Emotion Detection Using Feature Extraction Tools

Jacek Grekow(✉)

Faculty of Computer Science, Bialystok University of Technology,
Wiejska 45A, 15-351 Bialystok, Poland
j.grekow@pb.edu.pl

Abstract. This paper presents an analysis of the effect of features obtained from 3 different audio analysis tools on classifier accuracy during emotion detection. The research process included constructing training data, feature extraction, feature selection, and building classifiers. The obtained results indicated leaders among feature extraction tools used during classifier building for each emotion. An additional result of the conducted research was obtaining information on which features are useful in the detection of particular emotions.

Keywords: Music emotion recognition · Audio feature extraction · Audio analysis tools · Music information retrieval

1 Introduction

It cannot be denied that listening to music has an emotional character. Detecting the emotions contained in music is one of the main causes of listening to it [1]. In the era of the Internet, searching music databases for emotions has become increasingly important. Automatic emotion detection enables indexing files in terms of emotions [2].

This paper presents the use of 3 different audio analysis tools (Marsyas, jAudio and Essentia) during emotion detection. The positives of this experiment were: we gained experience using these tools; we learned their strengths and weaknesses; we got insight into their construction and terms of use; and we checked their usefulness in emotion detection. Another result of this experiment was we extracted information on which features are useful during the detection of each emotion.

Music emotion detection studies are mainly based on two popular approaches: categorical or dimensional. The categorical approach [3–5] describes emotions with a discrete number of classes - affective adjectives. In the dimensional approach [6,7], emotions are described as numerical values of valence and arousal.

There are several other studies on the issue of emotion detection with the use of different audio tools for musical feature extraction. Studies [4,8] used a collection of tools that use the Matlab environment called MIR toolbox [9]. Feature extraction library jAudio [10] was used in studies [5,7]. Feature sets

F. Esposito et al. (Eds.): ISMIS 2015, LNAI 9384, pp. 267–272, 2015.
DOI: 10.1007/978-3-319-25252-0_29

extracted from PsySound [11] were used in study [6], while study [12] used the Marsyas framework [13]. The Essentia [14] library for audio analysis was used in study [15].

There are also papers devoted to the evaluation of audio features for emotion detection within one program. Song et al. [4] explored the relationship between musical features extracted by MIR toolbox and emotions. They compared the emotion prediction results for four sets of features: dynamic, rhythm, harmony, and spectral features. A comprehensive review of the methods that have been proposed for music emotion recognition was prepared by Yang et al. [16].

2 Music Data Sets

In this research, we use four emotion classes: e1 (energetic-positive), e2 (energetic-negative), e3 (calm-negative), e4 (calm-positive). They cover the four quadrants of the two-dimensional Thayer model of emotion [17]. They correspond to four basic emotion classes: happy, angry, sad, and relaxed.

To conduct the study of emotion detection, we prepared two sets of data. One set was used for building one common classifier for detecting the 4 emotions, and the other data set for building four binary classifiers of emotion in music. Both data sets consisted of six-second fragments of different genres of music: classical, jazz, blues, country, disco, hip-hop, metal, pop, reggae, and rock. The tracks were all 22050 Hz mono 16-bit audio files in .wav format. Music samples were labeled by the author of this paper, a music expert with a university musical education. Six-second music samples were listened to and then labeled with one of the emotions (e1, e2, e3, e4). In the case when the music expert was not certain which emotion to assign, such a sample was rejected. In this way, each file was associated with only one emotion.

The first training data set for emotion detection consisted of 324 files, with 81 files for each emotion (e1, e2, e3, e4). We obtained the second training data from the first set. It consisted of 4 sets of binary data. For example, data set for binary classifier e1 consisted of 81 files labeled as e1 and 81 files labeled as not e1 (27 files each from e2, e3, e4). In this way, we obtained 4 binary data sets (consisting of examples of e and not e) for 4 binary classifiers e1, e2, e3, e4.

3 Feature Extraction Using Audio Analysis Tools

With the Marsyas [13], the following features can be extracted: Zero Crossings, Spectral Centroid, Spectral Flux, Spectral Rolloff, Mel-Frequency Cepstral Coefficients (MFCC), and chroma features. For each of these basic features, Marsyas calculates four statistic features (the mean of the mean, the mean of the standard deviation, the standard deviation of the mean, the standard deviation of the standard deviation).

The following features are implemented in jAudio [10]: Zero Crossing, Root Mean Square, Fraction of Low Amplitude Frames, Spectral Centroid, Spectral

Flux, Spectral Rolloff, Spectral Variability, Compactness, Mel-Frequency Cepstral Coefficients (MFCC), Beat Histogram, Strongest Beat, Beat Sum, Strength of Strongest Beat, Linear Prediction Coefficients (LPC), Method of Moments (Statistical Method of Moments of the Magnitude Spectrum), Area Method of Moments. jAudio also calculates metafeatures, which are the feature templates that automatically produce new features from existing ones. jAudio provides three basic metafeature classes (mean, standard deviation, and derivative), which are also combined to produce two more metafeatures (derivative of the mean and derivative of the standard deviation).

We used version 2.0.1 of Essentia [14], which contains a number of executable extractors computing music descriptors for an audio track: spectral, time-domain, rhythmic, tonal descriptors. Essentia also calculates many statistic features: the mean, geometric mean, power mean, median of an array, and all its moments up to the 5th-order, its energy and the root mean square (RMS). To characterize the spectrum, flatness, crest and decrease of an array are calculated. Variance, skewness, kurtosis of probability distribution, and a single Gaussian estimate are calculated for the given list of arrays.

The previously prepared, labeled by emotion, music data sets served as input data for 3 tools used for feature extraction. The obtained lengths of feature vectors, dependent on the package used, were as follows: Marsyas - 124 features, jAudio - 632 features, and Essentia - 471 features.

4 Results

4.1 The Construction of Classifiers

We built classifiers for emotion detection using the WEKA package [18]. During the construction of the classifier, we tested the following algorithms: J48, RandomForest, BayesNet, IBk (K-nn), SMO (SVM). The classification results were calculated using a cross validation evaluation CV-10.

The first important result was that during the construction of the classifier for 3 data sets obtained from Marsyas, jAudio and Essentia, the highest accuracy among all tested algorithms was obtained for SMO algorithm. SMO was trained using polynominal kernel. The second best algorithm was RandomForest.

Table 1. Accuracy obtained for SMO algorithm

	Marsyas	jAudio	Essentia
Accuracy before attribute selection	55.24 %	58.95 %	59.26 %
Accuracy after attribute selection	58.95 %	**67.90 %**	**64.50 %**

The results obtained for SMO algorithm are presented in Table 1. The result (classifier accuracy) improved after applying attribute selection (attribute evaluator: WrapperSubsetEval, search method BestFirst). The best results were

obtained after applying attribute selection for data from jAudio (67.90 %) and Essentia (64.50 %).

Table 2. The most important features obtained from jAudio and Essentia

Tool	Selected features
jAudio	Strongest Beat, MFCC, LPC (Linear Prediction Coefficients), Statistical Method of Moments of the Magnitude Spectrum
Essentia	Energy of the Erbbands, MFCC, Onset Rate, Beats Loudness Band Ratio, Key Strength, Chords Histogram

The most important features obtained from jAudio and Essentia are presented in Table 2. In both cases, such features as MFCC and those pertaining to rhythm confirmed their usefulness in emotion detection; they are present in the obtained features from jAudio as well as Essentia. What distinguishes Essentia's set of features is the use of tonal features (Key Strength, Chords Histogram); and what distinguishes the features selected from jAudio is the use of statistical moments of the magnitude spectrum.

4.2 The Construction of Binary Classifiers

Once again the best results were obtained for SMO algorithm. The results are presented in Fig. 1. The best results were obtained for emotion e2 (91 %) regardless of the type of audio analysis tools. It is difficult to unequivocally select the best audio analysis tools used for features extraction. For detection of emotion e1, the best results were obtained using Essentia (80.86 %). For detection of emotion e2, all tools had the same results (approx. 91 %). For detection of emotion e3, the best results were obtained using the tools jAudio and Essentia (87 %), and for emotion e4 - Marsyas (82.71 %).

Essentia achieved the best results since in three cases the obtained classifier accuracy was the highest (for e1, e2, e3); the remaining tools achieved the

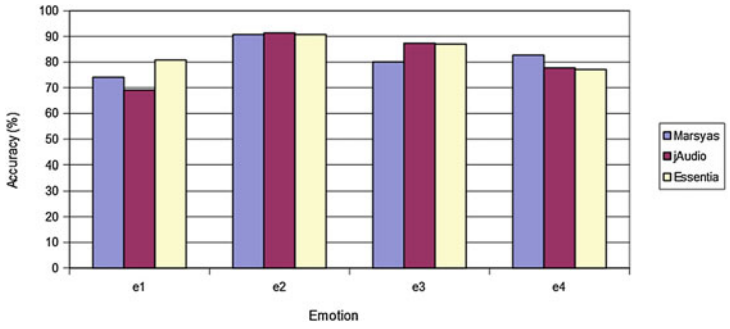

Fig. 1. Classifier accuracy for emotions e1, e2, e3, and e4 obtained for SMO

best results in two cases: Marsyas (e2, e4) and jAudio (e2, e3). The obtained binary classifier accuracy results were higher (15–24 percentage points) than the accuracy of one classifier recognizing four emotions. Table 3 presents the most important features obtained after feature selection for Essentia features for each emotion. In each features set, we had representatives of low-level, rhythm features, even though we had different sets for each emotion. Only in the case of classifier e4, tonal features were not used. The energies of the bands are important for e1, e2, and e4 classifiers, but they differ in to which bands they pertain: e1 - Barkbands, e2 - Erbbands, and Melbands, e4 - Barkbands, and Erbbands. High Frequency Content, which is characterized by the amount of high-frequency content in the signal is important for e3 and e4 classifiers. Beats Loudness Band Ratio (the beat's energy ratio on each band) is very important for emotion detection because it is used in all sets. Another important feature was the tonal feature: Chords Histogram, which was used by e2 and e3.

Table 3. Selected features used for building binary classifiers using Essentia

Classifier	Selected features
e1	Energy of the Barkbands, Onset Rate, Beats Loudness Band Ratio, Key Strength
e2	Average Loudness, Dissonance, Energy of the Erbbands and the Melbands, MFCC, Beats Loudness Band Ratio, Chords Changes, Chords Histogram
e3	High Frequency Content, Silence Rate, Spectral Energy Band Middle Low, Beats Loudness Band Ratio, Key Strength, Chords Histogram
e4	Energy of the Barkbands, Energy of the Erbbands, High Frequency Content, Pitch Salience, Beats Loudness Band Ratio

5 Conclusions

This paper presents an analysis of the effect of features obtained from different audio analysis tools on classifier accuracy during emotion detection. The research process included constructing training data, feature extraction, feature selection, and building classifiers. The collected data allowed comparing different tools during emotion detection. The obtained results indicated leaders among feature extraction tools used during classifier building for each emotion. Only the use of several different tools achieves high classifier accuracy (80–90 %) for all basic emotions (e1, e2, e3, e4). An additional result of the conducted research was obtaining information on which features are useful in the detection of particular emotions. The obtained results present a new and interesting view of the usefulness of different feature sets for emotion detection.

Acknowledgments. This paper is supported by the S/WI/3/2013.

References

1. Krumhansl, C.L.: Music: a link between cognition and emotion. Am. Psychol. Soc. **11**(2), 45–50 (2002)
2. Grekow, J., Raś, Z.W.: Emotion Based MIDI Files Retrieval System. In: Raś, Z.W., Wieczorkowska, A.A. (eds.) Advances in Music Information Retrieval. SCI, vol. 274, pp. 261–284. Springer, Heidelberg (2010)
3. Grekow, J., Raś, Z.W.: Detecting emotions in classical music from MIDI files. In: Rauch, J., Raś, Z.W., Berka, P., Elomaa, T. (eds.) ISMIS 2009. LNCS, vol. 5722, pp. 261–270. Springer, Heidelberg (2009)
4. Song, Y., Dixon, S., Pearce, M.: Evaluation of musical features for emotion classification. In: Proceedings of the 13th International Society for Music Information Retrieval Conference (2012)
5. Xu, J., Li, X., Hao, Y., Yang, G.: Source separation improves music emotion recognition. In: ACM International Conference on Multimedia Retrieval (2014)
6. Yang, Y.-H., Lin, Y.-C., Su, Y.-F., Chen, H.H.: A regression approach to music emotion recognition. IEEE Trans. Audio, Speech, Language Process. **16**(2), 448–457 (2008)
7. Lin, Y., Chen, X., Yang, D.: Exploration of music emotion recognition based on midi. In: Proceedings of the 14th International Society for Music Information Retrieval Conference (2013)
8. Saari, P., Eerola, T., Fazekas, G., Barthet, M., Lartillot, O., Sandler, M.: The role of audio and tags in music mood prediction: a study using semantic layer projection. In: Proceedings of the 14th International Society for Music Information Retrieval Conference (2013)
9. Lartillot, O., Toiviainen, P.: MIR in Matlab (II): A toolbox for musical feature extraction from audio. In: International Conference on Music Information Retrieval, pp. 237–244 (2007)
10. McKay C., Fujinaga I., Depalle P.: jAudio: a feature extraction library. In: Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 05), pp. 600–603 (2005)
11. Cabrera, D.: PSYSOUND: a computer program for psychoacoustical analysis. In: Proceedings of the Australian Acoustical Society Conference, pp. 47–54 (1999)
12. Grekow, J.: Mood tracking of radio station broadcasts. In: Andreasen, T., Christiansen, H., Cubero, J.-C., Raś, Z.W. (eds.) ISMIS 2014. LNCS, vol. 8502, pp. 184–193. Springer, Heidelberg (2014)
13. Tzanetakis, G., Cook, P.: Marsyas: a framework for audio analysis. Organized Sound **10**, 293–302 (2000)
14. Bogdanov, D., Wack N., Gomez E., Gulati S., Herrera P., Mayor O., Roma G., Salamon J., Zapata J., Serra X.: ESSENTIA: an audio analysis library for music information retrieval. In: Proceedings of the 14th International Conference on Music Information Retrieval, pp. 493–498 (2013)
15. Laurier, C.: Automatic Classification of Musical Mood by Content-Based Analysis. Ph.D. thesis, UPF, Barcelona, Spain (2011)
16. Yang, Y.-H., Chen, H.H.: Machine recognition of music emotion: a review. ACM Trans. Intell. Sys. Technol. **3**(3), 61 (2012). Article No. 40
17. Thayer, R.E.: The Biopsychology Arousal. Oxford University Press, Cambridge (1989)
18. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2005)