overwhelming visual clutter within the interface. Third, it is crucial to acknowledge that the mechanism prediction model employed in our study is trained using a manually labeled training dataset. However, it is essential to recognize that this training set might be limited in terms of sample size and potential biases, which could potentially impact the generalizability of the model's predictions. Fourth, the evaluation of the visualizations' effectiveness relies on the feedback obtained from participants involved in the user study. While none of the participants expressed significant concerns, we encourage future work to include quantitative experiments to further assess the efficacy of the visualization design.

8 Conclusion

In this research, we aim to enhance the music listening experience using music comments. To achieve this goal, we conducted a need-finding questionnaire and interviews to comprehend the challenges and requirements of music listeners in exploring songs and comments. Subsequently, we integrated two features into an existing music app for music exploration. We evaluated our design through a usage scenario and a user study. Our findings reveal that users demand a more personalized music listening experience, and providing additional comment-related information and well-designed data displays promotes text-based community sharing, communication, and interaction.

A The definition of comment style

There are 13 different kinds of comment in our observational study. Their definitions are as follows.

Contextual Background: This type of comment provides relevant background information related to the discussed topic, offering context and historical perspective to enhance the understanding of the discussion.

Expert Analysis: These comments are written by professionals or experts and aim to provide in-depth assessments and insights regarding specific topics, products, works, or services. They often include professional opinions and ratings.

Shared Emotions: This comment expresses the commenter's emotions or experiences that resonate with the discussed topic or work, emphasizing the emotional connection and shared feelings.

Trending Highlights: These comments highlight the current popularity and trends surrounding a particular topic, product, service, or work, often based on social media or internet trends.

Creative Team Insights: This comment type offers detailed insights into the creative team or authors behind a work, including their background, previous works, artistic style, and other relevant information.

Literary Assessment: These comments pertain to literary works such as novels, poetry, or plays, providing evaluations of the work's structure, themes, language, or style.

Creative Excellence: These comments focus on the creative and artistic aspects of the content, emphasizing its uniqueness and creative qualities.

Latest Updates: These comments provide information about the most recent developments or news regarding a specific topic, product, or event, offering insights into current events.

Personal Experiences: This comment includes the commenter's personal experiences or stories related to the topic or work, using personal narratives to support or explain their viewpoints.

Real-time Commentary: These comments are related to ongoing events, live broadcasts, or on-site activities, offering real-time commentary and viewpoints on current happenings.

Social Media Trends: These comments relate to trends, news, or hot topics on social media platforms, often including commentary and analysis of social media events.

Fan Sentiments: This comment type encompasses opinions and emotional expressions from both fans and critics regarding specific celebrities, works, teams, or products, highlighting the sentiments and reasons for their support or criticism.

Concise Remarks: These comments are brief and to the point, providing a succinct opinion or comment without detailed analysis or descriptions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12650-024-00966-2.

Acknowledgements We would like to express our gratitude to our domain experts and the anonymous reviewers for their insightful comments. This work is funded by grants from the National Natural Science Foundation of China (No. 62372298), the Shanghai Frontiers Science Center of Human-centered Artificial Intelligence (ShangHAI), and the Key Laboratory of Intelligent Perception and Human–Machine Collaboration (ShanghaiTech University), Ministry of Education.

References

Alper B, Yang H, Haber E, Kandogan E (2011) Opinionblocks: visualizing consumer reviews. In: IEEE VisWeek 2011 workshop on interactive visual text analytics for decision making

Barzilay R, Elhadad M (1999) Using lexical chains for text summarization. Adv Autom Text Summ:111–121

Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. Ann Math Stat 41(1):164–171

Beliga S, Meštrović A, Martinčić-Ipšić S (2015) An overview of graph-based keyword extraction methods and approaches. J Inf Organ Sci 39(1):1–20

Bharti SK, Babu KS (2017) Automatic keyword extraction for text summarization: a survey. arXiv:1704.03242

Brigl T (2018) Extracting reliable topics using ensemble latent Dirichlet allocation. PhD thesis, Technische Hochschule Ingolstadt

Byron L, Wattenberg M (2008) Stacked graphs-geometry & aesthetics. IEEE Trans Visual Comput Graphics 14(6):1245–1252

Chen S, Chen S, Lin L, Yuan X, Liang J, Zhang X (2017) E-map: a visual analytics approach for exploring significant event evolutions in social media. In: 2017 IEEE conference on visual analytics science and technology (VAST), pp 36–47. https://doi.org/10.1109/VAST.2017.8585638

Chen S, Chen S, Wang Z, Liang J, Yuan X, Cao N, Wu Y (2016) D-map: visual analysis of ego-centric information diffusion patterns in social media. In: 2016 IEEE conference on visual analytics science and technology (VAST), pp 41–50. https://doi.org/10.1109/VAST.2016.7883510

Chen S, Li S, Chen S, Yuan X (2019) R-map: A map metaphor for visualizing information reposting process in social media. IEEE Trans Visual Comput Graphics 26(1):1204–1214

Chen X (2018) Research on the characteristics and communication mode of interesting virtual community: take net ease cloud music as an example. In: 4th international symposium on social science (ISSS 2018). Atlantis Press, pp 408–412

Chuang Z-J, Wu C-H (2004) Multi-modal emotion recognition from speech and text, pp 45–62

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Cui W, Liu S, Tan L, Shi C, Song Y, Gao Z, Qu H, Tong X (2011) Textflow: towards better understanding of evolving topics in text. IEEE Trans Visual Comput Graphics 17(12):2412–2421

Cui Y, Che W, Liu T, Qin B, Wang S, Hu G (2020) Revisiting pre-trained models for Chinese natural language processing. In: Proceedings of the 2020 conference on empirical methods in natural language processing: findings. Association for Computational Linguistics, pp 657–668

Dörk M, Gruen D, Williamson C, Carpendale S (2010) A visual backchannel for large-scale events. IEEE Trans Visual Comput Graphics 16(6):1129–1138

Eckman P (1972) Universal and cultural differences in facial expression of emotion. In: Nebraska symposium on motivation, vol 19. University of Nebraska Press, pp 207–284

Ester M, Kriegel H.-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the second international conference on knowledge discovery and data mining, KDD'96. AAAI Press, pp 226–231

Graves A, Jaitly N, Mohamed A.-r (2013) Hybrid speech recognition with deep bidirectional lstm. In: 2013 IEEE workshop on automatic speech recognition and understanding. IEEE, pp 273–278

Graves A, Schmidhuber J (2005) Framewise phoneme classification with bidirectional lstm networks. In: Proceedings 2005 IEEE international joint conference on neural networks, 2005, vol 4. IEEE, pp 2047–2052

Grivet S, Auber D, Domenger J.-P, Melançon G (2006) Bubble tree drawing algorithm. In: Computer vision and graphics. Springer, pp 633–641

Grootendorst M (2020) Keybert: minimal keyword extraction with bert. https://doi.org/10.5281/zenodo.4461265

Hanada M (2018) Correspondence analysis of color-emotion associations. Color Res Appl 43(2):224–237

Hasan KS, Ng V (2014) Automatic keyphrase extraction: a survey of the state of the art. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (Volume 1: Long Papers), pp 1262–1273

Havre S, Hetzler E, Whitney P, Nowell L (2002) Themeriver: visualizing thematic changes in large document collections. IEEE Trans Visual Comput Graphics 8(1):9–20
Hruschka DJ, Schwartz D, St. John DC, Picone-Decaro E, Jenkins RA, Carey JW (2004) Reliability in coding open-ended data: lessons learned from hiv behavioral research. Field Methods 16(3):307–331
Hulth A (2003) Improved automatic keyword extraction given more linguistic knowledge. In: Proceedings of the 2003 conference on Empirical methods in natural language processing, pp 216–223
Juslin PN (2013) From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. Phys Life Rev 10(3):235–266. https://doi.org/10.1016/j.plrev.2013.05.008
Keogh E, Chu S, Hart D, Pazzani M (2004) Segmenting time series: a survey and novel approach. In: textitData mining in time series databases. World Scientific, pp 1–21
Kucher K, Paradis C, Kerren A (2018) The state of the art in sentiment visualization. In: Computer graphics forum, vol 37. Wiley Online Library, pp 71–96
Lafferty J, McCallum A, Pereira FC (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data
Leskovec J, Backstrom L, Kleinberg J (2009) Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, pp 497–506
Liu S, Zhou M. X, Pan S, Qian W, Cai W, Lian X (2009) Interactive, topic-based visual text summarization and analysis. In: Proceedings of the 18th ACM conference on information and knowledge management, pp 543–552
Marin MM, Bhattacharya J (2010) Music induced emotions: some current issues and cross-modal comparisons. Music Educ:1–38
Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. arXiv:cs/0205070
Ren Y, Harper FM, Drenner S, Terveen L, Kiesler S, Riedl J, Kraut RE (2012) Building member attachment in online communities: applying theories of group identity and interpersonal bonds. MIS Q:841–864
Scherer KR, Zentner MR (2001) Emotional effects of music: production rules
Shneiderman B (1996) The eyes have it: A task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE symposium on visual languages. IEEE, pp 336–343
Siddiqi S, Sharan A (2015) Keyword and keyphrase extraction techniques: a literature review. Int J Comput Appl 109(2)
Song G, Ye Y, Du X, Huang X, Bie S (2014) Short text classification: a survey. J Multimed 9(5):1–2
Sugiana D, Hafiar H (2018) Construction of self-identity and social identity of“koes plus’’ music fans. MIMBAR: Jurnal Sosial Dan Pembangunan 34(1):176–184
Uzun Y (2005) Keyword extraction using naive bayes. In: Bilkent University, Department of Computer Science, Turkey www. cs. bilkent. edu. tr/ ~ guvenir/courses/CS550/Workshop/Yasin_Uzun. pdf
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. arXiv:1706.03762
Wang H, Fu R (2020) Exploring user experience of music social mode-take netease cloud music as an example. In: International conference on applied human factors and ergonomics. Springer, pp 993–999
Wang Y, Haleem H, Shi C, Wu Y, Zhao X, Fu S, Qu H (2018) Towards easy comparison of local businesses using online reviews. In: Proceedings of computer graphics forum, vol 37. Wiley Online Library, pp 63–74
Wei F, Liu S, Song Y, Pan S, Zhou MX, Qian W, Shi L, Tan L, Zhang Q (2010) Tiara: a visual exploratory text analytic system. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, pp 153–162
Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Davison J, Shleifer S, von Platen P, Ma C, Jernite Y, Plu J, Xu C, Scao TL, Gugger S, Drame M, Lhoest QA (2020) State-of-the-art natural language processing, M. Rush. Huggingface’s transformers
Yatani K, Novati M, Trusty A, Truong KN (2011) Review spotlight: a user interface for summarizing user-generated reviews using adjective-noun word pairs. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 1541–1550
Zeng H, Shu X, Wang Y, Wang Y, Zhang L, Pong T-C, Qu H (2020) Emotioncues: Emotion-oriented visual summarization of classroom videos. IEEE Trans Vis Comput Graphics
Zeng H, Wang X, Wu A, Wang Y, Li Q, Endert A, Qu H (2019) Emoco: visual analysis of emotion coherence in presentation videos. IEEE Trans Visual Comput Graphics 26(1):927–937
Zhang Q, Wang Y, Gong Y, Huang X-J (2016) Keyphrase extraction using deep recurrent neural networks on twitter. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp 836–845
Zhao J, Gou L, Wang F, Zhou M (2014) Pearl: an interactive visual analytic tool for understanding personal emotion style derived from social media. In: Proceedings of 2014 IEEE conference on visual analytics science and technology (VAST). IEEE, pp 203–212