International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 05, May-2019

Support Vector Machine and Sounds

Anam Bansal

Assistant Professor, Central University of Punjab, Bathinda

Abstract— Sounds produced through different means can help in recognizing the context of sound generation or these sounds can help in various fields if classified or detected properly. Sounds, in other words, audio have been used since years in different applications and researchers have used various neural networks and machine learning models for recognizing and classifying sounds. Audio event detection and event classification have achieved a new phase using machine learning classifiers such as Support Vector Machines (SVM), K- Nearest Neighbour, Artificial Neural Networks etc. The most promising classifier that is used in field of audio is Support Vector Machine. It has achieved good performance in every application related to audio. In this paper, the use of SVM for various audio related tasks is discussed.

Keywords— Machine Learning, Support Vector Machines, Hyperplane, Kernels, Ensemble

I. INTRODUCTION

Audio can be described as mechanical waves in gases, liquids, and solids. Audio recognition and classification can be described as identification of sounds in various contexts and employing them for certain useful researches. Sounds or in other words, audio have been used in various disciplines like speech synthesis[1], speech recognition[2][3], music genre classification [4], discrimination of normal and pathological data in hospitals[5]. It has also been employed in aircraft industry [6][7]. Then various environmental events are detected using sounds [8]. Researchers used different techniques and classifiers for audios.

SVM has been massively used for sound related applications. SVM is a machine learning algorithm which is based on the Vapnik-Chervonenkis theory. It is powerful algorithm that generalizes the model to the new data. This property of SVM is called regularization. SVM was basically designed for solving supervised machine learning problems in which the training data is already given the labels of appropriate classification classes.

Describing SVM geometerically, it searches for the op- timal hyperplane which separates the two classes of the context. Initially SVM was designed for linear problems but since all problems were not linear, so researchers found a way to transform non linear problems to linear problems using kernels which is discussed in next section.

Following section discusses the framework of SVM, followed by the work of researchers who have used SVM in areas of audio.

II. SUPPORT VECTOR MACHINES

SVM is a supervised classifier of machine learning. It is widely used for classification. Given a training set, SVM finds a hyperplane which separates the data set into two classes. The items in training data are represented as points in the plane. Hyperplane, in simple words, if we think of a binary classification, is a line that divides the data set into two classes. Hyperplane is decided by taking into account the distance of data points from the hyperplane. The distance of the data points from the hyperplane should be maximum, while retaining the correct class. The distance between hyperplane and closest points of each class is called margin. More the margin, more the chances of accurate classification. When new test data comes,when the data is projected onto the plane, the side of the hyperplane in which the projections of items are, that is class of those items.

The points which decide the position of hyperplane are called support vectors. In other words, the data points of each class closest to the hyperplane are support vectors. They are critical elements, as if they are removed or altered, the position of hyperplane will alter. Figure 1 describes the hyperplanes, support vectors and margin. Since the hyper- plane B has the highest margin from support vectors of both classes, so hyperplane B is the correct one. Two dimensional classification using SVM is depicted in figure 2.

If the data to be classified is not linearly separable i.e. it is not possible to find a line or plane separating

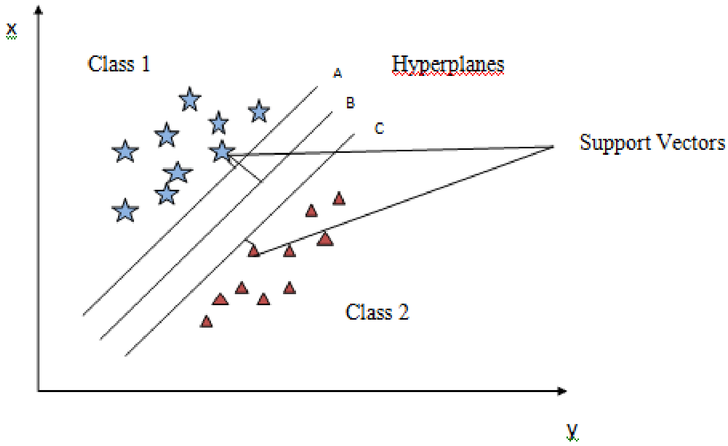

Fig. 1: Hyperplane in SVM

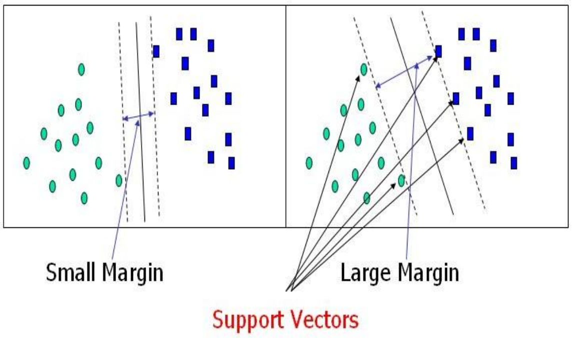

Fig. 2: Two dimensional Classification

the data points into classes (Figure 3), then the data should be mapped from low dimension to higher dimension using non-linear mapping already decided. The non-linear function used for mapping is also called kernel function. Commonly used and known kernel functions are Gaussian radial basis function (Figure 4), multilayer perceptron, linear (Figure 6), quadratic, polynomial (Figure 5)etc. A separating hyperplane is found to classify the higher dimensional data. Non-Linear mapping is depicted in Figure 7.

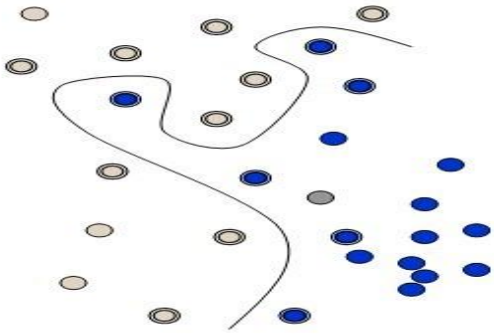

Fig. 3: Multidimensional Classification

SVM is a linear classifier basically but it becomes non- linear classifier when non-linear separable data is mapped to higher dimensions using non-linear mapping. SVM was initially developed only for two class problems. But if the problem involves multiple classes, then it is solved using n SVMs. It can be considered a set of binary classification subproblems. Each SVM out of n SVM divides one class from the other rest of the classes. The decomposition of the problem of multiclass problem into several binary problems is called binarization. There are two techniques employed to solve binarization problem:

1) **One-vs-all(OVA):** In this technique, using SVM one class is separated from all other classes by classifying the examples into two categories. One containing all the examples belonging to that particular class and other containing the examples not belonging to that class. So, there is one SVM for each class.

2) **One-vs-One(OVO):** In this technique, there is one SVM for each pair of classes. Each SVM performs classification between two classes. One class wins in case of each classification task. The final decision about the class of

the data item is taken through voting. The score of the winner class increases based on the decision of each classifier. At the end of the classification process,the class which has the maximum score is the class of the data item.

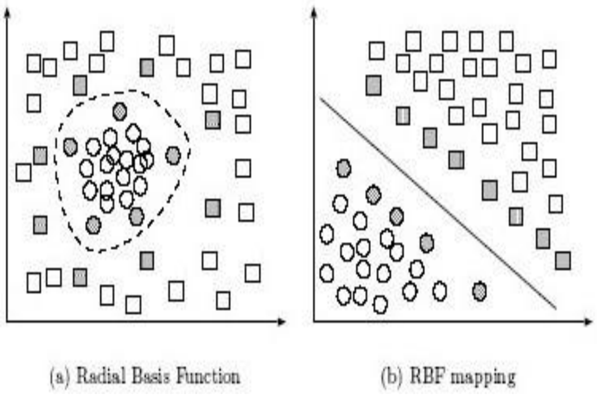

Fig. 4: Radial Basis Kernel function

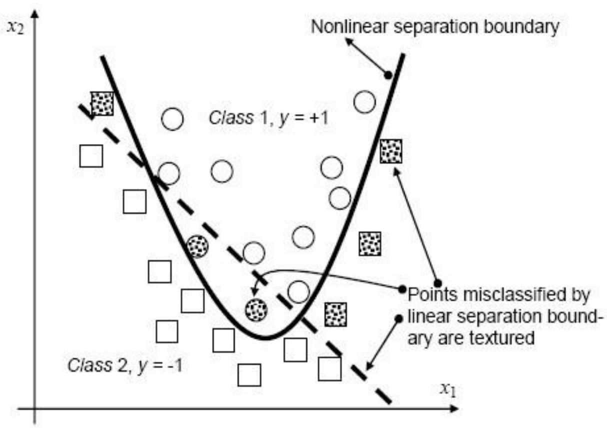

Fig. 5: Polynomial Kernel Function

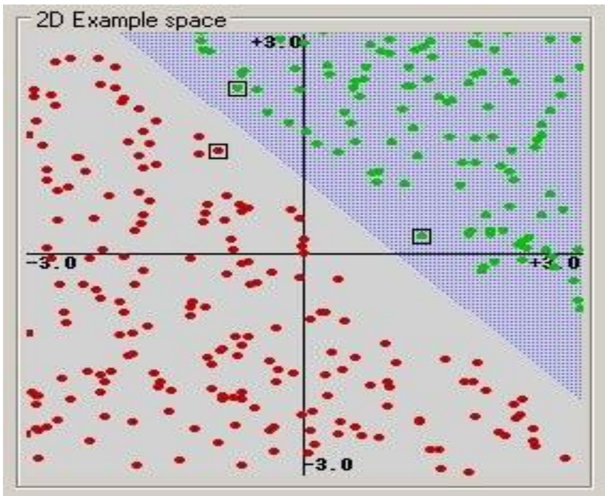

Fig. 6: Linear Kernel Function

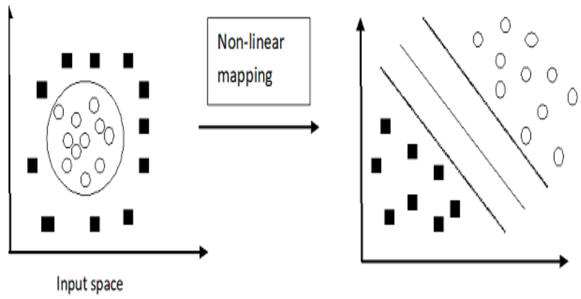

Fig. 7: Non-Linear Mapping

III. SVM IN SOUND RELATED APPLICATIONS

SVM has been widely employed for various sound related applications.
Music Genres are automatically classified using audio signal using SVM[9]. The kernel function used with SVM is polynomial kernel function. Accuracy if 78% and 81% is obtained with GTZAN dataset over the ten musical genres and the ISMIR 2004 genre dataset over the six musical genres respectively.

Music mood classsification is obtained using Audio power and audio harmonicity features[10]. SVM is used for classification of features. Accuracy obtained was 74.28% using Audio Power and Audio Harmonicity, 37.14% using Audio Spectrum Projection, and 28.57% using Audio Power, Au- dio Harmonicity and Audio Spectrum Projection. Another researcher used a set of 133 descriptors and a support vector machine classifier to predict the mood cluster in music [11]. In this research work, SVM is optimized using grid search algorithm.

For civil infrastructure projects, continuous tracking and monitoring of tasks performed by construction heavy equipment using SVM [12] [13]. It includes multiple steps- record- ing of audio signals from construction equipment using audio sensors, filtering of audio signals to reduce background noise, converting them into time frequency representations using Short-Time Fourier Transform(STFT), classifying these rep- resentations using SVM and window filtering the output of the classifier by setting proper thresholds.

For multimedia content analysis, audio classification and segmentation is obtained using ensemble of SVM and Artificial neural networks(ANNs) [14]. The ensemble segments a superimposed audio stream on the basis of its content into four main audio types: pure-speech,music, environment sound, and silence. Audio stream is classified, firstly, into speech and nonspeech segment by using bagged support vector machines; nonspeech segment is further classified into music and environment sound by using artificial neural net- works and lastly, speech segment is classified into silence and pure-speech segments on the basis of rule-based classifier. Misclassification rate is reduced using SVM.

Wavelets and SVMs are used to classify and categorize the audio data[15]. Wavelets are first applied to extract audio features. SVM classifies the features and classification erros are reduced from 16(8.1%) to six (3.0%).

Human sounds are classified using SVM using psycoa- coustic data. A scream classification model, with sounds of speech and screams indicating different acoustical character- istics, was investigated.

Classification and Segmentation of audio streams is done using SVM [16] [17][18]. An audio clip is classified into one of the five classes: pure speech, non-pure speech, music, environment sound, and silence. SVM achieves higher accuracy as compared to K-Nearest Neighbour and Guassian Mixture Model(GMM).

Human emotions are differentiated into four emotional states namely happy, sad, anger and neutral. SVM is used to recognize human emotions either from the speaker's utterances [19] [20] [21] or in response to audio music using brain signals [22].

Mixed type audio classification is accomplished using Support Vector Machines(SVM). SVM classify audio data into five types: music, speech, environment sound, speech mixed with music, and music mixed with environment sound [23]. The experimental results show that the proposed system outperforms other classification systems using k Nearest Neighbor (k-NN), Neural Network (NN), and Naive Bayes (NB).

In the field of sports, a novel audiovisual feature based framework is proposed for event detection in broadcast video of multiple different field sports. This is done through SVM [24].

A mobile phone identifier called Weighted Support Vector Machine with Weighted Majority Voting (WSVM-WMV) for a closed-set mobile phone identification task is proposed [25]. The proposed WSVM-WMV can be regarded as a generalization of the traditional SVM identifier. On using Mel-frequency Cepstral Coefficients (MFCC) and Linearfrequency Cepstral Coefficients (LFCC) as the feature vectors, the proposed identifier can improve the identification accuracy from 92.42% to 97.86% and from 90.44% to 98.33% respectively, as compared with the traditional SVM identifier in identifying a set of 21 mobile phone.

IV. CONCLUSION

Support Vector Machines have been used extensively in various sound related applications. SVM is a machine learning classifier that provides the best results as compared to other known machine learning classifiers in the field of sounds. It has been widely employed in sound source clas- sification, emotion recognition, context classification, sports event detection and heavy equipment event detection in civil infrastructure projects.

References

[1] Andrew J Hunt and Alan W Black. "Unit selection in a concatenative speech synthesis system using a large speech database". In: *1996 IEEE International Conference on Acoustics, Speech, and Signal Pro cessing Conference Proceedings*. Vol. 1. IEEE. 1996, pp. 373–376.

[2] Bhupinder Singh, Neha Kapur, and Puneet Kaur. "Speech recognition with hidden Markov model: a review". In: *International Journal of Advanced Research in Computer Science and Software Engineering* 2.3 (2012).

[3] Mathias De Wachter et al. "Template-based contin- uous speech recognition". In: *IEEE Transactions on Audio, Speech, and Language Processing* 15.4 (2007), pp. 1377–1390.

[4] George Tzanetakis and Perry Cook. "Musical genre classification of audio signals". In: *IEEE Transactions on speech and audio processing* 10.5 (2002), pp. 293– 302.

[5] Zulfiqar Ali et al. "Clinical informatics: mining of pathological data by acoustic analysis". In: *2017 International Conference on Informatics, Health & Technology (ICIHT)*. IEEE. 2017, pp. 1–8.

[6] Richard O Nielsen. "Acoustic detection of low flying aircraft". In: *2009 IEEE Conference on Technologies for Homeland Security*. IEEE. 2009, pp. 101–106.

[7] LP Dickinson and Neville H Fletcher. "Acoustic detec- tion of invisible damage in aircraft composite panels". In: *Applied Acoustics* 70.1 (2009), pp. 110-119

[8] Aki Harma, Martin F McKinney, and Janto Skowronek. "Automatic surveillance of the acoustic activity in our living environment". In: *2005 IEEE International Conference on Multimedia and Expo*. IEEE. 2005, 4–pp.

[9] YMD Chathuranga and KL Jayaratne. "Automatic mu- sic genre classification of audio signals with machine learning approaches". In: *GSTF Journal on Computing (JoC)* 3.2 (2018).

[10] Johanes Andre Ridoean et al. "Music mood clas- sification using audio power and audio harmonicity based on MPEG-7 audio features and Support Vector Machine". In: *2017 3rd International Conference on Science in Information Technology (ICSITech)*. IEEE. 2017, pp. 72–76.

[11] Cyril Laurier et al. "Audio music mood classification using support vector machine". In: *MIREX task on Audio Mood Classification* (2007), pp. 2–4.

[12] Chieh-Feng Cheng et al. "Activity analysis of con- struction equipment using audio signals and support vector machines". In: *Automation in Construction* 81 (2017), pp. 240–253.

[13] Chieh-Feng Cheng et al. "Audio signal processing for activity recognition of construction heavy equip- ment". In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construc- tion*. Vol. 33. Vilnius Gediminas Technical University, Department of Construction Economics fffdfffdfffd. 2016, p. 1.

[14] Saadia Zahid et al. "Optimized audio classification and segmentation algorithm by using ensemble meth- ods". In: *Mathematical Problems in Engineering* 2015 (2015).

[15] Chien-Chang Lin et al. "Audio classification and categorization based on wavelets and support vector machine". In: *IEEE Transactions on Speech and Audio Processing* 13.5 (2005), pp. 644–651.

[16] Lie Lu, Hong-Jiang Zhang, and Stan Z Li. "Content- based audio classification and segmentation by using support vector machines". In: *Multimedia systems* 8.6 (2003), pp. 482–492.

[17] Guodong Guo and Stan Z Li. "Content-based au- dio classification and retrieval by support vector ma- chines". In: *IEEE transactions on Neural Networks* 14.1 (2003), pp. 209–215.

[18] Makoto Kumon et al. "Sound source classification using support vector machine". In: *IFAC Proceedings Volumes* 40.13 (2007), pp. 465–470.

[19] MS Sinith et al. "Emotion recognition from audio signals using Support Vector Machine". In: *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*. IEEE. 2015, pp. 139–144.

[20] Jasmine Bhaskar, K Sruthi, and Prema Nedungadi. "Hybrid approach for emotion classification of audio conversation based on text and speech mining". In: *Procedia Computer Science* 46 (2015), pp. 635–643.

[21] Prajakta P Dahake, Kailash Shaw, and P Malathi. "Speaker dependent speech emotion recognition us- ing MFCC and Support Vector Machine". In: *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*. IEEE. 2016, pp. 1080–1084.

[22] Adnan Mehmood Bhatti et al. "Human emotion recog- nition and analysis in response to audio music using brain signals". In: *Computers in Human Behavior* 65 (2016), pp. 267–275.

[23] Lei Chen, Sule Gunduz, and M Tamer Ozsu. "Mixed type audio classification with support vector machine". In: *2006 IEEE International Conference on Multime- dia and Expo*. IEEE. 2006, pp. 781–784.

[24] David A Sadlier and Noel E O'Connor. "Event detec- tion in field sports video using audio-visual features and a support vector machine". In: *IEEE Transactions on Circuits and Systems for Video Technology* 15.10 (2005), pp. 1225–1233.

[25] Yuechi Jiang and Frank HF Leung. "Mobile phone identification from speech recordings using weighted support vector machine". In: *IECON 2016-42nd An- nual Conference of the IEEE Industrial Electronics Society*. IEEE. 2016, pp. 963–968.

[26] Jing Liu and Lingyun Xie. "Svm-based automatic clas- sification of musical instruments". In: *2010 Interna- tional Conference on Intelligent Computation Technol- ogy and Automation*. Vol. 3. IEEE. 2010, pp. 669–673.

[27] Jia-Ching Wang et al. "Robust environmental sound recognition with fast noise suppression for home automation". In: *IEEE Transactions on Automation Science and Engineering* 12.4 (2015), pp. 1235–1242.

