Volume 3, Issue 11, November-2016, pp. 580-584 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

A Survey on: Sound Source Separation Methods

[1]Ms. Monali R. Pimpale, [2]Prof. Shanthi Therese , [3]Prof. Vinayak Shinde,

[1]Department of Computer Engineering, Mumbai University,
Shree L.R. Tiwari College of Engineering and Technology,Mira Road, India.

[2]Department of Information Technology, Mumbai University,
Thadomal College of Engineering and Technology,Mumbai, India

[3]Department of Information Technology, Mumbai University,
Shree L.R. Tiwari College of Engineering and Technology,Mira Road, India

Abstract— now a day's multimedia databases are growing rapidly on large scale. For the effective management and exploration of large amount of music data the technology of singer identification is developed. With the help of this technology songs performed by particular singer can be clustered automatically. To improve the Performance of singer identification the technologies are emerged that can separate the singing voice from music accompaniment. One of the methods used for separating the singing voice from music accompaniment is non-negative matrix partial co factorization. This paper studies the different techniques for separation of singing voice from music accompaniment.

Keywords—singer identification, non-negative matrix partial co factorization

I.INTRODUCTION

The development of singer identification enables the effective management of large amounts of music data. With this singer identification technology, songs performed by a particular singer can be automatically clustered for easy management or searching. There are many algorithms which are used for singer identification which are based on the concept of feature extraction which identifies the appropriate singer from the obtained features. In popular music, singing voice is combined with music accompaniment. So those methods based on the features extracted directly from the accompanied vocal segments are difficult to acquire good performance when accompaniment is stronger or singing voice is weaker. To get better performance the techniques are emerged which separates the singing voice from music accompaniment. There are many sound source separation algorithms which separates the singing voice from music accompaniment. Sound source separation means the tasks of evaluating the signal produced by an individual sound source from a mixture signal consisting of multiple sources. This is a very fundamental problem in many audio signal processing tasks, since analysis and processing of isolated or single sources can be done with much better accuracy than the processing of mixtures of sounds. The term unsupervised learning is used to characterize algorithms which try to separate and learn the structure of sound sources in mixed data based on information-theoretical principles, such as statistical independence between sources, instead of highly sophisticated modeling of the source characteristics or human auditory perception. There are many unsupervised learning sound source separation algorithm some of them are independent component analysis (ICA), sparse coding, and non-negative matrix factorization, which has been tremendously used in source separation tasks in several application areas [1].

II.SOUND SOURCE SEPARATION

Source separation is process in which several signals are mixed together to form a combined signal and the objective of source separation is to obtain or recover the original component signals from the mixed or combined signal. This is fundamental problem in many audio signal processing tasks, because analysis and processing of isolated sources can be done with better accuracy than the processing of mixtures of sounds.

The well-known example of a source separation is the cocktail party problem, where a multiple people are talking simultaneously in a room (for eg, in cocktail party), and an individual who performs the role of listener is trying to follow one of the discussions. The human brain is capable to handle this type of auditory source separation problem, but it is very difficult problem to be solved in digital signal processing. Several approaches have been proposed for solving this problem but development is currently still very much in progress.

III.SUPERVISED LEARNING

Supervised leaning is the type of machine learning algorithm which uses known dataset which is also referred as training data for making the predictions. The training dataset includes input values and corresponding response values. Size of training dataset decides predictive power of the model.

In supervised learning model is prepared through a training process where the model is required to make predictions and is corrected when the predictions are wrong. The training process continues until the model achieves a desired level of accuracy on the training data. Supervised learning includes three categories of algorithm:

A. Classification

When the data are being used in the process of category prediction, supervised learning is also called classification. When there are only two choices, the learning is called **two-class** or **binomial classification**. When there are more than two categories, then this problem is known as **multi-class classification**.

B. Support Vector Machine (SVM)

The support Vector Machine (SVM) was mainly attracted a high degree of interest in the machine learning research community. Support Vector Machine is a supervised learning method used for classification. SVM simultaneously work on minimization of the imperial classification error and maximization of the geometric margin. For this reason SVM called maximum margin classifiers. Data is classified by using the hyperplane. Sample along the hyperplane is called Support Vector (SV). The separating hyperplane is defined as the hyerplane that maximize distance between the two parallel hyperplanes. If distance or margin between parallel hyperplane is better than SVM gives good classification.

Disadvantage: Most serious problem with SVMs is the high algorithmic complexity and the extensive memory requirements of the required quadratic programming in large tasks. They can be abysmally slow in test phase. This performs poorly on songs where much of the frequency distribution of the background is close to the vocal range [17].

C. Gaussian mixture model (GMM)

Gaussian mixture model (GMM) is used as a classifier for the classification of the voice and unvoiced signal. Gaussian mixture model (GMM) is a mixture of several Gaussian distribution and therefore represent different subclasses inside one big class. GMM to represent perfectly the data distribution: the most important thing for classification is to obtain a good separator between the classes. This process of classification was confirmed by considering discriminative training of GMMs for classification. Gaussian mixture model (GMM) is supervised learning which is best work on the maximum likelihood (ML) estimation using expectation maximization (EM). If we compare traditional GMM with pseudo GMM the nonlinear maps have better performance on nonlinear problems, while the computational complexity is almost the same as the Expectation-Maximization (EM) algorithm for traditional GMM according to the iteration procedures. In the training phase, a music database with manual vocal/nonvocal transcriptions is used to form two separate GMM: a vocal GMM and second nonvocal GMM. The expectation maximization (EM) algorithm is an iterative method for calculating maximum likelihood distribution parameter estimates from incomplete data. EM algorithm is high for two major reasons as similar to other kernel based methods, it have to calculate kernel function for each sample-pair over training set and in order to get the largest Eigen value [16].

Disadvantages -This requires large set of Gaussian functions with GMMs and also gives poor performance.

D. Regression.

When a value is being predicted the supervised learning is called regression. It is used to estimate real values such as cost of houses, total sales etc. based on continuous variables like total sale and stock prices etc.

E. Anomaly detection

In many cases the goal is to identify data points or categories that are simply unusual. The possible variations are so numerous and the training examples are so few, in such cases it is not feasible to learn what and how fraudulent activity looks like. The anomaly detection took the approach which simply learn how normal activity looks like (using a history non-fraudulent transactions) and identify the things that are significantly different

IV.UNSUPERVISED LEARNING

Unsupervised learning is type of machine learning algorithm to make presumption from dataset consisting of input data without responses. Unsupervised machine learning technique is not provided with accurate results during training data. It finds hidden clusters in input data sets which assist it in getting the right results. Unsupervised learning refers to the problem which finds hidden structure in unlabeled data.

A. Computational Auditory Stream Analysis (CASA):

CASA methods are based on the ability of humans to catch the sound and recognize individual sound sources in a mixture of sound are referred to as auditory scene analysis. Computational models of this function mainly consist of two main stages. In First stage, the mixture signal is decomposed into its elementary time-frequency components. Then in second stage, these time frequency components are organized and grouped to their respective sound sources. Our brain does not resynthesize or separate the acoustic waveforms of each source separately; still the human auditory system is a useful reference in the development of one-channel sound source separation systems, as it is the only existing system which can robustly separate sound sources in different circumstances.

Disadvantage: The performance of current CASA system is still limited by pitch estimation errors and residual noise.

B. Beamforming:

Beamforming achieves sound separation by using the principle of spatial filtering. The focus of beamforming is to magnify the signal coming from a specific direction by a suitable configuration of a microphone array at the same time the signals coming from other directions are rejected. As the number of microphones and the array length increases the amount of noise attenuation increases. With a properly configured array, beamforming can achieve high-quality separation.

Disadvantage: The amount of noise attenuation increases as the number of microphones.

C. ICA (Independent Component Analysis):

ICA is method for separating multivariate (multidimensional) signal into its subcomponents. ICA assumes thee subcomponents of multivariate signal are independent of each other and they are non-Gaussian signals. ICA has been usually used in various 'blind' source separation tasks, where no or little prior information is available about the source signals. ICA has two assumptions:

1.The sound source signal are independent of each other.

2.The values in each source signal have non Gaussian distribution[12][13].

Disadvantage: A key and primary issue of this method is before an effective source separation the system should estimate the number of unknown sources from the mixed signals.

D. Pitch Estimation and Tracking

Pitch is a perceptual property that allows the ordering of sounds on a frequency-related scale. The technical term for this property is fundamental frequency (f0).Along with duration, loudness, and timbre pitch is also a major auditory attribute of musical tones. Pitch may be quantified as a frequency, but pitch is not a purely objective physical property; it is a subjective psychoacoustical attribute of sound. Historically, the study of pitch and pitch perception has been a central problem in psychoacoustics (i.e. the scientific study of sound perception). The estimation of pitch is highly related to source separation in the field of music. Many music source separation methods use pitch estimation as a previous step. Additionally pitch estimation approaches often share the same techniques as those of source separation. In the field of pitch estimation several tasks are often differentiated. Monophonic pitch estimation consists in estimating the pitch line of an audio recording where a single pitched sound is present at any given time.

Predominant pitch, bass line or melody estimation often refers to the estimation of one of the pitch lines in a polyphonic recording, where the selection of the pitch line depends on the application. Multiple pitch estimation consists in extracting all the pitch lines in a polyphonic recording. These two last families of methods are the ones of interest in the field of source separation [6][7].

Disadvantages: Estimated fundamental frequency of singing is difficult to be very accurate because of the influence of accompaniment. Even if the estimated fundamental frequency is correct the extracted harmonics of singing voice are not completely pure because the some harmonics components of singing voice may be superimposed by pitched instrument.

E. Sparse coding

Sparse coding is class of unsupervised learning which represents a mixture signal in terms of a small number of active elements chosen out of a larger set. Sparse coding is an efficient approach for learning structures and separating sources from mixed data. Sparse coding is a basic task in many fields including signal processing, neuroscience and machine learning where the goal is to learn a basis that enables a sparse representation of one given set of data, if one exists.

Sparseness: The concept of sparse coding refers to representation method where only few units are effectively used to represent typical data vector. As result of this most of the units takes values close to zero while only few unit takes significantly non zero values.The degree of sparseness is decided based on the values of vector. If elements of vector are roughly equally active then degree of sparseness is at low level. If the most of elements take zero values on other hand few of them take significant values then the degree of sparseness is at high level [11][14].

F. Non-negative matrix factorization

Non negative matrix factorization (NMF) is a low-rank approximation method where a nonnegative input data matrix is approximated as a product of two non-negative factor matrices. NMF has been used in various applications, including image processing, brain computer interface, document clustering, collaborative predictions, and so on. NMF plays important role in the sound source separation. The algorithms based on non-negative matrix factorization are robust and efficient for sound source separation when the sources or components of signal are dependent. NMF gives two output matrices one contain the all vocal attribute and other matrix indicates musical activities (i.e. musical notes).

Recent advances in matrix factorization methods suggest collective matrix factorization or matrix co-factorization to incorporate side information, where several matrices (target and side information matrices) are simultaneously decomposed, sharing some factor matrices. Matrix co-factorization methods have been developed to incorporate label information, link information, and inter-subject variations [2][9].

Disadvantage – Imposes only the non-negativity constraint.

G. Non negative matrix partial co factorization:

Many algorithms based on the non-negative matrix factorization (NMF) were developed in applications for blind or semi-blind source separation and those NMF algorithms are efficient and robust for source separation when sources are statistically dependent under conditions that additional constraints are imposed such as non-negativity, sparsity, smoothness, lower complexity or better predictability. However, without any prior knowledge of a source signal, the standard NMF cannot separate specific source signal from the mixing signal. To tackle this problem, nonnegative matrix partial co-factorization (NMPCF) was introduced. NMPCF is a joint matrix decomposition integrating prior knowledge of singing voice and accompaniment, to separate the mixture signal into singing voice portion and accompaniment portion. Matrix co-factorizations can be served as a useful tool when side information matrices are available, in addition to the target matrix to be factorized. NMPCF was emerged from the concept of joint decomposition or collective matrix factorization, which make the multiple input matrices be decomposed into several factor matrices while some of them are shared, therefore, shows a greater potential in singing voice separation from monaural recordings[3] [4].

V.CONCLUSION

This paper presents different source separation methods and also included NMPCF as new method for source separation which gives better performance than existing methods of source separation. So the NMPCF can be used for singer identification with better performance.

REFERENCES

[1] Tuomas Virtanen ,"Unsupervised Learning Methods for Source Separation in Monaural Music Signals" Tuomas Virtanen

[2] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 3, pp. 1066–1074,Mar. 2007.

[3] J. Yoo et al., "Nonnegative matrix partial co-factorization for drum source separation," in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., 2010, pp. 1942 1945.

[4] M. Kim et al., "Nonnegative matrix partial co-factorization for spectral and temporal drum source separation," IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1192–1204, Dec. 2011.

[5] Y. Hu and G. Z. Liu, "Singer identification based on computational auditory scene analysis and missing feature methods," J. Intell. Inf. Syst., pp. 1–20, 2013.

[6] McAulay, Robert J., and Thomas F. Quatieri. "Pitch estimation and voicing detection based on a sinusoidal speech model." Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on. IEEE, 1990.

[7] T. Virtanen, A. Mesaros, and M. Ryynanen, "Combining pitch-based inferenceandnon-negative spectrogram factorization in separating vocals from polyphonic music," in Proc. ISCA Tutorial Res. Workshop Statist. Percept. Audit. (SAPA), 2008

[8] Zafar Rafii and Bryan Pardo, "REpeating Pattern Extraction Technique (REPET): A Simple Method for Music/Voice Separation", IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 1, pp. 71 – 82, January 2013.

[9] Ying Hu and Guizhong Liu, "Separation of Singing Voice Using Nonnegative Matrix Partial CoFactorization for Singer Identification", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 4, pp. 643 – 653, April 2015.

[10] Yipeng Li, DeLiang Wang, Separation of Singing Voice from Music Accompaniment for Monaural Recordings, IEEE Transactions on Audio, Speech, and Language Processing,v.15 n.4, p.1475-1487, May 2007.

[11] Virtanen, Tuomas. "Sound source separation using sparse coding with temporal continuity objective." Proc. ICMC. Vol. 3. 2003.

[12] ICASSP 2007 Tutorial - Audio Source Separation based on Independent Component Analysis Shoji Makino and Hiroshi Sawada (NTT Communication Science Laboratories, NTT Corporation)

[13] Makino, Shoji, et al. "Audio source separation based on independent component analysis." Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on. Vol. 5. IEEE, 2004.

[14] Virtanen, Tuomas. "Separation of sound sources by convolutive sparse coding." ISCA Tutorial and Researc Workshop (ITRW) on Statistical and Perceptual Audio Processing. 2004.

[15] Non-negative matrix factorization based compensation of music for automatic speech recognition, Bhiksha Raj, T. Virtanen, Sourish Chaudhure, Rita Singh, 2010.

[16] Reynolds, Douglas A., Thomas F. Quatieri, and Robert B. Dunn. "Speaker verification using adapted Gaussian mixture models." Digital signal processing 10.1 (2000): 19-41.

[17] Hochreiter, Sepp, and Michael C. Mozer. "Monaural separation and classification of mixed signals: A support-vector regression perspective." 3rd International Conference on Independent Component Analysis and Blind Signal Separation, San Diego, CA. 2001.