CREATING A SIMPLIFIED MUSIC MOOD CLASSIFICATION GROUND-TRUTH SET

Xiao Hu **Mert Bay** **J. Stephen Downie**
International Music Information Retrieval Systems Evaluation Laboratory
The Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
{xiaohu, mertbay, jdownie}@uiuc.edu

ABSTRACT

A standardized mood classification testbed is needed for formal cross-algorithm comparison and evaluation. In this poster, we present a simplification of the problems associated with developing a ground-truth set for the evaluation of mood-based Music Information Retrieval (MIR) systems. Using a dataset derived from `Last.fm` tags and the USPOP audio collection, we have applied a K-means clustering method to create a simple yet meaningful cluster-based set of high-level mood categories as well as a ground-truth dataset.

1 INTRODUCTION

The emotional component of music has attracted interest in the MIR community, and experiments have been conducted to classify music by mood (e.g., [3]). However, the lack of a standardized mood term set and associated audio datasets that are accessible to the community impedes comparisons among approaches. This research strives to construct a highly simplified, yet reasonable, music mood term set, and to associate this term set with a commonly used audio test collection that could be used as ground truth for a proposed "Audio Music Mood Classification" task in Music Information Retrieval Evaluation eXchange (MIREX)[1].

Deriving a mood term set from the real world practice of music information services (i.e. popular music websites and software) has the advantage of grounding analyses in the realistic social contexts of music seeking and consumption. Besides, such services usually associate tracks with mood related labels or tags which can be used to obtain a ground-truth set. Therefore, we take this approach, basing our analyses on a dataset derived from `Last.fm` and the USPOP collection.

2 USPOP and Last.fm

USPOP is a collection of audio tracks from ~700 CDs collected in 2002 by Dan Ellis of Columbia University. It has been used as in previous MIREX [1] evaluations (e.g., Audio Genre Classification). In order to use the USPOP collection for an Audio Mood Classification MIREX task, its tracks must be tagged with a well grounded set of mood labels.

`Last.fm` is a website collecting public users' input about music such as playlists, favourite artists, and free-text tags. Many of these tags describe feelings inspired by the music pieces, and thus can be used as mood label candidates. Furthermore, user tags on `Last.fm` are available to the public through their web services API.[2] Thus, the `Last.fm` tags, when brought together with the USPOP tracks, provide an opportunity for MIR researchers to derive a mood classification category set associated with an audio dataset.

3 BRINGING TAGS AND AUDIO TOGETHER

3.1 Top Tags on USPOP Tracks

Through the `Last.fm` web services, we obtained tags for tracks in USPOP for which entries on `Last.fm` exist. Because of the popularity of the USPOP tracks, they are covered quite well by the `Last.fm` system. Information regarding 8333 USPOP tracks (out of the total 8764) is available at `Last.fm`, with 6747 USPOP tracks having at least one tag. We collected the top 100 tags for each of these 6747 songs, and resulting in 10178 unique tags.

Adj.	#	Adj.	#	Adj.	#
electronic	27949	instrumental	3664	political	1446
mellow*	15609	progressive	3563	aggressive*	1368
industrial	14002	dark*	3271	powerful*	1365
sad*	9967	vocal	3157	male	1304
awesome	6692	heavy	2701	emotional*	1304
classic	5841	cool*	2246	soft*	1278
relaxing*	4562	slow	2222	sleek	1132
sexy*	4327	experimental	2207	energetic*	1128
upbeat*	4318	melancholy*	2172	classical	1061
romantic*	3828	funny*	1669	calm*	860
happy*	3683	angry*	1496	depressing*	804

Table 1. Most popular adjectives for USPOP tracks

3.2 Adjective Tags on USPOP Tracks

As mood is usually described by adjectives, we used an English part-of-speech (POS) tagger[3] to filter the tags associated with the USPOP tracks. A total of 782 single

[1] http://music-ir.org/mirexwiki
[2] http://www.audioscrobbler.net/data/webservices
[3] http://search.cpan.org/dist/Lingua-EN-Tagger/

word adjective tags remained. Table 1 shows the most popular adjectives and their counts in the dataset. Among them, some are genre terms (e.g., “electronic”, “classic”) and others are non-mood related (e.g., “instrumental”, “female”). We selected 19 terms from the list that we deemed to be associated with music moods (marked with * in Table 1).

3.3 Track Clustering Based on Selected Adjectives

There are 2554 USPOP tracks associated with at least one of the 19 adjectives. Each of the 2554 tracks then was represented by a 19-dimensional binary tag vector. Each dimension corresponds to one of the 19 adjectives and has value 1 if the track is tagged by that adjective or 0 otherwise. K-means clustering [2] using Hamming distance was performed on this space varying the number of clusters from 3 to 12. The resulting clusters partitioned the 2554 tracks into mutually exclusive groups that can serve as a ground-truth set with mood labels. To determine the optimal number of clusters, K-means was performed with 100 random seeds for each number of clusters and the maximum silhouette value for each case was calculated [2]. Among the experiments, the 3-cluster solution resulted in the highest maximum silhouette value, which convinced us to choose 3 clusters (Table 2). Basic statistics of the individual clusters are given in Table 3.

# of clusters	3	4	5	6
Max. silhouette value	0.348	0.316	0.324	0.326

Table 2. Silhouette values for varying cluster numbers

	C.1	**C. 2**	**C.3**
# of tracks	1219	853	482
Intra-cluster variance	8.002	10.402	3.155
Ave. Intra-cluster distance	0.067	0.078	0.069

Table 3. Statistics of the 3 clusters of USPOP tracks

C.1		**C. 2**		**C.3**	
aggressive	90.7%	mellow	98.4%	upbeat	90.0%
angry	89.2%	calm	62.9%	happy	79.3%

Table 4. Adjectives with the highest ratios

3.4 Representative Adjectives in Clusters

After separating the 2554 tracks into 3 clusters, for each adjective we calculated the percentage of the tracks tagged with that adjective in each cluster. The reason we used such percentages instead of raw counts is that the dataset is highly unbalanced such that some adjectives were tagged to nearly 900 tracks whereas some others were tagged to only 70 tracks. Table 4 shows the top adjectives in 3 clusters and their percentages.

The term combinations shown in Table 4 seem reasonable, and these top adjectives in each cluster together can well define the mood nature of the cluster. To further verify the closeness of such top terms in this sample space, we performed Principal Component Analysis (PCA) on the 19 mood adjectives using the 2554 tracks as observations. Figure 1 is the plot of these adjectives based on the first two principal components of this dataset. As it can be seen, top adjectives in each of the 3 clusters are close to each other, which further confirms the clustering results are reasonable.

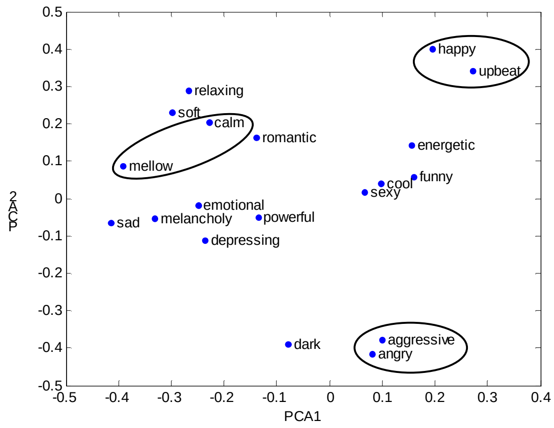

Figure 1. First two principal components of the dataset

4 CONCLUSION AND FUTURE WORK

The 3 mood clusters we derived from the associations between the `Last.fm` user tags and the USPOP audio collection provide a simplified mood ground-truth set that is rooted in the social context of the real world. This set is arguably over-simplified but it is a practical set nonetheless. As such, it should be seen as a starting point for further debates on the construction of mood-based evaluation tasks.

As future work, we will continue to examine datasets from other influential music services such as AllMusicGuide. We will also explore different clustering techniques and will investigate how our approach is generalizable to other contexts.

5 ACKNOWLEDGEMENT

This project is funded by the National Science Foundation (IIS-0327371) and the Mellon Foundation. We also thank the anonymous reviewers and Dr. David Dubin for his statistical consultations.

6 REFERENCES

[1] Downie, J. S. “The Music Information Retrieval Evaluation eXchange (MIREX)”, *D-Lib Magazine 2006, Vol. 12(12).*

[2] Kaufman, L. and Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis,* Wiley, 1990.

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