From Personalized to Hierarchically Structured Classifiers for Retrieving Music by Mood

Amanda Cohen Mostafavi[1], Zbigniew W. Raś[1,2], and Alicja Wieczorkowska[3]

[1] University of North Carolina, Dept. of Comp. Science, Charlotte NC 28223, USA,
[2] Warsaw University of Technology, Institute of Comp. Science, 00-665 Warsaw, Poland,
[3] Polish-Japanese Institute of Information Technology, 02-008 Warsaw, Poland

Abstract. With the increased amount of music that is available to the average user, either online or through their own collection, there is a need to develop new ways to organize and retrieve music. We propose a system by which we develop a set of personalized emotion classifiers, one for each emotion in a set of 16 and a set unique to each user. We train a set of emotion classifiers using feature data extracted from audio which has been tagged with a set of emotions by volunteers. We then develop SVM, kNN, Random Forest, and C4.5 tree based classifiers for each emotion and determine the best classification algorithm. We then compare our personalized emotion classifiers to a set of non-personalized classifiers. Finally, we present a method for efficiently developing personalized classifiers based on hierarchical clustering.

Keywords: Music information retrieval, classification, clustering

1 Introduction

With the average size of a person's digital music collection expanding into the hundreds and thousands, there is a need for creative and efficient ways to search for and index songs. This problem shows up in several sub-areas of Music Information Retrieval (MIR) such as genre classification, automatic artist identification, and instrument detection. Here we focus on indexing music by emotion, as in how the song makes the listener feel. This way the user could select songs that make him/her happy, sad, excited, depressed, or angry depending on what mood the listener is in (or wishes to be in). However, the way a song makes someone feel, or the emotions he associates with the music, varies from person to person for a variety of reasons ranging from personality and taste to upbringing and the music the listener was exposed to growing up. This means that any sort of effective emotion indexing system must be personal and/or adaptive to the user. This is so far a mostly unexplored area of MIR research, as many researchers that attempt to personalize their music emotion recognition systems do so from the perspective of finding how likely the song is to be tagged with certain emotions rather than finding a way to create a system that can be personalized.

We present a system through which we can build and train personalized user classifiers, which are unique for individual users. We built these classifiers based on user data accumulated through an online survey and music data collected via a feature extraction toolkit called MIRToolbox [1]. We then use four classification algorithms to determine the best algorithm for this data: support vector machines (SVM), k-nearest neighbors (kNN), random forest, and C4.5 trees. Based on the best algorithm, we build a broad non-personalized classifier to compare the personalized classifiers to. Finally, we present a more efficient method for building personalized classifiers with comparable accuracy and consistency. This method uses agglomerative clustering to group users based on background and mood states, then builds classifiers for each of these individual groups.

2 Related Work

There is some discussion as to the possible usefulness of creating a personalized music recommender system. On the one hand, [2] demonstrated that emotion in music is not so subjective that it cannot be modeled; on the other hand, the results from researchers who attempt to build personalized music emotion recommendation systems are very promising, suggesting personalization is at least a way to improve emotion classification accuracy. Yang et al. in [3] was one of the earliest to study the relationship between music emotion recognition and personality. The authors looked at users demographic information, musical experience, and user scores on the Big Five personality test to determine possible relationships and build their system. Classifiers were built based on support vector regression, and test regressors trained on general data and personalized data. The results were that the personalized regressors outperformed the general regressors in terms of improving accuracy, first spotlighting the problem of trying to create personalized recommendation systems for music and mood based on general groups. However, there has been continued work on collaborative filtering, as well as hybridizing personalized and group based preferences. Lu and Tseng in [4] proposed a system that combined emotion-based, content-based, and collaborative-based recommendation and achieved an overall accuracy of 90%. In [5], the authors first proposed the idea of using clustering to predict emotions for a group of users. The results were good, but some improvement was needed. The users were clustered into only two groups based on their answers to a set of questions, and the prediction was based on MIDI files rather than real audio. In this work, we propose creating personalized classifiers first (trained on real audio data), clustering users, creating representative classifiers for each cluster, and then allowing the classifiers to be altered based on user behavior.

Each of the possible classification algorithms has been used commonly in previous MIR research, with varying results. kNN had been evaluated previously in [6] and [7] for genre classification. [6] achieved a 90%-98% classification accuracy by combining kNN and Neural Network classifiers and applying them to MIDI files using a 2-level genre hierarchical system. On the other hand, [7] only achieved a 61% accuracy at the highest using real audio and k of 3. Random

Forest was used principally in [8] for instrument classification in noisy audio. Sounds were created with one primary instrument and artificial noise of varying levels added in incrementally. The authors found that the percentage error was overall much lower than previous work done with SVM classifiers on the same sounds up until the noise level in the audio reached 50%. They also observed that Random Forest could indicate the importance of certain attributes in the classification based on the structure of the resulting trees and the attributes used in the splitting. SVM classification is one of the more common algorithms used in MIR for a variety of tasks, such as [9] for mood classification, [10] for artist identification (compared with kNN and Gaussian Mixture Models), and [11] for mood tracking. It has also been evaluated beside other classifiers in [7] for genre identification. These evaluations have shown the SVM classifier to be remarkably accurate, particularly in predicting mood. Regarding C4.5 decision trees, the authors in [12] in a comparison of the J48 implementation of C4.5 to Bayesian network, logical regression, and logically weighted learning classification models for musical instrument classification found that J48 was almost universally the most accurate classifier (regardless of the features used to train the classifier). The classification of musical instrument families (specifically string or woodwind) using J48 ranged in accuracy from 90-92%, and the classification of actual instruments ranged from 60-75% for woodwinds and 60-67% for strings.

3 Data Composition and Collection

3.1 Music Data

Music data was collected from 100 audio clips 25-30 seconds in length culled from one of the author's personal music collection. These clips were split into 12-15 segments (depending on the length of the original clip) of roughly 0.8 seconds in order to allow for changes in annotation as the clip progresses, resulting in a total of 1440 clips. These clips originated from several film and video game sound tracks in order to achieve a similar effect to the dataset composed in [13] (namely a set composed of songs that are less known and more emotionally evocative). As such the music was mainly instrumental with few if any intelligible vocals. The MIRToolbox [1] collection was then used to extract musical features. MIRToolbox is a set of functions developed for use in MATLAB which uses, among others, MATLAB's Signal Processing toolbox. It reads .wav files at a sample rate of 44100 Hz. The following features were extracted using this toolbox.

- **Rhythmic Features** (fluctuation peak, fluctuation centroid, frame-based tempo estimation, autocorrelation, attack time, attack slope): Rhythmic features refer to the set of audio features that describe a song's rhythm and tempo, or how fast the song is, although features such as attack time and attack slope are better indicators of the rhythmic style of the audio rather than pure tempo estimation. Fluctuation based features are based on calculations to a fluctuation summary (calculated from the estimated spectrum with a Bark-band redistribution), while the rest of the rhythmic features are based on the calculation of an onset detection curve (which shows the rhythmic pulses in the song in the form of amplitude peaks for each frame).

- **Timbral Features** (spectral centroid, spectral spread, coefficient of spectral skewness, kurtosis, spectral flux, spectral flatness, irregularity, Mel-Frequency Cepstral Coefficients (MFCC) features, zero crossings, brightness): Timbral features describe a piece's sound quality, or the sonic texture of a piece of audio. The timbre of a song can change based on instrument composition as well as play style. Most of these features are derived from analysis of the audio spectrum, a decomposition of an audio signal. MFCC features are based on analysis of audio frequencies (based on the Mel scale, which replicates how the human ear processes sound). Brightness and zero crossings are calculated based on the audio signal alone.
- **Tonal Features** (pitch, chromagram peak and centroid, key clarity, mode, Harmonic Change Detection Function (HCDF)): Tonal features describe the tonal aspects of a song such as key, dissonance, and pitch. They are based primarily on a pitch chromagram, which shows the distribution of energy across pitches based on the calculation of dominant frequencies in the audio.

3.2 User Data

We have created a questionnaire so that individuals can go through multiple times and annotate different sets of music based on their moods on a given day. 68 users completed the questionnaire between 1 and 8 times, resulting in almost 400 unique user sessions.

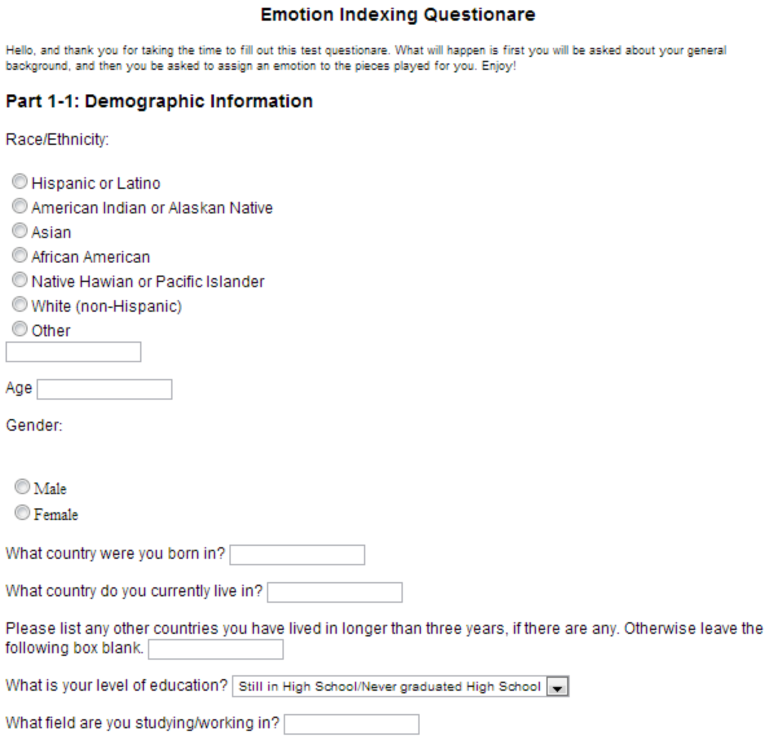
Questionnaire Structure. The Questionnaire is split into 5 sections:

- Demographic Information (where the user is from, age, gender, ethnicity),
- General Interests (favorite books, movies, hobbies),
- Musical Tastes (what music the user generally likes, what he listens to in various moods),
- Mood Information (a list of questions based on the Profile of Mood States),
- Music Annotation (where the user annotates a selection of musical pieces based on mood).

The demographic information section is meant to compose a general picture of the user (see Figure 1). The questions included ask for ethnicity (based on the NSF definitions), age, what level of education the user has achieved, what field they work or study in, where the user was born, and where the user currently lives. Also included is whether the user has ever lived in a country other than where he/she was born or where he/she currently lives for more than three years. This question is included because living in another country for that long would expose the user to music from that country.

The general interests section gathers information on the user's interests outside of music (see Figure 2). It asks for the user's favorite genre of books, movies, and what kind of hobbies he/she enjoys. It also asks whether the user enjoyed math in school, whether he/she has a pet or would want one, whether he/she believes in an afterlife, and how he/she would handle an aged parent. These questions are all meant to build a more general picture of the user.

The musical taste section is meant to get a better picture of how the user relates to music (Fig. 3). It asks how many years of formal musical training the user has had, his/her level of proficiency in reading/playing music if any, and what genre of music the user listens to when they are happy, sad, angry, or calm.



Emotion Indexing Questionare

Hello, and thank you for taking the time to fill out this test questionare. What will happen is first you will be asked about your general background, and then you be asked to assign an emotion to the pieces played for you. Enjoy!

Part 1-1: Demographic Information

Race/Ethnicity:

○ Hispanic or Latino
○ American Indian or Alaskan Native
○ Asian
○ African American
○ Native Hawian or Pacific Islander
○ White (non-Hispanic)
○ Other

Age

Gender:

○ Male
○ Female

What country were you born in?

What country do you currently live in?

Please list any other countries you have lived in longer than three years, if there are any. Otherwise leave the following box blank.

What is your level of education? Still in High School/Never graduated High School

What field are you studying/working in?

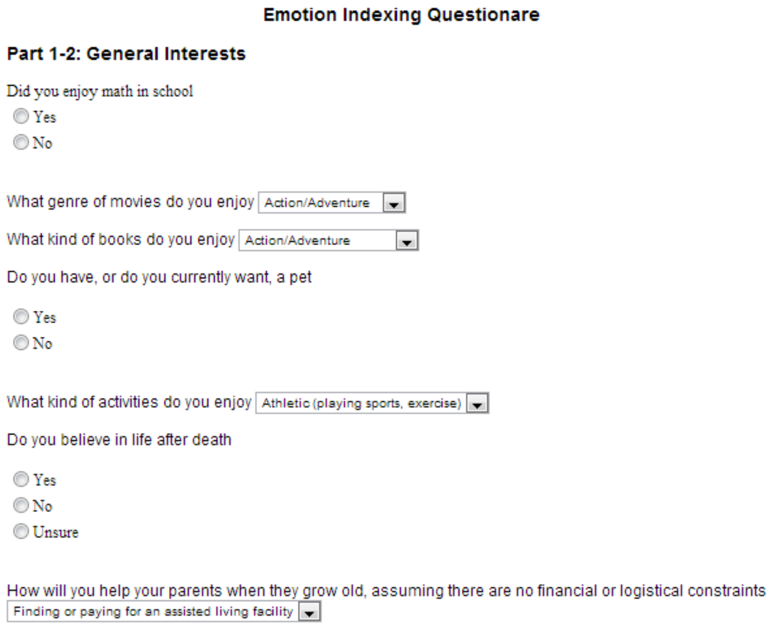
Fig. 1. The demographic information section of the questionnaire

The mood information section is a shortened version of the Profile of Mood States [14]. The Profile of Mood States asks users to rate how strongly he/she has been feeling a set of emotions over a period of time from the following list of possible responses:

- Not at all; A little; Moderately; Quite a bit; Extremely.

The possible emotions asked about in the mood information session are:

- Tense; Shaky; Uneasy; Sad; Unworthy; Discouraged; Angry; Grouchy; Annoyed; Lively; Active; Energetic; Efficient.



Emotion Indexing Questionare

Part 1-2: General Interests

Did you enjoy math in school

○ Yes
○ No

What genre of movies do you enjoy: Action/Adventure

What kind of books do you enjoy: Action/Adventure

Do you have, or do you currently want, a pet

○ Yes
○ No

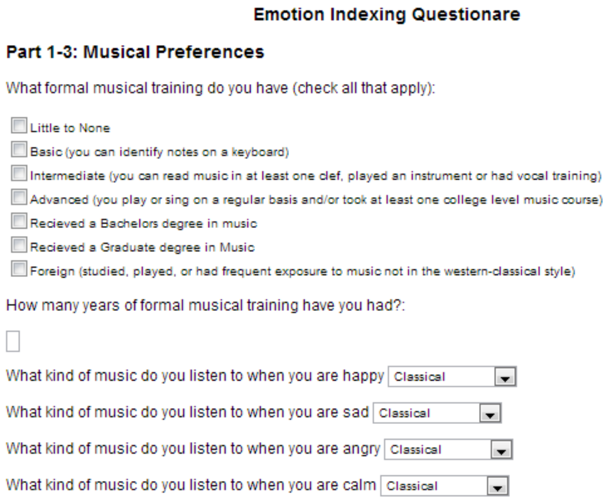
What kind of activities do you enjoy: Athletic (playing sports, exercise)

Do you believe in life after death

○ Yes
○ No
○ Unsure

How will you help your parents when they grow old, assuming there are no financial or logistical constraints: Finding or paying for an assisted living facility

Fig. 2. The general interests section of the questionnaire



Emotion Indexing Questionare

Part 1-3: Musical Preferences

What formal musical training do you have (check all that apply):

□ Little to None
□ Basic (you can identify notes on a keyboard)
□ Intermediate (you can read music in at least one clef, played an instrument or had vocal training)
□ Advanced (you play or sing on a regular basis and/or took at least one college level music course)
□ Recieved a Bachelors degree in music
□ Recieved a Graduate degree in Music
□ Foreign (studied, played, or had frequent exposure to music not in the western-classical style)

How many years of formal musical training have you had?:

□

What kind of music do you listen to when you are happy: Classical

What kind of music do you listen to when you are sad: Classical

What kind of music do you listen to when you are angry: Classical

What kind of music do you listen to when you are calm: Classical

Fig. 3. The musical taste/background information section of the questionnaire

A sample of these questions can be seen in Figure 4. This is the section that is filled out every time the user returns to annotate music, since their mood would affect how they annotate music on a given day. These answers are later converted into a mood vector for each session, which describes the user's mood state at the time of the session.

Mood Vector Creation. For each session the user is asked to select an answer describing how much he/she has been feeling a selection of emotions. Once this is finished, his/her answers are then converted to a numerical mood vector as follows: each answer is given a score based on the response, with 0 representing "Not at all", 1 – "A little", 2 – "moderately", 3 – "Quite a bit", and 4 – "Extremely". From here, sets of mood scores corresponding to different emotions are added together into a set of scores:

$$TA = Tense + Shaky + Uneasy \tag{1}$$

Where TA stands for Tension/Anxiety,

$$DD = Sad + Unworthy + Discouraged \tag{2}$$

Where DD stands for Depression/Dejection,

$$AH = Angry + Grouchy + Annoyed \tag{3}$$

Where AH stands for Anger/Hostility,

$$VA = Lively + Active + Energetic \tag{4}$$

Where VA stands for Vigor/Activity,

$$FI = WornOut + Fatigued + Exhausted \tag{5}$$

Where FI stands for Fatigue/Inertia,

$$CB = (Confused + Muddled) - Efficient \tag{6}$$

Where CB stands for Confusion/Bewilderment.

These scores are recorded, along with a total score calculated as follows:

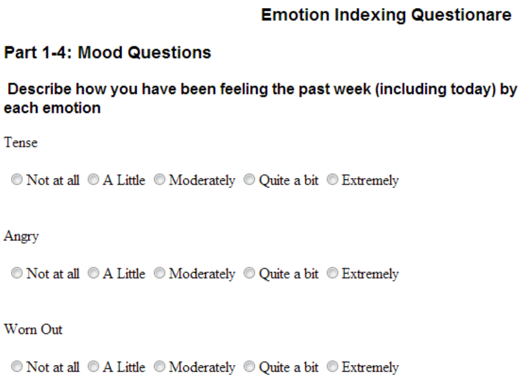
$$Total = (TA + DD + AH + FI + CB) - VA \tag{7}$$

The mood vector is then defined for user u and session s as

$$m(u, s) = (TA(u, s), DD(u, s), AH(u, s), FI(u, s), CB(u, s), VA(u, s), Total(u, s)) \tag{8}$$

Finally, the music annotation section is where users go to annotate a selection of clips. 40 clips are selected randomly from the set of 1440 clips mentioned in Section 3.1. The user is then asked to check the checkbox for the emotion he/she feels in the music, along with a rating from 1-3 signifying how strongly the user feels that emotion (1 being very little, 3 being very strongly). The user has a choice of 16 possible emotions to pick (to be specific, 12 emotions and 4 generalizations), based on a 2-D hierarchical emotional plane (see Fig. 6 for the emotion plane and Fig. 5 for a view of the questionnaire annotation section).

When the user goes through the questionnaire any time after the first time, he only has to fill out the mood profile and the annotations again. Each of these separate sections (along with the rest of the corresponding information) is treated as a separate user, so each individual session has classifiers trained for each emotion, resulting in 16 emotion classifiers for each user session.



Emotion Indexing Questionare

Part 1-4: Mood Questions

Describe how you have been feeling the past week (including today) by selecting an appropriate box after each emotion

Tense

○ Not at all ○ A Little ○ Moderately ○ Quite a bit ○ Extremely

Angry

○ Not at all ○ A Little ○ Moderately ○ Quite a bit ○ Extremely

Worn Out

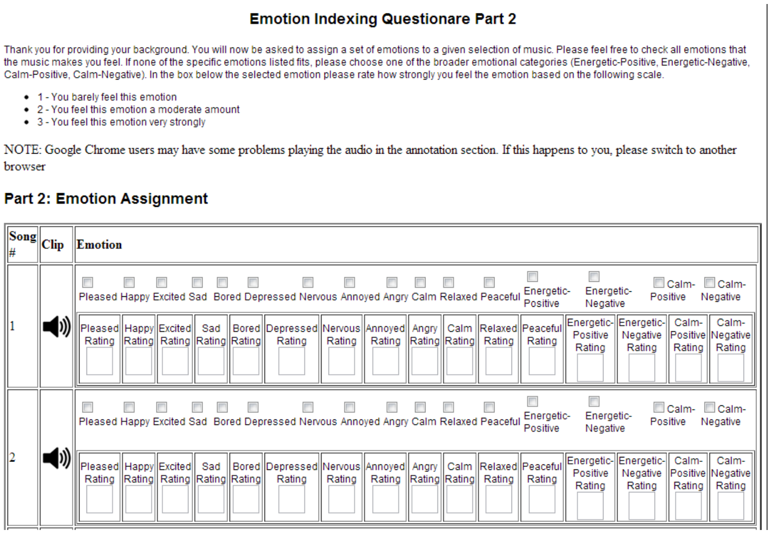
○ Not at all ○ A Little ○ Moderately ○ Quite a bit ○ Extremely

Fig. 4. Part of the mood state information section of the questionnaire. This section is filled out every time the user reenters the questionnaire (the user starts on this page once he/she has filled out the rest of the questionnaire once)

Emotion Model. This model was first presented in [5], and implements a hierarchy on the 2-dimensional emotion model, while also implementing discrete elements. The 12 possible emotions are derived from various areas of the 2-dimensional arousal-valence plane (based on Thayer's 2-dimensional model of arousal and valence [15]). However, there are also generalizations for each area of the plane (excited-positive, excited-negative, calm-positive, and calm-negative) that the users can select as well. This compensates for songs that might be more ambiguous to the user; if a user generally knows that a song is high-energy and positive feeling but the words excited, happy, or pleased do not adequately describe it, they can select the generalization of energetic-positive.

3.3 Classifier Development

Personalized classifiers were trained and tested using the classification algorithms listed previously (C4.5, SVM, Random Forest, kNN). The user annotation data



Emotion Indexing Questionare Part 2

Thank you for providing your background. You will now be asked to assign a set of emotions to a given selection of music. Please feel free to check all emotions that the music makes you feel. If none of the specific emotions listed fits, please choose one of the broader emotional categories (Energetic-Positive, Energetic-Negative, Calm-Positive, Calm-Negative). In the box below the selected emotion please rate how strongly you feel the emotion based on the following scale.

- 1 - You barely feel this emotion
- 2 - You feel this emotion a moderate amount
- 3 - You feel this emotion very strongly

NOTE: Google Chrome users may have some problems playing the audio in the annotation section. If this happens to you, please switch to another browser

Part 2: Emotion Assignment

Song #	Clip	Emotion
1	🔊	□ Pleased □ Happy □ Excited □ Sad □ Bored □ Depressed □ Nervous □ Annoyed □ Angry □ Calm □ Relaxed □ Peaceful □ Energetic-Positive □ Energetic-Negative □ Calm-Positive □ Calm-Negative Pleased Rating □ Happy Rating □ Excited Rating □ Sad Rating □ Bored Rating □ Depressed Rating □ Nervous Rating □ Annoyed Rating □ Angry Rating □ Calm Rating □ Relaxed Rating □ Peaceful Rating □ Energetic-Positive Rating □ Energetic-Negative Rating □ Calm-Positive Rating □ Calm-Negative Rating □
2	🔊	□ Pleased □ Happy □ Excited □ Sad □ Bored □ Depressed □ Nervous □ Annoyed □ Angry □ Calm □ Relaxed □ Peaceful □ Energetic-Positive □ Energetic-Negative □ Calm-Positive □ Calm-Negative Pleased Rating □ Happy Rating □ Excited Rating □ Sad Rating □ Bored Rating □ Depressed Rating □ Nervous Rating □ Annoyed Rating □ Angry Rating □ Calm Rating □ Relaxed Rating □ Peaceful Rating □ Energetic-Positive Rating □ Energetic-Negative Rating □ Calm-Positive Rating □ Calm-Negative Rating □

Fig. 5. The music emotion annotation section, also filled out every time the user goes through the questionnaire. The user clicks on a speaker to hear a music clip, then checks an emotion and supplies a rating 1 to 3

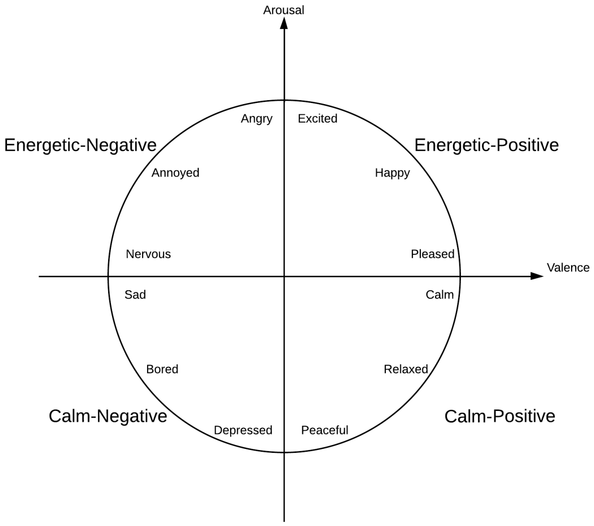


Fig. 6. A diagram of the emotional model

was first converted so that each annotation for each song was represented as a vector of 16 numbers with each number representing the emotion labeling. The numbers ranged from 0 to 3, with 0 representing an emotion that was not selected by the user and the remaining numbers being the strength the user entered with the annotation. These vectors for all the users were then linked with the feature data extracted from the corresponding music clips. From this resulting table all the annotations and music data linked with individual user IDs were separated and used to train and test personalized classifiers for each emotion. This resulted in each user having at most 16 personalized classifiers (depending on whether the user used a given emotion during the course of annotating), where for each classifier the class attribute was one of the 16 possible emotions. The classifiers were evaluated via Weka [16] using 10-fold cross validation. For the C4.5 classifier we used the J48 implementation in Weka and for kNN we used Weka's IBk. Analysis of the results indicates which classifier is most effective for personalized classification and, therefore, the most effective cluster-driven classifier.

4 Results

The results are listed in Table 1. All four classifiers achieved a relatively high average accuracy, above 80%. SVM achieved the lowest accuracy, 82.35%, while J48 trees achieved the highest accuracy, 86.62%. However, SVM as well as Random Forest achieved the highest average F-score (a combined measure of precision and recall). IBk on the other hand had the lowest F-score of 0.92. SVM was expected to have a higher accuracy as it works so well with music data, but our previous success with J48 means the high accuracy and F-score are not surprising.

The Kappa statistic reveals further insights into the effectiveness of each classifier. This statistic measures the agreement between a true class and the prediction, and the closer to 1 the statistic is the more agreement (1 represents complete agreement). None of the classifiers reaches higher than 0.1, although again IBk has the highest average Kappa (J48, again, the lowest). This suggests that while J48 is overall very accurate it is more inconsistent in terms of this particular set of data, while SVM is moderately accurate and very consistent.

Table 1. Table of classifier accuracies and F-scores

Classifier	Average Accuracy	Average F-Score	Average Kappa
SVM	82.35%	0.90	0.137658
IBk	85.7%	0.87	0.153468
J48	86.62%	0.89	0.076869
Random Forest	84.25%	0.90	0.133027

As it proved to be the most accurate classifier, we have chosen J48 as the algorithm to use to build the non-personalized classifiers for comparison. We again built 16 emotion classifiers, this time using all the user annotations to train and test rather than individual user annotations. The results compared to the personalized J48 classifiers are shown in Table 2.

Table 2. Comparison of classifier accuracies and F-scores between personalized and non-personalized classifiers

Classifier	Average Accuracy	Average F-Score	Average Kappa
Personalized J48	86.62%	0.89	0.076869
Non-personalized J48	87.29%	0.82	0.00015

The average accuracy does not change too much between personalized and non-personalized classifiers (only 0.7 percentage point). However, this was mainly due to the fact that several of the emotions were not used to the same extent as others when tagging (for example, the generalized emotions), and in that case all the classifier did was predict '0' (for emotions that were not selected). This raised the accuracy for those classifiers, but it is not nearly as indicative as to the quality of the classifier as the F-Score and Kappa, which showed a great deal of improvement in the personalized classifier. The average F-score for the non-personalized classifiers is 0.07 less than the average F-score for personalized classifiers, and the average Kappa for the non-personalized classifiers is far less than the personalized classifiers. These both signify a significant loss in classifier consistency once the classifiers are no longer personalized.

5 Hierarchical Cluster Driven Classifiers

We have also developed a more efficient method of developing personalized classifiers. Using agglomerative clustering, we have developed hierarchical cluster-based classifiers (see Figure 7). These clusters are built based on the user data gathered through the questionnaire (as described in Section 3.2).

We can now build a tree structure of classifiers where the leaf nodes of the tree are labeled by vectors representing individual users and the root of any subtree is labeled by a smallest generalization vector covering vector labels associated with all leaves of that subtree. Each node of the tree has its own set of 16 emotion classifiers, one for each possible emotion, based on the annotation data from the group of users assigned to that node. The lower the node on the tree, the more specialized its classifiers are. Now, if there is a need, we can assign a new user to a correct node of the tree structure which is the lowest one labeled by generalization vector containing the vector label of that user. Then, we can apply the classifiers associated with that node to annotate the music.

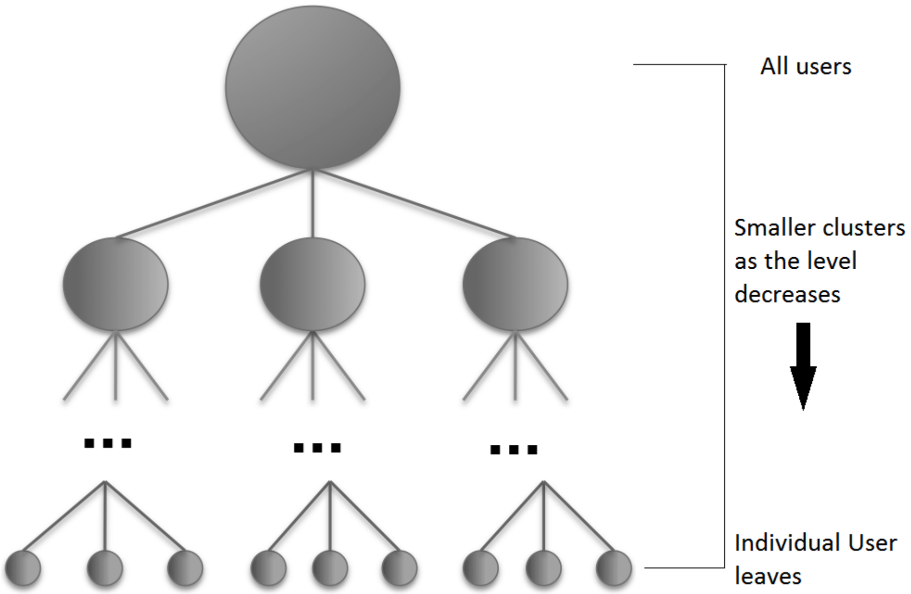


Fig. 7. A visual of the proposed classifier hierarchy

5.1 Data Storage

Each user's questionnaire answers are converted into three sets of data: a vector D (with an appropriate subscript) resulting from the questions the user answered in the first part of the questionnaire, the set of mood vectors M (with an appropriate subscript) built for each user's session based on the profile of mood states questions, and a set of classifiers C (with appropriate subscripts) extracted from decision tables associated with each mood vector.

5.2 User Generalization

Let us assume that we have a group of users where each one is represented by vector D_a, where a is a user identifier and D_a shows the answers from the first part of the questionnaire. We run an agglomerative clustering algorithm on this set of vectors and a tree T representing the outcome of the algorithm is created. For each node of T, we first create the smallest generalized description D_C of all the users C assigned to that node.

For example, assume that we have a cluster C with two users a and b. Vector D_C is created as the smallest generalization of vectors D_a and D_b such that both of them are included in D_C. The coordinate i of D_C is built from coordinates i of D_a and D_b as:

$$D_{Ci} = \{k : min(D_{ai}, D_{bi}) \leq k \leq max(D_{ai}, D_{bi})\} \tag{9}$$

The mood vectors and decision tables assigned to the nodes of T which are not leaves are generalized on a more conditional basis, using the distance between mood vectors. For example, let us assume that users $a1, a2$ end up in the same cluster C and their sessions are represented by mood vectors m[a1,8], m[a2,5], m[a1,2]. If the distance between any of these vectors is less than λ (a given threshold), then they are added together by following the same strategy we used for vectors D representing the first part of the questionnaire. It should be noted that mood vectors representing sessions of different users can be added together, whereas vectors representing the same user may remain separated. When the new mood vectors for a node of T are built, then the new decision table for each of these new mood vectors is built by taking the union of all decision tables associated with mood vectors covered by this new mood vector (Fig. 8).

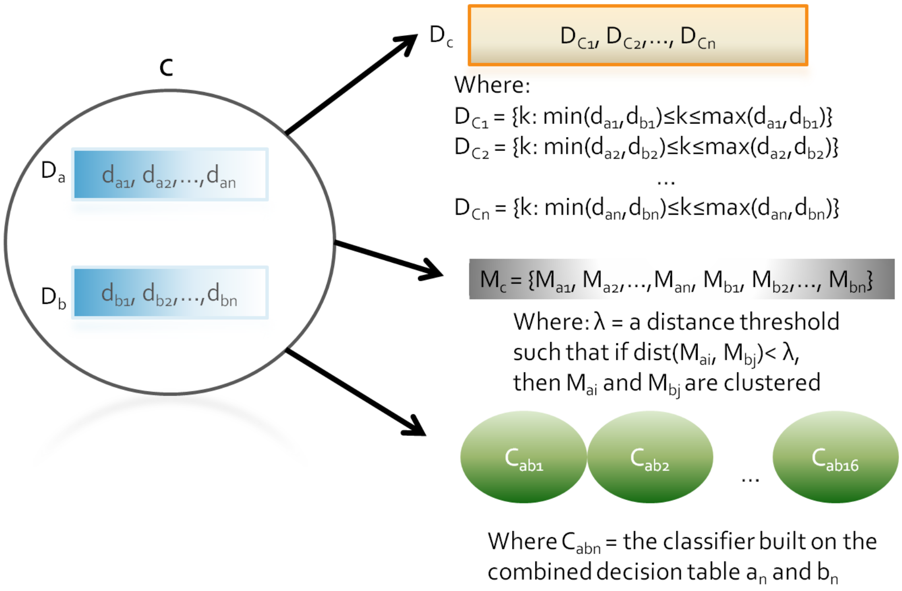

C
D_a
$d_{a1}, d_{a2},\dots, d_{an}$
D_b
$d_{b1}, d_{b2},\dots, d_{bn}$
D_c
$D_{C1}, D_{C2},\dots, D_{Cn}$
Where:
$D_{C1} = \{k: \min(d_{a1},d_{b1}) \le k \le \max(d_{a1},d_{b1})\}$
$D_{C2} = \{k: \min(d_{a2},d_{b2}) \le k \le \max(d_{a2},d_{b2})\}$
…
$D_{Cn} = \{k: \min(d_{an},d_{bn}) \le k \le \max(d_{an},d_{bn})\}$
$M_c = \{M_{a1}, M_{a2},\dots,M_{an}, M_{b1}, M_{b2},\dots, M_{bn}\}$
Where: λ = a distance threshold
such that if $dist(M_{ai}, M_{bj}) < \lambda$,
then M_{ai} and M_{bj} are clustered
C_{ab1}
C_{ab2}
…
C_{ab16}
Where C_{abn} = the classifier built on the
combined decision table a_n and b_n

Fig. 8. A visualization of the data stored for each cluster

5.3 New User Placement

When a new user, x, fills out the first part of the questionnaire his/her representative vector D_x is created. Then it is checked to see if this vector is equal to one of the representative vectors representing leaves of the tree structure. If this is not the case, then the representative sets of vectors belonging to the parents of these leaves are checked to see if one of them contains D_x. This is accomplished by comparing the individual column values of each column, D_{xi}, to each column range D_{Ci}. If every D_{xi} fits in the range of each D_{Ci}, then the user is assigned to that cluster. If D_x does not fit within any node on a given level, then it is compared to the parent of the node that D_x most closely matches (based on the number of column ranges in D_C that D_x does fit in). For example, if on the bottom level D_x cannot be assigned to any cluster, but all i in D_{xi} fit in the ranges for D_{Ci} except for one, this would make it the closest matched cluster, and then D_x would be compared to the parent of D_C. The goal is to assign the user to the cluster on the lowest level it can fit in order to utilize the most specialized classifiers built (since lower level classifiers are built on data from a more homogenized group of users, and therefore a more unified group of annotations). These classifiers will therefore be more accurate to the new user, solving the ”cold-start” issue inherent in collaborative recommender systems.

5.4 Test case: New user

To demonstrate the effectiveness of this system, we analyze one new user using this system for the first time. This user has filled out the same questionnaire questions, however they only have one mood vector which (along with their other questionnaire answers) is used to assign this user to a cluster. This user also annotated a different set of songs, which had the same feature data extracted from them as the initial training set.

Once this user is assigned to a node in our tree structure, they are given the set of emotion classifiers made for the cluster in that part of the tree. The user's new annotations were then used to evaluate the classifiers' effectiveness. The resulting statistics are shown in Table 3.

The accuracy and F-scores can stay relatively high for each classifier, although it is not consistent. This could be explained by the user not annotating songs with certain emotions, which would make the accuracies very difficult to judge. This could indicate also that certain emotions are easier to classify for songs than others. Observe that the lower accuracies (aside from Happy) are for emotion classifiers from the calm-positive quadrant of the arousal-valence plane (Calm, Relaxed, Peaceful), and Calm-Positive is the least accurate quadrant classifier (Energetic-Positive, Energetic-Negative, Calm-Positive, Calm-Negative). This could imply that emotions with a high valence and low arousal are particularly difficult to detect in music, but this would require further investigation.

6 Conclusion

We have presented a system through which we build personalized music emotion classifiers based on user data accumulated through an online survey. Using this data, we have built classifiers that are about as accurate as standard, non-personalized ones but far more consistent. In a real world situation, this would mean that music that accurately reflects the user's mood (or desired mood) would be recommended far more often than not. Future work would involve using these classifiers in a full music player, and improving classifiers by retraining

Table 3. Statistics for the classifiers built for a new user

Emotion	Accuracy	Average Precision	Average Recall	Average F-Score
Pleased	87.5	0.875	0.875	0.87
Happy	50	0.493	0.500	0.493
Excited	70.8333	0.675	0.708	0.691
Sad	100	1.000	1.000	1.000
Bored	91.6667	0.917	0.917	0.917
Depressed	95.8333	0.918	0.958	0.938
Nervous	79.1667	0.756	0.792	0.773
Annoyed	91.6667	0.917	0.917	0.917
Angry	100	1.000	1.000	1.000
Calm	83.3333	0.761	0.833	0.795
Relaxed	75	0.714	0.750	0.729
Peaceful	62.5	0.594	0.625	0.609
Energetic-Positive	87.5	0.911	0.875	0.892
Energetic-Negative	87.5	0.875	0.875	0.875
Calm-Positive	54.1667	0.574	0.542	0.557
Calm-Negative	87.5	0.915	0.875	0.894

them through usage and by ensuring the clusters are disjoint, using Michalski's STAR method applied in $AQ15$ to build disjoint D-vector representations [17].

Acknowledgments. This project was partially supported by the Research Center of PJIIT, supported by the Polish Ministry of Science and Higher Education.

References

1. Lartillot, O., Toiviainen, P., Eerola, T.: MIRtoolbox. University of Jyväskylä (2008)
2. Laurier, C., Herrera, P.: Automatic Detection of Emotion in Music: Interaction with Emotionally Sensitive Machines. In Vallverdu, D., Casacuberta, D., eds.: Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence. IGI Global (2009) 9–32
3. Yang, Y.H., Su, Y.F., Lin, Y.C., Chen, H.H.: Music Emotion Recognition: The Role of Individuality. In: Proc. International Workshop on Human-centered Multimedia 2007 (HCM'07), Augsburg, Germany, ACM (September 2007)
4. Lu, C.C., Tseng, V.S.: A novel method for personalized music recommendation. Expert Systems with Applications **36**(6) (August 2009) 10035–10044
5. Grekow, J., Raś, Z.W.: Detecting Emotions in Classical Music from MIDI Files. In: ISMIS'09, Prague, Czech Republic, Springer (2009) 261–270
6. Mckay, C., Fujinaga, I.: Automatic genre classification using large high-level musical feature sets. In: ISMIR 2004. (2004) 525 – 530
7. Silla Jr., C.N., Koerich, A.L., Kaestner, C.A.A.: A Machine Learning Approach to Automatic Music Genre Classification. J. Braz. Comp. Soc **14**(3) (2008) 7–18
8. Kursa, M., Rudnicki, W., Wieczorkowska, A., Kubera, E., Kubik-Komar, A.: Musical Instruments in Random Forest. LNCS **5722** (2009) 281–290
9. Laurier, C., Meyers, O., Marxer, R., Bogdanov, D., Serrà, J., Gómez, E., Herrera, P., Wack, N.: Music classification using high-level models. ISMIR 2009 (2010)
10. Mandel, M.I., Ellis, D.P.W.: Song-level features and support vector machines for music classification. In Reiss, J.D., Wiggins, G.A., eds.: ISMIR 2005. Volume 6., London, U.K. (2005) 594–599
11. Panda, R., Paiva, R.P.: Using Support Vector Machines for Automatic Mood Tracking in Audio Music. In: 130th Audio Engineering Society Convention. (2011)
12. Zhang, X., Ras, Z.: Differentiated Harmonic Feature Analysis on Music Information Retrieval for Instrument Recognition. In: IEEE GrC 2006, Atlanta, Georgia (2006) 578–581
13. Eerola, T., Vuoskoski, J.K.: A comparison of the discrete and dimensional models of emotion in music. Psychology of Music **39**(1) (August 2011) 18–49
14. McNair, D.M., Lorr, M., Droppleman, L.F.: Profile of Mood States (POMS) (1971)
15. Thayer, R.E.: The biopsychology of mood and arousal. Oxford University Press (1989)
16. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software. ACM SIGKDD Explorations Newsletter **11**(1) (November 2009) 10
17. Michalski, R., Mozetic, I., Hong, J., Lavarac, N.: The Multi-Purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains. In: AAAI-86 Proceedings, AAAI (1986) 1041–1045