AUTOMATIC MUSIC MOOD CLASSIFICATION VIA LOW-RANK REPRESENTATION

Yannis Panagakis and Constantine Kotropoulos

Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 54124, GREECE
email: {panagakis,costas}@aiia.csd.auth.gr

ABSTRACT

The problem of automatic music mood classification is addressed by resorting to low-rank representation of slow auditory spectro-temporal modulations. Recently, it has been shown that if each data class is linearly spanned by a subspace of unknown dimensions and the data are noiseless, the lowest-rank representation (LRR) of a set of test vector samples with respect to a set of training vector samples has the nature of being both dense for within-class affinities and almost zero for between-class affinities. Consequently, the LRR exactly reveals the classification of the data, resulting into the so-called *Low-Rank Representation-based Classification (LRRC)*. The performance of the LRRC is compared against three well-known classifiers, namely the Sparse Representations-based Classifier, Support Vector Machines, and Nearest Neighbor classifiers for music mood classification by conducting experiments on the MTV and the Soundtracks180 datasets. The experimental results validate the effectiveness of the LRRC among the classifiers that is compared to.

1. INTRODUCTION

The efficient organization of large music databases is of paramount importance in the era of Web 2.0, since millions of music recordings are nowadays available. The conventional approach employed for music organization and retrieval is based on artist and album information, while the musical genre is often adopted in order to infer semantic similarities between musical recordings. However, music has the ability to convey emotions. Thus, there are circumstances that humans need to access music that match their *mood*, associated to a specific activity, such as relaxing, being active and so on. Therefore, the annotation of music recordings in terms of mood becomes important. Annotating manually the music recordings is not an option, because it is a time consuming and expensive process not to mention the different perception of emotions for the same recording by humans [5, 20]. Consequently, recognizing the perceived emotional content of music automatically turns to be a promising means to enhance music organization, retrieval, and exploration. This task is commonly referred to as automatic music mood classification (AMC). A considerably volume of research in AMC have been done so far. The interest reader may refer to [7, 12]. Depending on the choice of the mood representation to be employed for ground-truth, the available AMC systems can be divided into *dimensional* and *discrete/categorical* [5, 12, 19]. Each model of mood representation is supported by a vast amount of studies in psycology [5, 12].

Dimensional mood models rely on the assertion that different mood states are represented by linear combinations of two or three basic moods. A popular dimensional model is the Thayer mood model, where moods are represented as points in the *Arousal-Valence* plane [21] as depicted in Figure 1 (a). Thayer's model has been recently employed in many AMC systems [4, 25, 26]. Discrete or categorical mood models describe the different mood states by lists of adjectives, usually organized in clusters and train classifiers to predict the overall emotion for a song [11, 15, 22]. Ekman [6] defined six universal emotions (namely anger, disgust, fear, happiness, sadness, and surprise) for facial expressions. However, some of them (e.g., disgust) may not be suitable for music [10]. From a music psychology perspective, Hevner describes the different music mood states by employing a list of 66 adjectives arranged in eight clusters [8] (Figure 1 (b)). Also, five adjective clusters have been employed for mood representation in the Music Information Retrieval Evaluation eXchange (MIREX) AMC task (Figure 1 (c)). A mapping of Hevner's mood clusters, onto Thayer's mood plane has been derived by experts, as depicted in Figure 1 (d) [20]. Although, there are common grounds between theoretical music mood models and listeners' mood perception, the theoretical models do not cover all mood categories emerged from social tagging in music [10]. The latter problem along with the lack of publicly available mood annotated datasets make the evaluation and the comparison of AMC systems hard [10, 11].

A variety of features have been employed in AMC systems. A common choice is to model music by the long-term statistical distribution of short-time features. Such features include timbral texture features, rhythmic features, pitch content, or their combinations and results into a *bag-of-features (BOF)* vector. However, mood is not completely encapsulated within the audio signal. Consequently, the audio features may be complemented by features derived by metadata associated to a music recording including information about artist, genre, and lyrics. Commonly used classifiers are Support Vector Machines (SVM), Nearest-Neighbor (NN) ones, or classifiers, which resort to Gaussian Mixture Models [12].

In this paper, we propose a framework for AMC that can be adapted to both discrete and categorical mood models. There is evidence that the initial perception of audio is performed in the primary auditory cortex, where the audio signal is encoded in terms of its spectral and temporal modulations [17]. Thus, it is reasonable to employ the auditory model proposed in [24] in order to map a given music recording to a four-dimensional (4D) representation of its slow spectral and temporal modulations, the so-called *cortical repre-*

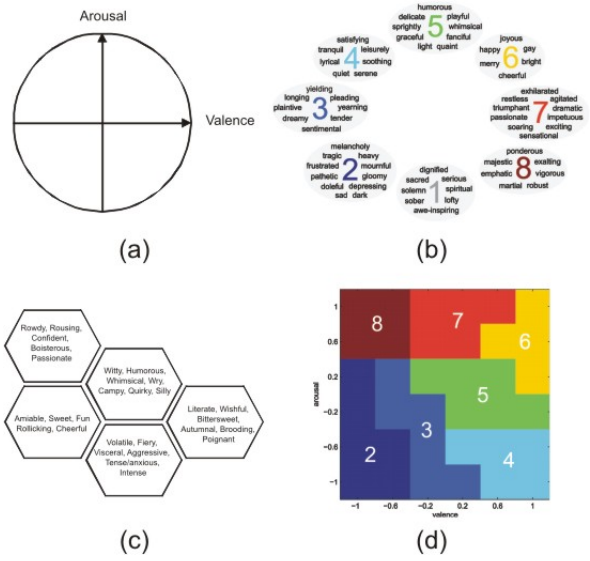

Figure 1: Different mood representations: (a) Thayer's mood plane; (b) Hevner's adjective-cluster model [20]; (c) MIREX adjective cluster model; (d) Mapping of Hevner's mood clusters onto Thayer mood plane [20].

sentation. Cortical representations have been proved a robust alternative to the conventional BOF approach for music genre classification [18]. If sufficient training music recordings are available for each mood class and assuming that each class is linearly spanned by the corresponding cortical representations it is possible to express any test cortical representation as a linear combination of the training representations, where it belongs to. This linear representation can be found by seeking the lowest-rank representation (LRR) of test samples with respect to training samples [14]. The LRR possesses both dense within-class affinities and almost zero between-class affinities. Thus, it reveals exactly the classification of the data, resulting into a novel classification scheme, the so-called *Low-Rank Representation-based Classification (LRRC)*.

The performance of the LRRC in music mood classification is assessed by conducting three sets of experiments in two datasets, namely the *MTV* [20] and the *Soundtracks180* [5] dataset. The MTV dataset is annotated by employing both Thayer's dimensional mood model and Hevner's categorical mood model. The Soundtracks180 dataset contains excerpts from film soundtracks equally distributed among six discrete mood categories. The LRRC is compared against three well-known classifiers, namely the Sparse Representations-based Classifier (SRC) [23], the SVM with a linear kernel, and the NN classifier with cosine distance metric. Experimental results, indicate that the LRRC exhibits the best performance with respect to the classification accuracy, among the classifiers that is compared to for music mood classification.

In summary, the contributions of this paper include:

- The proposal of a general purpose classifier (i.e., the LRRC) that resorts to the lowest-rank representation of the test feature vectors with respect to training feature vectors.
- The proposal of a novel automatic music mood classification framework. This framework resorts to cortical representations for music representation, while the LRRC is employed for music mood classification.

The paper is organized as follows. In Section 2, notation conventions are introduced. The computational auditory model and cortical representation of sound are briefly introduced in Section 3. The LRRC is detailed in Section 4. Experimental results are demonstrated in Section 5, and conclusions are drawn in Section 6.

2. NOTATIONS

Throughout the paper, matrices are denoted by uppercase boldface letters (e.g., $\mathbf{X}, \mathbf{Y}$), vectors are denoted by lowercase boldface letters (e.g., $\mathbf{x}$), and scalars by lowercase letters (e.g., i, μ, ε). The ith column of $\mathbf{X}$ is denoted as $\mathbf{x}_i$. The set of real numbers is denoted by $\mathbb{R}$, while the set of nonnegative real numbers is denoted by $\mathbb{R}_+$.

A variety of norms on vectors and matrices will be used. For example, $\|\mathbf{x}\|_2$ is the ℓ_2 norm of $\mathbf{x}$. The Frobenius norm and the nuclear norm of $\mathbf{X}$ (i.e., the sum of singular values of a matrix) are denoted by $\|\mathbf{X}\|_F$ and $\|\mathbf{X}\|_*$, respectively. The ℓ_∞ norm of $\mathbf{X}$, denoted by $\|\mathbf{X}\|_\infty$, is defined as the element of $\mathbf{X}$ with the maximum absolute value. The trace of $\mathbf{X}$ is denoted $\mathrm{tr}(\mathbf{X})$.

Let $\mathrm{span}(\mathbf{X})$ denote the linear space spanned by the columns of $\mathbf{X}$. Then, $\mathbf{y} \in \mathrm{span}(\mathbf{X})$ denotes that $\mathbf{y}$ belongs to $\mathrm{span}(\mathbf{X})$, and $\mathbf{Y} \in \mathrm{span}(\mathbf{X})$ denotes that all column vectors of $\mathbf{Y}$ belong to $\mathrm{span}(\mathbf{X})$.

3. COMPUTATIONAL AUDITORY MODEL AND CORTICAL REPRESENTATION OF SOUND

The computational auditory model proposed in [24] is inspired by psychoacoustical and neurophysiological investigations in the early and central stages of the human auditory system. An acoustic signal is analyzed by the human auditory model and a 4D representation of sound is obtained, the so-called *cortical representation*. The model consists of two basic stages. The first stage converts the acoustic signal into an auditory representation, the so-called *auditory spectrogram*. This representation is a time-frequency distribution along a logarithmic frequency axis. At the second stage, the spectral and temporal modulation content of the auditory spectrogram is estimated by multiresolution wavelet analysis. The multiresolution wavelet analysis is implemented via a bank of two-dimensional Gaussian filters, that are selective to different spectro-temporal modulation parameters ranging from slow to fast temporal rates (in Hertz) and from narrow to broad spectral scales (in Cycles/Octave), which results in a 4D representation of time, frequency, rate, and scale. Mathematical formulation and details about the auditory model and the cortical representation of sound can be found in [16].

Psychophysiological evidence justifies the choice of *scales* $\in \{0.25, 0.5, 1, 2, 4, 8\}$ (Cycles / Octave) as well as both positive and negative *rates* $\in \{\pm 2, \pm 4, \pm 8, \pm 16, \pm 32\}$ (Hz) to represent the sound spectro-temporal modulations. The cochlear model, employed in the first stage, has 128 filters with 24 filters per octave, covering $5\frac{1}{3}$ octaves along the tonotopic axis. For each music recording, the extracted 4D cortical representation is averaged along time and the average rate-scale-frequency 3D cortical representation is thus obtained. By vectorizing the 3D cortical representation, each music recording is finally represented by $\mathbf{x} \in \mathbb{R}_+^{7680}$.

4. CLASSIFICATION VIA LOW-RANK REPRESENTATION

In pattern analysis and machine learning, an underlying assumption is that the high-dimensional data have some type of intrinsic structure that enables their low-dimensional representation and efficient processing. For instance, Principal Component Analysis [9] and Robust Principal Component Analysis [3] are based on the assumption that data are approximately drawn from a *single* low-rank linear subspace. However in practice, it is more reasonable to assume that the data are drawn from a *mixtrure* or *union* of *several* low-rank linear independent subspaces. Such an assumption is valid in many real-world cases [3, 14, 23]. By adopting the aforementioned assumption and building on recent theoretical investigations on low-rank representation (LRR) [14], we propose a novel classification scheme that finds the lowest-rank representation of new (test) samples subject to a given matrix of training samples.

Let $\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2|\ldots|\mathbf{X}_k] \in \mathbb{R}^{m\times n}$ be a set of n training samples $\mathbf{x}_j \in \mathbb{R}^m$, $j = 1,2,\ldots,n$ that belong to k classes, exactly drawn form a union of k independent linear subspaces of unknown dimensions. The columns of $\mathbf{X}_i \in \mathbb{R}^{m\times n_i}$ correspond to the n_i training samples from the ith subspace. Furthermore, let us denote by $\mathbf{Y} = [\mathbf{Y}_1|\mathbf{Y}_2|\ldots|\mathbf{Y}_k] \in \mathbb{R}^{m\times p}$ the matrix that contains in its columns p new (test) samples, with the columns of $\mathbf{Y}_i \in \mathbb{R}^{m\times p_i}$ refers p_i test samples that belong to the ith class. By assuming that: 1) the data are drawn from independent linear subspaces (i.e., span$(\mathbf{X}_i)$ linearly spans the ith class data space, $i = 1,2,\ldots,k$), 2) $\mathbf{Y} \in$ span$(\mathbf{X})$, and 3) the data contain neither outliers nor noise, then each test vector sample that belongs to the ith class can be represented as a linear combination of the training samples in $\mathbf{X}_i$. That is, $\mathbf{Y}_i = \mathbf{X}_i\mathbf{Z}_i$ with $\mathbf{Z}_i \in \mathbb{R}^{n_i\times p_i}$. Accordingly, $\mathbf{Y} = \mathbf{X}\mathbf{Z}$, where $\mathbf{Z} = \text{diag}[\mathbf{Z}_1, \mathbf{Z}_2, \ldots \mathbf{Z}_k] \in \mathbb{R}^{n\times p}$ is a block-diagonal matrix. Therefore, the lth test sample can be represented as $\mathbf{y}_l = \mathbf{X}\mathbf{z}_l \in \mathbb{R}^m$, where $\mathbf{z}_l = [\mathbf{0}^T|\ldots|\mathbf{0}^T|\mathbf{z}_i^T|\mathbf{0}^T|\ldots|\mathbf{0}^T]^T \in \mathbb{R}^n$ is the augmented coefficient vector, whose elements are zero except those associated with the ith class. Consequently, having found such a block-diagonal matrix $\mathbf{Z}$ capturing both dense within-class affinities and zero between-class affinities, the classification of the data is exactly revealed.

Following [14], and under the aforementioned three assumptions, the block-diagonal matrix $\mathbf{Z} \in \mathbb{R}^{n\times p}$ is the lowest-rank representation of the test data $\mathbf{Y} \in \mathbb{R}^{m\times p}$ with respect to training data $\mathbf{X} \in \mathbb{R}^{m\times n}$ or equivalently the solution of the optimization problem:

$$\underset{\mathbf{Z}}{\text{argmin}} \quad \text{rank}(\mathbf{Z}) \quad \text{subject to } \mathbf{Y} = \mathbf{X}\,\mathbf{Z}. \tag{1}$$

The optimization problem (1) does not have a unique solution and it is difficult to be solved due to the discrete nature of the rank function. However, the rank function can be replaced by the nuclear norm resulting to the convex optimization problem:

$$\underset{\mathbf{Z}}{\text{argmin}} \quad \|\mathbf{Z}\|_* \quad \text{subject to } \mathbf{Y} = \mathbf{X}\,\mathbf{Z}. \tag{2}$$

Liu *et al.* [14] have proved that the optimal solution of (2) is *unique* and it is also a solution of (1). Although the solution of (2) can be obtained in closed form [14], it is not stable in many cases due to numerical issues. In order to overcome this problem, we propose to solve the convex problem (2) iteratively. In this paper, we choose the Augmented Lagrange Multiplier (ALM) [1, 13] method due to its simplicity and the good convergence properties. To this end, (2) is converted to the equivalent

$$\underset{\mathbf{Z},\mathbf{J}}{\text{argmin}} \quad \|\mathbf{J}\|_* \quad \text{subject to } \mathbf{Y} = \mathbf{X}\,\mathbf{Z},\ \mathbf{Z} = \mathbf{J}, \tag{3}$$

which can be solved by minimizing the augmented Lagrange function:

$$\begin{aligned} f(\mathbf{Z},\mathbf{J}) \quad &= \quad \|\mathbf{J}\|_* + \text{tr}\left(\mathbf{\Lambda}_1^T(\mathbf{Y}-\mathbf{X}\mathbf{Z})\right) + \text{tr}\left(\mathbf{\Lambda}_2^T(\mathbf{Z}-\mathbf{J})\right) \\ &+ \quad \frac{\mu}{2}\left(\|\mathbf{Y}-\mathbf{X}\mathbf{Z}\|_F^2 + \|\mathbf{Z}-\mathbf{J}\|_F^2\right), \end{aligned} \tag{4}$$

where $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2$ are the Lagrange multipliers and $\mu > 0$ is a penalty parameter. The minimization of (4) with respect to $\mathbf{Z}$ and $\mathbf{J}$ can be performed in an alternating fashion by first fixing $\mathbf{Z}$ and updating $\mathbf{J}$, then fixing $\mathbf{J}$ and updating $\mathbf{Z}$, and finally updating the Lagrange multipliers. The inexact ALM method for the minimization of (2) is outlined in Algorithm 1, which is a special case of Algorithm 1 in [14]. Step 3 of the Algorithm 1 can be solved via the Singular Value Thresholding operator [2]. The convergence of Algorithm 1 can be proved as in [13]. The computational cost of Algorithm 1 is comparable to that of a linear SVM, since the most demanding step of Algorithm 1 is the Step 3, which involves the computation of an SVD.

Algorithm 1 Solving (2) by inexact ALM

Input: Training matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ and test matrix $\mathbf{Y} \in \mathbb{R}^{m\times p}$.
Output: Matrix $\mathbf{Z} \in \mathbb{R}^{n\times p}$.

1: Initialize: $\mathbf{Z} = \mathbf{J} = \mathbf{0}$, $\mathbf{\Lambda}_1 = \mathbf{0}$, $\mathbf{\Lambda}_2 = \mathbf{0}$, $\mu = 10^{-6}$, $\varepsilon = 10^{-2}$.
2: **while** not converged **do**
3: Fix $\mathbf{Z}$ and update $\mathbf{J}$ by $\mathbf{J} = \text{argmin}\ \frac{1}{\mu}\|\mathbf{J}\|_* + \frac{1}{2}\|\mathbf{J} - (\mathbf{Z} + \mathbf{\Lambda}_2/\mu)\|_F^2$.
4: Fix $\mathbf{J}$ and update $\mathbf{Z}$ by $\mathbf{Z} = \left(\mathbf{I} + \mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{Y} + \mathbf{J} + (\mathbf{X}^T\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2)/\mu\right)$.
5: Update the Lagrange multipliers by $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_1 + \mu(\mathbf{Y} - \mathbf{X}\mathbf{Z})$, $\mathbf{\Lambda}_2 = \mathbf{\Lambda}_2 + \mu(\mathbf{Z} - \mathbf{J})$.
6: Update μ by $\mu = \max(\mu, 10^6)$.
7: Check convergence conditions $\|\mathbf{Y} - \mathbf{X}\mathbf{Z}\|_\infty < \varepsilon$ and $\|\mathbf{Z} - \mathbf{J}\|_\infty < \varepsilon$.
8: **end while**

The lth test sample $\mathbf{y}_l \in \mathbb{R}^m$ can be classified as follows. Ideally, the lth column of $\mathbf{Z}$ (i.e., $\mathbf{z}_l \in \mathbb{R}^n$) contains non-zero entries in positions associated with the columns of the training matrix $\mathbf{X}$ stemming from a single class so that we can easily assign $\mathbf{y}_l$ to that class. However, due to modeling errors, there are small non-zero entries in $\mathbf{z}_l$ that are associated to multiple classes. To cope with this problem, each test sample $\mathbf{y}_l$ is classified to the class that minimizes the ℓ_2 norm residual between $\mathbf{y}_l$ and $\hat{\mathbf{y}}_i = \mathbf{X}\,\delta_i(\mathbf{z}_l)$, where $\delta_i(\mathbf{z}_l) \in \mathbb{R}^n$ is a new vector whose nonzero entries are the entries in $\mathbf{z}_l$ that are associated to the ith class only. The procedure is outlined in Algorithm 2.

5. EXPERIMENTAL EVALUATION

In order to assess the performance of the LRRC, experiments were conducted by employing two mood annotated datasets.

Algorithm 2 Low-Rank Representation-based Classification

Input: Training matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ and test matrix $\mathbf{Y} \in \mathbb{R}^{m \times p}$.
Output: A class label for each column of $\mathbf{Y}$.

1: Solve (2) by employing Algorithm 1 and obtain $\mathbf{Z} \in \mathbb{R}^{n \times p}$.
2: **for** $l = 1$ to p **do**
3: **for** $i = 1$ to k **do**
4: Compute the residuals $r_i(\mathbf{y}_l) = \|\mathbf{y}_l - \mathbf{X}\,\delta_i(\mathbf{z}_l)\|_2$.
5: **end for**
6: class$(\mathbf{y}_l) = \text{argmin}_i\, r_i(\mathbf{y}_l)$.
7: **end for**

The first dataset contains 195 full music recordings with total duration 14.2 h from the *MTV Europe Most Wanted Top Ten* of 20 years (1981-2000), covering a wide variety of popular music genres. The ground-truth was obtained by five annotators (four males and one female) who were asked to make a forced binary decision according to the two dimensions in Thayer's mood plane (i.e., assigning either $+1$ or -1 for arousal and valence respectively) according their mood perception. Based on the mean arousal and the mean valence values, each recording can be described as single point in Thayer's mood plane, and thus can be assigned to a Hevner mood cluster [20] by employing the mapping depicted in Figure 1 (d). The dataset is abbreviated as *MTV* hereafter. The second dataset is a subset of the *Soundtracks* dataset [5] (abbreviated as the *Soundtracks180*) with 180 excerpts from film soundtracks. Excerpts have duration between 10 and 30 sec and are equally distributed among six discrete mood categories, such as happiness, sadness, fear, anger, surprise, and tenderness. Soundtracks180 was annotated by 12 expert musicologists who had all studied a musical instrument for at least 10 years.

All the recordings were converted to monaural wave format at a sampling frequency of 16 kHz and quantized with 16 bits. Moreover, the audio signals have been normalized, so that they have zero mean amplitude with unit variance in order to remove any factors related to the recording conditions. The cortical representations were extracted by the middle 30 sec of each recording in the MTV dataset and by employing the whole music excerpt for the Soundtracks180 dataset. The experimental results, presented below, were obtained by employing stratified 10-fold cross-validation. The LRRC is compared with three classifiers, namely the SRC, SVM with linear kernel, and NN with cosine distance metric. Due to the assumed subspace structure of cortical representations, both linear SVM, SRC, and NN are appropriate for separating features from different music recordings. Furthermore, the aforementioned classifiers are working in the same feature space with the LRRC, which makes possible compare their performance fair.

Two sets of experiments were conducted in the MTV dataset. In the first set, the ground-truth is obtained by employing the Thayer's mood model, while in the second one the Hevner's mood model has been adopted. The two dimensions in Thayer's mood plane, can be treated as being independent of each other. Therefore classification can reasonably be done independently [20] by making binary decisions between excitation and calmness on the arousal dimension and negativity and positivity in the valence dimension, respectively. By adopting Hevner's mood model, the task is to classify the music recordings into seven mood clusters. Following [20], and by allowing for slight variations in subjective perception, classifier predictions on the true mood cluster or its two direct neighbors are considered to be correct. The classification results are summarized in Table 1. The last three rows of Table 1 include the classification results obtained by Schuller *et al.* [20] by employing audio features (without feature selection) and an SVM with a linear kernel.

Table 1: Classification accuracies on the MTV dataset by employing Thayer's and Hevner's mood models.

Classifier/Reference	Mood Model	Accuracy (%)
LRRC	Thayer, Arousal	68.28
	Thayer, Valence	61.43
	Hevner	64.57
SRC	Thayer, Arousal	64.49
	Thayer, Valence	61.75
	Hevner	56.50
SVM	Thayer, Arousal	65.24
	Thayer, Valence	57.51
	Hevner	61
NN	Thayer, Arousal	61.94
	Thayer, Valence	59.08
	Hevner	59.65
[20]	Thayer, Arousal	71.80
	Thayer, Valence	60.50
	Hevner	65.40

In Table 2, the classification accuracies obtained by conducting experiments on the Soundtracks180 dataset are presented.

Table 2: Classification accuracies on the Soundtracks180 dataset.

Classifier	Accuracy (%)
LRRC	39.44
SRC	39.44
SVM	37.22
NN	33.88

By inspecting Tables 1 and 2, the LRRC clearly exhibits the best performance, with respect to the classification accuracy, among the classifiers that is compared to, with respect to stratified 10-fold cross-validation. Both the LRRC and the SRC exhibit better performance in Valence prediction compared to the system proposed in [20] while in Arousal prediction our results are inferior to that reported in [20]. The latter may attributed to the cortical representations which do not depend on extensive feature selection applied to the BOF employed in [20]. However, the classification accuracy obtained by the LRRC, when the Hevner's model has been adopted in annotation of the MTV dataset, is comparable to that reported in [20], motivating further research. Furthermore, the best classification accuracy on the Soundtracks180 dataset obtained the LRRC (i.e., 39.44 %) is quite acceptable since the cortical representations extracted by very short excerpts that may be not able to capture accurately the mood aspects of music. Currently, there are not other published results in AMC by employing this dataset.

Generally speaking, although our initial assumptions are quite restrictive, since should be the data drawn from a union of low-rank linear independent subspaces, when the subspaces are low-rank and the data vectors are high-dimensional, as in our case, the assumption of independence is roughly equal to the disjoint assumption. That is, the intersection of every two subspaces is the null set. The latter

assumption is more realistic and possibly justifies the success of the LRRC on the AMC.

6. CONCLUSIONS

A novel automatic music mood classification framework has been proposed. This framework resorts to cortical representations for music representation, while a novel classifier, namely the LRRC, has been proposed for music mood classification. The performance of the LRRC is assessed by conducting three sets of experiments on two datasets annotated by one dimensional (i.e., Thayer's model) and two categorical mood models. The LRRC exhibits the best performance, with respect to the classification accuracy, among the classifiers that is compared to, when applied to the music mood classification task.

ACKNOWLEDGMENT

This work has been supported by HRAKLEITOS II research project, co-funded by the European Social Fund and National Resources.

REFERENCES

[1] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, Belmont, MA, 2nd edition, 1996.

[2] J. F. Cai, E. J. Candes, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal Optimization*, 2(2):569–592, 2009.

[3] E. J. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *arXiv:0912.3599v1*, 2009.

[4] T. Eerola, O. Lartillot, and P. Toiviainen. Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In *Proc. 10th Int. Symposium Music Information Retrieval*, pages 621–626, Kobe, Japan, 2009.

[5] 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, 2011.

[6] P. Ekman. *Emotion in the Human Face*. Cambridge University Press, 2nd edition, 1982.

[7] Z. Fu, G. Lu, K. M. Ting, and D. Zhang. A survey of audio-based music classification and annotation. *IEEE Trans. Multimedia*, 2010.

[8] K. Hevner. Experimental studies of the elements of expression in music. *American Journal of Psychology*, 48(2):246–268, 1936.

[9] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(1933):417–441,498–520, 1933.

[10] X. Hu. Music and mood: Where theory and reality meet. In *Proc. 11th 5th iConference*, University of Illinois, Urbana-Champaign, USA, 2010.

[11] X. Hu, S. J. Downie, C. Laurier, M. Bay, and A. F. Ehmann. The 2007 MIREX audio mood classification task: Lessons learned. In *Proc. 9th Int. Symposium Music Information Retrieval*, Philadelphia, USA, 2008.

[12] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull. Music emotion recognition: A state of the art review. In *Proc. 11th Int. Symp. Music Information Retrieval*, pages 255–266, Utrecht, The Netherlands, 2010.

[13] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical Report UILU-ENG-09-221, 2009.

[14] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *arXiv:1010.2955v3*, 2010.

[15] L. Lu, D. Liu, and H. J. Zhang. Automatic mood detection and tracking of music audio signals. *IEEE Trans. on Audio, Speech, and Language Processing*, 14(1):5 – 18, 2006.

[16] N. Mesgarani, M. Slaney, and S. A. Shamma. Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations. *IEEE Trans. Audio, Speech, and Language Processing*, 14(3):920–930, May 2006.

[17] R. Munkong and J. Biing-Hwang. Auditory perception and cognition. *IEEE Signal Processing Magazine*, 25(3):98–117, 2008.

[18] Y. Panagakis and C. Kotropoulos. Music genre classification via topology preserving nonnegative tensor factorization and sparse representations. In *Proc. 2010 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pages 249–252, Dallas, Texas, 2010.

[19] B. Schuller, J. Dorfner, and G. Rigoll. Determination of nonprototypical valence and arousal in popular music: Features and performances. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010(2010):19 pages, 2010.

[20] B. Schuller, C. Hage, D. Schuller, and G. Rigoll. "Mister D.J., Cheer Me Up!': Musical and textual features for automatic mood classification. *Journal of New Music Research*, 39(1):13–34, 2010.

[21] R. E. Thayer. *The Biopsychology of Mood and Arousal*. Oxford University Press, Boston, USA, 1989.

[22] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas. Multilabel classification of music into emotions. In *Proc. 9th Int. Symp. Music Information Retrieval*, pages 325–330, Philadelphia, USA, 2008.

[23] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2009.

[24] X. Yang, K. Wang, and S. A. Shamma. Auditory representations of acoustic signals. *IEEE Trans. Information Theory*, 38(2):824–839, 1992.

[25] Y. H. Yang and H. H. Chen. Ranking-based emotion recognition for music organization and retrieval. *IEEE Trans. on Audio, Speech, and Language Processing*, 19(4):762–774, 2011.

[26] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. Chen. A regression approach to music emotion recognition. *IEEE Trans. on Audio, Speech, and Language Processing*, 16(2):448–457, 2008.