Creating Classifiers for a Personalized Music Recomendation System

Amanda Cohen Mostafavi, Cyril Laurier, Perfecto Herrera

Abstract The topic of music emotion recognition is emerging in the field of music information retrieval. Personalized recommendation of music is the next logical step within the topic of detecting emotion in music. While a program can eventually learn someone's taste and interpretation of music, being able to assign the user to a group based on similar tastes will allow the program to learn even faster. This paper will show how, using clustering techniques, classifiers can be personaized and then grouped together based on similar interpretations of music in relation to emotion.

Key words: music information retrieval, personalized classifiers, music emotion recognition

1 Introduction

With a person's individual music collection growing larger by the day, and even more music readily available online, there is a demand for new and interesting ways to organize and retrieve music. The field of Music Information Retrieval exists to fill this demand by developing new ways to programatically retrieve information from audio. One aspect of this field is automatic classification of music by emotions, and specifically by personalized emotions. Since music and emotions are both so

Amanda Cohen Mostafavi
University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte NC 28262, e-mail: acohen24@uncc.edu

Cyril Laurier
Music Technology Group at Universitat Pompeu Fabra, Roc Boronat, 138 08018 Barcelona e-mail: cyril.laurier@upf.edu

Perfecto Herrera
Music Technology Group at Universitat Pompeu Fabra, Roc Boronat, 138 08018 Barcelona e-mail: perfecto.herrera@upf.edu

subjective, it is necessary that an emotion classification system be personalized. The authors have sought to create a system where classifiers can learn the preferences and emotional profile of their user, and in the process have discovered that these classifiers can be clustered, resulting in a reverse-engineered social network based on musical preference.

This paper is organized as follows:

- Section 2 (Background): This section will discuss the current state of research in music emotion recognition and music personalization and recommendation. Previous research by the authors in this area will also be discussed.
- Section 3 (Tailored Tagger): This section will discuss the Tailored Tagger, the first step in creating personalized classifiers based on user behavior.
- Section 4 (Clustering Classifiers): This section will discuss how we intend to learn how to cluster personalized classifiers.

2 Background

This section will discuss the current state of related music information retrieval research. It has been divided into two subsections based on the two areas of music information retrieval most relevant to this topic: music emotion recognition and music personalization and recommendation.

2.1 Music Emotion Recognition

The process of detecting the emotion associated with a piece of music goes by several names: music emotion recognition (or MER for short), music mood detection, automatic indexing of music by emotion to name a few examples. Broadly, the goal of this area of research is to develop ways to detect the pervading emotion in a piece of audio. This is generally done by extracting audio features, although there has been previous work in determining emotion based on scalar music theory (see [8]).

There has been discussion first of all about how exactly to measure and model emotions. T. Eerola et al compare in [2] two of the most common ways to model emotions in relation to music: dimensional and discrete. In dimensional modelling, possible emotions are modelled on a 2 (or 3) dimensional plane with different areas representing emotions of varying positivity and negativity or varying levels of energy/intensity. Discrete modelling views emotions as a set of broad emotions or factors (usually giving a set of words to describe possible emotions). The authors found that either model was sufficient, although discrete models resulted in inconsistent ratings for music that was more ambiguous in the implied emotion. The work presented in this paper uses first a discrete model (in the case of the Tailored Tagger, which will be discussed further in the next section), and then a hybrid between the two model types.

Generally speaking, music emotion recognition algorithms are developed by using a ground truth of previously annotated music to train a set of classifiers. This ground truth is usually based on a ”wisdom of crowds” form of annotation (songs are annotated based on majority vote). The main difference is in what algorithms or classifiers are used. Two of the authors previously in [7] developed several in-depth emotion classification methods based on audio content analysis using support vector machine (SVM) classifiers. In [5], the authors tested several multi-label classification algorithms with SVM as a base classifier to solve the problem of multi-label emotion classification. Their results were generally accurate, achieving 73-87% accuracy.

2.2 Music Personalization and Recommendation

Although not always specific to emotion, personalized music recommendation is also an emerging field in music information retrieval. Very early work on this topic was demonstrated in 2000 in [1].

Although [4] demonstrated that emotion in music is not so subjective that it cannot be modelled, it is still the next logical step for music emotion recognition to be personalized to users as well. Yang et al in [13] 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 et al in [9] proposed a system that combined emotion-based, content-based, and collaborative-based recommendation and achieved an overall accuracy of 90%.

In [3], the author first proposed the idea of using clustering in order 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.

3 Tailored Tagger

The first step in the process of forming clusters of classifiers is to create personalized music classifiers for individual users. It was our additional goal to do this without a user having to annotate a large number of sample songs. To this end, the authors created a tool referred to as the Tailored Tagger. This tool interfaces with a user's music collection through either iTunes or Winamp and allows the user to annotate his or her own music collection. It also classifies music in the user's music collection, and the user can either correct that classification or keep it which is how the tool learns.

For the initial classification, a set of binary-labelled datasets were used to build a set of initial SVM classifiers. Four datasets were used to build four emotion classifiers, one for each of the possible emotions used by the tool; happy, sad, angry, or peaceful. These emotions were selected as representations of four main areas of emotion: high-energy and positive (happy), high-energy and negative(angry), low-energy and positive (peaceful), and low-energy and negative (sad). This dataset was provided by one of the authors, as was previously used in [6]. Once the initial classifiers are trained, the tool interfaces with the user's music collection and classifies the currently playing song based on the initial classifiers. From there the user can either press submit, indicating he agrees with the system's classification, or uncheck and recheck the boxes that he feels fits the song better and press submit. From there, depending on which boxes were checked or unchecked, the initial emotion classifiers are retrained with the currently playing song being added to the training set.

The audio feature extraction was done using internal libraries from the Music Technology Group at Universitat Pompeu Fabra [12]. The features extracted are low-level audio descriptors (such as MFCC), and high-level descriptors (such as pitch and tempo).

3.1 Example

Below is an example of a typical situation in which the Tailored Tagger comes across a song, classifies it one way, and the user in turn corrects the classification.

Figure 1 shows the system coming across the song "Hurt" by Nine Inch Nails. This is a song the user has not previously tagged, and its assumed that this is the first time the song has been played while the tagger is running. At this time, the song is analyzed and its feature information is extracted. The feature information is then sent to all four mood classifiers, which will classify the song as "[emotion]" or "not [emotion]". Any classifier that returns "[emotion]" is returned to the program, and the corresponding mood is highlighted. In this case, "Sad" and "Peaceful" were returned for this song.

Figure 2 shows the user wishing to add a tag to the song, in this case "Angry". At this point, the user would check the mood they wish to add to the song and hit the Submit button. In this case, once the user hits that button, the Angry classifier

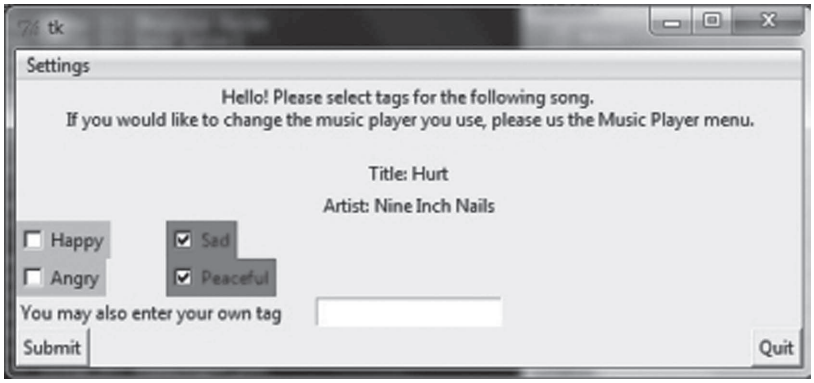

Fig. 1 The tagger comes across a song the user has not previously tagged. Dark highlighted emotions are tagged by the system

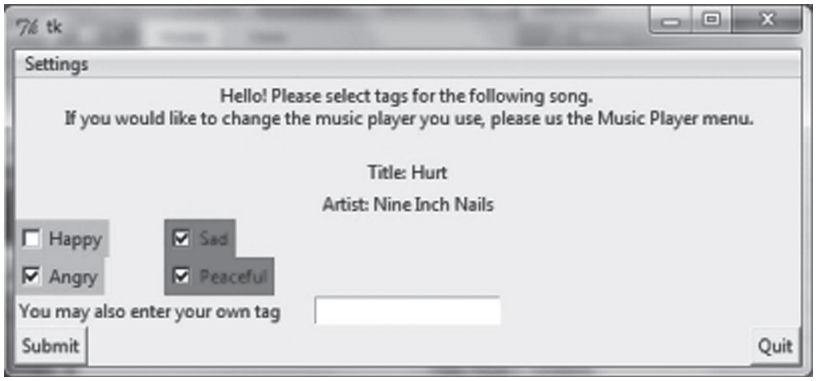

Fig. 2 The user wants to add an additional tag to the song

needs to be retrained. The song is added to the training set for ”Angry” and given the annotation ”Angry”, and then the SVM classifier is retrained. Now the ”Angry” classifier not only knows to recognize ”Hurt” as an angry song, but also to recognize similar songs as angry. The user can also remove tags that they don’t agree with. In this case, if the user wanted to remove ”Peaceful”, they would simply uncheck the box next to ”Peaceful”. Once that happens, the song is added to the training set for ”Peaceful” as was done with ”Angry”. However this song is given the annotation ”not Peaceful” before the classifier is retrained. Now the ”Peaceful” classifier knows not to label this song (or similar songs) as peaceful. When the user comes across the song again, the previous tag information is saved as the user tags (as signified by the light highlighting)

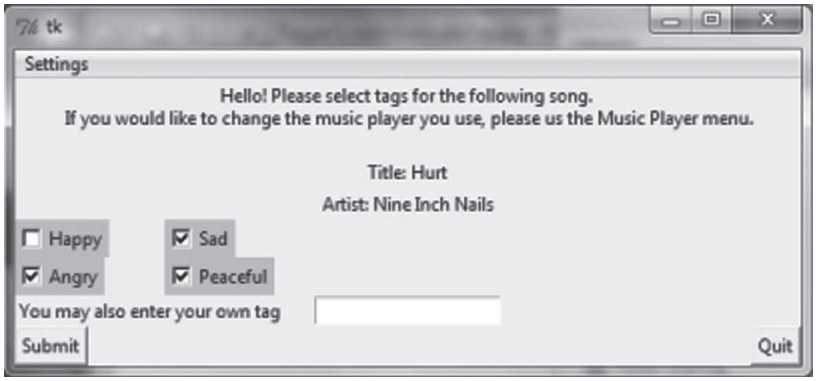

Fig. 3 The system plays the same song again, this time with the tags saved as user tags

4 Clustering Classifiers

We have been working to take what was developed as far as an adaptive classifier and use that to cluster personalized classifiers based on additional information. We have revised our previous questionnaire so that individuals can go through multiple times and annotate different sets of music based on their moods on a given day. This has given us almost 400 samples. Once we are done collecting data, we will create individual classifiers for each persons session (since annotations may differ based on the persons mood for that day). Our next step would be to cluster these individual classifiers and build classifiers for the whole cluster that a future user would be assigned to.

4.1 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. 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 they were born or where they currently live 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. It asks for the user's favorite genre of books, movies, and what kind of hobbies they enjoy. It also asks whether the user enjoyed math in school (since there is a defined connection between a person's math ability and how they interpret music), whether they have a pet or would want one, whether they believe in an afterlife, and how they would handle an aged parent. These questions are all meant to build a more general picture of the user.

The musical tastes section is meant to get a better picture of how the user relates to music. It asks how many years of formal musical training the user has had, as well as their level of proficiency in reading or playing music if any. It also asks what genre of music the user listens to when they're happy, sad, angry, and calm.

The mood information section is a shortened version of the Profile of Mood States [10]. The Profile of Mood States asks users to rate how strongly they have been feeling a set of emotions (from "Not at all" to "Extremely") over a period of time. 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.

Finally, the music annotation section is where users go to annotate a selection of songs. 40 songs are selected randomly from a set of 100 songs. 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, based on a 2-D hierarchical emotional plane.

When the user goes through the quesitonnaire 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 to be clustered.

4.2 Emotion Model

This model was first presented in [3], 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 [11]). 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 don't adequately describe it, they can select the generalization of energetic-positive.

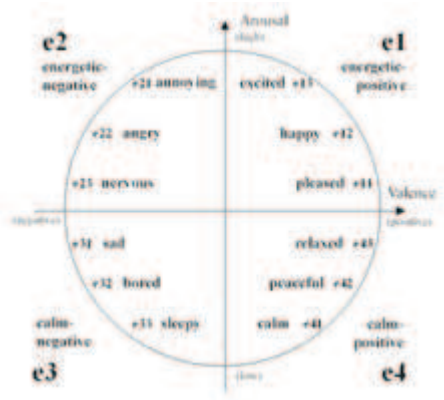

Fig. 4 A diagram of the emotional model to be used for classifier clustering

4.3 Annotation Normalization

It is concievable that different users will label the same song by two or more diametrically opposed emotions, and that different users will interpret this different ways. We therefore will normalize the user annotation data as follows.

First of all, we assume that each emotion has a corresponding opposite emotion in the 2-D plane. Generally, anything in diagonal quadrants is considered to be opposite (as in energetic-positive is directly opposed to calm-negative). The exact opposites are listed below:

- Happy/Sad
- Excited/Sleepy
- Pleased/Bored
- Nervous/Relaxed
- Angry/Peaceful
- Annoying/Calm

Additionally, the generalizations are opposed to each other (energetic-positive/calm-negative and calm-positive/energetic negative).

We then look at the annotations. Given two opposite emotions with weights A and B:

- If $A+B<3$, then $A=A+[(3-(A+B))/2]$ and $B=B+[(3-(A+B))/2]$
- If $A+B>3$, then $A=A-[((A+B)-3)/2]$ and $B=B-[((A+B)-3)/2]$

So for example, if someone annotated a song as happy with a weight of 1 and sad with a weight of 1, then the new weight of each emotion would be $1+[(3-2)/2]=1.5$. Likewise, if someone annotated a song as happy with a weight of 3 and sad with a weight of 2, then the new weight for happy would be $3-[(5-3)/2]=2$ and the new weight of sad would be $2-[(5-3)/2]=1$. This would however not apply in situations where one emotion was annotated and its opposite was not (e.g. happy with a weight of 2 and sad with a weight of 0), nor would normalization be used if neither an emotion nor its opposite were annotated (e.g. neither happy nor sad was annotated).

4.4 Classifier Clustering

We will be using an agglomerative clustering algorithm that incorporates manhattan distance to cluster individual user sessions based on their answers to the questions listed in the previous subsection. Each user session will be viewed as an individual vector. This will allow new users to join clusters that are formed not only on common background but common emotional state (since the answers to the profile of mood states questions change each user session).

The distance between each vector will be found as follows

4.5 Personalized Recommender System

Once the clusters are found, representative decision tables for each emotion will be formed based on the vectors contained in each cluster and SVM classifiers will be trained based on these tables, resulting in 16 representative emotion classifiers for each cluster. New users will then answer the same set of questions, which will determine which cluster the user belongs to and therefore which set of classifiers to use. The user will then query the system for a possible emotion, as well as a weight. If there is a song that matches the emotion query (based on the result of classification by each individual classifier), that song will be returned. If not, then the query will be extended to include all possible weights of the given emotion. Once the result is returned, the user will be able to agree or disagree with the result. If the user agrees, nothing will be altered. If the user disagrees, then they will be able to reannotate the song, and the associated emotion classifier will be retrained in the same way as was demonstrated in the Tailored Tagger.

5 Conclusion

This paper outlined the Tailored Tagger, a system which interfaces with a user's music player and learnes user preferences and behavior by retraining SVM classifiers for specific emotions. It also outlined a system of clustering users and their classifiers based on user data. Finally, it showed how these approaches can be combined so that each cluster has an associated set of classifiers that will also change and be retrained based on user behavior.

Future work will involve implementing this classifier system and evaluating the results based on individual usage. Improvements from that point will be based on the results, although one possible improvement would be finding a way to weight songs annotated by the current user so that the classifier learnes the user's behavior more quickly.

6 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. OISE-0730065

References

1. M Alghoniemy and A H Tewfik. Personalized music distribution. *2000 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings Cat No00CH37100*, pages 2433–2436, 2000.
2. T. Eerola and J. K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18–49, August 2010.
3. Jacek Grekow and Zbigniew W. Raś. Detecting Emotions in Classical Music from MIDI Files. pages 261–270, Prague, Czech Republic, 2009. Springer-Verlag.
4. P. Herrera and C Laurier. Automatic Detection of Emotion in Music: Interaction with Emotionally Sensitive Machines. pages 9–32. IGI Global, 2009.
5. Konstantinos Konstantinos, Grigorios Tsoumakas, George Kalliris, and Ioannis Vlahavas. {Multi-Label} Classification of Music into Emotions. In *Proceedings of {ISMIR} 2008*, pages 325–330, 2008.
6. C Laurier, O Meyers, J Serrà, M Blech, and P Herrera. Music Mood Annotator Design and Integration. In *7th International Workshop on {Content-Based} Multimedia Indexing*, Chania, Crete, Greece, 2009.
7. Cyril Laurier. *Automatic Classificatoin of Musical Mood by Content-Based Analysis*. PhD thesis, Universitat Pompeu Fabra, 2011.
8. W Jiang R Lewis and Z W Ras. Mining Scalar Representations in a {Non-Tagged} Music Database. *in Foundations of Intelligent Systems*, pages 819–824, 2007.
9. Cheng-Che Lu and Vincent S Tseng. A novel method for personalized music recommendation. *Expert Systems with Applications*, 36(6):10035–10044, 2009.
10. D M McNair, M Lorr, and L F Droppleman. Profile of Mood States (POMS), 1971.
11. R E Thayer. *The biopsychology of mood and arousal*. Oxford University Press, {USA}, 1989.

12. Nicolas Wack. Essentia and gaia: audio analysis and music matching c++ libraries developed by the music technology group, 2010.
13. Y.-H. Yang, Y.-F. Su, Y.-C. Lin, and H.-H. Chen. Music Emotion Recognition: The Role of Individuality. In *Proceedings International Workshop on Human-centered Multimedia 2007* {*(HCM'07)*}, Bavaria, Germany, September 2007. ACM.