Music Genre Classification using MFCC, SVM and BPNN

Gursimran Kour
M. Tech Scholar
RBIEBT, Kharar

Neha Mehan
Asst. Professor
RBIEBT, Kharar

ABSTRACT
In the field of musical information retrieval, genre categorization is a complicated mission. MFCC is one of the feature extraction method use in classification of musical genre that is based on short speech signals. Searching and organizing are the main characteristics of the music genre classification system these days. This paper describes a new technique that uses support vector machines to classify songs based on features using MFCC, BPNN and SVM classifier does not classify songs based on the short signals. So these categories a number of acoustic features that include Mel-frequency Cepstral coefficients are extracted to characterize the audio content. Support vector machines and BPNN classifies audio into their respective classes by learning from training data. The simulation is taken place in MATLAB by making experiments on different genres .The results obtained by this proposed technique are promising**.**

General Terms
Classification Algorithms, Pattern Recognition.

Keywords
SVM, MFCC, BPNN, training, classification, feature extraction.

1. INTRODUCTION
In the digital music library, music genre (Abu et. al, 2006) searching has become an essential need in today's era. Genre classification is mainly used in following places:

a. Developing automic playlists on MP3

b. Storing enormous number of songs online.

In most of the Genre classification MFCC is the most common feature extraction method that is based on the short duration frame.

In this paper, we will use MFCC for feature extraction of the music genre and SVM as well as BPNN for classification of the music genre classification.

The rest of the paper is organized as: section II will describe the literature survey, Section III will describe flowchart of proposed work, Section IV will describe the methodology, Section V will contain the results, and Finally Section VI will contain the conclusion of the proposed work.

2. RELATED WORK
In the past years, many techniques has been introduced for the feature extraction as well as classification of genre. But we will focus on the common techniques that has good results.

Lin, Chen, Truong, and Chang (2005), presents the wavelets method for classification of music. Wavelets are used for feature extraction purpose in which two features has been extracted i.e sub-band power and picth. This method uses SVM and MFCC for classification of the music genre based on the features obtained by wavelets.

Ajmera, McCowan, and Bourlard (2003), proposed two methods called artificial neural network (ANN) and hidden Markov model (HMM), on the basis of which classification of music is done in news.

Kiranyaz, Qureshi, and Gabbouj (2006), presents a generic audio classification technique.

Panagiotakis and Tziritas (2005), proposed a method for classification of genre that is based on root mean square and zero crossing technique.

Eronen et al. (2006), surveys the low dimensional feature vector against standard feature spaces.

Li, Sethi, Dimitrova, and McGee (2001), classifies the audio signal based on thr classifiers i.e MFCC and LPC. It has been shown that cepstral-based features gave a better classification accuracy.

(**Esmaili, Krishnan, & Raahemifar, 2004),** presents joint time–frequency analysis for classification of audio signal.

Umapathy, Krishnan, and Jimaa (2005), proposed an adaptive time frequency decomposition algorithm, for feature extraction and then LDF has been used to classify between six genre classifications.

Jiang, Bai, Zhang, and Xu (2005), presents support vector machine (SVM) for the classification of audio signals, which classifies audio clips into one of five classes.

McConaghy, Leung, Boss, and Varadan (2003), presents Radial basis function neural networks (RBFNN) for the classification of real life RADAR signals.

Guo and Li (2003), proposed a new technique called called distance-from-boundary (DFB), for the classification of music genre, it also uses SVM

Mubarak, Ambikairajah, and Epps (2005), proposed two methods called mel-frequency cepstral coefficient (MFCC) and GMM classifier, on the basis of which classification of music/speech is done.

Li (2000), proposes a new pattern classification method called the nearest feature line (NFL).

Huang and Hansen (2006), proposes two new extended-time features called variance of the spectrum flux (VSF) and variance of the zero-crossing rate (VZCR) that are used for pre-classification purpose then for post-classification purpose GMM has been used.

3. PROPOSED FLOWCHART
In this paper, automatic music feature extraction and classification approaches are presented. In order to discriminate the music genre, feature extraction (Duda, 2006) method MFCC has been used to characterize the audio content as shown in figure below Figure 1. Back Propagation Neural Network (BPNN) and Support vector machine (SVM)

are applied to obtain the classes by learning from training data.

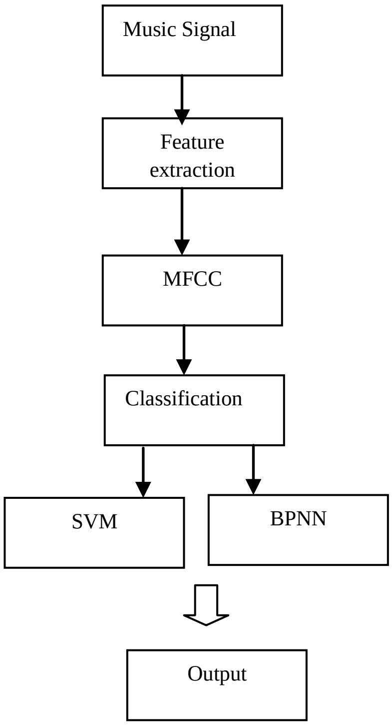

Figure. 1 Flowchart

4. PROPOSED WORK

4.1 Acquisition

First step of the proposed work is acquisition (Sorsa, 2006) of music signals. That can be done from any audio saver.

4.2 Feature Extraction Using MFCC

Acoustic features representing the audio information can be extracted from the speech signal initially. The mel-frequency cepstrum has best for recognizing structure of music signals as reviewed from previous papers and in molding the subjective pitch and frequency content of audio signals

9 P.S , 2001 ; R.L, 2001). Psychophysical studies have found the phenomena of the mel pitch scale and the critical band, and the frequency scale-warping to the mel scale has led to the cepstrum domain representation as shown in Figure. 2.

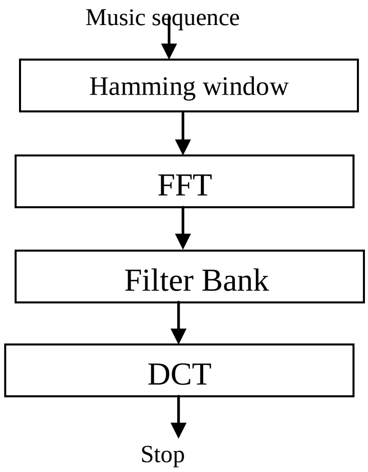

Figure. 2 MFCC algorithm

4.3 Classification using SVM and BPNN

The SVM (Vapnik, 1998), is a useful statistic machine learning technique that has been successfully applied in the pattern recognition area. Support vector machines are based on the Structural Risk Minimization theory from computational knowledge hypothesis. SVM (Lu, 2003) are independent of the dimensionality of the feature space, represented in Figure. 3. Characteristics of SVM:

- High dimensional input space
- Document vectors are sparse
- Few irrelevant features
- Mainly text classification problems are linear

To obtain the geometric distance from the hyperplane to a data point, we must normalize by the magnitude of w. This distance is simply:

$$d\,((w,b),x) = \frac{yi\,(xi \cdot w + b)}{\|w\|} \geq \frac{1}{\|w\|}$$

Intuitively, we want the hyperplane that maximizes the geometric distance to the closest data points.

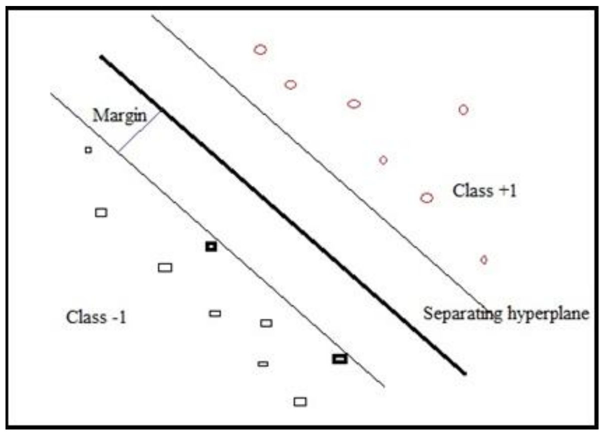

Figure.3 SVM working [14]

The use of neural network (Haykin, 2006) (BP NN) in classification has mainly two advantages: code simplification and accurate classification. Another main profit is the extensibility of the system i.e. ability to recognize more genre of music. Below Figure. 4 shows the block diagram of neural network.

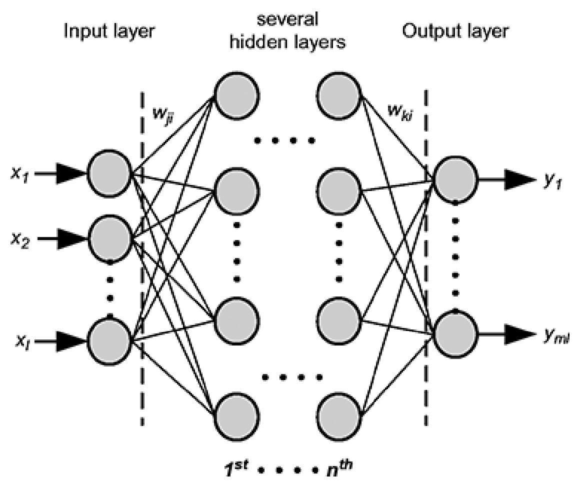

Figure. 4 BPNN Working [7]

5. RESULTS AND DISCUSSIONS

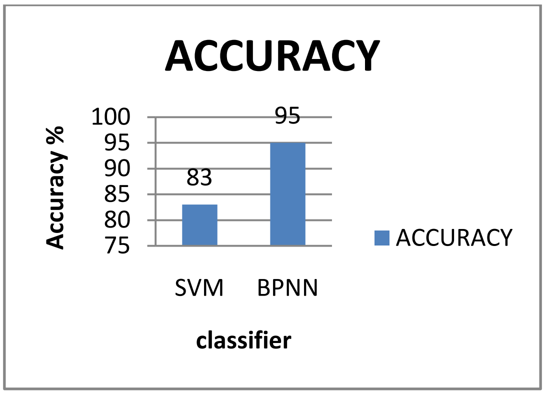

Figure.5 Performance Graph

The above Figure.5 shows that accuracy of classification of music genre for SVM is 83 % and for BPNN is 95%. The whole simulation has been taken place in MATLAB 7.10 environment using accuracy parameter.

6. CONCLUSION AND FUTURE SCOPE

In this paper, we have proposed an automatic music classification system using SVM and BPNN. MFCC is calculated as features to characterize audio content. Support vector machine (SVM) and Back Propagation Neural Network (BPNN) learning algorithm has been used for the classification of genre classes of music by learning from training data. Experimental results show that the proposed audio classification scheme is very effective and the accuracy rate is 95%. The performance was compared to SVM which showed an accuracy of 83%. Neural network and SVM has been used because of their good classification and training accuracy among machine learning algorithms.

Future work will consider alternative feature selection techniques better adapted to GMM classification. Furthermore, hierarchical classification wherein instruments are grouped into families will be envisaged. The recognition of typical instrumental ensembles (solos, duets, trios, etc.) will be introduced at a high level of taxonomy. As for classification, probabilistic outputs for GMM will be considered together with a time dynamic approach.

7. REFERENCES

[1] Abu-El-Quran, A. R., Goubran, R. A., & Chan, A. D. C. (2006). Security monitoring using microphone arrays and audio classification. IEEE Transactions on Instrumentation and Measurement, 55(4), 1025–1032.

[2] Ajmera, J., McCowan, I., & Bourlard, H. (2003). Speech/music segmentation using entropy and dynamism features in a HMM classification framework. Speech Communication, 40(3), 351–363.

[3] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification. New York: John Wiley-Interscience. Eronen, A. J., Peltonen, V. T., Tuomi, J. T., Klapuri, A. P., Fagerlund, S.

[4] Sorsa, T., et al.(2006). Audio-based context recognition. IEEE Transactions on Audio, Speech and Language Processing, 14(1), 321–329.

[5] Esmaili, S., Krishnan, S., & Raahemifar, K. (2004). Content based audio classification and retrieval using joint time–frequency analysis. In IEEE international conference on acoustics, speech and signal processing (pp. 665–668).

[6] Guo, G., & Li, S. Z. (2003). Content-based audio classification and retrieval by support vector machines. IEEE Transactions on Neural Networks, 14(1), 308–315.

[7] Haykin, S. (2001). Neural networks a comprehensive foundation. Asia: Pearson Education.

[8] Huang, R., & Hansen, J. H. L. (2006). Advances in unsupervised audio classification and segmentation for the broadcast news and NGSW corpora. IEEE Transactions on Audio, Speech and Language Processing, 14(3), 907–919.

[9] Jiang, H., Bai, J., Zhang, S., & Xu, B. (2005). SVM-based audio scene classification. Proceeding of the IEEE, 131–136.

[10] Kiranyaz, S., Qureshi, A. F., & Gabbouj, M. (2006). A generic audio classification and segmentation approach for multimedia indexing and retrieval. IEEE Transactions on Audio, Speech and Language Processing, 14(3), 1062–1081.

[11] Li, S. Z. (2000). Content-based audio classification and retrieval using the nearest feature line method. IEEE Transactions on Speech and Audio Processing, 8(5), 619–625.

[12] Lin, C.-C., Chen, S.-H., Truong, T.-K., & Chang, Y. (2005). Audio classification and categorization based on wavelets and support vector machine. IEEE Transactions on Speech and Audio Processing, 13(5), 644–651.

[13] Li, D., Sethi, I. K., Dimitrova, N., & McGee, T. (2001). Classification of general audio data for content-based retrieval. Pattern Recognition Letters, 22, 533–544.

[14] Lu, L., Zhang, H.-J., & Li, S. Z. (2003). Content-based audio classification and segmentation by using support vector machines. Multimedia Systems, 8, 482–492.

[15] McConaghy, T., Leung, H., Boss, E., & Varadan, V. (2003). Classification of audio radar signals using radial basis function neural networks. IEEE Transactions on Instrumentation and Measurement, 52(6), 1771–1779.

[16] Mubarak, O. M., Ambikairajah, E., & Epps, J. (2005). Analysis of an MFCC-based audio indexing system for efficient coding of multimedia sources. In IEEE international conference on acoustics, speech and signal processing (pp. 619–622).

[17] Palanivel, S. (2004). Person authentication using speech, face and visual speech. Ph.D thesis, Madras: IIT.

[18] Panagiotakis, C., & Tziritas, G. (2005). A speech/music discriminator based on rms and zero-crossings. IEEE Transactions on Multimedia, 7(1), 155–156.

[19] Rabiner, L., & Juang, B. (2003). Fundamentals of speech recognition. Singapore: Pearson Education.