
Survey on Features and Classification Techniques in Music Genre Classification

¹Swati A. Patil, ²K.Thirupathi Rao, ³Sonal Patil

¹Research Scholar, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India – 522502

²Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India – 522502, India

³Assistant Professor, G. H. Raisoni Institute of Engineering and Management, Jalgaon, India

Email: ¹swati.patil251@gmail.com, ²kthirupathirao@kluniversity.in, ³sonal.patil@raisoni.net

Received: 09th July 2018, Accepted: 14th August 2018, Published: 31st August 2018

Abstract

Music is an integral part of society and even human being. A human being listen music to make a mood and change a mood. He listens music in sad and happy moments even when he is getting bored. In early days only folk music or local music is the way for this purpose. But in digital era or due to entry of various cultures in the society people are listening music from various parts of the country or world. The choice of listening the music varies from people to people from mood to mood. Sometimes we want to listen silent music or rock or pop etc. and sometimes it depends on the occasions also. As far as mood is concern people listen sad or happy or romantic music to make the mood or change the mood. This type of classification of songs is called genres. Genre classification is one of the prominent fields in MIR. Most of Genre classification involves feature extraction and the classification techniques along with some preprocessing if needed. In this review paper we tried to discuss and compare various techniques of features extraction and classification methods. In the literature most widely used features are MFCC and timbre features and classification methods are KNN and SVM. Low level feature and high level features are taken into consideration and most suitable feature is identified.

Keywords

KNN, SVM, Low Level Feature, High Level Features, MIR, MFCC, Timbre Features, Genre Classification.

Introduction

Nowadays music or songs are widely disseminated data on the Internet and the amount of uploading or releasing of music/ songs are going just increasing and not stopping or reducing. Effective organization and process such large data as well as efficient indexing, searching and retrieval of such large amount of musical data is a big encounter. Most music classification methods are on music categorized on genre, mood, or multiple semantic tags [17].

Music Genre is an important aspect used to categorize the music/songs at music shops, broadcasting and searching. Automatic music genre classification is a fundamental component of music information retrieval systems [10].

For mankind it is not difficult to classify the music in different genres but to understand it by computers is a big challenge. There are different view to define genre like instrument, melody, tempo, rhythm to differentiate musical genres. [6]

Music Genre classification is mostly important in music database management system to search the music as per the genre. Genre classification is a highly independent form of labeling in and of itself [9].

Due to technical development and inventions of smart portable devices like mobile phones, iPods music has been reached to everyone [7]. We frequently come across the people with head phone or ear plug on the way while walking, driving even while studying, while doing exercise or yoga etc. As Internet is available at very low cost music has been reached to people easily and readily accessible. When we listen songs we never want to listen full play list. Many times we want to listen the songs which suites to our mood. So in this case it is important or beneficial to have a list of songs as per mood/genre. Si that user can easily search the music from the huge database.

It is vital task and studied profoundly by MIR community since 2002[1]. Basically it is developed from pattern recognition concept initially by Tzanetakis and Cook [2]. The Indian songs composition is categorized into three classes like classical, semiclassical, and Light. Classical music flows some strict pattern based on the ragas which has some fixed sequence of swaras and these swaras are frequently used in composition. In light music more focus is given on the lyrics and melodic lines are less complex where as in classical music melodic lines are more focused. Semi classical music lies in between classical and light music.[3]. In early twenty first century, online collection of music is one of the challenges to the industry. Here genre classification plays important role for searching. However for online databases the labeling of genre is done by experts as

well as laypersons which is time intense and costly. Also in much composition the genre lies on boundaries of other genre which make confusion in classification. No strict mathematical model or strategy is available for classification. But it is observed .that music in same genre share some common characteristics like rhythm, instrument, pitch.[3] . Many songs are unlabeled and everyday hundreds of new songs are released. The process of genre classification includes three tasks. Music modeling, Feature extraction and Classification. Music modeling means a short segment of a song, Feature extraction has three level high level means labeling genre, mood and artist, middle level features are rhythm and pitch. Lowest level features are timbral features. Third task is classification which involves supervised and unsupervised classification methods like SVM, KNN, GMM, MLP, BPNN, ANN etc.[4]

Topical concerns in MIR take account of automatic audio genre/mood classification, music similarity computation, audio artist identification, audio-to-score alignment, query-by-singing/humming, tracking, and so on[5]. All the classification methods in MIR contain feature extraction as a first step before classification. Each classification system consists of minimum two blocks: feature extractor and classifier [17]. Genre classification and identification is one of the most meaningful example in IR (Information retrieval) which includes extraction of features from given audio signal. In machine learning language it is called as Music Genre Classification. Since audio signal is dynamic in nature, extracting meaningful feature from signal is quite difficult. One more reason for this is that in some cases song listener does not agree with genre of a song. So to understand the genre by machine is somewhat state of art or tricky task [8]. Same instrument show different genre and mood.[17] If we compare old music/songs Vs new music /songs, In old music the tome of music is mostly same so that it would be easy to classify the old songs using classification algorithm. The existing system is giving more accurate results. But nowadays music is continuously changing. Impact of globalization is on music also. New generation likes to listen the fusion song which has east and western touch. No doubt old songs have their own space in the society, they have their own fans. New generation also listen the songs from 50s with same interest as they have in new songs. There are evergreen songs which may be melodic, sizzling songs which are listened and tuned by many people as far as India is concern.

Much work is there by researcher on Genre Classification. Since new songs have fusion of many culture all over the world, somewhere there is a chance of misclassification due to use of musical instrument, due to similar features in the music, no similarity of

rhythm or beats all over the song, there may be sudden change in frequencies, tempo, tone. Sad song contains rock music with high pitch and loudness, some segment of song contains rock music and some segment of music contains classical music which is generating misclassification in Music Genre. Here we will discuss various features used and classification algorithm with their performance and accuracy.

Features

In this section we are discussing various features or aspects for Genre Classification. There are two types of features low level and high level features. Major task in classification is feature extraction irrespective of what we are classifying. Feature extraction would play important role in content based image retrieval, video retrieval, audio retrieval, extraction many sort of information. For example in image various features are color, shape, texture. Without feature extraction we cannot classify anything. Features are nothing but distinctive attribute or property. The features can be extracted by two ways. Either directly from the time domain signal or from a transformation domain depending upon the choice of the signal analysis approach [12]. In [13] audio features are extracted at into two levels: frame-level and high-level for genre classification.

Any recording of music is shown by three levels features: namely the auditory cortical representations: it plays important role in hearing process, the MFCCs encode timbral properties and the Chroma features: harmonic features [17].

Signal parameters, or features, are assessed from the analysis of small window segments of the audio signal, and the analysis is done at homogeneously spaced intervals of time.

Whatever features extracted are feed to the classifier or pattern recognizer [11].

For rhythmic based musical genre classification beat histogram method is adapted by authors of [16]. Here three features used are Envelope, Tonal and spectral shape. Flux, MFCC, spectral centroid and spectral flatness ate used in spectral shape.

Envelop is determined by RMS and tonal features are determined by pitch chroma coefficients.

For music whose long time interval representation helps more for classification, the MFCC have a poor classification performance. Scattering transform is an extension of MFCC.

[11] Classified into physical features and perceptual features as shown in Table I. In physical features, distinctive characteristic of digital signal which are measurable are taken into consideration whereas in perceptual features, features which are recognized by auditory model are taken into consideration.

Physical Features	Perceptual Features
MFCC, zero crossing rate (ZCR), short time energy, Band-Level Energy, Spectral Centroid, Spectral Roll Off, Spectral Flux, Fundamental frequency.	Loudness, Pitch

Table 1: Physical and Perceptual Features

ZCR is number of times the signal crosses the zero that it changes its sign in the window.

Short time energy indicates temporal envelop of the signal and is calculated by taking mean squared value in the data frames.

Energy within quantified region of the wave form spectrum is called Band level Energy.

Spectral roll off and spectral centroid are the center of gravity of the magnitude spectrum. These are common descriptor of the signal and define the gross shape of the signal.

A measure of the local spectral rate of change is called Spectral Flux provides.

MFCC provide a compact representation of the short-time spectrum envelope. MFCC have been strongly applied and used in speech recognition and, much more recently, to music.

Loudness is a sensation of signal strength.

Pitch is closely correlated with the physical attribute of fundamental frequency (F0). Pitch is the major feature for identifying male/female voice.

Type of Features	Extracted From
Symbolic Feature	Scores
Text features	Song lyrics
Acoustic Features	Tonality, pitch, beat
Content based Acoustic Features	Timbre features(MFCC,Spectral centroid, Spectral roll off, ZCR,energy,LPC),Rhythmic contents, pitch content

Table 2: Features and Their Types [14]

In most of the papers Features extracted are acoustic features as mentioned above. In [15] spectrograms are generated from audio signal and then textural features are extracted from this pictorial illustration using Gabor filter and LPQ texture descriptor.

The genre classification is based on harmonicity of the spectrum and periodicity of the gains [18].These two features are the vectors of acoustic parameter.

Though MFCC proved to give best result among all the features, its main disadvantage is it is not robust against additive noise. So it should be normalized,

hence [23] proposed to lift up the log-mel-amplitudes to a suitable power (around 2 or 3) before calculating Discrete Cosine Transform, which weakens the influence of low-energy components.

Sr. no	Features	Mathematical Formula
1	RMS	$\sqrt{(x_1^2 + x_2^2 + \dots + x_n^2)/n}$
2	ZCR	$ZC = \frac{1}{T-1} \sum_{t=1}^T \text{func} \{ s_t s_{t-1} < 0 \}$
3	Spectral Centroid	$SC = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$
4	Flux	$\sum_{k=1}^{N/2} (x_r[k] x_{r-1}[k])^2$
5	Spectral roll Off	$\sum_{n=0}^k x(n) = 0.85 \sum_{n=0}^{N-1} x(n)$

Table 3: Audio Features and their formulae

MFCCs are commonly derived as follows: [24,25]

1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
3. Take the logs of the powers at each of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

Classification Methods:

Classification is used to classify the music as per their genre. Like pop, rock, jazz, blues etc. Classification also classifies the music based on emotions like happy, sad, romantic etc. In the existing work of genre classification the songs are classified as happy, sad, rock, pop, jazz, folk etc. The classification is based on the dataset used.

All the work up till now has used the readymade dataset of the songs like last.fm, Millionsong, Labrosa where sample of songs belonging to various genre are available. By features extraction techniques as mentioned above and applying classification algorithm such as SVM, KNN, MLP, GMM results are evaluated. In all the classification SVM performed well. In [19] SVM is trained with One Vs-All methodology, implemented in WEKA for the Magnatune 2004 and RTV music corpus. Rock and metal genre are misclassified. Identification accuracy of Jazz songs is low. Gaussian Mixture model is used for classification in [20]. GMM are sensitive to curse

of dimensionality. Hence Principal Component Analysis is used for dimensionality reduction. SVM, GMM and MNN with six layers and thirty two hidden layers are compared in [21] for Shrilankan Songs. The authors concluded that hierarchical classification is ideal for multiclass classification problem. Sparse Representation classification method for cortical representation which generated over complete dictionary of basis signal for music genres. Here Dataset used is GTZAN and ISMIR2004.

Conclusion:

In this paper we have survey different features of any audio signal for genre classification and various machine learning techniques to classify genres. Most popular feature of audio is MFCC since it is very close to human auditory system. SVM and KNN are most widely used classification techniques in Genre Classification. All the work is on some standard datasets which are in English songs.

References

- [1] Combining Visual And Acoustic Features For Music Genre Classification, Loris Nanni, Yandre M.G. Costa , Alessandra Lumini , Moo Young Kim, Seung Ryul Baek.
- [2] Tzanetakis, G., & Cook, P. (2002). Musical Genre Classification Of Audio Signals. IEEE 746 Transactions On Speech And Audio Processing, 10(5), 293–302.
- [3] Classification Of Different Indian Oongs Based On Fractal Analysis Atin Das, Pritha Das. Complex Systems, 15 (2005) 253–259; " 2005 Complex Systems Publications, Inc.
- [4] Automatic Genre Classification Of Indian Tamil And Western Music Using Fractional Mfcc Betsy Rajesh, D. G. Bhalke, Int J Speech Technol, Doi 10.1007/S10772-016-9347-3, Springer
- [5] Music Emotion Classification And Context-Based Music Recommendation Byeong-Jun Han , Eungmi Rho, Sanghoon Jun, Eenjun Hwang, Multimed Tools Appl (2010) 47:433–460, Springer.
- [6] Musical Genre Classification Using Support Vector Machines, Changsheng Xu, Namunu C. Maddage, Xi Shao, Fang Cao, Qi Tian, Icassp 2003, IEEE.
- [7] Audio Pattern Recognition And Mood Detection System, Priyanka Tyagi, Abhishek Mehrotra, Shanu Sharma, And Sushil Kumar, © Springer Science+Business Media Singapore 2016 M. Pant Et Al. (Eds.), Proceedings Of Fifth International Conference On Soft Computing For Problem Solving, Advances In Intelligent Systems And Computing.
- [8] Music Genre Classification Using Radial Basis Function Networks And Particle Swarm Optimization, Alex Alexandridis, Eva Chondrodima, Georgia

- Paivana, Marios Stogiannos, Elias Zois, Haralambos Sarimveis, 6th IEEE Computer Science And Electronic Engineering Conference (CEEC) [9]Genre Classification Using Graph Representations Of Music , Rachel Mellon, Dan Spaeth, Eric Theis
- [10] A Comparative Study On Content-Based Music Genre Classification, Tao Li, Mitsunori Ogihara, Qi Li
- [11] Audio Signal Processing Preeti Rao, Chapter In Speech, Audio, Image And Biomedical Signal Processing Using Neural Networks, Springer-Verlag, 2007.
- [12] Audio Signal Feature Extraction And Classification Using Local Discriminant Bases Karthikeyan Umapathy, Student Member, IEEE, Sridhar Krishnan, Senior Member, IEEE, And Raveendra K. Rao, Senior Member, IEEE, Transactions On Audio, Speech, And Language Processing, Vol. 15, No. 4, May 2007
- [13] Evaluation Of Different Audio Features For Musical Genre Classification, Babu Kaji Baniya, Deepak Ghimire, And Joonwhoan Lee, 2013 IEEE Workshop On Signal Processing Systems
- [14] Audio Feature Reduction And Analysis For Automatic Music Genre Classification, Babu Kaji Baniya, Joonwhoan Lee, Ze-Nian Li, 2014 IEEE International Conference On Systems, Man, And Cybernetics October 5-8, 2014, San Diego, Ca, Usa
- [15] Music Genre Recognition Using Gabor Filters And Lpq Texture Descriptors Yandre Costal Luiz Oliveira, Alessandro Koerich, Fabien Gouyon, Ciarp 2013, Part Ii, Lncs 8259, Pp. 67–74, 2013. Springer-Verlag Berlin Heidelberg 2013
- [16] Beat Histogram Features For Rhythm-Based Musical Genre Classification Using Multiple Novelty Functions, Athanasios Lykartsis, Alexander Lerch, 18th Int. Conference On Digital Audio Effects (Dafx-15), Trondheim, Norway, Nov 30 - Dec 3, 2015.
- [17] Music Classification By Low-Rank Semantic Mappings, Yannis Panagakis* And Constantine Kotropoulos, Eurasip Journal On Audio, Speech And Music Processing 2013,
- [18] Influence Of Low-Level Features Extracted from Rhythmic And Harmonic Sections On Music Genre Classification Aldona Rosner, Felix Weninger, Björn Schuller, Marcin Michalak, Bozena Kostek, Man-Machine Interactions Advances In Intelligent Systems And Computing, Doi: 10.1007/978-3-319- 02309-0_51, _C Springer International Publishing Switzerland 2014.
- [19] Feature Engineering for Automatic Music Genre Classification, Paolo Annesi, Robert Basili, Raffaele Gitto, Riccardo Petitti.
- [20] From Multi-labeling to multidomain labeling : A novel two dimensional approach to music Genre Classification, Hanna Lukashevich, Jakob Abeßer, Christian Dittmar, Holger Grossmann, 10th

International Society for Music Information Retrieval Conference (ISMIR 2009).

[21] A SVM based classification approach to musical Audio, Namunu Chinthaka Maddage, Changsheng

[22] Music Genre Classification using locally pressing in negative tensor factorization and sparse representation, Yannis Panagakis, Constantine Kotropoulos, Gonzalo R. Arce, 10th International Society for Music Information Retrieval Conference (ISMIR 2009)

[23] V. Tyagi and C. Wellekens (2005), On desensitizing the Mel-Cepstrum to spurious spectral components for Robust Speech Recognition, in Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on, vol. 1, pp. 529–532.

[24] Min Xu; et al. (2004). "HMM-based audio keyword generation". In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh. Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia (PDF). Springer. ISBN 3-540-23985-5. Archived from the original (PDF) on 2007-05-10.

[25] Jump up^ Sahidullah, Md.; Saha, Goutam (May 2012). "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition". *Speech communication*. 54 (4): 543–565. doi:10.1016/j.specom.2011.11.004