# MIREX 2012: MOOD CLASSIFICATION TASKS SUBMISSION

**Renato Panda**
CISUC
University of Coimbra
panda@dei.uc.pt

**Rui Pedro Paiva**
CISUC
University of Coimbra
ruipedro@dei.uc.pt

**ABSTRACT**

In this work, three audio frameworks – Marsyas, MIR Toolbox and PsySound3, were used to extract audio features from the audio samples. These features are then used to train several classification models, resulting in the different versions submitted to MIREX 2012 mood classification task.

## 1. INTRODUCTION

The classification system is built on research results obtained during the current year [1-2]. First, several audio features are extracted from the existing dataset. To this end, Marsyas [3] framework as well as MIR Toolbox [4] and PsySound3 [5] are employed. The remaining process of training a classifier and predicting labels for the test sets are conducted in MATLAB, using Support Vector Machines with the libSVM library [6].

The first classification approach uses a SVM model trained to differentiate between the five existing clusters. A second approach was submitted, using five different classifiers (regression), each one trained to identify a specific cluster. Finally, another approach extends this strategy, adding a second level of classification to the system.

A variation of these systems uses a previously obtained subset of features, instead of selecting the entire feature space that was extracted.

## 2. FEATURE EXTRACTION

Over the decades, several authors in other fields have studied the relations between music and emotion. Still, many of these relations are still unclear and further research is needed to implement computation models able to capture such characteristics.

In this work, three audio frameworks were used to extract features from the existent audio files: Marsyas, a fast framework coded in C++; MIR Toolbox, an integrated set of functions written in MATLAB, that are specific to the extraction of musical features and provide a high number of both low and high-level audio features; PsySound3, a MATLAB toolbox for the analysis of sound recordings using physical and psychoacoustical algorithms.

The process results in a total of 312 features: 124 obtained with Marsyas, 177 with MIR Toolbox (using statistics such as mean, standard deviation, kurtosis and skewness) and 11 with PsySound3, based on the 15 best features identified in [7]. A brief description of these features is presented in Table 1.

| Framework | Features |
|---|---|
| Marsyas (124) | Centroid, rolloff, flux, Mel frequency cepstral coefficients (MFCCs), Peak Ratio – Chroma. |
| MIR Toolbox (177) | Among others: Root mean square (RMS) energy, rhythmic fluctuation, tempo, attack time and slope, zero crossing rate, rolloff, flux, high frequency energy, Mel frequency cepstral coefficients (MFCCs), roughness, spectral peaks variability (irregularity), inharmonicity, pitch, mode, harmonic change and key. |
| PsySound3 (11) | Loudness, sharpness, timbral width, spectral and tonal dissonances, pure tonalness, multiplicity. |

**Table 1.** Used frameworks and respective features.

## 3. EMOTION CLASSIFICATION

The emotion classification problem was approached with three different solutions: a single classifier; an ensemble using five classifiers; and a hierarchical strategy using three regressors in the first level as well as a second level where a simple classifier is used. All these use support vector machines based on the libSVM implementation.

### 3.1 Single SVM Classifier

A single classifier is trained using the radial basis function kernel (RBF) to predict one of the five existent clusters, as studied in [2].

### 3.2 Classifiers Ensemble

A more complex solution was submitted, since it achieved better results in previous studies [1], [8]. The process consists in using one regression model (SVR) for each cluster in a total of five. Each model is trained with the songs from his cluster, labeled as 1, and an equal amount of songs from the other clusters, labeled as 0. During the testing phase, each new song is passed to the system and the regressor/cluster with the highest result is selected.

### 3.3 Hierarchical Classification Model

A third strategy tries to reduce the semantic and acoustic ambiguity problems between clusters 1-5 and 2-4 that were previously identified [9]. To this end, a hierarchical system is employed, with the first level using an ensemble of three regression models to select between groups of clusters 1-5, 2-4 and cluster 3. In cases where groups 1-5 and 2-4 are selected, a second level of classification is used to distinguish between the two using a single SVM classifier.

### 3.4 Feature Selection and SVM parameters

In addition to the usage of the entire feature set to train each model, a subset of features were also built, obtained with ReliefF and RReliefF [10] selection algorithm and a "MIREX-like" dataset, organized in a similar way [1-2]. Furthermore, better parameters for each SVM model (cost, and $\gamma$, as well as $\varepsilon$ for regression) are obtained using grid search.

## 4. ACKNOWLEDGMENTS

This work was supported by the MOODetector project (PTDC/EIA-EIA/102185/2008), financed by the Fundação para Ciência e a Tecnologia (FCT) and Programa Operacional Temático Factores de Competitividade (COMPETE) – Portugal.

## 5. REFERENCES

[1] R. Panda and R. P. Paiva, "Music Emotion Classification: Analysis of a Classifier Ensemble Approach," in *5th International Workshop on Machine Learning and Music (in ICML)*, 2012, pp. 1–2.

[2] R. Panda and R. P. Paiva, "Music Emotion Classification: Dataset Acquisition and Comparative Analysis," in *15th International Conference on Digital Audio Effects (DAFx-12)*, 2012.

[3] G. Tzanetakis, "Manipulation, Analysis and Retrieval Systems for Audio Signals," Science and Technology. Princeton University, 2002.

[4] O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio," in *Proc. 10th Int. Conf. on Digital Audio Effects*, 2007, pp. 237–244.

[5] D. Cabrera, S. Ferguson, and E. Schubert, "'Psysound3': Software for Acoustical and Psychoacoustical Analysis of Sound Recordings," in *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, 2007, pp. 356–363.

[6] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *Computer*. pp. 1–30, 2001.

[7] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen, "A Regression Approach to Music Emotion Recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 448–457, Feb. 2008.

[8] J. Wang, H. Lo, and S. Jeng, "Mirex 2010: Audio Classification Using Semantic Transformation And Classifier Ensemble," in *Proc. of The 6th International WOCMAT & New Media Conference (WOCMAT 2010)*, 2010, pp. 2–5.

[9] C. Laurier and P. Herrera, "Audio music mood classification using support vector machine," in *MIREX task on Audio Mood Classification*, 2007, pp. 2–4.

[10] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1–2, pp. 23–69, 2003.