ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Survey on Classification Techniques for Music Mood

Rasika Sonawane, Dr. Prachi Joshi

M.E. Student, Dept. of Computer Engineering, Savitribai Phule Pune University, Pune, India [1]

Assistant Professor, Dept. of Computer Engineering, Savitribai Phule Pune University, Pune, India [2]

ABSTRACT: The popularity of internet, downloading and purchasing music from online music shops is growing dramatically. As an intimate relationship presents between music and human emotions, we often choose to listen a song that suits our mood at that instance. Often, automatic methods are needed to classify music by moods even from the uploaded music files in social networks. The paper discusses about the techniques and features used for the classification task. In this survey we will see some of the work done by the researchers on the same techniques.

KEYWORDS: timbre features, modulation features, temporal features, SVM Classifier, K-NN Classifier, GMM Classifier

I. Introduction

In the past few years, research in Music Information Retrieval has been very active. It has produced automatic classification methods in order to deal with the amount of digital music available. A relatively recent problem is the automatic mood classification of music consisting in a system taking the waveform of a musical piece and outputting text labels describing the mood in the music (as happy, sad, etc...). It has already been demonstrated that audio-based techniques can achieve satisfying results to a certain extent. Using a few simple mood categories and carefully checking for reliable agreements between people, automatic classification based on audio features gives promising results. Psychological studies have shown initially at the Music Technology Group that part of the semantic information of songs resides exclusively in the lyrics. This means that lyrics can contain relevant emotional information that is not included in the audio.

Music can be concordant or discordant; this is known from the physics of wave propagation, and the note systems which have emerged across the world reflect this. Very discordant sounds are perceived negatively by any human and, indeed, by some other. Expert knowledge can say only so much since expertise tends to be for a particular genre of music, or for Western music theory, much of which does not apply to music from other parts of the world.

Music information retrieval is a growing field with focus on automatically extracting information from musical sources for analysis. The musical source comes in many formats including written score as well as audio. A variety of machine learning and statistical analysis techniques are applied. Work in the field of music information retrieval has discovered features for predicting genre, determining key and tempo of music, distinguishing instruments, analyzing the similarity of music, transcribing to score from audio and eliciting musical information from written scores.

As it is a well established fact that music indeed has an emotional quotient attached with it, it is very essential to know what are the intrinsic factors present in music which associate it with a particular mood or emotion. A lot of research has been done and still going on in capturing various features from the audio file based on which we can analyze and classify a list of audio files. Audio features are nothing but mathematical functions calculated over the audio data, in order to describe some unique aspect of that data. In the last decades a huge number of features were developed for the analysis of audio content.

II. Related work

Liu D et al. [1] presented a high-accuracy audio classification algorithm based on SVM-UBM using MFCCs as classification features. MFCCs are extracted in frame level, then Gaussian Mixture Model (UBM) is employed to integrate sequences of frame-level MFCCs within a clip to form the clip-level feature and audio classification is done using SVM with these clip-level features.

Ruijie Zhang et al. [2] presented a high-accuracy audio classification algorithm based on SVM-UBM using MFCCs as classification features. MFCCs are extracted in frame level, then Gaussian Mixture Model (UBM) is employed to integrate

DOI: 10.15680/IJIRCCE.2016. 0411132

sequences of frame-level MFCCs within a clip to form the clip-level feature and audio classification is done using SVM with these clip-level features. Russel (1980) proposed the circumplex model of affect based on the two dimensional model. These two dimensions are denoted as "pleasant", "Unpleasant" and "arousal sleep". There are 28 affect words in Russel's circumplex models and are shown in Figure 1 Later Thayer(1989) adapted Russel's model using the two dimensional energy stress model. Different researchers used their own taxonomies which are the subsetsof Russel's taxonomy.

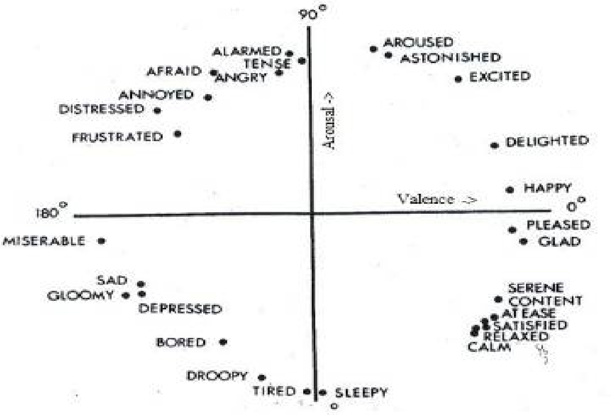


Figure 1. Russell's circumplex model of 28 affects words

C.-H. Lee et al. [3] proposed an automatic music genre classification approach based on long-term modulation spectral analysis of spectral (OSC and MPEG-7 NASE) as well as cepstral (MFCC) features. Modulation spectral analysis of every feature value will generate a corresponding modulation spectrum and all the modulation spectra can be collected to form a modulation spectrogram which exhibits the time-varying or rhythmic information of music signals. Each modulation spectrum is then decomposed into several logarithmically-spaced modulation sub-bands. The modulation spectral contrast (MSC) and modulation spectral valley (MSV) are then computed from each modulation sub-band. Effective and compact features are generated from statistical aggregations of the MSCs and MSVs of all modulation sub-bands.

X Sun et al. [4] proposed method using a framework named information cell mixture models (ICMM) to automate the task of music emotion classification. Designed system has potential application in both unsupervised concept learning as well as supervised classification learning. The system is acceptable for music mood classification because emotion is a vague concept and has a cognitive structure. The application of ICMM is also suitable for music emotion classification.

P. Dunker et al. [5] authors given a technical solution for automated slideshow generation by extracting a set of high-level features from music, like beat grid, mood and genre and intelligently combining this set with image high-level features. An advantage of this high-level concept is to enable the user to incorporate his preferences regarding the semantic aspects of music and images. For example, the user might request the system to automatically create a slideshow, which plays soft music and shows pictures with sunsets from the last 10 years of his own photo collection. The high-level feature extraction on both, the audio and the visual information is based on the same underlying machine learning core, which processes different audio- and visual- low- and mid-level features. Authors also described the technical realization and evaluation of the algorithms with suitable test databases.

E. E. P. Myint et al. [6] designed a self-colored music mood segmentation and a hierarchical framework based on new mood taxonomy model to automate the task of multi-label music mood classification. The developed mood taxonomy model combines Thayer's 2 Dimension (2D) models and Schubert's Updated Hevner adjective Model (UHM) to mitigate the probability of error causing by classifying upon maximally 4 class classification from 9. The verse and chorus parts approximately 50 to 110 sec of the whole songs is exerted manually as input music trims in this system. Consecutive self colored mood is segmented by the image region growing method. The extracted feature sets from these segmented music pieces are ready to inject the Fuzzy Support Vector Machine (FSVM) for classification. One-against-one (O-A-O) multi-class classification method are used, for 9 class classification upon updated Hevner labeling.

Y. Panagak et al. [7] addressed the automatic mood classification problem by resorting the low rank representation of slow auditory spectra 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.

M. Barthet et al. [8] described state of the art studies on music and emotions from different disciplines including psychology, musicology and music information retrieval. Based on these studies, propose new insights to enhance automated music emotion recognition models.

Y. Songet al. [9] collected a ground truth data set of 2904 songs that have been tagged with one of the four words “happy”, “sad”, “angry” and “relaxed”, on the Last.FM web site. An excerpt of the audio is then retrieved from 7Digital.com, and various sets of audio features are extracted using standard algorithms. Two classifiers are trained using support vector machines with the polynomial and radial basis function kernels, and these are tested with 10-fold cross validation. Results show that spectral features outperform those based on rhythm, dynamics, and, to a lesser extent, harmony.

M. S. Y. Aw et al.[10] given a way in which music can be displayed for the user based on similarity of the acoustic features. By translating all songs in the music library onto a two-dimensional feature space, the user can better understand the relationship between the songs, with the distance between each song reflecting its acoustic similarity. The proposed approach avoids the need to depend on contextual data (such as metadata) and other collaborative filtering methods. With the song space visualizer, the user can make song choices or allow the system to automate the song selection process given a seed song.

B. K. Baniya et al. [11] developed a method which considers the various kinds of audio features. From each feature’s frame, a bin histogram has been computed to save all needed data related with it. The histogram bins of every feature are made use for calculating the similarity matrix, Therefore, there are 59 similarity matrixes from the corresponding same amount of audio features. The intra and inter similarity matrix are utilized to computed the intra-inter similarity ratio Among them, some of the selected similarity ratios are ultimately used as prototypes from each feature and are used for classification by designing the nearest multi-prototype classifier.

Ren et al. [12] proposed the use of a two-dimensional representation of acoustic frequency and modulation frequency to extract joint acoustic frequency and modulation frequency features. Long-term joint frequency features, such as acoustic-modulation spectral contrast/valley (AMSC/AMSV), acoustic modulation spectral flatness measure (AMSFM), and acoustic-modulation spectral crest measure (AMSCM), are then computed from the spectra of each joint frequency sub-band and classification is done with help of SVM.

A. AUDIO FEATURES

Feature extraction and classifier learning are two important parts of a classification system. Feature extraction tackles with the problem of how to represent the examples to be classified in terms of feature vectors or pair wise similarities. We can divide audio features into three groups timbre, rhythm and pitch. Each classification tries to capture audio features from different viewpoint. We can divide audio features into two levels, low-level and mid-level features from the viewpoint of music understanding. Low-level features can be further separated into two classes of temporal feature and timber features. Where timber features capture the tonal quality of sound which is related to various instrumentation and temporal features capture the variation and evolution of timbre over time[13].

a. TIMBRE FEATURES

The majority of the features listed in Table 1 are timbre features. As a basic element of music, timbre is a term describing the quality of a sound. Different timbres are produced by different types of sound sources, like different voices and musical instruments.

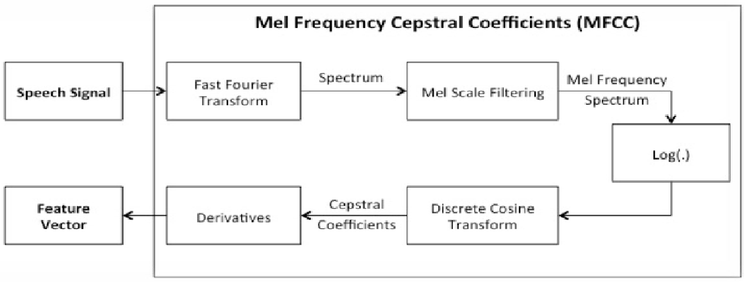


Fig 2. Mel-Frequency Cepstral Coefficient

We can define some summary features such as spectral centroid (SC), spectral rolloff (SR), spectral flux (SF), and spectral bandwidth (SB) capturing simple statistics of the spectra. Hereafter, we term the collection of these features as short time Fourier transform (STFT) features. It is possible to extract more powerful features such as MFCC, OSC, DWCH, and MPEG-7 audio descriptors like SFM, SCF, and ASE. [13] The most widely used acoustic features in speech and audio processing. MFCCs are essentially a low dimensional representation of the spectrum warped according to the mel-scale, which reacts the nonlinear frequency sensitivity of the human auditory system.

b. TEMPORAL FEATURES

Temporal features form another significant class of low-level features that capture the temporal evolution of the signal.Temporal features are generally constructed on top of timbre features. Some of the simplest types of temporal features are statistical moments such as mean, variance, co-variance and with no distinct rhythm pattern. In simple words we can saysad songs have a slow rhythm, whereas angry songs usually have a fast rhythm. Rhythm is the most extensively used mid-level feature in audio-based music classification. It describes how certain patterns occur and reappear in the music.Beat and tempo (beat-per-minute, BPM) are two important indications that describe rhythmic content of the music which have been used in music classification.

Low-Level Features	Timber Features	spectral centroid (SC) spectral rolloff(SR) spectral flux (SF) spectral bandwidth (SB) Mel-frequency ceptrum Coefficient(MFCC) Linear Predictive ceptrum Coefficient(LPCC) Octave-based spectral contrasts(OSC)
	Temporal Features	Statistical moments(SM) Amplitude modulation(AM) Auto-regressive modelling(ARM
Mid-Level Features	Rhythm	Beat histogram (BH). Beat-per-minute (BPM) Pitch
	Pitch	Pitch histogram(PH) Pitch class profile (PCP) Harmony
	Harmony	chord sequences (CS)

Table I. Different Features For Mood Classification

The auto-correlation of the time-domain envelop signal is determined.[13] The peaks of the auto-correlation function are then identified which correspond to probable regularity of the music under analysis. The beat histogram represents the distributions of the regularities showed in the envelop signal, where rhythmic features can be obtained such as magnitudes and locations of dominant peaks and BPM. As the mood of a song is extremely correlated with rhythm, these features have good experimental performance.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

c. *PITCH AND HARMONY*

Pitch and harmony are also important components of music. Pitch is defined as most fundamental frequency of the sound determined by what the ear judges. However, a pitch is not equal to the fundamental frequency because the perception of pitch is completely subjective while frequency measurement is objective.Other reasons like differences in timbre, loudness, and musical context also affect pitch.

B. CLASSIFIERS

In standard classification, we are presented with a training data set where each example comes with a label. The purpose is to propose a classification rule that can best predict the labels for unseen data. K-nearest neighbour (K-NN) , support vector machine (SVM) and GMM classifier are most popular choices for classifiers.

a. *SVM CLASSIFIER*

VM is the high-tech binary classifier based on the large margin principle. Given labelled instances from two different classes, SVM classifier finds the optimal separating hyper plane which maximizes the distance between support vectors and the hyper plane. Cyril Laurier, Perfecto Herrera[14] uses Support Vector Machine classifier to predict the mood cluster. They uses a set of 133 descriptors. The features are spectral, temporal, tonal but also describe loudness ability. The features were selected beforehand according to experiments on annotated databases. A grid search algorithm is used to optimize SVM.

b. *K-NN CLASSIFIER*

K-NN is one of the most accepted classifiers used for both general classification problems and in mood based music classification as well. K-NN uses training data directly for the classification of testing data. We can predict label of thetesting instance by majority voting on the labels of the nearest instances in the training set. Homer H. Chen[15] uses fuzzy K-NN as classifier. Fuzzy k-NN classifier is a combination of fuzzy logic and k-NN classifier. It contains two steps: fuzzy labelling that computes the fuzzy vectors of the training samples and fuzzy classification that computes the fuzzy vectors of the input samples.

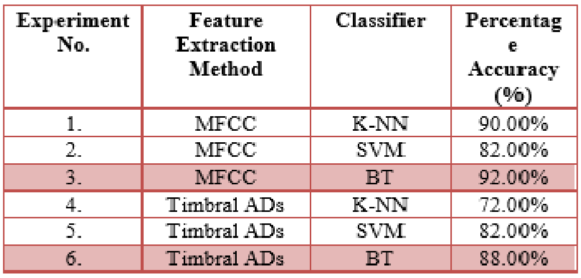
c. *GMM CLASSIFIER*

For the GMM classifier, we fit the Gaussian mixture model over the distributions of features in each and every class. With the class conditional probability distribution, labelling of testing example can be done according to the Bayes rule,

$f(x)=\arg\max P(y=k|x)$ $P(y=k|x) = P(x|y=k)P(y=k) / \sum P(x|y=k)p(y=k)$

The decision based on the maximizer of the posterior probability identifies the labels, data and the conditional probability of example for class label estimated from the training data using GMM. George Tzanetakis [16] explains GMM classifier and the EM algorithm.

Zhouyu Fu, Guojun Lu[13] conducted experiment in which they made all possible combinations of feature extraction methods and classifiers. In this manner total six experiments are done for different music.



Experiment No.	Feature Extraction Method	Classifier	Percentag e Accuracy (%)
1.	MFCC	K-NN	90.00%
2.	MFCC	SVM	82.00%
3.	MFCC	BT	92.00%
4.	Timbral ADs	K-NN	72.00%
5.	Timbral ADs	SVM	82.00%
6.	Timbral ADs	BT	88.00%

Table II. Experiment Performed With different types of music

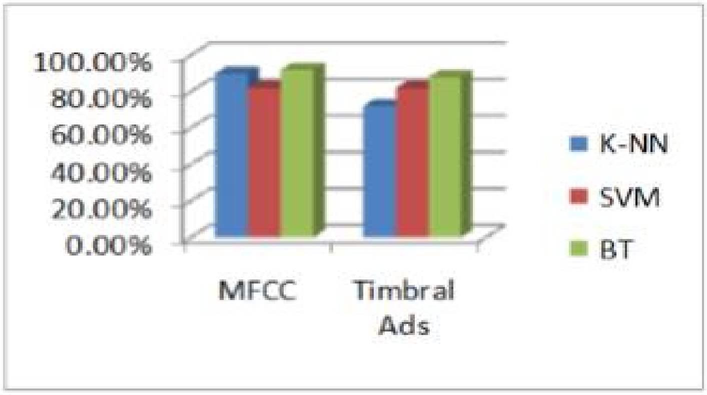


Fig 3. Percentage accuracies obtained for different type of music.

III. CONCLUSION

In this survey we have studied the some of the techniques developed by the researchers on Mood Classification In Music and the survey also provided current discussion of audio features used for mood based music classification. It describe the difference in the features and the types of classifiers used for different mood based classification systems also states how much accuracy can be achieved with particular classifier. If multiple features are available, we can combine those features in some way for music classification. Feature combination from different sources is an effective way to improve the performance of mood based music classification system.

REFERENCES

1. Liu, Dan, Lie Lu, and HongJiang Zhang. "Automatic mood detection from acoustic music data." In *ISMIR*, pp. 81-87. 2003.Ruijie Zhang, Bicheng Li, Tianqiang Peng, "Audio Classification Based on SVM -UBM", ICSP2008 Proceedings
2. Ruijie Zhang, Bicheng Li, Tianqiang Peng, "Audio Classification Based on SVM -UBM", ICSP2008 Proceedings
3. C.-H. Lee, J.-L. Shih, K.-M. Yu, and H.-S. Lin, "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features." IEEE Transactions on Multimedia, vol. 11, no. 4, pp. 670–682, 2009
4. X. Sun and Y. Tang, "Automatic Music Emotion Classification Using a New Classification Algorithm," Computational Intelligence and Design, 2009. ISCID '09. Second International Symposium on, Changsha, 2009, pp. 540-542.
5. P. Dunker, C. Dittmar, A. Begau, S. Nowak and M. Gruhne, "Semantic High-Level Features for Automated Cross-Modal Slideshow Generation," *2009 Seventh International Workshop on Content-Based Multimedia Indexing*, Chania, 2009, pp. 144-149.
6. E. E. P. Myint and M. Pwint, "An approach for mulit-label music mood classification," *Signal Processing Systems (ICSPS), 2010 2nd International Conference on*, Dalian, 2010, pp. V1-290-V1-294.
7. Y. Panagakis and C. Kotropoulos, "Automatic music mood classification via low-rank representation," in Proc, 2011, pp. 689–693, 2010.
8. M. Barthet, G. Fazekas, and M. Sandler, "Multidisciplinary perspectives on music emotion recognition: recommendations for content- and context-based models." Proc. CMMR, pp. 492–507, 2012
9. Y. Song, S. Dixon, and M. Pearce, "Evaluation of musical features for emotion classification," in Proceedings of the 13th International Society for Music Information Retrieval Conference, Porto, Portugal, October 8-12 2012, pp. 523–528.
10. M. S. Y. Aw, C. S. Lim and A. W. H. Khong, "SmartDJ: An interactive music player for music discovery by similarity comparison," Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific, Kaohsiung, 2013, pp. 1-5.
11. B. K. Baniya, Choong Seon Hong and J. Lee, "Nearest multi-prototype based music mood classification," Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on, Las Vegas, NV, 2015, pp. 303-306.
12. Ren, Jia-Min, Ming-Ju Wu, and Jyh-Shing Roger Jang. "Automatic music mood classification based on timbre and modulation features." IEEE Transactions on Affective Computing 6.3 (2015): 236-246.
13. Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang, "A Survey of Audio-Based Music Classification and Annotation", IEEE Transactions on Multimedia, Vol. 13, No. 2, April 2011.
14. C. Laurier and P. Herrera . Audio music mood classification using support vector machine. 2007.
15. Yang, Yi-Hsuan, and Homer H. Chen. "Machine recognition of music emotion: A review." *AC Transactions on Intelligent Systems and Technology (TIST)* 3.3 (2012): 40.
16. Tzanetakis, George, and Perry Cook. "Musical genre classification of audio signals." *IEEE Transactions on speech and audio processing* 10.5 (2002): 293-302.
17. K.WestandS.Cox,"Featuresandclassifiersfortheautomaticclassification of musical audio signals," in Proc. Int. Conf. Music Information Retrieval, 2004.
18. K. Umapathy, S. Krishnan, and S. Jimaa, "Multigroup classification of audio signals using time-frequency parameters," IEEE Trans. Multimedia, vol. 7, no. 2, pp. 308–315, Apr. 2005.
19. Learning and Applications San Diego, California (USA) December 2008.
20. Dr.M.Hemalatha, N.Sasirekha, S.Easwari, N.Nagasaranya "An Empirical Model for Clustering and Classification of Instrumental Music using Machine Learning Technique", 2010 IEEE International Conference on Computational Intelligence and Computing Research
21. Cyril Laurier , Jens Grivolla Fundaci´o , Perfecto Herrera," Multimodal Music Mood Classification using Audio and Lyrics", International Conference on Machine Learning.
22. S. H. Deshmukh and S. G. Bhirud, "Analysis of Audio Descriptor Contribution in Singer Identification Process," International Journal of Emerging Technology and Advanced Engineering, vol. 4, no. 2, February 2014.
23. H. Fletcher, "Loudness, Pitch and Timber of Musical Tones and their Relations to the Intensity, the Frequency and the Overtone Structure," JASA, vol. 6, no. 2,2011.