ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 1, January 2020

Speech/music Change Point Detection using PLP and SVM

R. Thiruvengatanadhan

Assistant Professor, Department of Computer Science and Engineering,
Annamalai University, Annamalainagar, Tamilnadu, India. E-mail: thiruvengatanadhan01@gmail.com

Abstract: An audio clip can consist of several classes. It can consist of speech and music which is typical in radio broadcasting. A human listener can easily distinguish audio signals into these different audio types by just listening to a short segment of an audio signal. Changes in audio signal characteristics help in detecting the category change point between different categories. In this paper, Perceptual Linear Prediction (PLP) features are extracted which are used to characterize the audio data. Support Vector Machine (SVM) is used to detect change point of audio. The results achieved in our experiments illustrate the potential of this method in detecting the change point between speech and music changes in audio signals.

Keywords: Speech, Music, Feature Extraction, PLP, SVM.

I. Introduction

Category change points in an audio signal such as speech to music, music to advertisement and advertisement to news are some examples of segmentation boundaries. Systems which are designed for classification of audio signals into their corresponding categories usually take segmented audios as input. However, this task in practice is a little more complicated as these transitions are not so obvious all the times [1]. Category change point detection of acoustic signals into significant regions is an important part of many applications. Systems which are developed for speech/music classification, indexing and retrieval usually take segmented audios rather than raw audio data as input. A first content characterization could be the categorization of an audio signal as one of speech, music, or silence [2], [3]. Changes in audio signal characteristics help in detecting the category change point between different categories. In speech/music change point detection the audio signal can be segmented into speech and music regions. A human listener can easily distinguish audio signals into these different audio types by just listening to a short segment of an audio signal.

II. Acoustic Feature Extraction

Acoustic feature extraction plays an important role in constructing an audio change point detection system. The aim is to select features which have large between-class and small within-class discriminative power.

A. Perceptual Linear Prediction (PLP)

Hermansky developed a model known as PLP. It is based on the concept of psychophysics theory and discards unwanted information from the human pitch [4]. It resembles the procedure to extract LPC parameters except that the spectral characteristics of the speech signal are transformed to match the human auditory system.

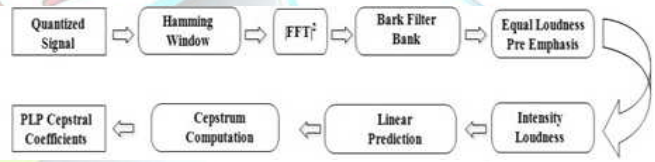

Fig. 1. PLP Parameter Computations.

PLP is the approximation of three aspects related to perceptron namely resolution curves of the critical band, curve for equal loudness and the power law relation of intensity loudness. The process of PLP computation is shown in Fig. 1. The audio signal is hamming windowed to reduce discontinuities. The Fast Fourier Transform (FFT) transforms the windowed speech segment into the frequency domain [5].

The auditory warped spectrum is convolved with the power spectrum of the simulated critical-band masking

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 1, January 2020

curve to simulate the critical-band integration of human hearing. Critical band is the frequency bandwidth created by the cochlea, which acts as an auditory filter. The cochlea is the hearing sense organ in the inner ear. Bark scale corresponds to 1 to 24 critical bands. The power spectrum of the critical band masking curve and auditory warped spectrum are convoluted to simulate the human hearing resolution. The equal loudness pre-emphasis needs to compensate the unequal perception of loudness at varying frequencies. An all pole model normally applied in Linear Prediction (LP) analysis is used to approximate the spectral samples. Either the coefficients can be used as such for representing the signal or they can further be transformed to Cepstral coefficients. In this work, a 9th order LP analysis is used to approximate the spectral samples and hence obtained a 9-dimensional feature vector for a speech signal of frame size of 20 milliseconds is obtained.

III. Support Vector Machine (SVM)

In machine learning, support vector machines (SVMs, also called support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier [6]. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on [7]. In addition to performing linear classification, SVMs can efficiently perform non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces [8], [9].

Kernel methods have received major attention, particularly due to the increased popularity of the Support Vector Machines. Kernel functions can be used in many applications as they provide a simple bridge from linearity to non-linearity for algorithms which can be expressed in terms of dot products. Some common Kernel functions include the linear kernel, the polynomial kernel and the Gaussian kernel [10].

IV. Experimental Results

A. The database

Performance of the proposed audio change point detection system is evaluated using the Television broadcast audio data collected from Tamil channels, comprising different durations of audio namely speech and music from 5 seconds to 1 hour. The audio consists of varying durations of the categories, i.e. music followed by speech and speech in between music etc., Audio is sampled at 8 kHz and encoded by 16-bit.

B. Acoustic feature extraction

9 PLP features are extracted a frame size of 20 ms and a frame shift of 10ms of 100 frames as window are used. Hence, an audio signal of 1 second duration results in 100 × 9 feature vector. The classifier SVM model is trained to map the distribution of the feature vectors in the left and right half of the window over the hyper plane, then the misclassification rate of the left and right half feature vectors of the window are used for testing.

C. Category change point detection

The sliding window is initially placed at the left end of the signal. SVM model is trained to capture the distribution of the feature vectors in the left half of the window and the feature vectors in the right half of the window are used for testing. A major change detected based on a threshold indicates that the characteristics of the signal in the right half of the window is different from the signal in the left half of the window and hence the middle of the window is a category change point. The performance of the proposed speech/music change point detection system is shown in Fig. 2 for SVM.

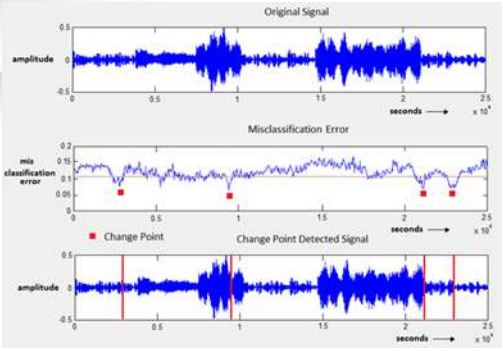

Fig. 2. Snapshot of Speech/Music Change Point Detection Systems Using SVM.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 1, January 2020

The performance of the speech/music change point detection system using SVM to detect the change point in terms of the various measures is shown in Fig. 3.

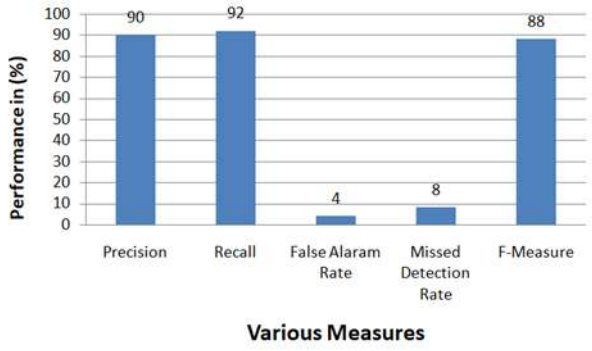

Fig. 3. Performance of detect the change point in terms of the various measures using SVM.

V. CONCLUSION

In this paper we have proposed a method for detecting the category change point between speech/music using Support Vector Machine (SVM). The performance is studied using 9 dimensional PLP features. SVM based change point detection gives a better performance of 88% F-measure is achieved.

REFERENCES

[1]. Reda Elbarougy. Speech Emotion Recognition based on Voiced Emotion Unit. International Journal of Computer Applications 178(47):22-28, September 2019.

[2]. D. Li, I. K. Sethi, N. Dimitrova, and T. Mc Gee, "Classification of General Audio Data for Content Based Retrieval," *Pattern Recognition Letters*, vol. 22, no. 1, pp. 533-544, 2001.

[3]. P. Dhanalakshmi, S. Palanivel and M. Arul, "Automatic Segmentation of Broadcast Audio Signals using Autoassociative Neural Networks," *ICTACT Journal on communication technology*, vol. 1, Issue 04, 2010.

[4]. Peter M. Grosche, Signal Processing Methods for Beat Tracking, Music Segmentation and Audio Retrieval, Thesis, Universit¨at des Saarlandes, 2012.

[5]. PetrMotlcek, Modeling of Spectra and Temporal Trajectories in Speech Processing, PhD thesis, Brno University of Technology, 2003.

[6]. Vishal Deshwal and Mukta Sharma. Breast Cancer Detection using SVM Classifier with Grid Search Technique. International Journal of Computer Applications 178(31):18-23, July 2019.

[7]. Manisha Prajapati and Archit Yajnik and. POS Tagging of Gujarati Text using VITERBI and SVM. *International Journal of Computer Applications* 181(43):32-35, March 2019.

[8]. Ayobami I Ojelabi, Oluwabusayo I Omotosho and Olajide A Oladejo. Classification and Detection of Citrus Disease using Feature Extraction and Support Vector Machine (SVM). *International Journal of Computer Applications* 177(17):17-25, November 2019.

[9]. S. Nilufar, Edmonton, N. Molla, and K. Hirose. Spectrogram based features selection using multiple kernel learning for speech/music discrimination. kernel learning for speech/music discrimination. Acoustics, Speech and Signal Processing (ICASSP), pages 501–504, March 2012.

[10]. Lim and Chang. Enhancing support vector machine-based speech/music classification using conditional maximum a posteriori criterion. Signal Processing, IET, 6(4):335–340, June 2012.