

Article

A Novel Emotion-Aware Hybrid Music Recommendation Method Using Deep Neural Network

Shu Wang ^{1,2,3} , Chonghuan Xu ² , Austin Shijun Ding ⁴ and Zhongyun Tang ^{5,6,*}¹ School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China; eragon@zjgsu.edu.cn² School of Business Administration, Zhejiang Gongshang University, Hangzhou 310018, China; talentxch@zjgsu.edu.cn³ Modern Business Research Center, Zhejiang Gongshang University, Hangzhou 310018, China⁴ Sobey School of Business, Saint Mary's University, Halifax, NS B3H 3C3, Canada; Shijun.Ding@smu.ca⁵ School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China⁶ School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310014, China

* Correspondence: tangzy@zjgsu.edu.cn

Abstract: Emotion-aware music recommendations has gained increasing attention in recent years, as music comes with the ability to regulate human emotions. Exploiting emotional information has the potential to improve recommendation performances. However, conventional studies identified emotion as discrete representations, and could not predict users' emotional states at time points when no user activity data exists, let alone the awareness of the influences posed by social events. In this study, we proposed an emotion-aware music recommendation method using deep neural networks (emoMR). We modeled a representation of music emotion using low-level audio features and music metadata, model the users' emotion states using an artificial emotion generation model with endogenous factors exogenous factors capable of expressing the influences posed by events on emotions. The two models were trained using a designed deep neural network architecture (emoDNN) to predict the music emotions for the music and the music emotion preferences for the users in a continuous form. Based on the models, we proposed a hybrid approach of combining content-based and collaborative filtering for generating emotion-aware music recommendations. Experiment results show that emoMR performs better in the metrics of *Precision*, *Recall*, *F1*, and *HitRate* than the other baseline algorithms. We also tested the performance of emoMR on two major events (the death of Yuan Longping and the Coronavirus Disease 2019 (COVID-19) cases in Zhejiang). Results show that emoMR takes advantage of event information and outperforms other baseline algorithms.

Keywords: emotion-aware; hybrid approach; music recommendation; deep neural network; acoustic clue; circadian rhythm; event-related



Citation: Wang, S.; Xu, C.; Ding, A.S.; Tang, Z. A Novel Emotion-Aware Hybrid Music Recommendation Method Using Deep Neural Network. *Electronics* **2021**, *10*, 1769. <https://doi.org/10.3390/electronics10151769>

Academic Editors: Mehdi Elahi, Amin Beheshti and Mohammad Sina Kiarostami

Received: 23 June 2021

Accepted: 22 July 2021

Published: 24 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development of the modern Internet has witnessed the thriving of personalized services that exploit users' preference data to help them navigate through the enormous amount of heterogeneous content on the Internet. Recommender systems are such services developed to help users filter out useful personalized information [1–3]. In the field of music recommendations, the large number of digital music content provides a huge opportunity for recommenders to suggest music content that meets users' preferences and reduces the search cost for users to find their favorite music [4]. Existing music recommenders leverage the user-item matrix containing both the users and music items traits [5], utilizing acoustic metadata, editorial metadata, cultural metadata [6], and even location-based [7], context-based [8] information to make better recommendations. However, as a kind of emotional stimulus, music has the power to influence human emotion cognition [9].

Thus, music recommendations based on the impact on human emotion has received research interest in both academic and commercial sectors. To make music recommendations emotion-aware, studies focus mainly on the aspects of music emotion recognition [10] and user's emotional or affective preference for music [11].

Neurobiological studies and clinical practices have revealed the human brain has particular substrates to process emotional information in music [12], making recognizing music emotions critical in utilizing complex human emotion information. Several studies have proposed methods to recognize music emotions to help make recommendations. Deng et al. constructed the relational model of the acoustic (or low-level) features of music to the emotional impact of the music. They utilized the users' historical playlist data to generate emotion-related music recommendations [13]. Rachman et al. used music lyrics data to build a psycholinguistic model for music emotion classification [14].

In addition, studies have proved that individuality has significant impacts on music emotion recognition [15]. Thus, from the individual perspective of user preferences, studies have also offered insights into user emotion state representation and emotion preference recognition. Some studies tried to get the emotional state of the users directly from the users' biophysical signals. Hu et al. build the link of music-induced emotion with biophysical signals for emotion-aware music information retrieval [16]. Ayata et al. detected user emotions for music recommendations by obtaining biophysical signals from the users via wearable physiological sensors [17]. However, the biophysical data can be expensive to obtain for a music recommender.

There is strong evidence that the brain mechanisms mediating musical emotion recognition are also dealing with the analysis and evaluation of the emotional content of complex social signals [18]. Modern social media services provide valuable information about user preferences and psychological traits [19]. It is not surprising that social media footprints, such as likes, posts, comments, social events, etc., are being exploited to infer individual traits. Researchers have turned to rich user activity and feedback data on social media [20,21] for music emotion preference modeling. Park et al. proposed an emotion state representation model by extracting social media data to recommend music [22]. Shen et al. proposed a model to represent user emotion state as short-term preference and used data from social media to provide music recommendations based on the model [23]. Other improvements in emotion-aware music recommendations including the context-based method [24], hybrid approaches [25], incorporation of deep learning methods [26], etc.

Though huge progress has been made to emotion-aware music recommendations, there are still problems in existing models and methods. Firstly, existing studies tend to represent music emotion, and user emotion states in a discrete manner, such as the six Ekman basic emotions used by Park et al. [22] and Polignano et al. [27]. However, studies have proved that aside from the six basic emotions, there are more categories of emotions [28], and it is better to model emotion in a continuous way [29]. Secondly, user emotion state extracted from social media data can only represent user emotion state or the music emotion preference of the user at some certain time points, as stated by Shen et al. [23] emotion accounts for a short-term preference that drifts across time [30]. While the recommender should generate recommendations whenever the user enters the system, there is still a need for continuous emotion state recognition across time. Thirdly, it has been proven that events, especially major social events, regulate human emotions [31] while existing studies fail to utilize the information of events to improve emotion state representation.

In this paper, we proposed an emotion-aware music recommendation method (emoMR) that utilized a deep neural network. Representing music emotions continuously, generating user emotion states when no user activity data presents, and capturing event-related information to refine the emotion states are the three main issues this paper deals with. We proposed a music emotion representation model by exploiting the low-level features of the music audio signals and the music metadata to predict the music emotion representations in a continuous form based on the emotion valence-arousal model [32]. We also proposed

an emotion state representation model based on the artificial emotion generation model. Event-related information and implicit user feedback from social media were used as the exogenous factors of the model. The endogenous factors of the model were constructed using the human circadian rhythm model. The two proposed models were trained via a deep neural network architecture (emoDNN) and can generate accurate predictions. We took a hybrid approach that combines the advantage of both a content-based approach and collaborative filtering approach to generate recommendations. The results of the experiments show that emoMR outperforms the other baseline algorithms in selected metrics. Although the length of the recommendation list reduces the performance of the compared methods, emoMR still has an advantage over other emotion-aware music recommendation methods. The results also suggest that our method is able to use event-related information to improve the quality of emotion-aware music recommendations.

The main contributions are summarized as follows:

1. The proposed models represent music emotions and users' music emotion preferences under certain emotion states in a continuous form of valence-arousal scores allowing for complex emotion expressions in music recommendations;
2. Trained models with a deep neural network (emoDNN) that enable the rapid processing of both music emotions and user emotion states even on time points when no user activity data exists;
3. Incorporated event-related information to allow the hybrid approach to generate music recommendations considering the influences posed by events on emotions.

The rest of this paper is arranged as follows: Section 2 overviews the related works. Section 3 explains that the recommendation models consists of the music emotion representation model, the emotion state representation model, and the deep neural network framework. Section 4 demonstrates the experiments and results of comparing emoMR with baseline algorithms. Section 5 concludes the whole paper.

2. Related Works

Compared to popular recommendation models currently used in music recommendations, emotion-aware music recommenders offer higher quality recommendations by introducing an additional preference of emotions since music itself has the ability to regulate human emotions [33]. The striking ability of music to induce human emotions has invited researchers to retrieve emotional information from the music and exploit emotion-related data from users for music recommendations. Novel approaches have been forged to utilize emotion-related data, such as exploiting user interactions to make emotion-aware music recommendations [34]. In this section, we discuss several works on improving music recommendations by exploiting emotion-related information.

The first factor of exploiting emotional information in music recommendations is the representation of emotions. The basic idea behind music emotion representation is using acoustical clues to predict music emotions. Acoustical clues or low-level audio features of music can be used to predict human feelings [35]. Barthelet et al. proposed an automated music emotion recognition model using the acoustical music traits of both frequencies and temporal to aid context-based music recommendations [36]. Several other models such as the emotion triggering low-level feature model [37], musical texture and expressivity features [38], and acousticvisual emotion Gaussians model [39] have been proposed for music recommendations with emotion information. Other studies leveraged the fruit of music information retrieval [40] and combine low-level audio features with music metadata such as lyrics [41] and genre [42] to represent music emotions. While acoustical clues are widely used in music emotion recognition, new types of music have been invented every year. Combining both the low-level audio features and the descriptive metadata is more promising in music emotion representation.

The second factor of exploiting emotional information in music recommendations is the representation of the user's emotion state. Basically, there are two kinds of approaches to representing the user's emotion state. One is to get physiological emotion

metrics directly from the users, and the other is predicting the user emotion state by the activity data generated by the user. Although recommenders or personalized systems can directly ask their users for their current emotion state, it is hard for a user to describe the emotional status. Therefore, studies have used biophysical signals or metrics (such as electromyography and skin conductance [43]) to recognize user emotions. Some studies exploit the technologies used in cognitive neuroscience to represent user emotion states through electroencephalography (EEG) [44,45] in music-related activities. Ayata et al. even built an emotion-based music recommendation system by tracking user emotion states with wearable physiological sensors [17]. The problem with the direct approach is that it is expensive to obtain the biophysical metrics data from the users, and existing biophysical signal acquisition equipment might interfere with user experience. The other approach, however, exploits the user activity data and involves no expensive data acquisition sessions, which is regarded as more promising. Activities like the operations and interactions on music systems can be used to represent user emotion states [46,47]. Activities on social media are a fruitful source of user emotion state representations. Rosa et al. used the lexicon-based sentiment metric with a correction factor based on the user profile to build an enhanced sentiment metric to represent user emotion states in music recommendation systems [48]. Deng et al. modeled user emotions with user activity data crawled from the Chinese Twitter service, Weibo in music recommendations [49]. A recent study used implicit user feedback data from social networks to build an emotion-aware music recommender based on hybrid information fusion [50]. Another study used an affective coherence model to build an emotion-aware music recommender, utilizing data from social media to compute an affective coherence score and predict the user's emotion state [27]. Among the studies that utilized activity data from social media, sentiment analysis is the dominant technology used to extract and build the emotion state representations.

The third factor of exploiting emotion information in music recommendations is the recommendation model used for recommendations. As a special kind of recommendation item, the models used for emotion-aware music recommendations vary according to the specific research scenarios. Lu et al. used a content-based model to predict music emotion tags in music recommendations [51]. Deng et al. used a collaborative filtering model in music recommendations after representing the user emotion state through data from Weibo [49]. Kim et al. employed a tag-based recommendation model to recommend music after semantically generating emotion tags [52]. Han et al. employed a context-based model after classifying music emotion in music recommendations [53]. Besides these models, hybrid models that combine the advantages of multiple recommendation models have also been applied in emotion-aware music recommendations, especially those with the help of deep learning [54]. New information processing technologies such as big data analysis technologies [55] and machine learning have accelerated music recommendations in many ways [56]. Therefore, compared to the other models, hybrid approaches can come up with better performances. Deep learning technology helps improve not only the recommendation models [57], but also the emotion recognition (representation) [58].

To sum up, although conventional studies exploit both the acoustical clues and meta-data of music, few have investigated the approach of representing music emotions in a continuous manner. Existing studies have not proposed a way to represent users' emotion states at time points when no user activity data exists. Conventional studies also tend to ignore the influences posed by events on human emotions. All of these ask for further investigation.

3. Model Construction

Recommending music regarding the emotional preference of the users calls for the need to exploit emotional information, which entails identifying the emotional characters of the music and recognizing the emotional preference of the user under a certain emotional state. In this section, we first proposed a music emotion representation model, then we proposed an emotion state representation model. Based on the two models we describe,

the hybrid approach took by our recommendation process. The general architecture of the proposed recommendation method (emoMR) is illustrated in Figure 1.

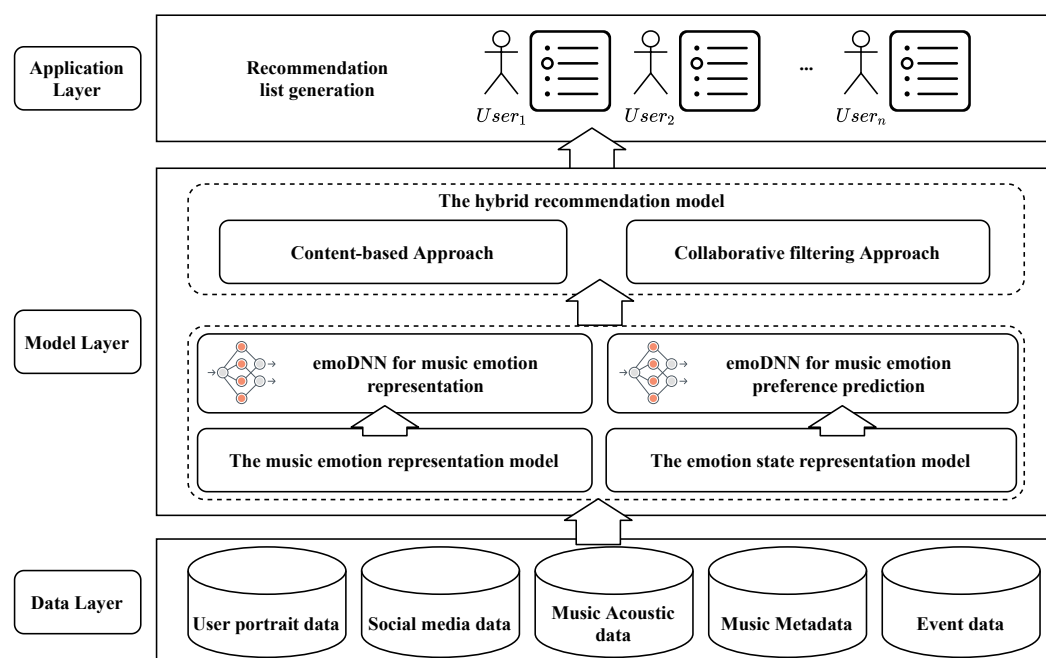


Figure 1. The general architecture of emoMR.

The general architecture is composed of 3 layers: The data layer, the model layer, and the application layer from bottom up. The data layer provides data form the model layer and the application layer generates the recommendation lists upon the model layer. The details for each of the layers are described as:

1. The data layer. This layer offers 5 kinds of data for further processing. The user portrait data contains information inherited from conventional recommendation systems, such as the user profile from the registration system. The social media data contains the users' activity data from social media, which is used as implicit user feedback in our recommendation method. The music acoustic data contains the audio signal data of the music. The music metadata contains descriptive data of the music, such as the genre, artist, lyrics, etc. The event data contains public opinion data on certain events.
2. The model layer. In the model layer, the music emotion representation model and the emotion state representation model generate the music emotion representation and emotion state representation using data from the data layer. The models are trained to predict music emotions for the music and the music emotion preferences for users with a deep neural network (emoDNN). The hybrid recommendation model combining the content-based and collaborative filtering recommendation approaches uses the data generated by the trained models to make recommendations.
3. The application layer. In this layer, the proposed method generates music recommendation lists for users.

As emotional stimulus, it can be pretty complex to classify the music into a certain emotion kind [42]. Things like one piece of music that sound like *Joy* to one person but *Sad* to another, which happens ubiquitously. Perhaps a horrible event had just happened to the latter. The human cognitive process of emotion is typically complex. A person may be in several emotional states simultaneously under a certain situation [59]. For example, one might be in a complex emotional state of both *Joy* and *Fear* while playing horror video games. To express the music emotion characters and human emotion states quantitatively, we model the emotion using the valance-arousal model [32] which quantifies emotions by expressing emotions as points in a two-dimensional plane. The model is illustrated in

Figure 2. The plane's horizontal axis is depicted by valence ranging from unpleasant to pleasant, and its vertical axis is depicted by arousal, which indicates the activation or energy level of the emotion. Therefore, this paper proposes a music emotion representation model and an emotion state representation model based on the emotion valence-arousal model.

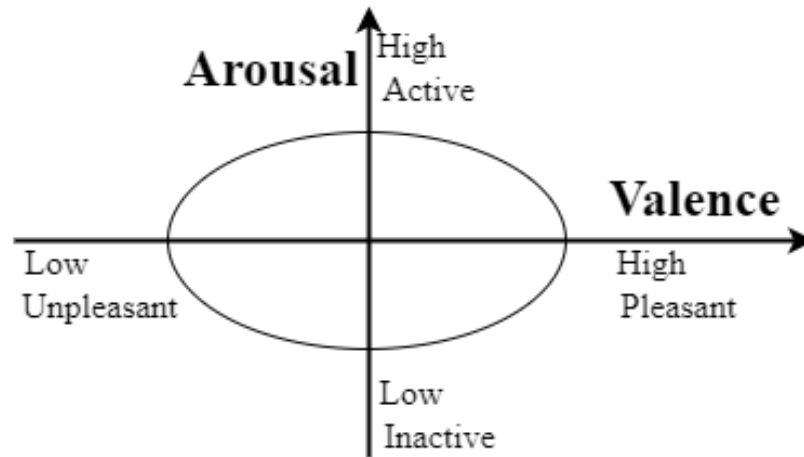


Figure 2. The emotion valence-arousal model.

3.1. The Music Emotion Representation Model

The emotion of a piece of music is the emotional response of the person who listens to it. However, the actual emotional response is hard to obtain. Thus, our music emotion representation model takes the approach of generating quantitative representations of the music emotion by utilizing the features of the music to predict the emotional response. Aside from other studies which tend to make classifications of music emotion, we represent it as a tuple $E_i = \langle s_{valence}, s_{arousal} \rangle$ in which $s_{valence}$ indicates the emotion valence score and $s_{arousal}$ indicates the emotion arousal score, E_i indicates the emotion of music i .

For a piece of music, there are 2 sets of features that are exploited:

1. Low-level audio features, which extracted from the audio data;
2. Music descriptive metadata, such as genre, year, artist, lyrics, etc.

3.1.1. Low-Level Audio Features Extraction

Studies have shown that the low-level audio features of pitch, Zero-Crossing Rate (ZCR), Log-Energy (LE), Teager Energy Operator (TEO), and Mel-Frequency Cepstral Coefficients (MFCC) can determine the emotional state of music audio signals [60–63]. The extraction methods for these features are illustrated as follows:

1. Pitch: Pitch extraction calculates the distances between the peaks of a given segment of the music audio signal. Let Sig_i denote the audio segment, k denotes the pitch period of a peak, and Len_i denotes the window length of the segment, and the pitch feature can be obtained using Equation (1):

$$Pitch(k) = \sum_{i=0}^{Len_i - k - 1} Sig_i Sig_{i+k}. \quad (1)$$

2. ZCR: The Zero-Crossing Rate describes the rate of sign changes of the signal during a signal frame. It counts the times the signal changes across positive and negative values. The definition of ZCR is shown in Equation (2),

$$ZCR(i) = \frac{1}{2Len_w} \sum_{n=1}^{Len_w} Len_w |sf(Sig_i(n) - sf(Sig_i(n-1)))| \quad (2)$$

where Len_w is the length of the signal windows and $sf(\cdot)$ is the sign function, and is defined in Equation (3):

$$sf(Sig_i(n)) = \begin{cases} 1, & Sig_i(n) \geq 0, \\ -1, & Sig_i(n) < 0. \end{cases} \quad (3)$$

3. LE: This feature estimates the energy of the amplitude of the audio signal. The calculation can be formulated as Equation (4):

$$LE(Sig_i) = \log_{10} \left(\sum_{i=0}^{Len_w} Sig_i^2 \right). \quad (4)$$

4. TEO: This feature links to the energy of the audio signal as well, but from a nonlinear perspective. The TEO of a signal segment can be calculated using Equation (5):

$$TEO(Sig_i) = Sig_i^2 - Sig_{i+1}Sig_{i-1}. \quad (5)$$

5. MFCC: The MFCC is derived from a mel-scale frequency filter-bank. The calculation of MFCC can be obtained by first segmenting audio signals into Len_w frames and then applying a Hamming Window (HW) defined by Equation (6) to each frame:

$$HW(n) = 0.54 - 0.46 \cos(2\pi \frac{n}{Len_s}), 0 \leq n \leq Len_s. \quad (6)$$

Len_s is the number of samples in a given frame. For an input signal Sig_i , the output signal O_i is defined as Equation (7):

$$O_i = Sig_i \times HW(i). \quad (7)$$

The output signal is then converted into the frequency domain by Fast Fourier Transform (FFT). A weighted sum of spectral filter components is then calculated with triangular filters. The Mel spectrum is then obtained by Equation (8):

$$Mel(i) = 2595 \times \log_{10} \left(1 + \frac{i}{700} \right). \quad (8)$$

The features reveal emotions due to their connections to the polarity and energy of the emotions. Pitch and ZCR show great discrimination across pleasant and unpleasant emotions, Log-Energy and TEO measure the energy of different emotions, and MFCC compacts signal energy into its coefficients. The features describe the music emotion from both the emotional valence and arousal perspective.

3.1.2. Music Metadata Exploitation

Aside from the low-level audio features, the music also has some important metadata that suggests additional information about the music. Unfortunately, in the common practice of the music industry, there is no unified metadata structure standard. Therefore, we refined the metadata structure in the work [64] and applied it to the music emotion representation model. We eliminate the notation, attention-metadata, and usage from the structure, because they are within the user domain and hard to obtain when training the model. The metadata of rights and ownership, publishing, production, record-info, carrier, and website are also eliminated due to their irrelevance to the music emotion. Before the training process, 6 classes of metadata are kept in our structure. They are listed in Table 1.

Table 1. Music metadata structure.

Class	Name	Description
M_1	Musical information	Musical properties of the audio signal, e.g. duration, key
M_2	Performance	Descriptions of the people that are involved in a musical performance, e.g. artists
M_3	Descriptor	Describing the musical work, e.g. title
M_4	Playback rendition	Information useful for rendering of the music file, e.g. relative volume
M_5	Lyrics	Text of the musical work and related information, e.g. translated lyrics text
M_6	Instrumentation & arrangement	Information about the used instruments and orchestration, e.g. instrument type

The metadata is then represented by $Meta = \{M_i | i = 1, 2, \dots, 6\}$. The labels and categorical metadata are converted into numerical representations. Unlike the other metadata, the lyrics (M_5) contain blocks of long texts instead of labels. To leverage the texts in the lyrics and represent (M_5) as labels, the lyrics texts are processed by standard Natural Language Processing (NLP) technologies, such as stop-words removal, tokenization, and Term Frequency-Inverse Document Frequency (TF-IDF) score generation. The emotion of the lyrics can then be obtained by sentiment analysis. Described in the work [27], sentiment analysis generates scores for 6 Ekman basic emotions $EE = \{Joy, Anger, Sadness, Surprise, Fear, Disgust\}$. Let $S(\cdot)$ denote the score for a particular basic emotion, then $S_t = \{S_t(emo) | emo \in EE\}$ is the emotion vector of item t . To fit the emotion vector of the lyrics (the item is lyrics) into our valence-arousal based model that is $M_5 = \langle Valence, Arousal \rangle$, and is converted as follows:

1. Normalization: The min-max normalization is used to convert the 6 elements in S_T to values within the range of $[0, 1]$. Let $\tilde{S}_i(emo)$ denote the emo score of the i th lyrics after normalization, and can be described by Equation (9):

$$\tilde{S}_i(emo) = \frac{s_i(emo) - \min_{1 \leq j \leq n} (s_j(emo))}{\max_{1 \leq j \leq n} (s_j(emo)) - \min_{1 \leq j \leq n} (s_j(emo))}, emo \in EE. \quad (9)$$

2. Calculate valence and arousal: The emotions of $E_{po} = \{Joy, Anger, Surprise\}$ are treated as positive while $E_{ne} = \{Sadness, Fear, Disgust\}$ are treated as negative. The valence and arousal values of s_i can be obtained by Equation (10):

$$\begin{cases} Valence(s_i) = \frac{1}{3} \cdot \left(\sum_{e_{po} \in E_{po}} \tilde{S}_i(e_{po}) - \sum_{e_{ne} \in E_{ne}} \tilde{S}_i(e_{ne}) \right), \\ Arousal(s_i) = \frac{1}{3} \cdot \sum_{emo \in EE} \tilde{S}_i(emo) - 1. \end{cases} \quad (10)$$

The music emotion representation model predicts the emotional response, which is a $\langle Valence, Arousal \rangle$ vector using the music feature vector. The music feature vector is composed of the low-level audio features of the music and the metadata of the music. The vector V_{music} is shown as:

$$V_{music} = \langle Pitch, ZCR, LE, TEO, MFCC, M_1, M_2, \dots, M_6 \rangle.$$

Posts and comments of a piece of music reflect the affection and emotional response of users to the music [50]. Label data can be obtained by exploiting the implicit user feedback from social network posts and comments using text sentiment analysis and obtain the label data of $\langle Valence, Arousal \rangle$ vectors using the conversion method described above to train the model.

3.1.3. The Deep Neural Network for the Music Emotion Representation Model

The works by Tallapally et al. [65] and Zarzour et al. [66] provided an approach to use a deep neural network to predict user ratings, scores, and preferences. The main idea of building the music emotion representation model with a deep neural network is using the music feature vector V_{music} to predict the emotional response it brings about. The response which indicates the user's emotion-related preference can be quantified by the $\langle Valence, Arousal \rangle$ vector. We proposed an approach of emoDNN to predict the vector. The general architecture is illustrated by Figure 3.

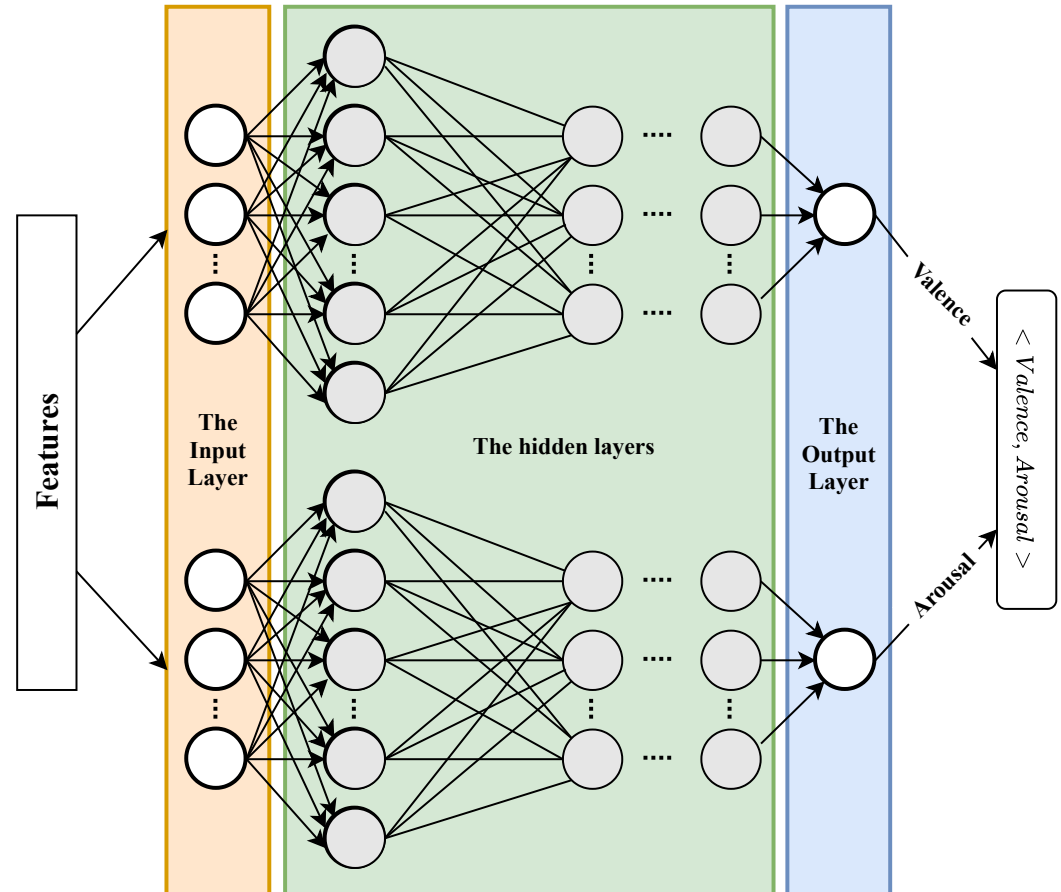


Figure 3. The general architecture of emoDNN.

The architecture used two deep neural networks to predict the *Valence* and *Arousal* respectively. Each network consists of an input layer, hidden layers, and an output layer. The input layer consists of the vector V_{music} within which the features are converted to numerical representations as mentioned above. The hidden layers can be customized to investigate the feature-emotion interactions. The output layer is the predicted *Valence* and *Arousal*. The total quantity of nodes in the hidden layers is set by Equation (11) to ensure the accuracy and training speed of emoDNN [67]:

$$N_{hidden} = \sqrt{N_{input} + N_{output}} + \tau. \quad (11)$$

N_{hidden} is the number of nodes in the hidden layers, N_{input} and N_{output} indicate the nodes in the input layer and output layer respectively, and τ is a constant number between [1, 10]. The loss function is set to Mean Squared Error (MSE) shown in Equation (12):

$$LOSS(X, Y) = \frac{1}{n} \sum_{i=1}^n (E(x_i) - y_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - wx_i - b)^2. \quad (12)$$

$E(x_i)$ is the i th actual output while y_i is the i th expected output. w is the weight and b is the bias. To handle over-fitting and gradient disappearance, the activation function is set to Rectified Linear Unit (ReLU) as shown in Equation (13):

$$\text{ReLU}(z) = \max(0, z) = \begin{cases} z, & z \geq 0 \\ 0, & z < 0. \end{cases} \quad (13)$$

3.2. The Emotion State Representation Model

The emotion state representation model aims to identify the music emotion preference of the user in a certain emotion state at arbitrary time points. Recent emotion-aware recommendation studies use discrete user-generated content from a social network to identify user emotion [26] or transform emotion to context-aware [8]. These studies fail to represent a consecutive user emotion state, thus, unable to represent the user emotion state at an arbitrary time point. For example, a user only interacts with the system and social media in the morning, traditional methods can not represent the emotion state of the user at 15:00, because no user feedback data can be utilized at 15:00.

To represent the emotion state of the users at arbitrary time points, an emotion generation model was built based on the cognitive theory of emotion [68] and the cognitive appraisal theory of emotion [69]. The theories suggest that music stimulates the human cognitive process and evokes emotion which is a cognitive process that consists of several spontaneous components that happen simultaneously. The model is used to generate emotions in a consecutive manner [70] artificially. We utilized the model to generate the emotion state of the user at arbitrary time points. A simplified version of the model is depicted by Figure 4. We defined the 4 modules in the model according to our need to generate the emotion state at arbitrary time points:

1. The environment module: It utilizes the information that affects user emotion in the environment. The information creates a context that can be identified from social networks and public opinion. This module influences the emotion evaluation of the user and contributes to the exogenous factors of the emotion state;
2. The experience module is controlled by the characters of the user and contributes to the endogenous factors of the emotion state;
3. The evaluation module synthesis the output of the exogenous and endogenous factors and generate the emotion state at arbitrary time points by evaluation;
4. The emotion module identifies the music emotion preference user the emotion state generated by the evaluation module.

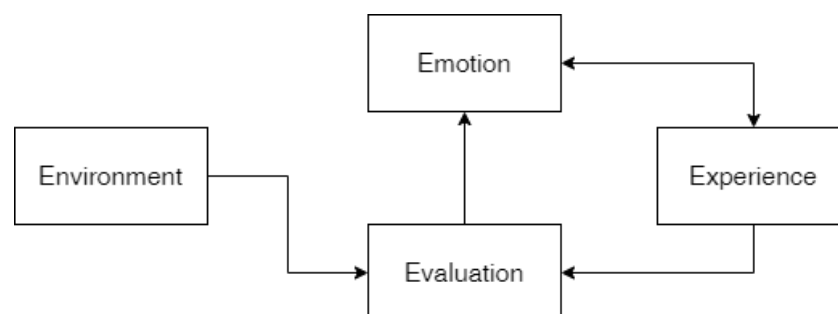


Figure 4. The artificial emotion generation model.

Therefore, building the emotion state representation model can be divided into four tasks: (1) Model the exogenous factors of the emotion state, (2) model the endogenous factors of the emotion state, (3) represent the emotion state, and (4) identify the music emotion preference.

3.2.1. The Exogenous Factors of the Emotion State

Information from the environment that affects the user emotion has long been investigated by studies that raise context as an influential factor [8]. The information acts as a kind of social stimuli and influences user emotion. The influence has been proven by cognitive attention studies [71] and neuroscience [72]. Among the many types of data, implicit user feedback information such as rating, review, and clicks were introduced to address the influences of information from the environment on user emotion [50]. Based on the social network data used by Polignano et al. [27], we propose three types of data as the exogenous factors of the emotion state:

1. Events: Events influence user emotions by raising public opinions. Thus, events can be identified by analyzing the sentiments of the raised public opinions;
2. Posts: Users react to certain kinds of emotions on social networks by posting messages. Sentiment analysis on posts within a certain time range can utilize posts to describe user emotion;
3. Comments: Comments within a certain time range reflect user emotional responses. The sentiments of these comments help distinguish user emotion states.

Let T_{estat} be the time point for estimating the user emotion state. Given TH^{range} as the time range. The events, posts, and comments from the social network (e.g., Weibo, Facebook, Twitter) are collected and processed by sentiment analysis and converted into valence-arousal vectors. Especially, the public opinions for the events can be collected by trends from the social networks (e.g., Weibo top trends). The process has been described in Section 3.1.2. The exogenous factors at T_{estat} ($EXF_{T_{estat}}$) can be described as:

$$EXF_{T_{estat}} = \{sen_{type}\}, type \in \{events, posts, comments\}$$

where $sen_{type} = \langle Valence, Arousal \rangle$ is a two-element vector, consists of the valence and arousal scores of the texts for the $types$. Since trends on social networks tend to have a longer duration than individual social network activities. The exogenous factors use the event information neglected in other studies can help decide the user emotion state where traditional social stimuli such as the implicit user feedback of user rating, review, and clicks fail to present.

3.2.2. The Endogenous Factors of the Emotion State

Endogenous factors relate to the inherent mechanism of human emotions. Studies have shown that emotions are regulated by the bio-physical or bio-chemical mechanisms of the human body [73,74]. The performances of these mechanisms vary over time and can be described as the circadian rhythm [75]. The circadian rhythm acts as an internal timing system that has the power of regulating human emotion [76]. Much of human beings' physiological indicators vary according to the circadian rhythm, such as the hormones, blood sugar, body temperature, etc. [77]. Thus, it causes oscillation in human emotion. Figure 5 shows the mRNA and protein abundance oscillation across a certain period of circadian time (in hours) obtained by the study of Olde Scheper et al. [78].

The circadian rhythm as an endogenous cycle of the human biological process that recurs in approximately 24-hr intervals can be treated as a smooth rhythm with added noise. It can be modeled with data from known periods (24 hr each). A study proposed a cosine fit of the circadian rhythm curve [79]. The cosine model is shown in Equation (14).

$$CR_i = MESOR + A \cdot \cos\left(\frac{2\pi t_i}{\lambda} + \varphi\right) + \sigma_i \quad (14)$$

where CR_i is the value of the circadian rhythm model at time i , $MESOR$ is the Midline Estimating Statistic of Rhythm [80], A is the amplitude of the oscillation, t_i is the time point (sampling time), λ is the period, φ is the acrophase, and σ_i is an error term.

We use the cosine model to represent the endogenous factors of the user emotion state. The parameters can be obtained by utilizing the users' extracted emotional information

data proposed by Qian et al. [50] across a 24-hr period. Thus, the endogenous factors of the emotion state can be modeled as:

$$ENF_{T_{estat}} = \{CR_i\}, i = T_{estat}.$$

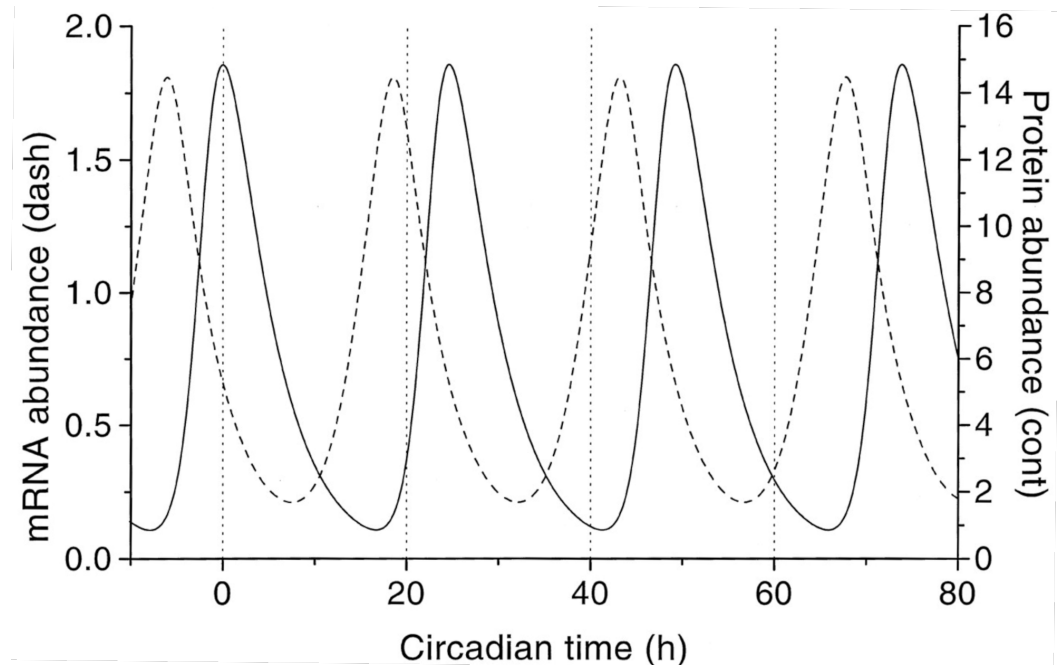


Figure 5. The mRNA and protein abundance oscillation across a certain period of circadian time.

3.2.3. The Emotion State Representation

In the above artificial emotion generation model, the evaluation module is used to represent the emotion state by synthesizing the exogenous and endogenous factors of the emotion state. However, user properties such as age, gender influence not only users' perception and appraisal of exogenous factors [81] but also the endogenous factors of users [82]. Thus, to represent the emotion state of a user at a certain time, we combine the exogenous and endogenous factors together with the user properties UP_{user} that are used in conventional music recommendation studies [83]. The emotion state representation is shown as:

$$V_{estat} = \langle EXF_{T_{estat}}, ENF_{T_{estat}}, UP_{user} \rangle.$$

3.2.4. The Music Emotion Preference Identification Using Deep Neural Network

Identifying the music emotion preference of a user at a certain time is the main purpose for constructing the emotion state representation model. We use the same deep neural network structure (emoDNN) in Section 3.1.3 to learn the model. Section 3.1 has proposed a music emotion representation model. The emotion of the music can be represented by a valence-arousal vector. The label values for training the emotion state representation model is obtained by utilizing the emotion (valence-arousal vector generated by the music emotion representation model) of the music liked (post, repost, like, positive comment) by the user at T_{estat} .

The nodes in the hidden layers are tuned to fit the situation for the input vector V_{estat} . The loss function and activation function are the same as those for the music emotion representation model. However, as the endogenous factors introduce a cosine function, the model cannot be treated as a linear one. Thus, we update the model and add a higher-order polynomial to fit the model. The model is shown in Equation (15):

$$Z(x) = \sum_{i=0}^m x_i^i w_i + \sum_j^n x_j w_j + b. \quad (15)$$

The high order polynomial uses the parameters of the cosine fit model of the endogenous factors to bring the oscillations of the circadian rhythm of the user to the model.

3.3. The Recommendation Process

The emotion-aware music recommendation method proposed in this paper (emoMR) consists of the following steps in the recommendation process: Music emotion representation generation, user preference for the music emotion calculation, similarity calculation, and recommendation list generation. The method is a hybrid recommendation method that takes both a content-based method and collaborative filtering method into the recommendation process. The process is depicted in Figure 6.

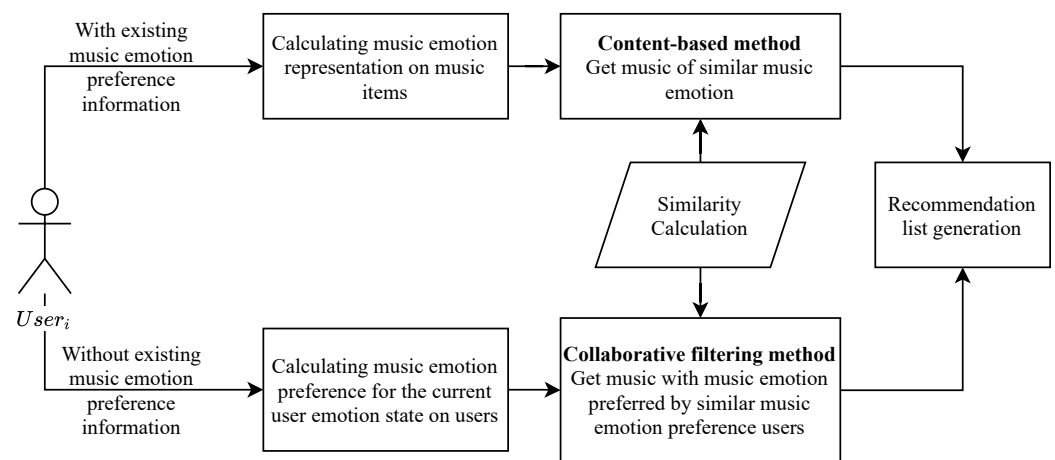


Figure 6. The recommendation process of emoMR.

When a user $User_i$ enters the recommendation process at time T_{estat} , the method first decides whether there the music emotion preference of the user exists in the system. The process can then be described as:

1. If the user's music emotion preference information exists in the system, the music emotion representation is calculated using the music emotion representation model. The emoDNN for the music emotion representation model can be trained, and its parameters can be saved for future use. The emotion representation for the music can also be stored to accelerate future recommendation cycles.
2. If the user's music emotion preference information does not exist in the system, the music emotion preference for the user's current emotion state is calculated using the emotion state representation model. The parameters for the trained emotion state representation model emoDNN and the calculated user music emotion preference can be stored to accelerate future recommendation cycles.
3. Content-based method is used to get a list of music of similar music emotion to the music emotion preference at the given time T_{estat} . Add the music emotion vector to the music feature vector used by existing content-based music recommenders [84]. Calculate the similarities of the music to the emotion preferred music and rank the similarities to generate a list of music.
4. Collaborative filtering method is used to get the music with the music emotion preferred by similar music emotion preference users at the given time T_{estat} . Add the music emotion preference vector of the user's feature vector. Calculate the similarities between $User_i$ and other users. Rank the users according to the similarities. The preferred music lists of the users in the similar user list are ranked by the music emotion preference of the users. Get Top-K music items from the lists of the users.

- Using the generated music list as the recommendation list.

The similarities in the process are calculated by cosine similarity [85] depicted in Equation (16). Music emotion preference is a significant factor that affects music recommendations. However, conventional music preferences are also significant. Thus, we combined the music emotion preference and conventional music preferences by adding the music emotion vectors to the existing feature vectors of the music and adding the music emotion preference vectors to the existing feature vectors of the users. Therefore, the results generated by our method take advantage of both conventional preferences and music emotion preferences:

$$Sim = Cos(\vec{V}_i, \vec{V}_j) = \frac{\vec{V}_i \cdot \vec{V}_j}{\|\vec{V}_i\| \cdot \|\vec{V}_j\|} = \frac{\sum x_i \cdot x_j}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum x_j^2}}, x_i \in \vec{V}_i, x_j \in \vec{V}_j. \quad (16)$$

4. Experiments

In this section, the performance of the proposed emotion-aware music recommendation method emoMR is evaluated experimentally. The performance is compared against alternative approaches. We introduce the dataset used in the experiments, the experiment designs, the model training, and the evaluation. We implemented a toy system to control the experiment procedures and collect metrics data.

4.1. Datasets

According to the design of our method, the following data need to be fed to our method to train the models and generate recommendations: The audio file of the music, the metadata of the music, the activity data of the users within at least a 24-hr period, and social media data including posts, comments, and events. Existing datasets such as Netflix, GrooveShark, Last.fm [86] come up with no emotion-related characters. To train the models proposed in our work and test the performance of emoMR against other approaches, we used a hybrid dataset consists of:

- Data from the myPersonality [27] dataset. This dataset comes with information about 4 million Facebook users and 22 million posts over their timeline. By filtering through this dataset, we extracted 109 users who post music links and 509 music posted in the form of music links. This dataset provides not only the user tags but also music information and implicit user feedback information to aid the training of the music emotion representation model and the emotion state representation model.
- Data acquired from music social platform. The platform is a Chinese music streaming service (NetEase Cloud Music, <https://music.163.com> (accessed on 12 July 2021)) with an embedded community for users to share their lives with music. We scraped the music chart with a web crawler (<https://github.com/MiChongGET/CloudMusicApi> (accessed on 12 July 2021)) and got 500 high-ranking songs. The metadata and audio file of the 500 songs and the 509 ones in the myPersonality dataset were acquired. The music in the myPersonality dataset contained almost English songs and was relatively old, while the 500 high-ranking songs were mostly Chinese songs. Therefore, the two datasets have 38 songs in common. We also searched through the community and acquired 200 users related to the 971 songs and 116,079 user activities, including user likes, posts, reposts, and comments.
- Data acquired from the social network (Weibo, <https://weibo.com> (accessed on 13 July 2021)). We acquired 105,376 event-related data from Weibo. The data were posted within the time window of the user activities of the music social platform data. We use this data to insert event-related public opinion information to emoMR.

The ratio of the data for training against testing was around 3:1. The text information of music metadata and user activities were processed with text sentiment analysis services to obtain the emotion representation and construct the label data. The text sentiment analysis services for the English contents were processed by the IBM Tone Analyzer (<https://www.ibm.com/cloud/antw>).

ibm.com/watson/services/tone-analyzer (accessed on 17 July 2021)), and the Chinese contents were processed by Baidu AI (https://ai.baidu.com/tech/nlp/emotion_detection (accessed on 17 July 2021)). In particular, the GDPR regulation [87] has placed restrictions on the exploitation of user data in order to limit illegal use. The third-party dataset and the datasets acquired from the music streaming and social network services are guaranteed to be anonymized by the corresponding APIs of the services. The datasets will be used for our research purpose only and will not be used for commercial purposes.

4.2. Model Training

To train the music emotion representation model and the emotion state representation model, we use 70% of the training data to train the emoDNNs and 30% of the training data to test the trained models. We referred to a novel music recommendation system using deep learning [88] for the hyper-parameter settings and the training process. The hyper-parameters for the two emoDNNs are listed in Tables 2 and 3. The number of the hidden layers of the two emoDNNs were set to 5 according to Equation (11) and the experience derived from Liu et al. [67].

Table 2. Hyper-parameters for the music emotion representation model emoDNN.

Parameter	Setting
Training epoch	300
Batch	20
Optimizer	Adam
Learn rate	0.05

Table 3. Hyper-parameters for the emotion state representation model emoDNN.

Parameter	Setting
Training epoch	500
Batch	10
Optimizer	Adam
Learn rate	0.01

The training process uses the above hyper-parameter settings to learn the weights for the two models. As shown in Figure 7a, the loss function value for the train and validation sets plunged as the epoch reaches 50. When the epochs reach 300, the loss function values drop to 0.017 for training and 0.007 for validation. Figure 7b shows the accuracy changes over different epochs. When the epochs reach 300, the accuracy values reach 0.938 for training and 0.964 for validation. The output accuracy on the testing set is 97.46%.

As shown in Figure 8a, the loss function value for the train and validation sets plunged as the epoch reaches 50. When the epoch reaches 300, the loss function values drop to 0.0027 for training and 0.0029 for validation. Figure 8b shows the accuracy changes over different epochs. When the epochs reach 300, the accuracy values reach 0.871 for training and 0.813 for validation. The output accuracy on the testing set is 88.81%.

The trained models' accuracy is 97.46% for the music emotion representation model and 88.81% for the emotion state representation model. The accuracy is enough for generating representations and predicting emotion preferences for music in the testing dataset. Due to the nonlinear nature of the emotion state representation model, which is introduced by the human circadian rhythm and the stochastically-occurring events, the accuracy of the emotion state representation model emoDNN is significantly lower than that of the music emotion representation model. This leaves room for future improvements.

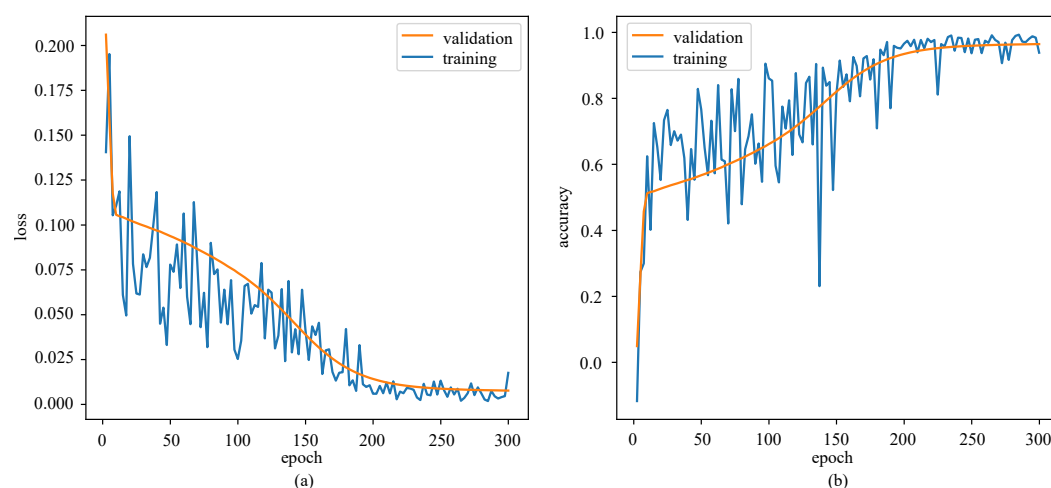


Figure 7. The loss function changes (a) and accuracy changes (b) over different epoches in training the music emotion representation model.

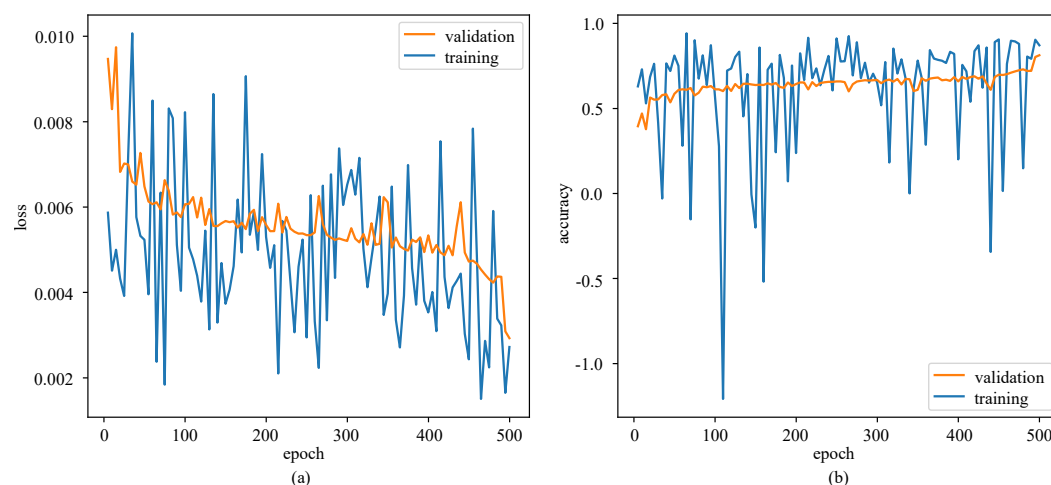


Figure 8. The loss function changes (a) and accuracy changes (b) over different epoches in training the emotion state representation model.

4.3. Baseline Algorithms and Metrics

The main purpose of the experiment is to compare the proposed emoMR with alternative approaches. For the comparison, we selected four alternative algorithms as the baseline algorithms, they are listed below:

1. A typical Content-Based music recommendation algorithm (CB) [89];
2. The Social Content-Based Collaborative Filtering algorithm (SCBCF) [90];
3. The Emotion-Aware Recommender System (EARS) [50];
4. The EMotion-aware REcSys (EMRES) [27].

The four baseline algorithms contain two approaches that do not take emotion information into account and two approaches that also utilize emotion information.

CB and SCBCF are alternative approaches that do not take emotion information into account. They are item similarity-based algorithms that share an easy way of computing similarity scores and are widely used in music recommendations.

EARS and EMERS are the two algorithms that also utilize emotion information. They demonstrate not only the ability to use item similarities but also the ability to use social network information and hybrid approaches. The two algorithms incorporate social network data to exploit the implicit user feedback in music recommendations and are emotion-aware.

To test the performance of emoMR against the baseline algorithms, we leveraged a variety of metrics commonly used in music recommendation method performance comparisons [8]. The metrics are: *Precision*, *Recall*, *F1*, and *HitRate* [27] listed by Equations (17)–(20). *Precision* is the fraction of recommended music pieces that are relevant. *Recall* is the fraction of relevant music pieces that are recommended. *F1* is the harmonic mean of precision and recall. *HitRate* is the fraction of hits.

$$Precision = \frac{|M_{rec} \cap M_{pick}|}{|M_{rec}|} \quad (17)$$

$$Recall = \frac{|M_{rec} \cap M_{pick}|}{|M_{pick}|} \quad (18)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (19)$$

$$HitRate = \frac{N_{hits}}{TOP - N} \quad (20)$$

where M_{rec} is the music recommended in the recommendation list and M_{pick} is the music actually picked by the user. If a piece of music is contained in a recommendation list for $User_i$ then it is a *hit*, N_{hits} is the number of *hits*, and $TOP - N$ is the number of Top-N recommendations.

4.4. Results

To assess the performance of our emoMR, we compared the *Precision*, *Recall*, *F1*, and *HitRate* of emoMR against those of the baseline algorithms. As the recommendation task was set to Top-N recommendations to suit the output of the above methods, we believe the N in Top-N has a significant influence on the performances of the methods. Therefore, the comparison was first given by a top-10 recommendation task which is a common practice among the recommendation systems of various music streaming services. Then, the performance variations were compared by changing the N from 5 to 20 to test the influence of N in Top-N. Finally, to test the event-related emotion utilization ability of emoMR, we compared the performances of the methods on two top-10 recommendation tasks using data within the time range of two major events on Chinese social media platforms.

4.4.1. Performance Comparison in the Top-10 Recommendation Task

We compared the performance of emoMR against the baseline algorithms of CB, SCBCF, EARS, and EMRES in the 4 metrics on a top-10 recommendation task. The results are shown in Figure 9.

The results show that emoMR outperforms the other baseline algorithms in all 4 metrics. The performance of the plain Content-Based algorithm (CB) is the worst. The SCBCF algorithm outperforms the CB algorithm due to its ability to exploit social network data and the hybrid approach it incorporates. The EARS, EMRES, and emoMR exploit not only the social network data but also information related to emotions. However, the EARS cannot utilize the low-level audio features, and the MERES takes a content-based-like approach. Thus, they fail to fuse more information than emoMR.

The results suggest that compared to the second-highest performance algorithm EMERS, emoMR improved 4.06% of *Precision*, 15.95% of *Recall*, 11.82% of *F1*, and 37.36% of *HitRate*. Compared to EARS, emoMR's *Precision*, *Recall*, *F1*, and *HitRate* were 10.81%, 32.92%, 25.25%, and 105.00% higher than those of EARS, respectively. The *Precision*, *Recall*, *F1*, and *HitRate* of emoMR were 16.47%, 78.68%, 57.09%, and 257.14% higher than those of SCBCF, respectively. Compared to CB, emoMR improved 203.97% of *Precision*, 165.85% of *Recall*, 144.02% of *F1*, and 400.00% of *HitRate*.

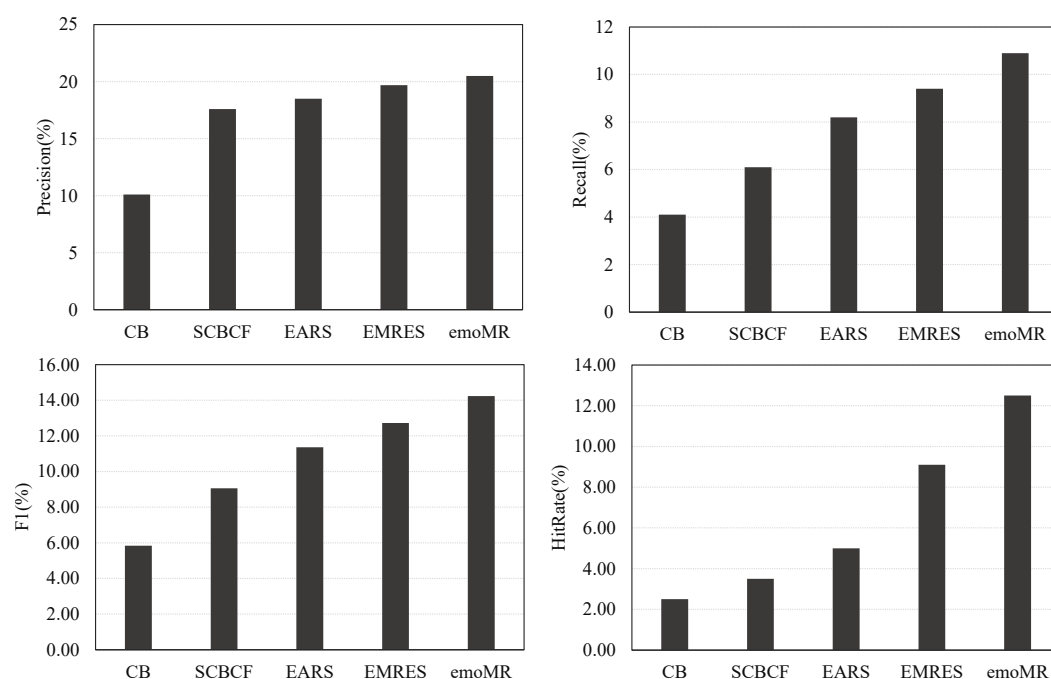


Figure 9. The performance comparison of the methods in the top-10 recommendation task.

4.4.2. Performance Comparison in the Top-N Recommendation Tasks

The performance of the five methods was also evaluated in Top-N tasks where N varies from 5 to 20 to reveal the influence of N on the performance. The results are shown in Figure 10.

The results show that as N varies from 5 to 20, the *Precisions* of the five methods decrease, whilst the *Recalls*, *F1s*, and *HitRates* increase. As the lengths of recommendation lists increase, and the precision drop, which means the tail items of the recommendation list contribute less to the quality of the recommendation. The precision of the emotion-aware algorithms also outperforms those non-emotion-aware algorithms. The increment in *Recalls* and *F1s* all suggest that increase the length of the recommendation list increases the advantages of emoMR over the other four algorithms. However, the *HitRates* suggests that EMRES outperforms emoMR on top-5 recommendations and that top-10 recommendations observe the greatest advantage of emoMR on EMRES in *HitRate*.

The comparisons under the top-5 recommendation task suggest the following results: The *Precision* of emoMR is 3.53%, 12.50%, 18.78%, and 90.24% higher than those of EMRES, EARS, SCBCF, and CB, respectively. The *F1* of emoMR is 0.61%, 13.70%, 19.39%, and 101.76% higher than those of EMRES, EARS, SCBCF, and CB, respectively. The *Recall* of emoMR has no significant advantage over EMRES in this situation, while it is 13.95%, 19.51%, and 104.16% higher than those of EARS, SCBCF, and CB, respectively. The *HitRate* of emoMR is 2.53% lower than EMRES while it is 71.11%, 140.62%, and 305.26% higher than those of EARS, SCBCF, and CB, respectively. Besides, emoMR still has an advantage over MERES in *HitRate* as the length of the recommendation list increases. Meanwhile, emoMR keeps leading the performance as the N in Top-N goes beyond 5.

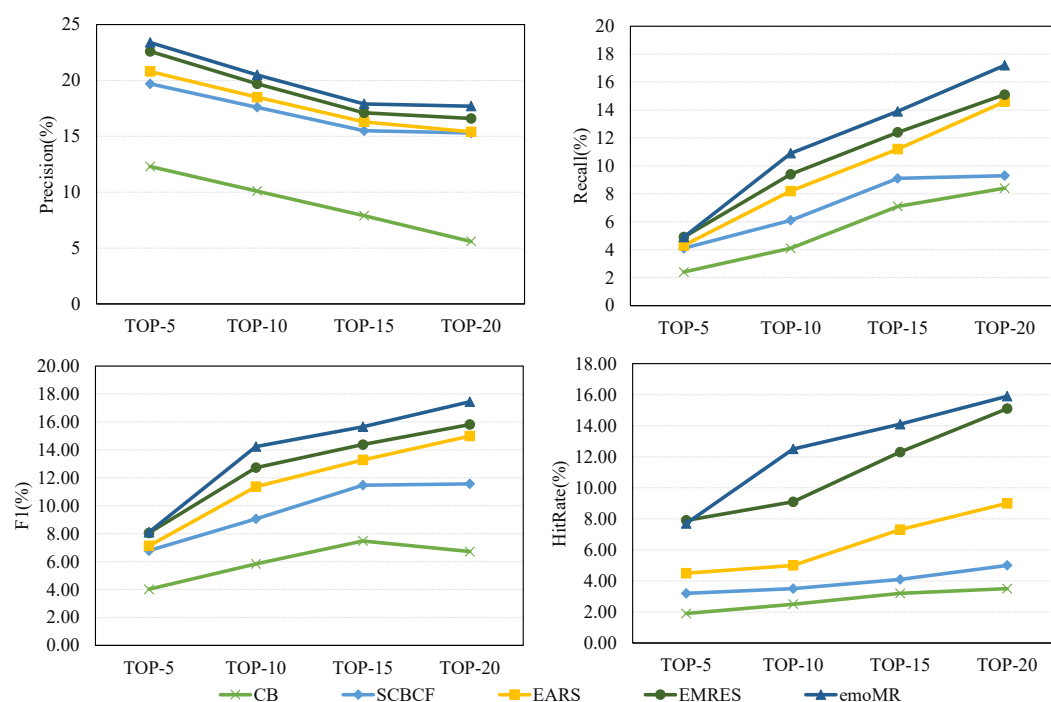


Figure 10. The performance comparison of the methods in the Top- n recommendation tasks.

4.4.3. Performance Comparison in Recommending Top-10 Items during Event-Related Time Ranges

The two major events are: (1) EVT1—the death of Yuan Longping “Father of hybrid rice” (22–25 May 2021) and (2) EVT2—imported Coronavirus Disease 2019 (COVID-19) cases in Zhejiang (9–12 June 2021). We grab the data from Weibo trends during the events’ time ranges. The performance results are shown in Figure 11.

The results show that the performance of emoMR significantly outstands the baseline algorithms in *Precision*, *Recall*, *F1*, and *HitRate* when recommending within certain event time ranges. The advantages of emoMR over other algorithms are even bigger compared to the top-10 recommendation performance, unaware of the major events.

EVT1 caused a general sad atmosphere in the Chinese social media platform of Weibo as the breaking news of Yuan Longping’s death started to go viral. For EVT1: Compared to the second highest performance algorithm EMERS, emoMR improved 7.03% of *Precision*, 18.55% of *Recall*, 14.51% of *F1*, and 48.64% of *HitRate*. Compared to EARS, emoMR’s *Precision*, *Recall*, *F1*, and *HitRate* were 13.90%, 35.29%, 27.79%, and 157.01% higher than those of EARS, respectively. The *Precision*, *Recall*, *F1*, and *HitRate* of emoMR were 23.83%, 94.91%, 69.99%, and 310.44% higher than those of SCBCF, respectively. Compared to CB, emoMR improved 115.15% of *Precision*, 194.87% of *Recall*, 116.92% of *F1*, and 461.22% of *HitRate*.

EVT2 brought public panic emotion to social media as people were worried about a new round of the outbreak of the pandemic. For EVT2: Compared to the second highest performance algorithm EMERS, emoMR improved 10.25% of *Precision*, 18.94% of *Recall*, 15.95% of *F1*, and 48.63% of *HitRate*. Compared to EARS, emoMR’s *Precision*, *Recall*, *F1*, and *HitRate* were 13.75%, 36.14%, 28.43%, and 159.04% higher than those of EARS, respectively. The *Precision*, *Recall*, *F1*, and *HitRate* of emoMR were 27.97%, 98.24%, 74.03%, and 318.46% higher than those of SCBCF, respectively. Compared to CB, emoMR improved 119.38% of *Precision*, 175.60% of *Recall*, 156.24% of *F1*, and 433.33% of *HitRate*.

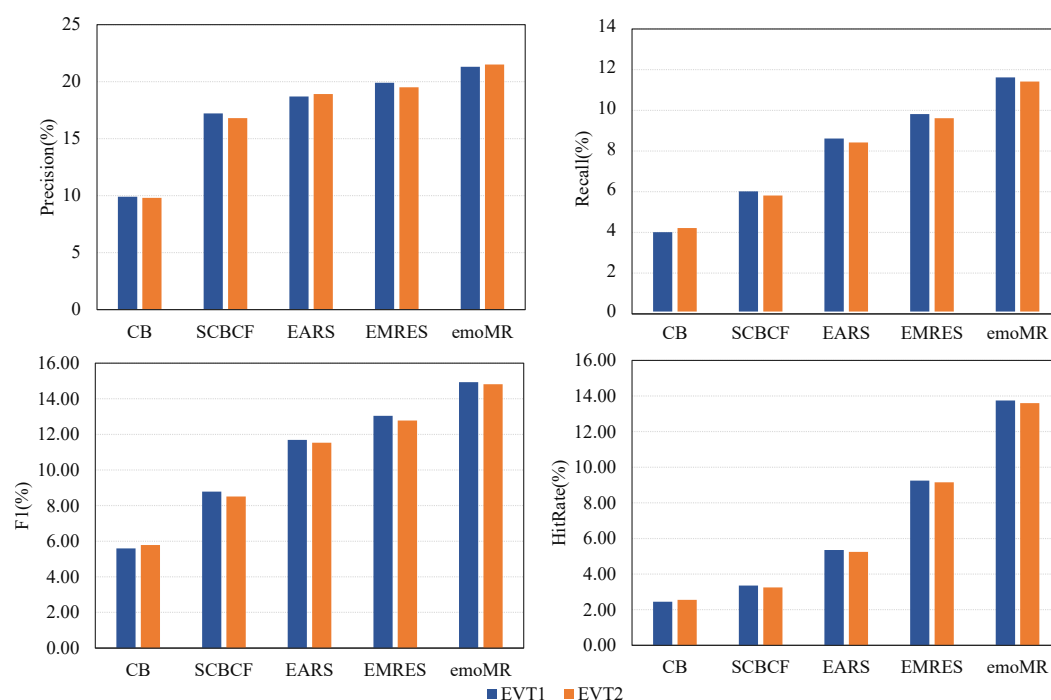


Figure 11. The performance comparison of the methods in recommending top-10 items during event-related time ranges.

4.5. Discussions

The results suggest that emoMR outperforms all the other baseline algorithms in the top-10 recommendation task. As studies suggest that more information fusion helps improve recommendation performance of Top-N tasks [16], we believe that is why emoMR outperforms the other algorithms. The results of the significant performance gaps between emotion-aware methods (emoMR, EMERS, and EARS) and emotion-unaware methods (SCBCF and CB) also indicate that the ability to utilize emotion-related information helps recommendation methods to gain advantages in music recommendations.

Our method outperforms the other baseline algorithms as the length of the recommendation list expands. Although the *HitRate* of EMERS exceeded emoMR in top-5 recommendations, emoMR still has its advantages in *HitRate* back in top-5 through top-20 recommendations. We believe the advantages of emoMR are gained by allowing more emotion variations to show up in the recommendation list. The music of similar music emotions and the users of similar music emotion preferences are ranked by the corresponding similarities. A longer recommendation list means a wider range of similarities which will include more emotion variations.

The advantages of emoMR in recommending top-10 items during event-related time ranges mark the unique ability of emoMR to utilize events information in emotion-aware music recommendations. As studies suggest that the brain mechanisms mediating musical emotion recognition are also dealing with the analysis and evaluation of the emotional content of complex social signals [18], social events are able to trigger the emotion cognition process. The events bring emotion-related information by public opinion. The sentiments of event-related public opinions were a critical part of the exogenous factors of the emotion state representation model. The results suggest that emoMR is able to utilize public opinion information during major events to refine the emotion state representation model by offering more information to the exogenous factors.

5. Conclusions

The development of personalized services on the Internet has called for new technologies to improve service quality. To improve the recommendation quality of the music recommenders, researchers have started introducing new information that has not been

exploited to make better suggestions. This paper proposed an emotion-aware hybrid music recommendation method using deep learning (emoMR) to introduce emotion information into music recommendations. Experiment results show that emoMR performs better in the metrics of *Precision*, *Recall*, *F1*, and *HitRate* than the other baseline algorithms in the top-10 recommendation task. Meanwhile, emoMR keeps its advantage over other baseline algorithms as the length of the recommendation list increases. We also tested the performance of emoMR on two major events (the death of Yuan Longping and the COVID-19 cases in Zhejiang). Results show that emoMR takes advantage of the events information and outperforms the other baseline algorithms. Our work contributes to the development of recommenders that include novel aspects of information. Our work can also be applied to a wide range of fields [91] to promote user satisfaction. Our innovative works in this paper are as listed as follows:

1. The proposed method predicts music emotions in a continuous form, enabling a more precise and flexible representation of music emotions than traditional discrete representations. By modeling the music emotion representation into a $\langle \textit{Valence}, \textit{Arousal} \rangle$ vector based on the emotion valence-arousal model using low-level audio features and music metadata, we build a deep neural network (emoDNN) to predict the emotion representations for the music items. Compared to the discrete representations of music emotions in other studies, our model predicts the music emotion in a continuous form of valence and arousal scores.
2. The proposed method predicts users' music emotion preference whenever it needs while traditional methods can only generate predictions at time points when user feedback data are present. By modeling the users' emotion states using an artificial emotion generation model with endogenous factors generated by the human circadian rhythm model and exogenous factors consisted of events and implicit user feedback, we use the emoDNN to predict the users' music emotion preferences under the emotion states represented by the model. Benefiting from the continuity of the human circadian rhythm, the model is able to predict continuously across time regardless of the absence of user feedback data at a given time point.
3. The proposed method can utilize event information to refine the emotion generation, which provides a more accurate emotion state description than traditional artificial emotion generation models. With the introduction of events in the exogenous factors of the user emotion state representation model, emoMR is able to express the influence of events (especially major social events) on user emotions.
4. The proposed method employs a hybrid approach of combining the advantages of both content-based and collaborative filtering approaches to generate the recommendations. Theoretically, our findings will contribute to the theory of emotion recognition by reflecting the cognitive theory of emotion [68] and the cognitive appraisal theory of emotion [69] with social signals.

We acknowledge some limitations in our study that allows for opportunities for future research and list them as: (1) Although the accuracy of the emotion state representation model is acceptable, there is still room for improvements compared to the accuracy of the music emotion representation model. Future works can benefit from improving the accuracy of the emotion state representation model. (2) Limited by the research paradigm and the designed model, our method is unable to predict the emotion of a person at a given time point. Future works can benefit from designing an intermediate output of emotion prediction to aid the recommendation process and validate the accuracy of the recommendation. (3) Limited by the approaches of the study and the dataset used, our study did not take users with minority personal traits such as bipolar traits into consideration which is a compelling choice. Future works can benefit from taking humanistic care for minorities. (4) The proposed method utilized the human circadian rhythm model to generate emotion state representations that need at least user activity data of a 24-hr time period. Future works can benefit from exploiting other models that require fewer data.

Author Contributions: S.W. and C.X. built the framework of the whole paper. S.W. and Z.T. carried out the collection and preprocessing of the data. S.W. designed the whole experiment and method. Z.T. implemented the experiment. C.X. and A.S.D. provided analytical and experimental tools. S.W. and C.X. wrote the manuscript. C.X., A.S.D. and Z.T. revised the manuscript. All authors read and approved the final manuscript.

Funding: This research is supported by the Project of China (Hangzhou) Cross-border E-commerce College (no. 2021KXYJ06), the Philosophy and Social Science Foundation of Zhejiang Province (21NDJC083YB), the National Natural Science Foundation of China (61802095, 71702164), the Natural Science Foundation of Zhejiang Province (LY20G010001), and the Scientific Research Projects of Zhejiang Education Department (Y201737967). Contemporary Business and Trade Research Center and Center for Collaborative Innovation Studies of Modern Business of Zhejiang Gongshang University of China (2021SMYJ05LL).

Institutional Review Board Statement: The presented research was conducted according to the code of ethics of Zhejiang Gongshang University.

Data Availability Statement: The data are available from the corresponding author on request.

Conflicts of Interest: The authors claim that there is no conflict of interest.

References

1. Xu, C. A novel recommendation method based on social network using matrix factorization technique. *Inf. Process. Manag.* **2018**, *54*, 463–474. [\[CrossRef\]](#)
2. Xu, C. A big-data oriented recommendation method based on multi-objective optimization. *Knowl. Based Syst.* **2019**, *177*, 11–21. [\[CrossRef\]](#)
3. Xu, C.; Ding, A.S.; Zhao, K. A novel POI recommendation method based on trust relationship and spatial-temporal factors. *Electron. Commer. Res. Appl.* **2021**, *48*, 101060. [\[CrossRef\]](#)
4. Song, Y.; Dixon, S.; Pearce, M. A survey of music recommendation systems and future perspectives. In Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval, London, UK, 19–22 June 2012; Volume 4, pp. 395–410.
5. Schedl, M.; Knees, P.; McFee, B.; Bogdanov, D.; Kaminskas, M. Music recommender systems. In *Recommender Systems Handbook*; Springer: Berlin, Germany, 2015; pp. 453–492.
6. Xie, Y.; Ding, L. A survey of music personalized recommendation system. In Proceedings of the 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018), Chongqing, China, 26–27 May 2018; Atlantis Press: Paris, France, 2018; pp. 852–855.
7. Cheng, Z.; Shen, J. On effective location-aware music recommendation. *ACM Trans. Inf. Syst. (TOIS)* **2016**, *34*, 1–32. [\[CrossRef\]](#)
8. Wang, D.; Deng, S.; Xu, G. Sequence-based context-aware music recommendation. *Inf. Retr. J.* **2018**, *21*, 230–252. [\[CrossRef\]](#)
9. Krumhansl, C.L. Music: A link between cognition and emotion. *Curr. Dir. Psychol. Sci.* **2002**, *11*, 45–50. [\[CrossRef\]](#)
10. Yang, Y.H.; Chen, H.H. Machine recognition of music emotion: A review. *ACM Trans. Intell. Syst. Technol. (TIST)* **2012**, *3*, 1–30. [\[CrossRef\]](#)
11. Mahsal Khan, S.; Abdul Hamid, N.; Mohd Rashid, S. Music and its familiarity affection on audience decision making. *J. Messenger* **2019**, *11*, 70–80. [\[CrossRef\]](#)
12. Carrington, S.J.; Bailey, A.J. Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum. Brain Mapp.* **2009**, *30*, 2313–2335. [\[CrossRef\]](#)
13. Deng, J.J.; Leung, C. Emotion-based music recommendation using audio features and user playlist. In Proceedings of the 2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012), Taipei, Taiwan, 23–25 October 2012; pp. 796–801.
14. Rachman, F.H.; Sarno, R.; Faticah, C. Music Emotion Classification based on Lyrics-Audio using Corpus based Emotion. *Int. J. Electr. Comput. Eng.* **2018**, *8*, 1720. [\[CrossRef\]](#)
15. Yang, Y.H.; Su, Y.F.; Lin, Y.C.; Chen, H.H. Music emotion recognition: The role of individuality. In Proceedings of the International Workshop on Human-Centered Multimedia, Augsburg, Bavaria, Germany, 28 September 2007; pp. 13–22.
16. Hu, B.; Shi, C.; Zhao, W.X.; Yang, T. Local and global information fusion for top-n recommendation in heterogeneous information network. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 1683–1686.
17. Ayata, D.; Yaslan, Y.; Kamasak, M.E. Emotion based music recommendation system using wearable physiological sensors. *IEEE Trans. Consum. Electron.* **2018**, *64*, 196–203. [\[CrossRef\]](#)
18. Omar, R.; Henley, S.M.; Bartlett, J.W.; Hailstone, J.C.; Gordon, E.; Sauter, D.A.; Frost, C.; Scott, S.K.; Warren, J.D. The structural neuroanatomy of music emotion recognition: evidence from frontotemporal lobar degeneration. *Neuroimage* **2011**, *56*, 1814–1821. [\[CrossRef\]](#)

19. Malhotra, A.; Totti, L.; Meira, W., Jr.; Kumaraguru, P.; Almeida, V. Studying user footprints in different online social networks. In Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, 26–29 August 2012; pp. 1065–1070.
20. Rivas, A.; González-Briones, A.; Cea-Morán, J.J.; Prat-Pérez, A.; Corchado, J.M. My-Trac: System for Recommendation of Points of Interest on the Basis of Twitter Profiles. *Electronics* **2021**, *10*, 1263. [\[CrossRef\]](#)
21. Huang, X.; Liao, G.; Xiong, N.; Vasilakos, A.V.; Lan, T. A Survey of Context-Aware Recommendation Schemes in Event-Based Social Networks. *Electronics* **2020**, *9*, 1583. [\[CrossRef\]](#)
22. Park, T.; Jeong, O.R. Social network based music recommendation system. *J. Internet Comput. Serv.* **2015**, *16*, 133–141. [\[CrossRef\]](#)
23. Shen, T.; Jia, J.; Li, Y.; Ma, Y.; Bu, Y.; Wang, H.; Chen, B.; Chua, T.S.; Hall, W. Peia: Personality and emotion integrated attentive model for music recommendation on social media platforms. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 206–213.
24. Ma, Y.; Li, X.; Xu, M.; Jia, J.; Cai, L. Multi-scale context based attention for dynamic music emotion prediction. In Proceedings of the 25th ACM international conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1443–1450.
25. Wu, D. Music personalized recommendation system based on hybrid filtration. In Proceedings of the 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China, 12–13 January 2019; pp. 430–433.
26. Abdul, A.; Chen, J.; Liao, H.Y.; Chang, S.H. An emotion-aware personalized music recommendation system using a convolutional neural networks approach. *Appl. Sci.* **2018**, *8*, 1103. [\[CrossRef\]](#)
27. Polignano, M.; Narducci, F.; de Gemmis, M.; Semeraro, G. Towards Emotion-aware Recommender Systems: an Affective Coherence Model based on Emotion-driven Behaviors. *Expert Syst. Appl.* **2021**, *170*, 114382. [\[CrossRef\]](#)
28. Cowen, A.S.; Keltner, D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7900–E7909. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Gunes, H.; Schuller, B.; Pantic, M.; Cowie, R. Emotion representation, analysis and synthesis in continuous space: A survey. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA, 21–25 March 2011; pp. 827–834.
30. White, C.; Ratcliff, R.; Vasey, M.; McKoon, G. Dysphoria and memory for emotional material: A diffusion-model analysis. *Cogn. Emot.* **2009**, *23*, 181–205. [\[CrossRef\]](#)
31. Newman, D.B.; Nezlek, J.B. The Influence of Daily Events on Emotion Regulation and Well-Being in Daily Life. *Personal. Soc. Psychol. Bull.* **2021**, 0146167220980882. [\[CrossRef\]](#)
32. Basu, S.; Jana, N.; Bag, A.; Mahadevappa, M.; Mukherjee, J.; Kumar, S.; Guha, R. Emotion recognition based on physiological signals using valence-arousal model. In Proceedings of the 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, India, 21–24 December 2016.
33. Paul, D.; Kundu, S. A Survey of Music Recommendation Systems with a Proposed Music Recommendation System. In *Emerging Technology in Modelling and Graphics*; Springer: Berlin, Germany, 2020; pp. 279–285.
34. Jazi, S.Y.; Kaedi, M.; Fatemi, A. An emotion-aware music recommender system: bridging the user's interaction and music recommendation. *Multimed. Tools Appl.* **2021**, *80*, 13559–13574. [\[CrossRef\]](#)
35. Coutinho, E.; Cangelosi, A. Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion* **2011**, *11*, 921. [\[CrossRef\]](#)
36. Barthet, M.; Fazekas, G.; Sandler, M. Music emotion recognition: From content-to context-based models. In *International Symposium on Computer Music Modeling and Retrieval*; Springer: Berlin, Germany, 2012; pp. 228–252.
37. Yoon, K.; Lee, J.; Kim, M.U. Music recommendation system using emotion triggering low-level features. *IEEE Trans. Consum. Electron.* **2012**, *58*, 612–618. [\[CrossRef\]](#)
38. Panda, R.; Malheiro, R.; Paiva, R.P. *Musical Texture and Expressivity Features for Music Emotion Recognition*; ISMIR: Paris, France, 2018; pp. 383–391.
39. Wang, J.C.; Yang, Y.H.; Jhuo, I.H.; Lin, Y.Y.; Wang, H.M. The acousticvisual emotion Guassians model for automatic generation of music video. In Proceedings of the 20th ACM international conference on Multimedia, Nara, Japan, 29 October–2 November 2012; pp. 1379–1380.
40. Aljanaki, A. Emotion in Music: Representation and Computational Modeling. Ph.D. Thesis, Utrecht University, Utrecht, The Netherlands, 2016.
41. An, Y.; Sun, S.; Wang, S. Naive Bayes classifiers for music emotion classification based on lyrics. In Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24–26 May 2017; pp. 635–638.
42. Kim, Y.E.; Schmidt, E.M.; Migneco, R.; Morton, B.G.; Richardson, P.; Scott, J.; Speck, J.A.; Turnbull, D. Music emotion recognition: A state of the art review. *Proc. Ismir* **2010**, *86*, 937–952.
43. Nakasone, A.; Prendinger, H.; Ishizuka, M. Emotion recognition from electromyography and skin conductance. In Proceedings of the 5th International Workshop on Biosignal Interpretation, Citeseer, Palmerston North, New Zealand, 13–15 December 2004; pp. 219–222.
44. Lin, Y.P.; Wang, C.H.; Jung, T.P.; Wu, T.L.; Jeng, S.K.; Duann, J.R.; Chen, J.H. EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1798–1806.
45. Hsu, J.L.; Zhen, Y.L.; Lin, T.C.; Chiu, Y.S. Affective content analysis of music emotion through EEG. *Multimed. Syst.* **2018**, *24*, 195–210. [\[CrossRef\]](#)

46. Dornbush, S.; Fisher, K.; McKay, K.; Prikhodko, A.; Segall, Z. XPOD-A human activity and emotion aware mobile music player. In Proceedings of the International Conference on Mobile Technology, Applications and Systems, IET, Guangzhou, China, 15–17 November 2005.
47. Sarda, P.; Halasawade, S.; Padmawar, A.; Aghav, J. Emusic: emotion and activity-based music player using machine learning. In *Advances in Computer Communication and Computational Sciences*; Springer: Berlin, Germany, 2019; pp. 179–188.
48. Rosa, R.L.; Rodriguez, D.Z.; Bressan, G. Music recommendation system based on user's sentiments extracted from social networks. *IEEE Trans. Consum. Electron.* **2015**, *61*, 359–367. [[CrossRef](#)]
49. Deng, S.; Wang, D.; Li, X.; Xu, G. Exploring user emotion in microblogs for music recommendation. *Expert Syst. Appl.* **2015**, *42*, 9284–9293. [[CrossRef](#)]
50. Qian, Y.; Zhang, Y.; Ma, X.; Yu, H.; Peng, L. EARS: Emotion-aware recommender system based on hybrid information fusion. *Inf. Fusion* **2019**, *46*, 141–146. [[CrossRef](#)]
51. Lu, C.C.; Tseng, V.S. A novel method for personalized music recommendation. *Expert Syst. Appl.* **2009**, *36*, 10035–10044. [[CrossRef](#)]
52. Kim, H.H. A semantically enhanced tag-based music recommendation using emotion ontology. In *Asian Conference on Intelligent Information and Database Systems*; Springer: Berlin, Germany, 2013; pp. 119–128.
53. Han, B.; Rho, S.; Jun, S.; Hwang, E. Music emotion classification and context-based music recommendation. *Multimed. Tools Appl.* **2010**, *47*, 433–460. [[CrossRef](#)]
54. Wang, X.; Wang, Y. Improving content-based and hybrid music recommendation using deep learning. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 627–636.
55. Awan, M.J.; Khan, R.A.; Nobanee, H.; Yasin, A.; Anwar, S.M.; Naseem, U.; Singh, V.P. A Recommendation Engine for Predicting Movie Ratings Using a Big Data Approach. *Electronics* **2021**, *10*, 1215. [[CrossRef](#)]
56. Schedl, M. Deep learning in music recommendation systems. *Front. Appl. Math. Stat.* **2019**, *5*, 44. [[CrossRef](#)]
57. Nam, J.; Choi, K.; Lee, J.; Chou, S.Y.; Yang, Y.H. Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from bach. *IEEE Signal Process. Mag.* **2018**, *36*, 41–51. [[CrossRef](#)]
58. Hossain, M.S.; Muhammad, G. Emotion recognition using deep learning approach from audio–visual emotional big data. *Inf. Fusion* **2019**, *49*, 69–78. [[CrossRef](#)]
59. Aldao, A.; Nolen-Hoeksema, S. One versus many: Capturing the use of multiple emotion regulation strategies in response to an emotion-eliciting stimulus. *Cogn. Emot.* **2013**, *27*, 753–760. [[CrossRef](#)]
60. Nwe, T.L.; Foo, S.W.; Silva, L. Detection of stress and emotion in speech using traditional and FFT based log energy features. In Proceedings of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia, Proceedings of the 2003 Joint, Singapore, 15–18 December 2004.
61. Hui, G.; Chen, S.; Su, G. Emotion Classification of Infant Voice Based on Features Derived from Teager Energy Operator. In Proceedings of the 2008 Congress on Image and Signal Processing, Sanya, China, 27–30 May 2008.
62. Jothilakshmi, S.; Ramalingam, V.; Palanivel, S. Unsupervised speaker segmentation with residual phase and MFCC features. *Expert Syst. Appl.* **2009**, *36*, 9799–9804. [[CrossRef](#)]
63. Hasan, M.; Rahman, M.S.; Shimamura, T. Windowless-Autocorrelation-Based Cepstrum Method for Pitch Extraction of Noisy Speech. *J. Signal Process.* **2013**, *16*, 231–239. [[CrossRef](#)]
64. Corthaut, N.; Govaerts, S.; Verbert, K.; Duval, E. *Connecting the Dots: Music Metadata Generation, Schemas and Applications*. In Proceedings of the 9th International Conference on Music Information, Philadelphia, PA, USA, 14–18 September 2008; pp. 249–254.
65. Tallapally, D.; Sreepada, R.S.; Patra, B.K.; Babu, K.S. User preference learning in multi-criteria recommendations using stacked auto encoders. In Proceedings of the 12th ACM Conference on Recommender Systems, Vancouver, BC, Canada, 2 October 2018; pp. 475–479.
66. Zarzour, H.; Al-Sharif, Z.A.; Jararweh, Y. RecDNNing: a recommender system using deep neural network with user and item embeddings. In Proceedings of the 2019 10th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 11–13 June 2019; pp. 99–103.
67. Liu, Z.; Guo, S.; Wang, L.; Du, B.; Pang, S. A multi-objective service composition recommendation method for individualized customer: hybrid MPA-GSO-DNN model. *Comput. Ind. Eng.* **2019**, *128*, 122–134. [[CrossRef](#)]
68. Bever, T.G. A cognitive theory of emotion and aesthetics in music. *Psychomusicol. J. Res. Music Cogn.* **1988**, *7*, 165. [[CrossRef](#)]
69. Garcia-Prieto, P.; Scherer, K.R. Connecting social identity theory and cognitive appraisal theory of emotions. In *Social Identities: Motivational, Emotional, Cultural Influences*; Psychology Press: Hove, UK, 2006; pp. 189–207.
70. Kim, H.R.; Kwon, D.S. Computational model of emotion generation for human–robot interaction based on the cognitive appraisal theory. *J. Intell. Robot. Syst.* **2010**, *60*, 263–283. [[CrossRef](#)]
71. Manera, V.; Samson, A.C.; Pehrs, C.; Lee, I.A.; Gross, J.J. The eyes have it: The role of attention in cognitive reappraisal of social stimuli. *Emotion* **2014**, *14*, 833. [[CrossRef](#)] [[PubMed](#)]
72. Blechert, J.; Sheppes, G.; Di Tella, C.; Williams, H.; Gross, J.J. See what you think: Reappraisal modulates behavioral and neural responses to social stimuli. *Psychol. Sci.* **2012**, *23*, 346–353. [[CrossRef](#)]
73. Pribram, K.H.; Melges, F.T. Psychophysiological basis of emotion. *Handb. Clin. Neurol.* **1969**, *3*, 316–341.

74. Takahashi, K. Remarks on emotion recognition from bio-potential signals. In Proceedings of the 2nd International conference on Autonomous Robots and Agents, Citeseer, Palmerston North, New Zealand, 13–15 December 2004; pp. 186–191.
75. Hofstra, W.A.; de Weerd, A.W. How to assess circadian rhythm in humans: a review of literature. *Epilepsy Behav.* **2008**, *13*, 438–444. [[CrossRef](#)] [[PubMed](#)]
76. Cardinali, D.P. The human body circadian: How the biologic clock influences sleep and emotion. *Neuroendocrinol. Lett.* **2000**, *21*, 9–16.
77. Panda, S. Circadian physiology of metabolism. *Science* **2016**, *354*, 1008–1015. [[CrossRef](#)] [[PubMed](#)]
78. olde Scheper, T.; Klinkenberg, D.; Pennartz, C.; Van Pelt, J. A mathematical model for the intracellular circadian rhythm generator. *J. Neurosci.* **1999**, *19*, 40–47. [[CrossRef](#)]
79. Refinetti, R.; Cornélissen, G.; Halberg, F. Procedures for numerical analysis of circadian rhythms. *Biol. Rhythm Res.* **2007**, *38*, 275–325. [[CrossRef](#)]
80. Mitsutake, G.; Otsuka, K.; Cornelissen, G.; Herold, M.; Günther, R.; Dawes, C.; Burch, J.; Watson, D.; Halberg, F. Circadian and infradian rhythms in mood. *Biomed. Pharmacother.* **2000**, *55*, s94–s100. [[CrossRef](#)]
81. Hansen, C.H.; Hansen, R.D. How rock music videos can change what is seen when boy meets girl: Priming stereotypic appraisal of social interactions. *Sex Roles* **1988**, *19*, 287–316. [[CrossRef](#)]
82. Turek, F.W.; Penev, P.; Zhang, Y.; Van Reeth, O.; Zee, P. Effects of age on the circadian system. *Neurosci. Biobehav. Rev.* **1995**, *19*, 53–58. [[CrossRef](#)]
83. Zheng, E.; Kondo, G.Y.; Zilora, S.; Yu, Q. Tag-aware dynamic music recommendation. *Expert Syst. Appl.* **2018**, *106*, 244–251. [[CrossRef](#)]
84. Soleymani, M.; Aljanaki, A.; Wiering, F.; Veltkamp, R.C. Content-based music recommendation using underlying music preference structure. In Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy, 29 June–3 July 2015; pp. 1–6.
85. Metcalf, L.; Casey, W. Chapter 2—Metrics, similarity, and sets. In *Cybersecurity and Applied Mathematics*; Metcalf, L., Casey, W., Eds.; Syngress: Boston, MA, USA, 2016; pp. 3–22.
86. Vigliensoni, G.; Fujinaga, I. The Music Listening Histories Dataset. In Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR17), Suzhou, China, 23–28 October 2017; pp. 96–102.
87. Goddard, M. The EU General Data Protection Regulation (GDPR): European regulation that has a global impact. *Int. J. Mark. Res.* **2017**, *59*, 703–705. [[CrossRef](#)]
88. Fessahaye, F.; Perez, L.; Zhan, T.; Zhang, R.; Fossier, C.; Markarian, R.; Chiu, C.; Zhan, J.; Gewali, L.; Oh, P. T-recsys: A novel music recommendation system using deep learning. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 11–13 January 2019; pp. 1–6.
89. Bogdanov, D.; Haro, M.; Fuhrmann, F.; Gómez, E.; Herrera, P. Content-based music recommendation based on user preference examples. *Copyr. Inf.* **2010**, *33*, 1–6.
90. Su, J.H.; Chang, W.Y.; Tseng, V.S. Effective social content-based collaborative filtering for music recommendation. *Intell. Data Anal.* **2017**, *21*, S195–S216. [[CrossRef](#)]
91. Xiang, K.; Xu, C.; Wang, J. Understanding the relationship between tourists' consumption behavior and their consumption substitution willingness under unusual environment. *Psychol. Res. Behav. Manag.* **2021**, *14*, 483–500. [[CrossRef](#)]