Multimedia Tools and Applications (2021) 80:13559–13574
https://doi.org/10.1007/s11042-020-10386-7

An emotion-aware music recommender system: bridging the user's interaction and music recommendation

Saba Yousefian Jazi[1] **· Marjan Kaedi**[1] **· Afsaneh Fatemi**[1]

Received: 26 July 2019 / Revised: 20 October 2020 / Accepted: 22 December 2020 /
Published online: 15 January 2021

Abstract

In emotion-aware music recommender systems, the user's current emotion is identified and considered in recommending music to him. We have two motivations to extend the existing systems: (1) to the best of our knowledge, the current systems first estimate the user's emotions and then suggest music based on it. Therefore, the emotion estimation error affects the recommendation accuracy. (2) Studies show that the pattern of users' interactions with input devices can reflect their emotions. However, these patterns have not been used yet in emotion-aware music recommender systems. In this study, a music recommender system is proposed to suggest music based on users' keystrokes and mouse clicks patterns. Unlike the previous ones, the proposed system maps these patterns directly to the user's favorite music, without labeling its current emotion. The results show that even though this system does not use any additional device, it is highly accurate compared to previous methods.

Keywords Emotion-aware music recommender · Keystroke pattern · Mouse click pattern · Collaborative filtering

✉ Marjan Kaedi
kaedi@eng.ui.ac.ir

Saba Yousefian Jazi
saba.yousefian@mehr.ui.ac.ir

Afsaneh Fatemi
a_fatemi@eng.ui.ac.ir

[1] Faculty of Computer Engineering, University of Isfahan, Hezar Jerib Street, Isfahan 8174673441, Iran

1 Introduction

Recommender systems have been developed to tackle the information overload problem. They help users find their desired items in the shortest possible time through the massive amount of information. There are different recommender algorithms, and "collaborative filtering" is one of the most common and successful ones [13, 15].

In collaborative filtering, users whose preferences are similar to the current user are selected, and then their favorite items are recommended to the current user [11, 41, 44]. One of the most common usages of recommender systems is to recommend the user's favorite music.

According to [4],"The music choice of a user is not only dependent on the historical preferences or music contents but also dependent on the mood of that user." Therefore, to be able to improve the accuracy of the music recommender systems, it is necessary to consider the user's emotions in recommending music to him [7, 14, 28, 40]. Thus, users receive recommendations that are closer to their instant emotions. So they are more satisfied with the suggested music.

User's emotion can be identified explicitly or implicitly. Explicit methods use direct inquiry from the user. Electrocardiogram sensors, body temperature sensors, eye pupil motion detectors [3], or facial muscle analyzers [8, 9] are used for implicit emotion identification. Also, these methods may use the user's sitting style [20], user's voice signals [22], user's interaction patterns with the keyboard and mouse [8, 24, 31, 35], and so on for emotion recognition. Several implicit methods are reviewed in Section 2.1.

In previous studies, emotion-aware music recommender systems have suggested music to users based on their current emotions (these studies will be reviewed in Section 2.2). As shown in Section 2.2, in all these studies, the user's emotion is not estimated based on their interaction with input devices (e.g., keyboard and mouse). However, it is shown that the user's current emotion is reflected in his/her interaction with input devices [20, 23]. On the other hand, to the best of our knowledge, in all the emotion-aware music recommender systems, the user's emotions are first estimated. Then the music is suggested to the user. So, the emotion estimation error decreases the music recommendation accuracy. The recommender system suggests music based on the users' keystroke and mouse click patterns in the proposed approach. The music is recommended directly, without labeling the user's emotion, so that the error of estimating the user's emotion does not negatively affect the recommendation accuracy. This system is based on collaborative filtering.

In the rest of the paper, the related works are reviewed in Section 2. In Section 3, the proposed method for recommending music based on the keystroke pattern and user mouse click is presented. The proposed approach is evaluated in Section 4. Finally, the paper approach is concluded in Section 5 with a summary.

2 Related works

As said before, the style of user's interaction with devices such as mouse and keyboard is one of the implicit methods of user's emotion recognition. In [19], a field study and a controlled experimental study with computer users were conducted. They gathered the users' interaction data with mouse and keyboards, and the users reported their mood during the work simultaneously. A significant correlation was found between the emotional data and the interaction data. They showed that the computer users' interaction behavior with mouse and keyboard could be used to predict mood.

On the other hand, the music evokes different visual emotions in the listener, but it can sometimes include a persona. This character's emotions are expressed in music and convey these emotions and even more complex mixed emotions to the listener. This allows the listener to experience a psychological journey, and this experience is usually accompanied by the arousal of real physiological states and practical tendencies in him. In this way, music can indirectly lead to a set of different emotions in the listener [29]. Music flows in every culture and has shown its unique effects on people's lives. An important reason for the global appeal of music is the unique emotional rewards it offers its listeners. In 2008, a domain-specific emotional model, GEMS, was proposed to determine the musical emotion [43]. Some other researchers used the models such as GEMS to study the emotion induction from music [2]. Some others tried to prove the relation between music genre and the felt emotions [17, 18]. These researches show that music can be transposed to emotions. "Music has been reported to evoke the full range of human's emotion: from sad, nostalgic, and tense, to happy, relaxed, calm, and joyous" [29].

So, the music, emotions, and user's interactions may be interrelated. In this study, we aim to use mouse/keyboard interactions to predict the user's emotion and use the predicted emotion in music recommendation.

In this section, the studies that identify the user's emotion based on his interaction with input devices are reviewed. Then the available emotion-aware music recommender systems are introduced.

2.1 Emotion estimation based on the user's interaction with input devices

In [20], the user's emotion is explored and analyzed using keyboard usage patterns and then has been categorized into three positive, negative, and neutral categories. The results have shown that this system has been more successful in estimating the users' negative emotions.

In [23], two categories of data are used to detect the user's emotions. The first category is the keystroke dynamic properties, and the second one is the text pattern properties. Seven different emotions have been tried, including anger, hatred, guilt, fear, joy, discomfort, and embarrassment, and the results have shown that among these seven emotions, the estimation of the user's enjoyment emotion has been more successful.

In [21], the user's keystrokes patterns, along with text semantics and heart rate, were used to estimate the user's emotion. The emotions, including joy, discomfort, fear, anger, surprise, love, hatred, fatigue, and disgust, were focused and analyzed.

In [35], collecting data of mouse movement and keyboard usage patterns (when typing an arbitrary text) and using a variety of clustering algorithms (including KNN, KStar, random committee, random forest, and a new clustering algorithm called bounded K-means), the users' emotions were categorized into five groups. These categories include happiness, inspiration, compassion, disgust, and fear.

In 2017, Pentel used movie clips to induce emotions in some users, and then emotions of the users were gathered using the "Think-Aloud" protocol. Meanwhile, some data was also collected from users' interactions through two experiments: In the first experiment, just the mouse's data was collected: In the first experiment, just the data about the mouse was collected. In the second experiment, in addition to the mouse data, keyboard data was also gathered. Using this data, supervised machine learning models were developed to estimate the user's emotions. The best models could predict the user's emotion with an accuracy of more than 90% [24].

In another study in 2017, [8] extracted some students' keyboard and mouse usage patterns while the students were programming. Students' emotional state was identified during solving programming problems based on their facial expressions and were recorded as videos. Then the patterns were mapped into emotional states. The results were 70.25% accurate.

As shown in the studies reviewed in this section, the user's keystroke and mouse click patterns may be used successfully to recognize the user's emotions. The recognized emotions can be essential inputs for recommender systems.

2.2 Emotion-aware music recommender systems

In [36], first, the user's speech signals are analyzed, and his emotional state is recognized in terms of arousal and valence. The five emotional states of anger, surprise, sadness, boredom, and happiness are considered here. Each emotional state is considered as a point in the two-dimensional coordinate system. After that, the target emotional state the user wants to reach is also considered an issue. Songs are also considered as points in this coordinate system, based on their emotional states. A hypothetical line is then drawn between the user's current emotion and the target emotion, and the songs on the theoretical line are recommended to the user.

In [9], users' emotions were extracted based on the data collected from wearable physiological sensors. The statistical properties of these data are given as inputs to the clustering algorithms. The clustering algorithms estimate the user's emotions in terms of arousal and valence. Then the collaborative filtering recommender system recommends music tailored to the user's emotions.

Xing et al. [40] have considered 580 music tracks and have asked 50 volunteers to label them in a fuzzy manner with vigorous, dignified, sad, dreaming, soothing, graceful, joyous, and exciting emotions. Then, with the help of machine learning methods, the music emotions have been estimated based on the characteristics of the music. The same experience has also been done on 500 images. Finally, a music-image exploring system is designed to offer the music and the image tailored to the user's emotion after the user enters his emotion and its intensity to the system.

In [14], a music recommender system is provided to suggest music to teens on the cellphone, which uses devices and equipment such as a microphone (for analyzing user tone), camera (for user's face analysis), and 3D accelerometer and gravity sensors of smartphones (to analyze user's behavior). The system consists of two parts. A part is installed on smartphones to collect the sensors' data and examine the user's emotion and the features of different music tracks. The second part is implemented on cloud space, analyzes the big data, and adapts the music to the individuals based on collaborative filtering.

In 2018, an interactive recommender system was being developed that explicitly asks the users' current emotions and then suggests them music based on their present emotions. The system uses a music knowledge graph (which is based on music, album, singer, and genre of music) and the user's knowledge graph (which is based on gender characteristics, user's emotions, and preferences). The results have shown that this system presents real-time and rich suggestions [45].

In [16], another system was developed. The characteristics of the music were analyzed. Accordingly, the music signals were classified into four groups: sad, happy, surprise, and anger. Then features are extracted from the user's image captured using a webcam. Only features of eyes and mouth are considered. With the SVM (Support Vector Machine) classifier's help, the users' emotions are classified into one of the categories of sad, happy, surprise, and anger. Then the music is suggested to the user according to his emotion.

In [22], an intelligent system for music selection is presented that analyzes user voice signals. It extracts 12 features. These features are classified using GMM (Gaussian Mixture Model) and SVM, and the emotional states are categorized into five emotions of anger, anxiety, boredom, happiness, and sadness. Then, suitable music is selected for the user to make him happy.

In [1], the correlation between the user data (e.g., location, time of the day, music listening history, and emotion) and the music is calculated using the deep convolutional neural networks, and the weighted feature approaches. Then, the music's user ratings were calculated based on the term-frequency and inverse document frequency (TF-IDF) approach. Finally, music is recommended to users based on the estimated user ratings.

In addition to the studies reviewed above, other studies used the users' interaction with mouse and/or keyboard for different purposes and applications. For example, in [12, 37], mouse clicks/keystrokes were analyzed to recognize the user's interest in different online store items.

As reviewed in this section, previous emotion-aware music recommenders estimate the user's emotion based on the input data. Then the estimated emotion is used to recommend music to the user. To the best of the authors' knowledge, none of them applied user's interaction data with the keyboard and mouse to estimate the user's emotion.

3 The proposed approach

The studies reviewed in Section 2 reveal that the estimation of the user's emotion is not accurate, depending on the algorithms and the data used. On the other hand, to recommend music following the user's emotion, it is also tried to determine which music is appropriate to that emotion. Even if the user's emotion has been accurately detected, finding music under any emotion is erroneous, let alone when the users' estimated emotion is inaccurate and erroneous itself. This study suggests that instead of mapping the implicit input data to the user's emotion and then the user's emotion to the proper music, the input data is directly mapped to appropriate music. The architecture of the proposed EMA-based EMotion-Aware (EMA-EMA) music recommender system is shown in Fig. 1.

The implicit input data in this study consists of the keystrokes and user mouse click patterns. Collecting these data does not require any specific hardware or sensor, and various studies [8, 20, 21, 23, 24, 35] have also shown that these data can reflect the user's current emotion. However, to the best of our knowledge, the user's keystrokes and mouse click patterns have not been used so far to recommend music.

The EMA-EMA method is based on collaborative filtering. In collaborative filtering, the users' similarity is used to recommend items. In our approach, the users' similarity is measured based on how they interact with the mouse and keyboard. So, intermittently and at intervals of T-second, four user's interaction features with the keyboard and mouse are implicitly extracted. These four features for the user u at i^{th} time interval include:

- $I_{u,i}^{1}$: the number of keystrokes,
- $I_{u,i}^{2}$: the average duration that each key remains pressed,

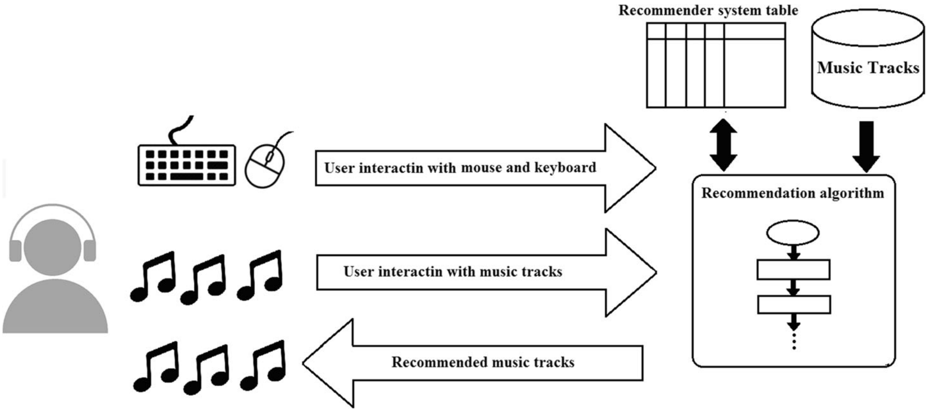

Fig. 1 The proposed system architecture

- $I^3_{u,i}$: the number of mouse clicks,
- $I^4_{u,i}$: the average duration that each mouse button remains held down.

The reason for choosing these four features is discussed in Section 4.1. These features are normalized and arranged in a vector, which is called the interaction vector of the user u at i^{th} time interval ($IV_{u,\ i}$), as follows:

$$IV_{u,i} = \left[I^1_{u,i}\ \ I^2_{u,i}\ I^3_{u,i}\ \ I^4_{u,i} \right] \tag{1}$$

This vector is the basis for two main tasks of the EMA-EMA method: data collection and recommendation. The rest of this section introduces these two tasks in detail. Figure 3 shows the flowchart of these tasks.

3.1 Data collection

Human's emotions change over time. New events also have an impact on this change. The user's reactions and interactions mostly represent his/her recent emotions, and they are less affected by his/her older emotions. In this research, the exponential moving average (EMA) (H. [5]; S. [6, 42]) is used to show this fact. EMA is a moving average that applies more weight to the most recent data.

EMA of interaction vectors of the user u is calculated for any continuous-time slot and is considered an indicator of the emotional state of the user u at that interval. To do this, the following equation is used:

$$EMA_{u,i} = \begin{cases} IV_{u,1} & i = 1 \\ \alpha.IV_{u,i} + (1-\alpha).EMA_{u,i-1} & i > 1 \end{cases} \tag{2}$$

α is a value between 0 and 1, called the smoothing constant. $EMA_{u,\ i}$ and $EMA_{u,\ i-1}$ are the EMA of user u interaction vectors at time intervals of i and $i-1$, respectively. When user u starts a session on the website, i (i.e., the time interval counter) is initialized by 1. After each T-second time interval, i is increased by 1 ($i = i + 1$).

In the calculation of $EMA_{u,\ i}$, the weight of each old interaction vector of the user u exponentially decreases. Passing the time, i is increased, and the EMA is updated. When the user is not working with the mouse and keyboard, no new interaction vector is obtained. Therefore, i is not increased and the EMA is not updated.

Since $EMA_{u,\ i}$ is obtained from averaging the interaction vectors, it is a vector of four elements:

$$EMA_{u,i} = \left[E^1_{u,i}\ \ E^2_{u,i}\ E^3_{u,i}\ \ E^4_{u,i}\right] \tag{3}$$

Each element corresponds to the EMA of one of the four elements of the interaction vector.

The contribution of each interaction vector in calculating $EMA_{u,\ i}$ becomes less at each consecutive time interval. Thus, the latest interaction vectors of user u have a higher impact on $EMA_{u,\ i}$.

In EMA-EMA method, $EMA_{u,\ i}$ is considered as the emotional state of the user u at time interval i which has been reflected in his interaction with the mouse and keyboard. As mentioned before, to reduce the error, the emotion is not labeled but the EMA vector is considered directly as the users' emotional state.

When the user u selects music M on the website at the i^{th} time interval, his latest EMA vector is considered as the description of his emotional state during selecting that music. Hereafter, this vector is named $EMA^M_{u,i}$. Similar to $EMA_{u,\ i}$, $EMA^M_{u,i}$ also has four elements: $E^{1,M}_{u,i}$, $E^{2,M}_{u,i}$, $E^{3,M}_{u,i}$, and $E^{4,M}_{u,i}$.

Furthermore, the user's interest in that music is calculated as a value in the range of 1 to 5 as follows: if the user clicks on the music and goes to the music details page, it gets 1 point. If he listens to the music online, one more point is added, and if the music is downloaded, three more points will be added. Eventually, the summation of these points is considered the estimated rating of the user u to music M ($Rating\ (u, M)$). This rating is a value between 1 and 5. The user's EMA vector ($EMA^M_{u,i}$), along with the music id (M), and the estimated rating of the user to that music ($Rating\ (u, M)$) constitute a record. This record is named the user-music record and is sent to the server and appended to the server's RS table. The structure of each record in the recommender system Table (RS table) is shown in Fig. 2.

Data is collected during the interaction of users with the website, and the recommender system Table (RS table) is gradually filled on the server. The RS table is the primary source for preparing recommendations to the users. As the number of data records in the RS table increases, the recommendations will be more accurate. The data collection process is a part of Fig. 3.

3.2 Recommendation

To recommend music to the user v at a specific time, the proper music for his current emotion should be chosen. For this purpose, the latest EMA vector of user v (called here as $EMA_{v,\ current}$) sent to the server is considered the indicator of his current emotional state. $EMA_{v,\ current}$ has the following elements:

Record id	$EMA^M_{u,i}$: EMA vector of user u in i^{th} time interval (when he was selecting music M)				M	$Rating\ (u, M)$ (Estimated rating of user u to the music M)
	$E^{1,M}_{u,i}$	$E^{2,M}_{u,i}$	$E^{3,M}_{u,i}$	$E^{4,M}_{u,i}$		

Fig. 2 The record structure in RS table

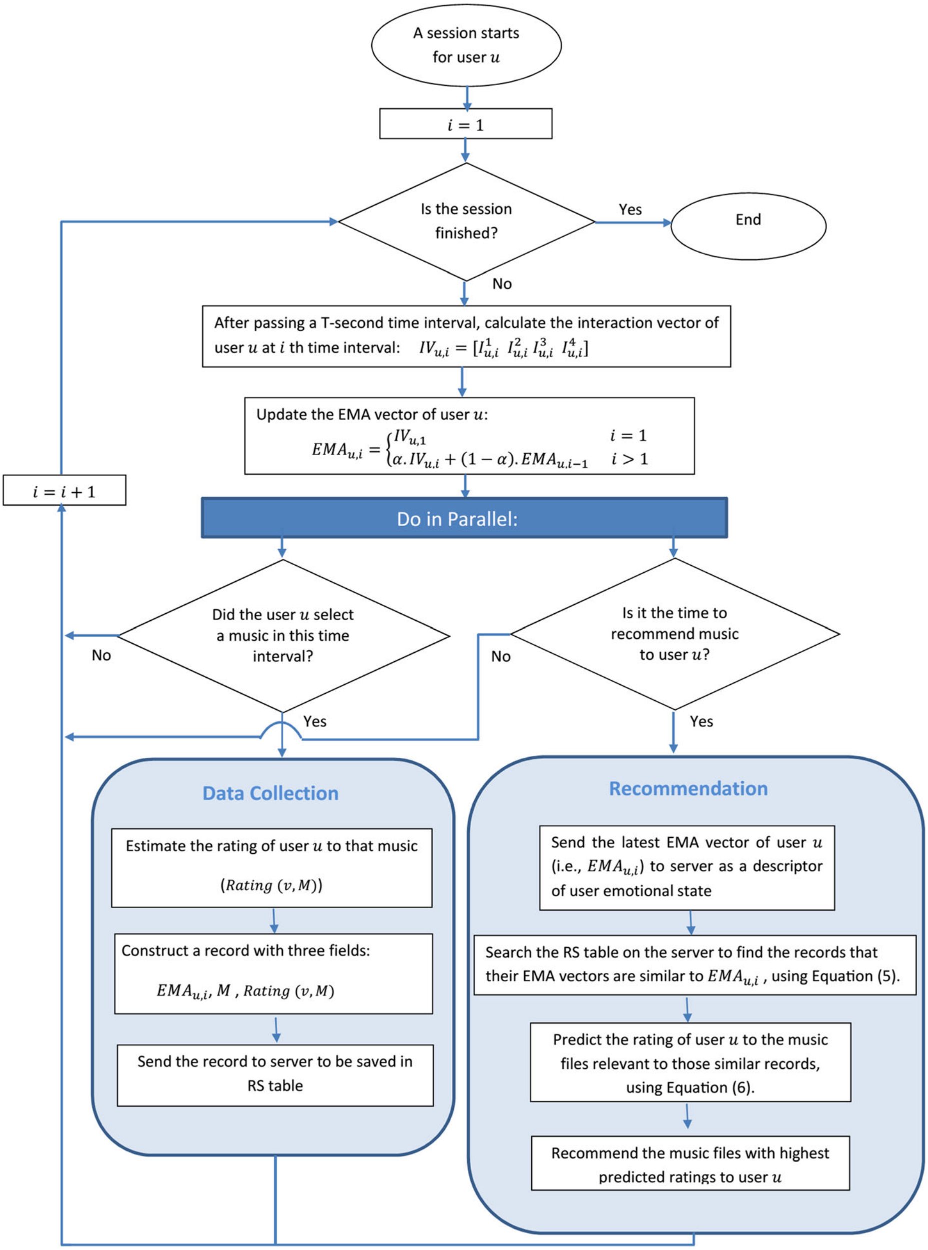

Fig. 3 The flowchart of the EMA-EMA method

$$EMA_{v,current} = \left[E^1_{v,current}\ \ E^2_{v,current}\ E^3_{v,current}\ \ E^4_{v,current} \right] \tag{4}$$

Then, the similarity of $EMA_{v,\ current}$ and EMA vector of every record in the RS table (e.g., $EMA^M_{u,i}$) is calculated based on collaborative filtering using adjusted cosine measure (Sarwar, Karypis, Konstan, & Riedl, 2001):

$$sim\left(EMA_{v,current}, EMA_{u,i}^{M}\right) = \frac{\sum_{k=1}^{4}\left(E_{v,current}^{k} - \overline{E_{v}^{k}}\right)\left(E_{u,i}^{k,M} - \overline{E_{u}^{k}}\right)}{\sqrt{\sum_{k=1}^{4}\left(E_{v,current}^{k} - \overline{E_{v}^{k}}\right)^{2}}\sqrt{\sum_{k=1}^{4}\left(E_{u,i}^{k,M} - \overline{E_{u}^{k}}\right)^{2}}} \quad (5)$$

Where, $E_{v,current}^{k}$, is the k^{th} element of latest EMA vector of user v. $E_{u,i}^{k,M}$ is the k^{th} element of EMA vector of user u (extracted from the RS table) that reflects the user u's emotion at the moment of choosing music M. $\overline{E_{v}^{k}}$ and $\overline{E_{u}^{k}}$ are respectively the averages of k^{th} element of EMA vectors of user v and user u in RS table. $sim\left(EMA_{v,current}, EMA_{u,i}^{M}\right)$ is the cosine similarity of two EMA vectors.

Afterward, the n EMA vectors that are the most similar ones are distinguished from the RS table. For each music M related to these records, the rating of the user v to music M is predicted using the weighted sum method [13, 32, 39] through Eq. (6):

$$Pred(u, M) = \frac{\sum_{u \,\epsilon\, usersRatedto(M)} sim\left(EMA_{v,current}, EMA_{u,i}^{M}\right) \times Rating\ (u, M)}{\sum_{u \,\epsilon\, usersRatedto(M)} sim\left(EMA_{v,current}, EMA_{u,i}^{M}\right)} \quad (6)$$

Where *usersRatedto*(M) is the set of users who have interacted with music M, and their EMA records exist in the RS table. *Rating* (u, M) represents the estimated rating of the user u to the music M and $sim\left(EMA_{v,current}, EMA_{u,i}^{M}\right)$ is calculated through Eq. (5).

After that, N music files with the highest predicted ratings are selected and recommended to user v in a descending order based on their predicted ratings. The flowchart of the recommendation process is presented in Fig. 3.

4 Evaluation

As mentioned in Section 3, four features are extracted from user's interaction with the keyboard and mouse. These four features, which have been selected among various mouse and keyboard features, are used as the basis for creating $IV_{u,\ i}$ and $EMA_{u,\ i}$ vectors. In this section, at first, the feature selection process is discussed. Then the proposed method is evaluated based on various criteria.

4.1 Evaluation of features

At first, a website is implemented to present music to the users, and 220 different music tracks are uploaded to the website database. For the feature selection process, an experiment was conducted; 88 users were asked to visit the website and interact with it in three separate sessions. During the user's interaction, his interaction features were extracted and his interaction vectors were recorded in 60 s intervals. But instead of four features mentioned in Section 3, each interaction vector consisted of 11 features of "the number of keystrokes," "the average duration that each key remains pressed", "the mouse velocity," "the mouse traveled distance," "the number of mouse hovers," "the mouse hover total time," "the number of mouse dwells," "the mouse dwells total time," "the number of mouse clicks," "the average duration that each mouse button remains held down," and "the scroll speed".

After 5 min, the user was asked to choose his current emotion among the emotions of anger, anxiety, boredom, happiness, surprise, and sadness, plus neutral expressions. Then, the EMA of the user's interaction vectors was calculated. The obtained vector, together with the user's current emotion, was sent to the server. After removing the outliers and incomplete data, 248 data records were obtained and stored on the server. After data collection, each feature's correlation coefficient with each emotion (plus neutral expression) was calculated based on the Pearson correlation coefficient. Then, the average absolute value of the correlations (AAC) was calculated for each feature. From this experiment, it was concluded that among 11 features, seven features are relatively correlated to user's emotion, with the AAC of higher than 0.3; therefore, they can be used to describe the user's emotion. However, recording all of these features during the user's interaction slows down the system. For this reason, only a few features that are most strongly correlated to the user's emotion were chosen. The four features of "the number of keystrokes," "the average duration that each key remains pressed," "the number of mouse clicks," and "the average duration that each mouse button remains held down," are the most correlated ones, with the AAC of higher than 0.4.

4.2 Evaluation of results

According to [33], evaluating the recommender systems is one of their most critical challenges. The related fields of machine learning and information retrieval introduce some of the evaluation metrics of recommender systems. In this research, we have used the accuracy and the RMSE[1] from the accuracy-related measures and the precision from the standard information retrieval measures.

The so-called quantitative measures make the evaluation results reproducible. However, there is still another essential factor, the user experience, in evaluating the recommender systems. The user's willingness through receiving useful and graceful recommendations is another primary objective of these systems. In this research, we have used the user satisfaction and emotion relevance from the user-centric measures.

The EMA-EMA recommender system is embedded in the implemented website, which presents 220 different music tracks are to users. Then, 174 users, from 14 to 52 years old, 96 males and 78 females, are asked to interact with the website. While interacting with the music website, the user engaged in normal activities such as browsing the web pages; entering the desired music category, singer name, track name, or relevant keywords in the website search field to find them; choosing his favorite category among the various music categories; browsing through the music tracks in each category; selecting his desired music tracks by clicking on them, and playing his favorite music tracks.

The EMA of users' interaction vectors is sent to the server and saved in the RS table in 60 s intervals. Entirely, 7870 data records are collected from these interactions and saved as the structure showed in Fig. 2. In this stage, the recommender tool was inactive. After cleaning the collected data from outliers, 7120 data records remained for system analysis and evaluation. The training set and the test set respectively consist of 5000 and 2120 records.

[1] Root Mean Square Error

4.2.1 RMSE measure

First, the ratings related to test data were predicted using the EMA-EMA method. The prediction experiments were repeated for various values of α (0.6, 0.7, 0.8, and 0.9) and T (120, 180, 240, and 300) to investigate these parameters' impact on the prediction error.

Whatever the α is greater, it means the most recent user's interactions impact the user EMA vector calculation, and the effect of older interactions is dampening more quickly.

The T parameter also influences the records used to predict the music ratings. Furthermore, it is influential in selecting the set of records, which is the basis for describing the user's emotion for musical advice. Remembering that the user's interactions were registered in the RS table in T seconds' intervals $T = 120$ means that the training data records that are 180 s apart should be used to make predictions The records between them should be discarded.

Then, the prediction error for various values of T and α was calculated using the RMSE measure (Avazpour, Pitakrat, Grunske, & Grundy, 2014). Equation (7) calculates this measure. In this equation, m is the number of predicted ratings, p_i is the i^{th} predicted rating, and r_i is the i^{th} real rating.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} (p_i - r_i)}{m}} \tag{7}$$

Figure 4 shows the RMSE for different T values. As shown, the lowest error value is obtained by $T = 240$ and $\alpha = 0.8$.

4.2.2 Precision measure

With a set of 7120 records collected in the RS table, the recommender system was activated. Thirty-three users were asked to work with the system, and their feedbacks about the suggested music were recorded. In these experiments, the parameters were set to the optimal values obtained in previous evaluations ($T = 240$ and $\alpha = 0.8$); therefore, the user's interaction vectors were sent to the server every 240 s. Afterward, precision was calculated according to Eq. (8)

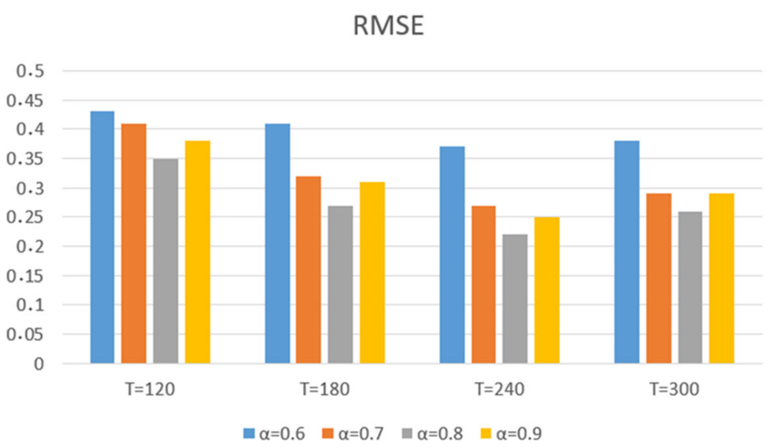

Fig. 4 RMSE for different values of α and T

[10, 13], where $N_{relevant}$ is the number of relevant recommended music tracks and N_{total} is the total number of music tracks recommended to users.

$$precision = \frac{N_{relevant}}{N_{total}} \tag{8}$$

The precision value indicates users are targeting how many of the suggested music tracks. To investigate the effect of the parameter S (recommendation list size) on precision, the precision was calculated for different values. Figure 5 shows the results. As shown, the best precision value is obtained for $S = 5$.

4.2.3 Evaluating the user-centric measures

A questionnaire was used to check how much the recommended music is appropriate for the user's emotion. After each suggestion, the user was asked about the relevance of the recommended music to his emotion. In these experiments, the parameters were set to their optimal values obtained in evaluations in Sections 4.1 and 4.2 ($N = 5$, $T = 240$, $\alpha = 0.8$). Table 1 summarizes the results compared to the results of the emotion-aware music recommender proposed by Lukose and Upadhya [22].

As shown in Table 1, our proposed method is better than the Lukose and Upadhya [22] for the emotions of Happiness, Boredom, and Anger. In [22], the user's emotion is first estimated, and then the appropriate music is recommended. The precision in [22] indicates the precision of user emotion estimation, and so the precision of proper music selection is assumed to be 100%. In other words, Lukose and Upadhya defined the recommender system precision as user's emotion detection. This is while the proper music selection is also associated with some error. So the actual precision of the music recommendation in [22] is less than the reported amount.

After interacting with the system, users were asked to rate their overall satisfaction with the system's suggestions. Figure 6 shows the results.

4.2.4 Evaluation against former methods

In addition to the evaluations conducted, the proposed recommender system is compared with baseline methods and some state-of-the-art music recommenders in terms of average precision and RMSE. The baseline methods include the user-based collaborative filtering recommender

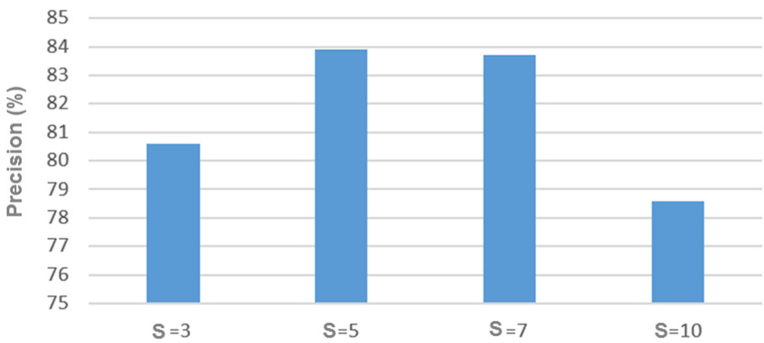

Fig. 5 Precision for different values of S

Table 1 The relevance degree of recommended music and user's emotion in the proposed method (EMA-EMA) compared to a former emotion-aware music recommender

	Anger	Anxiety	Boredom	Happiness	Sadness
Method of Lukose & Upadhya [22]	89%	83%	75%	69%	93%
EMA-EMA method	90%	79%	78%	94%	92%

(UBCF) [15, 34, 38], and the item-based collaborative filtering recommender (IBCF) [15, 34, 38], with adjusted cosine similarity measure and a recommendation list size of 5. The state-of-the-art methods include a context-aware recommender which recommends music based on the user's situational context [25, 26], a context-aware recommender which exploits information about the current situation and musical preferences of users [27], an emotion-aware music recommender which works based on the emotions explicitly stated by the users [1], and a session-sensitive music recommender which uses the clustered music files [30]. The comparison results are depicted in Figs. 7 and 8.

As shown, the EMA-EMA recommender system can reduce the error, increase the recommendation accuracy, and adjust the suggested music with the user's emotions. Furthermore, it tries to overcome the cold start problem [33] in music recommendation. This feature is since this method only uses the user's interactions in the current session to recommend him the music. So, the method does not require historical user ratings to identify his favorite music tracks. As a guideline for future works, the EMA-EMA method can also be used along with other music recommender systems to alleviate the cold start problem.

In the end, the proposed method, unlike many other emotion-aware recommenders [9, 14, 16, 22], does not require additional hardware (such as sensors, camera, or microphone) to identify the emotion; It collects data only based on user's regular interactions.

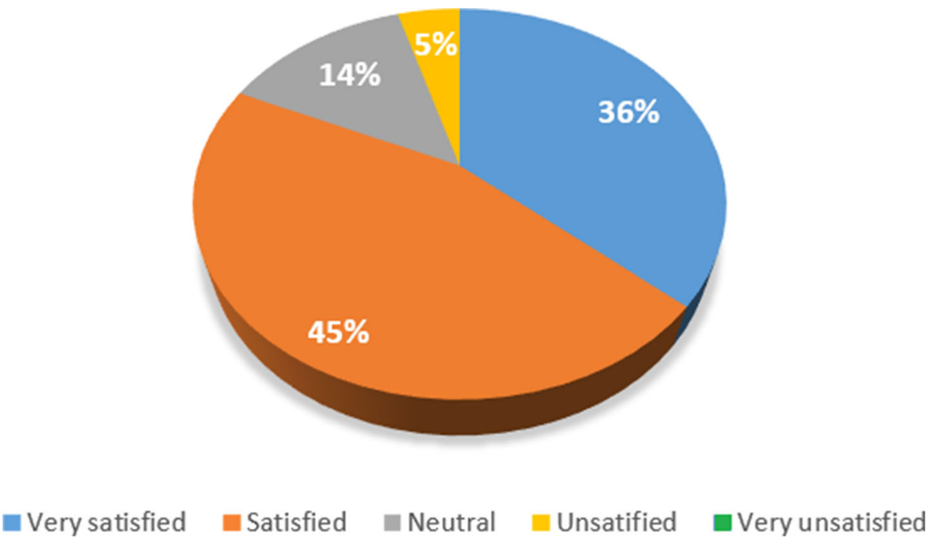

Fig. 6 Users' overall satisfaction from the EMA-EMA recommender system (0% very unsatisfied users)

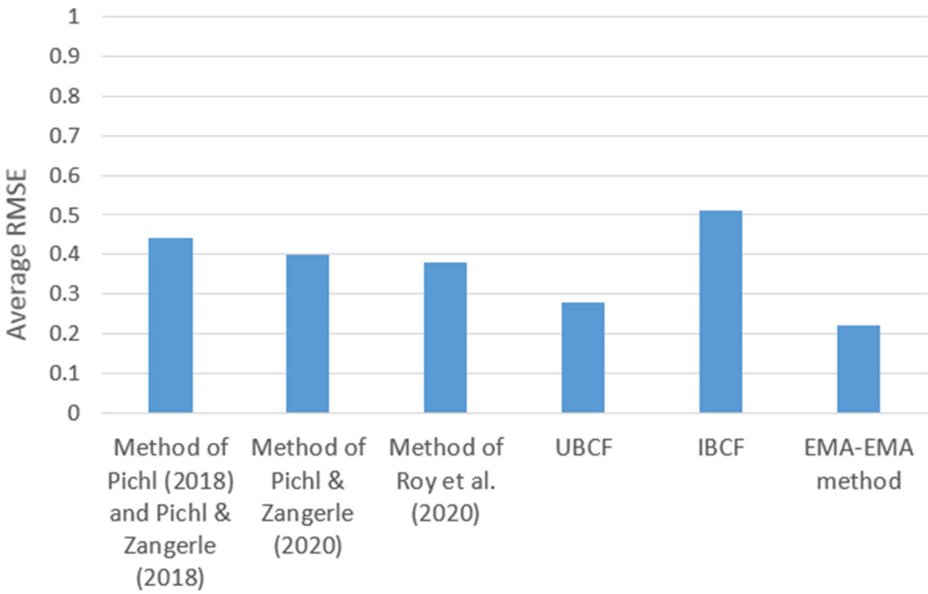

Fig. 7 The average RMSE of the proposed method (EMA-EMA) compared to former music recommenders

5 Conclusion

The user's interaction with the keyboard and mouse can reflect his emotions. In this study, a collaborative filtering recommender system has been proposed for recommending music to the user, which suggests music based on the user's keystrokes and mouse clicks patterns. Unlike the previous studies, the users' emotions are not labeled in this method. Still, the user's interaction patterns are directly mapped to the user's favorite music to reduce the error and

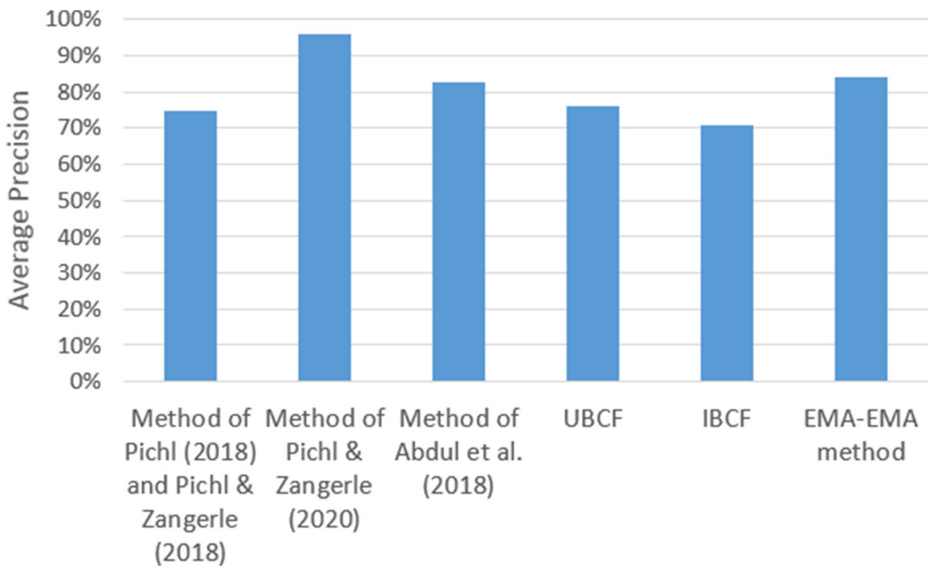

Fig. 8 The average precision of the proposed method (EMA-EMA) compared to former music recommenders

increase the accuracy. The evaluations indicate that the proposed method has high accuracy in finding favorite music for users. Also, it is shown that the performance of the proposed method is higher for happiness emotion.

In future studies, the accuracy of the recommendations for other emotions will be improved. For example, in addition to the four features extracted from the users' interaction with the mouse and the keyboard, more features (e.g., mouse movement, etc.) can be used to increase the system accuracy. Other user features such as age, gender, personality traits, and users' favorite music styles can also be considered in recommending music to the user. The effect of the proposed method in alleviating the cold start problem can be investigated as well.

References

1. Abdul A, Chen J, Liao H-Y, Chang S-H (2018) An emotion-aware personalized music recommendation system using a convolutional neural networks approach. Appl Sci 8:1103
2. Aljanaki A, Wiering F, Veltkamp RC (2016) Studying emotion induced by music through a crowdsourcing game. Inform Process Manag 52(1):115–128
3. Aracena C, Basterrech S, Snáel V, Velásquez J (2015) Neural networks for emotion recognition based on eye tracking data. Paper presented at the 2015 IEEE International Conference on Systems, Man, and Cybernetics
4. Ayata D, Yaslan Y, Kamasak ME (2018) Emotion based music recommendation system using wearable physiological sensors. IEEE Trans Consumer Electron 64(2):196–203
5. Burns H, Burns S (2015) Moving Averages 101: Incredible Signals That Will Make You Money in the Stock Market: CreateSpace Independent Publishing Platform
6. Burns S, Burns H (2017) 5 Moving Average Signals That Beat Buy and Hold: Backtested Stock Market Signals: CreateSpace Independent Publishing Platform
7. Deng S, Wang D, Li X, Xu G (2015) Exploring user emotion in microblogs for music recommendation. Expert Syst Appl 42(23):9284–9293
8. Estrada J, Buhia J, Guevarra A, Forcado MR (2017) Keyboard and mouse: tools in identifying emotions during computer activities. Paper presented at the International Conference on Big Data Technologies and Applications
9. Gavrilescu M (2015) Recognizing emotions from videos by studying facial expressions, body postures and hand gestures. Paper presented at the 2015 23rd Telecommunications Forum Telfor (TELFOR)
10. Geng B, Li L, Jiao L, Gong M, Cai Q, Wu Y (2015) NNIA-RS: A multi-objective optimization based recommender system. Physica A: Statistic Mech Appl 424:383–397
11. Goldberg D, Nichols D, Oki BM, Terry D (1992) Using collaborative filtering to weave an information tapestry. Commun ACM 35:61–70
12. Gu YH, Yoo SJ, Piao Z, No J, Jiang Z, Yin H (2016) A smart-device news recommendation technology based on the user click behavior. Paper presented at the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory
13. Hamedani EM, Kaedi M (2019) Recommending the long tail items through personalized diversification. Knowledge-Based Syst 164:348–357
14. Hu X, Bai K, Cheng J, Deng J-q, Guo Y, Hu B, . . . Wang F (2017) MeDJ: multidimensional emotion-aware music delivery for adolescent. Paper presented at the Proceedings of the 26th International Conference on World Wide Web Companion
15. Jannach D, Zanker M, Felfernig A, Friedrich G (2011) Recommender systems, an introduction. Hardback (November 2010)
16. Kabani H, Khan S, Khan O, Tadvi S (2015) Emotion based music player. Int J Eng Res Gen Sci *3*(1):2091–2730
17. Kawakami A, Furukawa K, Katahira K, Kamiyama K, Okanoya K (2012) Relations between musical structures and perceived and felt emotions. Music Perception: Interdisciplin J 30(4):407–417
18. Kawakami A, Furukawa K, Katahira K, Okanoya K (2013) Sad music induces pleasant emotion. Frontiers Psychol 4:311
19. Khan IA, Brinkman WP, Hierons R (2013) Towards estimating computer users' mood from interaction behaviour with keyboard and mouse. Front Comput Sci 7(6):943–954

20. Khanna P, Sasikumar M (2010) Recognising emotions from keyboard stroke pattern. Int J Comput Appl 11(9):1–5
21. KM AK, Kiran B, Shreyas B, Victor SJ (2015) A multimodal approach to detect user's emotion. Procedia Comput Sci 70:296–303
22. Lukose S, Upadhya SS (2017) Music player based on emotion recognition of voice signals. Paper presented at the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)
23. Nahin ANH, Alam JM, Mahmud H, Hasan K (2014) Identifying emotion by keystroke dynamics and text pattern analysis. Behav Information Technol 33(9):987–996
24. Pentel A (2017) Emotions and user interactions with Keyboard and Mouse. Paper presented at the 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)
25. Pichl M (2018) Multi-Context-Aware Recommender Systems: A Study on Music Rfecommendation. University of Innsbruck
26. Pichl M, Zangerle E (2018) Latent feature combination for multi-context music recommendation. Paper presented at the 2018 International Conference on Content-Based Multimedia Indexing (CBMI)
27. Pichl M, Zangerle E (2020) User models for multi-context-aware music recommendation. Multimed Tools Appl. https://doi.org/10.1007/s11042-020-09890-7
28. Poirson E, Da Cunha C (2018) A recommender approach based on customer emotions. Expert Syst Appl
29. Robinson J, Hatten RS (2012) Emotions in music. Music Theory Spectrum 34(2):71–106
30. Roy S, Biswas M, De D (2020) iMusic: A session-sensitive clustered classical music recommender system using contextual representation learning. Multimed Tools Appl 79:24119–24155
31. Salmeron-Majadas S, Santos OC, Boticario JG (2014) An evaluation of mouse and keyboard interaction indicators towards non-intrusive and low cost affective modeling in an educational context. Procedia Comput Sci 35:691–700
32. Sarwar BM, Karypis G, Konstan JA, Riedl J (2001) Item-based collaborative filtering recommendation algorithms. Www 1:285–295
33. Schedl M, Zamani H, Chen C-W, Deldjoo Y, Elahi M (2018) Current challenges and visions in music recommender systems research. Int J Multimed Information Retriev 7(2):95–116
34. Shakirova E (2017) Collaborative filtering for music recommender system. Paper presented at the 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)
35. Shikder R, Rahaman S, Afroze F, Al Islam AA (2017) Keystroke/mouse usage based emotion detection and user identification. Paper presented at the 2017 International Conference on Networking, Systems and Security (NSysS)
36. Song K-T, Cervantes C (2017) Music playing system and music playing method based on speech emotion recognition. In: Google Patents
37. Sulikowski P, Zdziebko T, Turzyński D, Kańtoch E (2018) Human-website interaction monitoring in recommender systems. Procedia Comput Sci 126:1587–1596
38. Sunitha M, Adilakshmi T (2018) Music recommendation system with user-based and item-based collaborative filtering technique. In Networking Communication and Data Knowledge Engineering (pp 267–278): Springer
39. Tso K, Schmidt-Thieme L (2006) Attribute-aware collaborative filtering. In From data and information analysis to knowledge engineering (pp 614–621): Springer
40. Xing B, Zhang K, Sun S, Zhang L, Gao Z, Wang J, Chen S (2015) Emotion-driven Chinese folk music-image retrieval based on DE-SVM. Neurocomput 148:619–627
41. Yusefi Hafshejani Z, Kaedi M, Fatemi A (2018) Improving sparsity and new user problems in collaborative filtering by clustering the personality factors. Electron Commerce Res 18(4):813–836
42. Zakamulin V (2017) Market Timing with Moving Averages: The Anatomy and Performance of Trading Rules: Springer
43. Zentner M, Grandjean D, Scherer KR (2008) Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8(4):494
44. Zhang R, Liu Q-d, Chun-Gui, Wei J-X, Huiyi-Ma (2014) Collaborative filtering for recommender systems. Paper presented at the Second International Conference on Advanced Cloud and Big Data (CBD), Huangshan
45. Zhou C, Jin Y, Zhang K, Li JYS, Wang X (2018) MusicRoBot: Towards Conversational Context-Aware Music Recommender System. In: Paper presented at the International Conference on Database Systems for Advanced Applications. Gold Coast, Australia