Mr. Emo: Music Retrieval in the Emotion Plane
Yi-Hsuan Yang, Yu-Ching Lin, Heng-Tze Cheng, Homer Chen
National Taiwan University
1 Roosevelt Rd. Sec.4, Taipei, 10617, Taiwan R.O.C.,
affige@gmail.com, vagante@gmail.com, mikejdionline@gmail.com, homer@cc.ee.ntu.edu.tw

ABSTRACT
This technical demo presents a novel emotion-based music retrieval platform for organizing and browsing music collections, called Mr. Emo. Unlike conventional approaches which quantize emotions into classes, Mr. Emo regards emotions as continuous variables defined by arousal and valence values (AV values) and employs regression models to predict them. Associated with the AV values, each music sample becomes a point in the arousal-valence plane, so a user can easily retrieve music samples of certain emotion(s) by specifying a point or drawing a trajectory in the displayed emotion plane. As being content centric and functionally powerful, such emotion-based retrieval amplifies the power of traditional keyword- or artist-based retrieval, e.g., one can browse the songs of an artist according to emotion. The demo shows the effectiveness of predicting AV values and numerous novel music retrieval methods in the emotion plane.

Categories and Subject Descriptors
H.5.5 [Sound and Music Computing]: systems

General Terms
Algorithms, performance, design, human factors

Keywords
Music information retrieval, emotion recognition, emotion plane

1. INTRODUCTION
Due to the fast growth of digital music collection and media playback on portable devices, effective retrieval and management of music is needed in the digital era. Music classification and retrieval by emotion is a plausible approach, for it is content-centric and functionally powerful.

Various research results have been reported in the field of music emotion recognition (MER) for recognizing the affective content (or evoking emotion) of music signals [1]. A typical approach is to categorize emotions into a number of classes (e.g., happy, angry, sad and relaxing), and apply machine learning techniques to train a classifier. This approach, though widely adopted, brings up the issue of granularity when it comes to practical usage. Obviously, classifying emotions into only a handful of classes cannot meet user demand for easy and effective information access. Using a finer granularity would also face the difficulty to describe emotion in a universal way since language is ambiguous and the description for the same emotion varies from person to person.

In light of the above observations, we propose to view emotions from a continuous perspective and define emotions in terms of arousal (how exciting or calming) and valence (how positive or negative). Therefore, MER becomes the prediction of the arousal and valence values (AV values), which correspond to a point in the two-dimensional arousal-valence emotion plane. A user can then retrieve music samples of certain emotion(s) by specifying a point or drawing a trajectory in the displayed emotion plane, as shown in Fig. 1. In this way, the granularity and ambiguity issues associated with emotion classes or adjectives can be successfully avoided since no categorical classes are needed, and numerous novel emotion-based music organization, browsing and retrieval methods can be easily realized.

In this demo we present such emotion-based music retrieval platform, called Mr. Emo. The critical task of predicting the AV values is accomplished by regression techniques, which has sound theoretical basis and yields satisfying prediction accuracy in our evaluation. We apply the trained regression models to 1000 music samples and design numerous novel retrieval methods in the emotion plane.

2. SYSTEM ARCHITECTURE
The system consists of two main parts as illustrated in Fig. 2: (1) the prediction of AV values using regression models; (2) the emotion-based visualization and retrieval of the music samples.
2.1 Emotion Prediction

Viewing arousal and valence as real values ranging from [-1, 1], we can formulate the prediction of AV values as a regression problem and train two regression models for arousal and valence respectively. Given \( x_i \), \( y_i \) for \( i = 1, \ldots, N \), where \( x_i \) is a feature vector for the \( i \)th input sample, and \( y_i \) is the real value to be predicted, a regression model (regressor) \( R(\cdot) \) is trained to minimize the mismatch (i.e., mean squared difference) between the predicted and ground truth value.

In our implementation, we adopt support vector regression [3] to train regressors since it yields the best prediction accuracy in our previous test [1]. The training set is composed of 60 English pop songs, whose AV values are annotated by 40 participants using the software AnnoEmo [2] in a subjective test. For feature extraction, we apply the toolkit Marsyas [4] to generate 64 timbral feature vectors (spectral centroid, spectral rolloff, spectral flux and MFCC) and 192 MPEG-7 features (spectral flatness and timbral texture features). Feature extraction and AV values prediction are efficient and takes less than five seconds per song. We demonstrate three novel retrieval methods that can be easily realized by Mr. Emo.

Query-by-emotion-point (QBEP). The user can retrieve music of a certain emotion by specifying a point in the emotion plane. The system would then return the music samples whose AV values are closest to the point. This retrieval method is functionally powerful since our criterion of selecting a kind of music to be listened is often related to our emotion state at the moment. In addition, a user can easily discover previously unfamiliar songs which is now organized and browsed according to emotion.

Query-by-emotion-trajectory (QBET). We can also generate a playlist by drawing a free trajectory and indicating the sequence of emotions along the trajectory. For example, as the trajectory in Fig. 1 goes from the first quadrant down to the third quadrant, the emotions of the songs in the playlist also vary accordingly.

Query-by-artist-and-emotion (QBAE). Associated with artist metadata, we can combine the emotion-based retrieval with the conventional artist-based one. As shown in Fig. 3, we can easily visualize the distribution of the music samples of an artist and browse them. With QBAE, we can learn that Sex Pistol usually sings songs of the second quadrant, or retrieve those sad songs sung by Beatles. In addition, QBAE and QBAE can be used in a cooperative way: we can select a song and browse the other songs sung by the same artist by QBAE, or select a song and browse the other songs that sound similar to it using QBEP. We can also recommend similar artists by modeling the distributions of music emotions as Gaussians and measuring similarity by KL distance.

3. SYSTEM DEMONSTRATION

Our music collection consists of 1000 pop songs sung by 52 artists. Feature extraction and AV values prediction are efficient and takes less than five seconds per song. We demonstrate three novel retrieval methods that can be easily realized by Mr. Emo.

Fig. 2. System architecture of Mr. Emo.

Fig. 3. Distributions of the music samples of three famous artists in the emotion plane.

4. REFERENCES


---

1 \( R^2 \) is a standard measurement for regression models. An \( R^2 \) of 1.0 means the model perfectly fits the data, while a negative \( R^2 \) means the model is worse than simply taking the sample mean.

2 Fig. 3 also shows the accuracy of Mr. Emo. The distributions match our common understanding of the styles of these artists.