Hierarchical Clustering of Music towards Human Mood

Karpagalakshmi R. C.
Lecturer,
Gnanamani College of Technology,
Namakkal(DT) India

D. Rajesh*
Assistant Professor - SITE,
VIT University, Vellore
TamilNadu, India.
extrajesh@rediffmail.com

Abstract- This paper presents the methods for mining the music, based on mood dimension. Mood is an emerging metadata type and access point in music digital libraries (MDL) and online music repositories. There is a growing interest in developing and evaluating Music Information Retrieval (MIR) systems that can provide automated access to the mood dimension of music. Music is nice thing to all. Mood as a music access feature that is not well understood as well as not standardized. To better understanding we develop method to evaluate automated mood access techniques. This paper explore the relationships that mood has with genre, artist and usage metadata. There is an important consistency within the genre-mood and artist-mood relationships. These consistencies lead to us to develop a cluster based approach by creating a relatively small set of data derived. The emotional component of music has been recognized as the most important factor. Music information behavior studies have also identified music mood as an important criterion used by people in music. Music evokes various human emotions or creates music moods through low level musical features. In fact, typical music consists of one or more moods and this can be used as an important factor for determining the similarity between music. In this paper, we propose a new music retrieval scheme based on the mood change pattern.

Keywords: music mood classification, audio features, mood labels.

I. INTRODUCTION

A. Background

General music perception – i.e. how we think and talk about music – is heavily influenced by emotions and context. Consequently, users' music information seeking behavior also reflects the importance of opinion/mood and theme associations for music songs. Searching for music usually is an exploratory and social process, in which people make use of collective knowledge, as well as the opinions and recommendations of other people [1]. Several existing papers aim at automatically inferring additional information from available content as well as (user generated) metadata.

Lu et al. [1] discussed that automatic mood classification can be criticized because the emotional meaning in music is highly subjective. However, they also stressed that there is a certain agreement on the music's mood and they showed that mood classification is possible. In addition, there is a strong application-oriented interest in mood classification: music download services [2] or audio players [3] allow music collection browsing using .mood. as one search criterion. Automatic mood classification could decrease the effort in providing the necessary metadata.
In music psychology and education, the emotional component of music has been recognized as the most strongly associated with music expressivity.

Music information behavior studies have also identified music mood as an important criterion used by people in music seeking and organization. Several experiments have been conducted to classify music by mood. However, a consistent and comprehensive understanding of the implications, opportunities and impacts of music mood as both metadata and content-based access points still eludes the MIR community. Since mood is a very subjective notion, there has yet to emerge a generally accepted mood taxonomy that is used within the MIR research and development community. For example, each of aforementioned studies used different mood categories, making meaningful comparisons between them difficult. In this paper we aim at bridging this gap between users' information needs and indexed music features by developing algorithms for classifying music songs by moods and themes.

There is a growing interest in tackling mood issues in the MIR community as evidenced by the ongoing discussions to establish an "Audio Mood Classification" (AMC) task at the Music Information Retrieval Evaluation eXchange (MIREX), this lack of common understanding is inhibiting progress in developing and evaluating mood-related Access mechanisms. Huron points out that since the preeminent functions of music are social and psychological, the most useful characterization would be based on four types of information: the style, emotion, genre, and similarity [8].

Thus, this paper is intended to general understanding of music mood issues by formally exploring the relationships between: 1) mood and genre; 2) mood and artist; and, 3) mood and recommended usage. It is also intended to contribute more specifically to the MIREX community by providing recommendations on how to proceed in constructing a possible method for conducting an "AMC" task.

Lu et al. [1] set up a mood classification system which defined four mood categories, which were derived from a two dimensional model of affect [4]. For the track selection, Lu et al. followed an expert-based approach: A music excerpt (western classical music) was appended to the ground truth only if three experts agreed on the mood.

In another study, Leman et al. [5] used 15 bipolar adjective pairs as mood descriptions selected by literature scan and trial experiments. Using a factor analysis on the gathered subjective data, they identified an underlying three dimensional space. Then they projected subjective mood assessments of music tracks onto that space and used linear

regression in order to predict these projections with audio features computed from the corresponding music excerpts. Leman et al. used a larger set of mood labels than Lu et al. but they did not try to directly predict the mood labels, but rather their projections in the found three-dimensional space. With respect to track selection, Leman's approach was user based: 20 people were asked to propose music in which they recognize an emotional affect and to describe it, given no constraints about musical style.
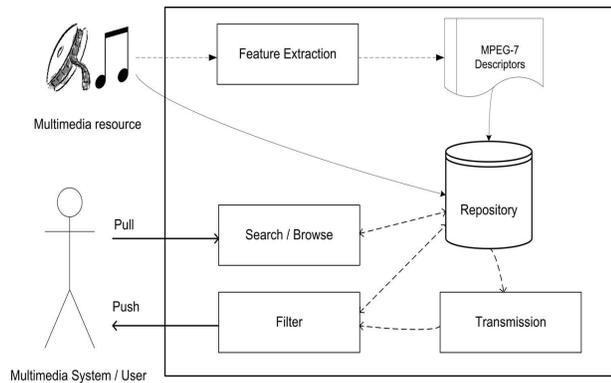

Figure1: Multimedia database architecture

Typically, in a multimedia database system one can distinguish two query scenarios: *pull* and *push*. In a *pull* scenario, a user submits queries to the system and receives a set of descriptions satisfying the constraints of the query. On the other hand, in a *push* scenario, a software agent selects descriptions and performs a set of actions afterwards.

B. Mood Categories

There are 179 mood labels in AMG where moods are defined as "adjectives that describe the sound and feel of a song, album, or overall body of work" 2 and include such terms as "happy", "sad", "aggressive", "stylish", "cheerful", etc. These mood labels are created and assigned to music works by professional editors. Each mood label has its own list of representative "Top Albums" and its own list of "Top Songs". The distribution of albums and songs across these mood lists is very uneven. Some moods are associated with more than 100 albums and songs while others have as few as 3 albums or songs. This creates a data sparseness problem when analyzing all 179 mood labels. There are 3 datasets.
(a). Whole Set: Comprises the entire 179 AMG mood label set. Its "Top Album" lists include 7134 album mood pairs. Its "Top Song" lists include 8288 song mood pairs.
(b). Popular Set: Comprises those moods associated with more than 50 albums and 50 songs. This resulted in 40 mood labels and 2748 album-mood and 3260 song-mood pairs.
(c). Cluster Set: Many albums and songs appear in multiple mood label lists. This overlap can be exploited to group similar mood labels into several mood clusters. Clustering condenses the data distribution and gives us a more concise, higher level view of the mood "space". The set of albums and songs assigned to the mood labels in the mood clusters forms our third dataset set.

II. MOOD CLUSTERING

In order to retrieve the music in effective manner, the clustering algorithm plays the important role. **Clustering** is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters). In our example, the whole music/song/album can be classified as five clusters. In this paper I am using hierarchical clustering algorithm for mood classification. An agglomerative hierarchical Clustering procedure using Ward's criterion [6] was applied to the similarity data.
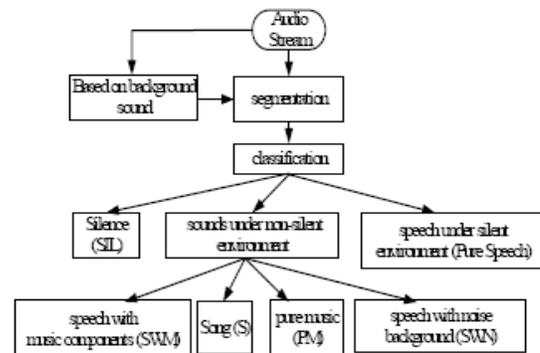
Segmentation and Classification

Figure2: The flowchart of segmentation and classification Algorithm

Figure2 shows the flowchart of proposed audio segmentation and classification algorithm. It is a hierarchical structure. In the first level, a long audio stream can be segmented into some audio clips according to the change of background sound by MBCR based histogram modeling. Then a two level classifier is adopted to hierarchically put the segmented audio clips into six pre-defined categories in terms of discriminative background sounds, which is pure speech (PS), pure music (PM), song (S), speech with music (SWM), speech with noise (SWN) and silence (SIL).

As for audio classification, most studies are focused on speech/music/silence/others separation [11,22]. Scheirer and Slaney [11] proposed to use thirteen features in time, frequency, and cestrum domains and model-based (MAP, GMM, KNN, etc.) classifier, which achieved an accuracy rate over 90% on real-time discrimination between speech and music. Further classification of audio data may take other sounds into consideration besides speech and music. Srinivasan, et al [13] proposed an approach to detect and classify audio that consists of mixed classes such as combinations of speech and music together with environment sounds.

In order to obtain robust and more meaningful clustering results, The AMG dataset provides two views: "Top Albums" and "Top Songs". Thus, we performed the following clustering methods independently on both the "Top Albums" and the "Top Songs" mood list data of the *Popular Set*. First, a co-occurrence matrix was formed such that each cell of the matrix was the number of albums (or songs) shared by two of the 40 "popular" mood labels specified by the coordinates of the cell. Second, an agglomerative hierarchical clustering procedure using Ward's criterion was applied to the similarity data. Third, the resultant two cluster sets (derived from album-mood and song-mood pairs respectively) were examined and found to

have 29 mood labels out of the original 40 that were consistently grouped into 5 clusters at a similar distance level. Table presents the resultant 5 mood clusters along with their constituent mood terms ranked by the number of associated albums.

Table: I

Cluster1	Cluster 2	Cluster3	Cluster4	Cluster5
Rowdy	Amiable /	Literate	Witty	Volatile
Rousing	Good natured	Wistful	Humorous	Fiery
Confident	Sweet	Bittersweet	Whimsical	Visceral
Boisterous	Fun	Autumnal	Wry	Aggressive
Passionate	Rollicking	Brooding	Campy	Tense/anxious
	Cheerful	Poignant	Quirky	Intense
			Silly	

We analyze the relationship of mood to genre, artist and usage using our three datasets. We focus on the "Top Album" lists from each of these sets rather than their "Top Song" lists.

The basic process of hierarchical clustering is this:

A. Start by assigning each item to a cluster, so that if you have N items, you now have N clusters, each containing just one item. Let the distances (similarities) between the clusters the same as the distances (similarities) between the items they contain.
B. Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one cluster less.
C. Compute distances (similarities) between the new cluster and each of the old clusters.
D. Repeat steps 2 and 3 until all items are clustered into a single cluster of size N.

Assume that there are *h* levels in the clustering hierarchy with level 1 being the lowest level and level *h* being the highest. In this clustered environment, communicate the gathered data to level-1 cluster heads (CHs).The level-1 CHs aggregate this data and communicate the aggregated data or estimates based on the aggregated data to level-2 CHs and so on. Finally, the level-h CHs communicate the aggregated data or estimates based on this aggregated data to the processing center.

The algorithm works in a bottom-up fashion. The algorithm first elects the level-1 cluster heads, then level-2 cluster heads, and so on. The level-1 cluster heads are chosen as follows. Each decides to become a level-1 CH with certain probability $p1$ and advertises itself as a cluster head Level-1 CHs then elect themselves as level-2 CHs with a certain probability $p2$

III. MUSIC MOODS AND GENRES

Each album in each individual "Top Album" list is associated with only one genre label. However, an album can be assigned to multiple "Top Album" moods lists. Thus, our genre-mood sample space is all existing combinations of genre and mood labels with each sample being the pairing of one genre and one mood label.

A. All Moods and Genres

There are 3903 unique albums in 22 genres in the *Whole Set*. This set contains 7134 genre-mood pairs, but their distribution across the 22 genres is very skewed with 4564 of them involving the "Rock" genre.

B. Popular Moods and Genres

The 40 mood labels in the *Popular Set* involve 2748 genre-mood pairs. Again, many of the pairs are in the "Rock" genre

C. Mood Clusters and Genres

In the *Cluster Set*, there are 1991 genre-mood cluster combinations, covering 20 genres. Among them, "Rock" albums again occupy a large portion of samples,

IV. MUSIC MOODS AND ARTISTS

Each album on AMG has a "Title" and an "Artist" field. For albums combining tracks by multiple artists, the "Artist" field is filled with "Various Artists". In the following analyses, we eliminated "Various Artists" as this label does not signify a unique analytic unit.

A. All Moods and Artists

There are 2091 unique artists in our *Whole Set*. Some artists contribute as many as over 30 artist-mood pairs each while 871 artists only occur once in the dataset and thus each of them only relates to one mood.

B. Popular Moods and Artists

The *Popular Set* contains 1142 unique artists. 29 of them appear in at least 9 artist-mood pairs, and together contribute 372 artist-mood pairs that form the testing sample space.

C. Mood Clusters and Artists

The *Cluster Set* contains albums by 920 unique artists. Among them, 24 artists who have no less than 8 artist mood pairs form a testing space of 248 artist-mood pairs.

V. MUSIC MOODS AND USAGES

Hu et al. [7] identified interesting relations between the recommended usage labels and music genres and artists as well as relations among the usages themselves. In this section, we explore possible relations between mood and usage. The following usage-mood analyses are based on intersections between our three AMG datasets and our earlier epinions.com dataset which contains 2800 unique albums and 5691 album-usage combinations [7].

A. All Moods and Usages

By matching the title and artist name of each album in our *Whole Set*As each album may have more than one mood label and more than one usage label, we count each combination of existing mood and usage labels of each album as one usagemood sample. There were 1440 usage-mood samples involving 140 mood labels.

B. Popular Moods and Usages

There are 84 common albums in the *Popular Set* and the epinions.com dataset, which yields 527 usage-mood pairs.

C. Mood Clusters and Usages

The usage-mood relationship appears to be much less stable than the genre-mood and artist-mood relationships.

Only 6 of the 11 usages have significant cluster relationships.

VI. CONCLUSION

Previous attempts to associate mood labels to music songs often rely on lyrics or audio information for clustering or classifying song corporas. Using our algorithm, music also becomes searchable by associated themes and
Moods, providing a first step towards effectively searching music by textual, descriptive queries. Automatic mood detection from music has two main benefits. Firstly, having the knowledge of mood in advance can allow for possible enhancement of the music experience (such as mood-based Visualizations) and secondly it makes 'query by mood' from music data-banks possible.

Mood plays an important role in MIR. It is very useful to retrieve the music in effective manner by using hierarchical algorithms. This algorithm will classify the songs in hierarchical order which is related to each other. But there are some drawbacks in hierarchical clustering. The drawbacks are they do not scale well and they can never undo what was done previously. In future we have to rectify this problem by using alternate solutions.

For future work, some of the promising ideas to be further investigated refer to refinements of the moods and themes clusters, as well as to other possible combinations of the audio and tag-based classifiers, i.e. meta classifiers. Can be applied or focused.

VII. REFERENCES

[1] L. Lu, D. Liu, H.-J. Zhang, Automatic Mood Detection and Tracking of Music Audio Signals, IEEE transactions on audio, speech, and language processing, Vol. 14(1), 5-18, 2006.

[2] http://www.allmusic.com

[3] http://www.moodlogic.com

[4] J.A. Russell, A circumplex model of affect, J. Personality & Social Psychology, Vol. 39, 1161-1178, 1980.

[5] M. Leman, V. Vermeulen, L. De Voogdt, D. Moelants, M. Lesaffre, Prediction of Musical Affect Using a Combination of Acoustic Structural Cues, J. of New Music Research, Vol. 34(1), 39-67, 2005.

[6] Berkhin, P. Survey of Clustering Data Mining Techniques, Accrue Software, 2002.

[7] Hu, X., Downie, J. S. and Ehmann, A. F. "Exploiting recommended usage metadata: Exploratory analyses", Proceedings of the 7th International Conference on Music Information Retrieval, ISMIR'06, Victoria, Canada.

[8] D. Huron. Perceptual and cognitive applications in music information retrieval. In International Symposium on Music Information Retrieval, 2000.

[9] Li, T. and Ogihara, M. "Detecting emotion in music", Proceedings of the 4th International. Conference on Music Information Retrieval ISMIR,Washington, D.C. 2003.

[10] Lu, L., Liu, D. and Zhang, H. "Automatic Mood Detection and Tracking of Music Audio Signals", IEEE Transaction on Audio, Speech, and Language Processing, Vol. 14 (1) 2006.

[11] Scheirer E., Slaney M. "Construction and Evaluation of a Robust Multifeature Music/Speech Discriminator", Proc. of ICASSP97, vol II,pp 1331-1334, 1997.

[12] Williams G. and Ellis D. (1999), "Speech/music discrimination based on posterior probability features", Proc European Conf. on Speech Communication and Technology, Budapest, Hungary, pp.687-690,1999.

[13] Srinivasan, S., Petkovic, D., Ponceleon, D., "Towards robust features for classifying audio in the Cue Video System". Proc. of the 7th ACM Intel. Conf. on Multimedia'99, pp.393-400, 1999.