TOWARDS BI-DIMENSIONAL CLASSIFICATION OF SYMBOLIC MUSIC BY AFFECTIVE CONTENT

António Pedro Oliveira and Amílcar Cardoso
Centre for Informatics and Systems
University of Coimbra

ABSTRACT

The work described in this paper is part of a project that aims to implement and assess a computer system that can control the affective content of the music output, in such a way that it may express an intended emotion. In this system, music classification is done with the help of a knowledge base with weighted mappings between continuous affective dimensions (valence and arousal) and music features (e.g., rhythm and melody) grounded on results from works of Music Psychology.

The system starts with the segmentation of MIDI music to increase the probability to obtain music that express only one kind of affective content. Then, feature extraction algorithms are applied to label these segments with music metadata (e.g., rhythm and melody). The mappings of the knowledge base are used to label music with affective metadata. According to the prediction results of listeners' affective answers, the subsets of features (and their weights) present in the knowledge base are being refined.

1. INTRODUCTION

Music has been widely accepted as one of the languages that convey affective content. The possibility to select music with an appropriate affective content can be helpful to adapt music to our emotional interest. However, only recently scientists have tried to quantify and explain how music expresses emotions. As a result of this, mappings are being established between emotions conveyed by the music and musical features [12] [8].

Our work intends to design a system that may select music with appropriate affective content by taking into account a knowledge base with mappings of that kind. Automated classification using machine learning approaches has the advantage of allowing one to perform classifications in a faster and more reliable way than manual classifications. So, we intend to improve the knowledge base by selecting prominent features and by defining appropriate weights. This is done, respectively, by using feature selection and linear regression algorithms.

The automatic selection of music according to an emotional description has a great application potential, namely in entertainment and healthcare. On the one hand, this system can be used in the selection of soundtracks for movies, arts, dance, deejaying, theater, virtual environments, computer games and other entertainment activities. On the other hand, it can be used in music therapy to promote psychological and physical healing. The next section makes a review of some of the most relevant contributions from Music Psychology and related works from Music Information Retrieval. Section 3 gives an overview of the system for classification. Section 4 presents the details of the experiments. Section 5 shows the experimental results, and finally section 6 makes some final remarks.

2. RELATED WORK

This work entails an interdisciplinary research involving Music Psychology and Music Information Retrieval. This section makes a review of some of the most relevant contributions for our work from these areas.

2.1. Music Psychology

Schubert [12] studied relations between emotions and musical features (melodic pitch, tempo, loudness, texture and timbral sharpness) using a 2 Dimensional Emotion Space. This study was focused on how to measure emotions expressed by music and what musical features have an effect on arousal and valence of emotions. Likewise, Korhonen [4] tried to model people perception of emotion in music. Models to estimate emotional appraisals to musical stimuli were reviewed [12] [6] and system identification techniques were applied. Livingstone and Brown [8] provided a summary of relations between music features and emotions, in a 2 Dimensional Space, based on some research works of Music Psychology. Gabrielsson and Lindstrom [3] is one of these works where relations between happiness and sadness, and musical features are established. Lindstrom [7] analysed the importance of some musical features (essentially melody, but also rhythm and harmony) in the expression of appropriate emotions.

2.2. Music Information Retrieval

The selection of the classifier model and the feature set is crucial to obtain good results in the detection of emotions in music. Van de Laar [13] compared 6 emotion detection methods in audio music based on acoustical feature analysis. Four central criteria were used in this comparison: precision, granularity, diversity and selection. There are also methods to extract segments of audio music with specific emotional expressions[15]. The method designed by Wu and Jeng consisted in 3 steps: collection of subject responses, data processing and segments extraction. This method associates emotional content to musical fragments, according to some musical features like pitch, tempo and mode.

Muyuan and Naiyao [10] made an emotion recognition system to extract musical features from MIDI music. Support Vector Machines were used to classify music in 6 types of emotions (e.g., joyous and sober). Both statistical (e.g., pitch, interval and note density) and perceptual (e.g., tonality) features were extracted from the musical clips. There are also models to recommend MIDI music based on emotions [5]. The model of Kuo et al., based on association discovery from film music, proposed prominent musical features according to a queried emotional description. These features were compared with features extracted from a music database (chord, rhythm and tempo). Then, music was ranked and a music list was recommended according to 15 groups of emotions.

3. SYSTEM DESCRIPTION

The system described in this paper is part of a project that has the objective of implementing and assessing a computer system that can control the affective content of the music output, in such a way that it may express an intended emotion. The system uses a database of pre-composed music represented at a symbolic level. We intend to accomplish the mentioned objective in 2 stages. The first consists in the selection / classification of music by affective content and it is the focus of this paper. The second stage will deal with the transformation of the affective content of selected music to approximate as far as possible its affective content to the intended emotion. Transformations of mode from minor to major to increase valence or increase of tempo to increase arousal may occur. Figure 1 presents an overview of different stages that compose our system for music classification. We will describe each of them in the following paragraphs.

3.1. Music segmentation

The expression of emotions in music varies as a function of time [4]. To facilitate classification, it is very important to obtain segments of music that express only one kind of affective content. Our segmentation module uses the Local Boundary Detection Model (LBDM) [1][2] to obtain weights based on the strength of music variations (pitch, rhythm and silence). These weights establish plausible points of segmentation between segments with different musical features that may reflect different affective content. We define a threshold to reduce the search space among the LBDM weights. This threshold is equal to 1.30*mean(LBDM weights)+1.30*standard deviation(LBDM weights). We obtain music chunks of different length with a minimum of notes *MinN* and a maximum of notes *MaxN*. To segment, we start at the beginning of the MIDI file and look for a plausible point of segmentation that corresponds to the maximum weight between the beginning of MIDI file+*MinN* and the beginning of MIDI file+*MaxN*. This process is repeated starting from the last point of segmentation until we come to the end of the MIDI file.

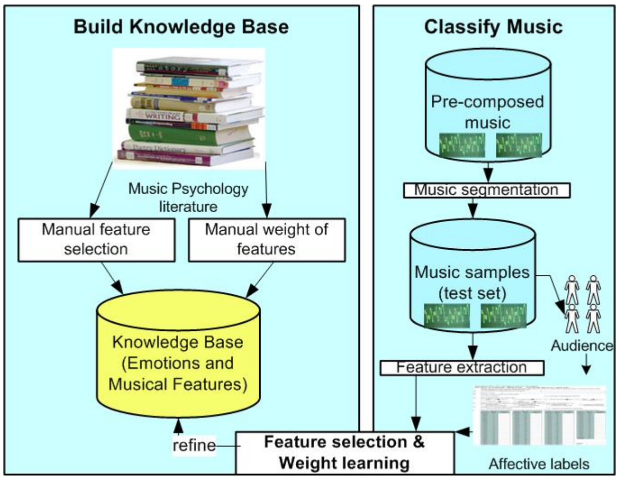

Figure 1. System overview

3.2. Music features extraction

The extraction of features was constrained by the available features in third party software: JSymbolic [9] and MIDI toolbox [2]. At this moment a set of 106 uni-dimensional features is being analysed. These features can be categorized in 7 groups: melody, rhythm, instrumentation, harmony, dynamics, pitch and texture. This metadata is used to label MIDI files. Special attention was devoted to music features considered important for emotional expression in works of Music and Emotions Psychology [11].

3.3. Knowledge base

The Knowledge Base (KB) comprises mappings between emotions and musical features grounded on research works of Music Psychology (section 2.1.). A table with mappings between two dimensions of affective states (valence and arousal) and high level musical features (instrumentation, dynamics, rhythm, melody and harmony) proposed in [11] is being used in the process of selecting the features and defining respective weights in the knowledge base. This is done separately for each affective dimension. Features are selected according to the importance given in the literature [11][8]. A positive or negative tentative weight is defined according to the positive or negative effect and degree of influence of each of the features in each of the dimensions. For instance, considering the weight x in $\Re$: $x \in [-1;1]$, register has a direct relationship with the valence of music, so a weight of 0.5 is given; tempo has a great direct relationship with the arousal of music, so a weight of 1.0 is given; mode (from minor to major) has a great direct relationship with the valence of music, so a weight of 1.0 is given.

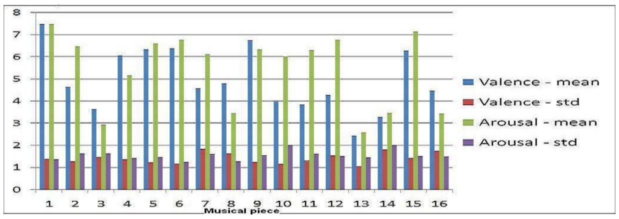

Figure 2. Mean and standard deviations of the emotional responses for valence and arousal

The emotional output of each music is calculated through a weighted sum of the features, with the help of a vector of weights for each affective dimension:

$$Valence = \sum_{i=0}^{n} valenceWeight_i * feature_i \qquad (1)$$

$$Arousal = \sum_{i=0}^{n} arousalWeight_i * feature_i \qquad (2)$$

4. DETAILS OF THE EXPERIMENT

For the experiments with listeners, we used a test set of 16 musical pieces. These pieces were of western tonal music (pop and r&b) and last, approximately, from 20 seconds to 1 minute. 53 different listeners were asked to label online the test set [1]. The obtained affective labels were used to refine the sets of features and corresponding weights derived from music psychology literature with experimentally-derived weighted mappings. This was done separately for the valence and arousal.

Concerning feature selection, several subsets of features from a set of 106 features were analysed for the affective dimensions with the help of the affective labels obtained for the test set and information from the literature [11]. This was done by applying feature selection algorithms: genetic search, best-first and greedy stepwise [14]. With the subsets of features selected, some algorithms of linear regression were used to refine the weights of each feature. Linear regression, SMO regression and SVM regression [14] were tested. The results of the next section were obtained with SVM regression, because it was, generally, the approach that gave us the best results.

5. RESULTS

Figure 2 shows the mean and standard deviation for emotional responses obtained in the online questionnaire [1]. The Cronbach's Alpha for these responses is 74.81%. Answers from listeners distant more than the mean ± 2*standard deviation (considered as outliers) were discarded.

We analysed the importance of specific subset of features: instrumentation (18 features), texture (14 features), rhythm (28 features), dynamics (4 features), pitch (22 features), melody (17 features) and harmony (2 features). The importance of individual features in each of these subsets was also analysed to have any idea of what are the most important features, as well as to avoid wrong inferences because of the lack (e.g., harmony and dynamics) or excess (e.g., rhythm and pitch) of features per group.

Features	Valence Cor. Cf.	Valence Det. Cf.	Valen. P-Val.	Arousal Cor. Cf.	Arousal Det. Cf.	Arous. P-Val.
Instrument.	94.33%	88.98%	0.001	80.07%	64.11%	0.001
Texture	74.57%	55.61%	0.001	80.30%	64.48%	0.001
Rhythm	98.26%	96.55%	0.001	99.70%	99.40%	0.001
Dynamics	31.70%	10.05%	0.2	59.44%	35.33%	0.02
Pitch	66.80%	44.62%	0.005	85.30%	72.76%	0.001
Melody	74.79%	55.94%	0.001	77.16%	59.53%	0.001
Harmony	37.21%	13.85%	0.2	42.91%	18.41%	0.1

Table 1. Prediction by groups of features - valence and arousal

Features	Cr. Coef.	Dt. Coef.	P-Val.
String Ensemble Fraction (-)	50.00%	25.00%	0.05
Saxophone Fraction (+)	36.45%	13.28%	0.2
Electric Guitar Fraction (+)	30.85%	9.52%	0.2
Avg. Number of Independent Voices (+)	39.84%	15.87%	0.1
Var. Number of Independent Voices (+)	38.25%	14.63%	0.1
Variability Note Duration (-)	65.70%	43.16%	0.005
Note Density (+)	56.68%	32.13%	0.02
Polyrhythms (-)	53.56%	28.69%	0.04
Average Note Duration (-)	52.05%	27.09%	0.04
Average Time Between Attacks (-)	52.02%	27.06%	0.04
Variation of Dynamics Each Voice (+)	22.00%	4.84%	0.4
Relative Strength of Top Pitches (+)	36.70%	13.47%	0.2
Relative Strength of Top Pitch Classes (+)	34.94%	12.21%	0.2
Importance of Middle Register (+)	32.80%	10.76%	0.2
Melodic Tritones (+)	47.08%	22.17%	0.05
Common Melodic Interval Prevalence (-)	37.30%	13.91%	0.2
Relative Strength Common Intervals (+)	37.00%	13.69%	0.2
Key mode (-)	23.31 %	5.43%	0.4

Table 2. Best features of each group - valence

5.1. Valence

Tables 1 and 2 present prediction results by groups of features for valence. From this, we can infer that rhythmic (e.g, variability of note duration, note density and polyrhythms), instrumentation (e.g., string ensemble fraction), melodic (e.g., melodic tritones) and texture features (e.g., average number of independent voices) are relevant to the valence of music.

Using the best features, the correlation and determination coefficients for training on the whole set were, respectively, 80.30% and 64.48%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 75.98% and 57.73%. The best features (rhythmic) and their weights were: -0.45*avg. time between attacks + 0.11*note density - 0.54*variability of note duration.

5.2. Arousal

Tables 1 and 3 present prediction results by groups of features for arousal. From this, we can infer that rhyth-

[1] http://student.dei.uc.pt/%7Eapsimoes/PhD/Music/icmc08/index.html

Features	Cr. Cf.	Dt. Cf.	P-Val.
Number of Unpitched Instruments (+)	59.89%	35.87%	0.02
Percussion Prevalence (+)	53.00%	28.09%	0.04
Range of Highest Line (-)	51.24%	26.26%	0.04
Var. Number Independent Voices (+)	44.93%	20.19%	0.1
Average Time Between Attacks (-)	73.43%	53.92%	0.001
Average Note Duration (-)	71.94%	51.75%	0.002
Note Density (+)	68.02%	46.27%	0.005
Variability of Time Between Attacks (-)	65.11%	42.39%	0.005
Strength Strongest Rhythmic Pulse (-)	61.14%	37.38%	0.01
Variation of Dynamics (+)	49.10%	24.11%	0.05
Avg. Note2Note Dynamics Change (+)	45.50%	20.70%	0.1
Importance of High Register (-)	54.92%	30.16%	0.02
Primary Register (-)	50.31%	25.31%	0.05
Stepwise Motion (-)	39.57%	15.66%	0.1
Most Common Melodic Interval (+)	35.79%	12.81%	0.2
Key mode (-)	13.02%	1.69%	0.5

Table 3. Best features of each group - arousal

mic (e.g., avg. note duration and note density), dynamics (e.g., variation of dynamics), instrumentation (e.g., numb. of unpitched instruments) and pitch features (e.g., importance of high register) are relevant to the arousal of music.

Using the best features, the correlation and determination coefficients for training on the whole set were, respectively, 91.67% and 84.03%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 81.85% and 66.99%. The best features and their weights were: -0.56*avg. note duration - 0.31*avg. time between attacks - 0.45*high register importance + 0.06*note density + 0.05*dynamics variation.

6. CONCLUSION

We presented a preliminary work that undertake music emotion classification as a regression problem. SVM regression obtained the best results in the prediction and classification of the dimensions of valence and arousal. Validation results using the coefficient of determination showed that the prediction/classification of arousal (84.03% /66.99%) is easier than the prediction/classification of valence (64.48%/57.73%). Rhythmic features proved to be very important to valence and arousal (e.g., time between attacks, note density and avg./variation of note duration). Dynamics (e.g., loudness variation) and pitch features (high register) were also important to predict the arousal.

Regarding the instrumentation not too much can be concluded because of the lack of musical pieces with similar instruments. Moreover, more instrumentation features are needed (e.g., chromatic percussion and analysis of the frequency spectrum of samples). It is also important to implement some features of dynamics (e.g., avg. loudness) for arousal prediction and harmony (e.g., consonance and chords) for valence prediction. Concerning the texture and melodic features there is the need of more tests. Therefore, it is our objective to extend this study to a statistical significant number of music files.

7. REFERENCES

[1] Cambouropoulos, E. ”The local boundary detection model (lbdm) and its application in the study of expressive timing.”, *Int. Computer Music Conf.*, 2001.

[2] Eerola, T., Toiviainen, P. ”Mir in matlab: The midi toolbox.”, *International Conference on Music Information Retrieval*, 2004.

[3] Gabrielsson, A., Lindström, E. ”The Influence Of Musical Structure On Emotional Expression.” *Music and emotion: Theory and research.* Oxford University Press, 2001, 223–248.

[4] Korhonen, M. ”Modeling continuous emotional appraisals of music using system identification.” *Master's thesis*, University of Waterloo, 2004.

[5] Kuo, F., Chiang, M., Shan, M., Lee, S. ”Emotion-based music recommendation by association discovery from film music.” *ACM International Conference On Multimedia*, 2005, 507–510

[6] Li, T., Ogihara, M. ”Detecting emotion in music.” *Int. Conf. Music Information Retr.*, 2003, 239–240

[7] Lindstrom, e. ”A Dynamic View of Melodic Organization and Performance.” *PhD thesis*, Acta Universitatis Upsaliensis Uppsala, 2004.

[8] Livingstone, S., Brown, A. ”Dynamic response: real-time adaptation for music emotion.” *Australasian Conference On Interactive Entertainment*, 2005, 105–111.

[9] McKay, C., Fujinaga, I. ”jsymbolic: A feature extractor for midi files.” *International Computer Music Conference*, 2006.

[10] Muyuan, W., Naiyao, Z., Hancheng, Z. ”User-adaptive music emotion recognition.” *International Conference on Signal Processing*, 2004.

[11] Oliveira, A., Cardoso, A. ”Towards affective-psychophysiological foundations for music production.” *Affective Computing and Intelligent Interaction*, 2007, 511–522.

[12] Schubert, E. ”Measurement and Time Series Analysis of Emotion in Music.” *PhD thesis*, University of New South Wales, 1999.

[13] van de Laar, B. ”Emotion detection in music, a survey.” *Twente Student Conference on IT*, 2006.

[14] Witten, I., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S. ”Weka: Practical machine learning tools and techniques with java implementations.” *International Conference on Neural Information Processing*, 1999, 192–196.

[15] Wu, T., Jeng, S.: ”Extraction of segments of significant emotional expressions in music.” *Workshop on Computer Music and Audio Technology*, 2006.