International Journal of Electronics and Communication Engineering (IJECE)
ISSN(P): 2278-9901; ISSN(E): 2278-991X
Vol. 2, Issue 5, Nov 2013, 73-80
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

UNDERWATER ACOUSTIC TARGET CLASSIFICATION SYSTEM USING SVM

SHERIN B. M, SUPRIYA M. H & P. R. SASEENDRANPILLAI

Department of Electronics, Cochin University of Science and Technology, Cochin, India

ABSTRACT

Underwater target classification is often a demanding task. Underwater acoustic target classification systems can identify the acoustic target pertaining to their characteristic acoustic signature captured by hydrophones. The performance of any classification system depends on the classification algorithm being used and the feature extraction method used to identify the acoustic signature, of which the Support Vector Machine (SVM) method features better generalization ability when dealing with high-dimensional data. SVM's are nonparametric models that rely on Structural Risk Minimization (SRM) principle in which model complexity is chosen according to data complexity. This paper presents a study on SVM based underwater acoustic target classification. Classification of 4 classes of acoustic targets using the nonlinear multiclass SVM algorithm is discussed. The acoustic features are extracted using Mel Frequency Cepstral Coefficients (MFCC), which has been extensively used in target recognition from acoustic signals. After preprocessing the acoustic data, the MFCC coefficients are extracted frame-by-frame and the feature vector at each frame is clustered using the k-means algorithm to form the acoustic signature. The acoustic signatures are labelled and fed to the SVM algorithm for classification. Performance of the proposed classifier is evaluated using the cross-validation technique. The proposed SVM classifier shows good generalization ability with an error rate of only 9% when evaluated with 10-fold cross-validation.

KEYWORDS: Feature Extraction, *k*- Means Clustering, Mel Frequency Cepstral Coefficients, Support Vector Machines

INTRODUCTION

Underwater target classification is a difficult and challenging task. Underwater target activity is reflected by acoustic events with each target having its own 'acoustic signature'. In its very basic form, an underwater target classification system is an acoustic recognition system. But the heavy distortion imposed by underwater propagation effects together with the noise introduced by the sea makes underwater target classification a cumbersome task. Fast and accurate underwater target recognition in the interference filled ocean is often required for military applications.

The classification problem concerns the construction of a procedure that will be applied to a variety of acoustic signals, in which each new signal is assigned to one of a set of pre-defined classes on the basis of observed features[1]. Automatic recognition, description and classification have become an important problem in a variety of scientific applications. Many classification algorithms exist which includes classical statistical methods such as Discriminant analysis, Mixture models, Naive Bayes classifiers, Decision tree and Rule based methods. Modern techniques include Artificial Neural Networks (ANN) and Support Vector Machines (SVM) based classification.

SVM was suggested by Vapnik in early 1990's. Currently, an outbreak of interest towards SVM's has emerged due to paramount advantages it offers. SVM's are nonparametric models, wherein parameters are not predefined and their number depends on the training data used. As opposed to Empirical Risk Minimization (ERM) followed by traditional methods, SVM relies on Structural Risk Minimization (SRM) in which parameters that define the capacity of the model are data driven in such a way as to match the model capacity to data complexity[2]. SVM's minimize true or expected risk thus minimizing generalization error. Furthermore, SVM's do not suffer from the curse of dimensionality.

The proposed underwater classification system is composed of the four modules.1. Signal processing – In this module, acoustic feature vectors which characterize the properties of individual acoustic files are extracted.2. Acoustic Modeling – This module performs a reduction of extracted feature vectors by modeling their distributions to form codewords.3. Acoustic Database - The codewords are updated and stored in a codebook which forms the acoustic database.4. Classifier – An SVM based classifier is employed to classify the acoustic signals.

The advantage of using SVM based classification technique is that SVM's are relatively easy to implement , very robust due to its sound theoretical background and do not suffer from the curse of dimensionality. Any dimensional problem is easily solvable with SVM keeping the model complexity relatively low when compared to other approaches. In SVM model capacity matches the training data complexity which resolves problems like over-fitting and under-fitting. The SVM creates a model with minimized VC dimension, which results in a low expected probability of error and thus good generalization performance. The code for the proposed classifier is developed in the MATLAB software package and performs classification satisfactorily.

THEORY AND DEFINITIONS

Statistical Learning Theory

The main goal of statistical learning theory (SLT) suggested by Vapnik is to provide a framework for studying the problem of inference, making predictions, constructing models from a set of data in a statistical framework [3]. The traditional concept of ERM is to find a function f(x) that minimizes the average risk on the training set. Vapnik's procedure is that for a given amount of data the hypothesis which minimizes the true risk must be chosen. Vapnik introduced a guaranteed true risk with the probabilistic confidence η, $0\leq\eta<1$. The upper bound of this risk

$$R(g) \leq R_{emp}(g) + \varphi;\ \varphi = \sqrt{\frac{h\left(log\left(\frac{2L}{h}\right)+1\right)- log\left(\frac{\eta}{4}\right)}{L}} \tag{1}$$

Where $R_{\text{emp}}(g)$ is the empirical risk and ϕ is a confidence interval. The confidence interval ϕ is proportional to VC dimension h, which is a measure of the capacity of a statistical classification algorithm and inversely proportional to number of training data L. As the ratio L/h gets smaller, the VC confidence becomes larger and the true risk diverges from the empirical risk. Thus for small amount of data ERM may not minimize the true risk, thus giving rise to generalization errors. Structural risk minimization minimizes the upper bound on true risk. Since the upper bound is independent of the underlying probability distribution p (x,y), this is valid for all possible p (x,y) and for any number of training data. The algorithm for SVM is an actual implementation of the SRM principle.

SUPPORT VECTOR MACHINES

SVM's are powerful tools for data classification that has been highly successful in a variety of applications. SVM's stemmed from the theory of Structural Risk Minimization and is a statistical classification method designed for binary classification.

SVM relies on three key ideas[4]. The first idea is to map the data to a high dimensional space, which may convert complex classification problems into simpler problems utilizing linear classifiers in this space. The second idea is to use only the training patterns that are near the decision surface for classification. These training patterns are called Support Vectors. The third key idea is to find the hyperplane that separates the data with the largest margin. This hyperplane is called Optimal Separating Hyperplane (OSH). Such a maximal margin classifier will have good generalization characteristics.

The learning problem setting for SVM is as follows: there is some unknown and nonlinear dependency (mapping, function) $y = f(x)$ between some high-dimensional input vector x and output y. There is no a-priori information about the probability distribution of the input and so a distribution free learning must be performed. The only information available is a training data set $\{\chi = [x(i),y(i)] \epsilon \Re m \times \Re,\ i=1,.....n\}$ where n stands for the number of the training data pairs and is therefore equal to the size of the training data set χ.

Linear SVM

Consider the problem of separating the training vectors belonging to two linearly separable classes.

$$(x_1,y), (x_2,y),....................., (x_n,y)\ ;\ x \in \Re^n,\ y \in \{+1,-1\} \tag{2}$$

A linear discrimination function/hyperplane is $d(x) = w^T.x + b$,where$w \in \Re^n$ and (.) denotes the dot product. w and b are the variables of the optimization problem,with feasible set defined by all possible separation Hyperplane, which is represented as

$$y_i(w^T x_i + b) \geq 1\ i = 1, \dots \dots \dots, n \tag{3}$$

Let H_0 be the region of vectors which satisfy the equation $d(x) = 0$ and H_1 and H_{-1} are two hyperplanes parallel to H_0 and defined by $d(x) = 1$ and $d(x) = -1$ respectively. The distance separating the H_1 and H_{-1}hyperplane which is the margin is $\frac{2}{\|w\|}$.

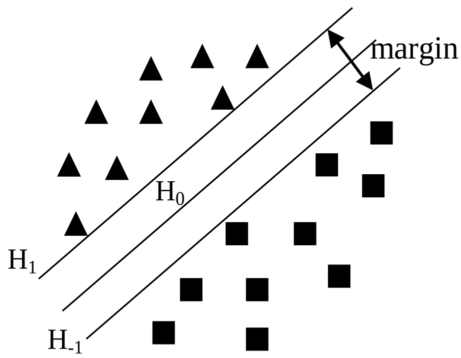

Figure 1: Binary Linear Classification

The problem of finding the optimal separating hyperplane is to maximize the margin $\frac{2}{\|w\|}$ or minimize$\frac{\|w\|^2}{2}$. This is a quadratic optimization problem with linear constraints defined by inequalities i.e: Find w and b such that is $\frac{\|w\|^2}{2}$ minimized for all (x_i, y_i) and

$$y_i(w^T x_i + b) \geq 1\ i = 1, \dots \dots \dots, n \tag{4}$$

The solution to the quadratic optimization problem involves constructing a dual problem where a Lagrange multiplier α_i associated with every constraint in the primary problem. Such an optimization problem is solved by the saddle point of the Lagrange functional.

$$L_p(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i \{y_i[w^T + b] - 1\} \tag{5}$$

where α_i are Lagrange multipliers. The search for an optimal saddle point (w_0, b_0, α)involves minimizing Lagrangian L_p*w.r.t.* w and b and maximizing *w.r.t.* nonnegative α_i [5]. Classical Lagrangian duality enables the primary solution in equation (5) to be transformed to its dual problem which is easier to solve. The dual problem is given by

$$maxL_d(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i\, y_j\, \alpha_i \alpha_j\, \mathrm{x}_i^T \mathrm{x}_j \tag{6}$$

s.t. $\alpha_i \geq 0, i = 1, \dots \dots \dots, n$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$

This classic quadratic optimization problem has a unique solution. According to Kuhn-Tucker theorem of optimization, the optimal solution satisfies

$$\alpha_i[y_i(w.x_i + b) - 1] = 0,\ i = 1,\dots\dots\dots\dots,n \tag{7}$$

The equation (7) will have non-zero Lagrange multipliers only when the points x_i satisfy

$$y_i(w.x_i + b) = 1 \tag{8}$$

These points are the support vectors which determine the Optimal Separating Hyperplane (OSH).

Soft Margin SVM

The learning procedure for linear SVM's is valid only when the training data sets are overlapping. In practice, the data may overlap due to noisy measurements or outliers. If the outliers are also accounted, the solution may not generalize well and so they are left on the 'wrong' side of the decision boundary. A soft margin is allowed by introducing a slack positive variable ξ_i for each training vector and all data inside this margin are neglected. Equation (2) can be modified for soft margin SVM as

$$y_i(w^T x_i + b) \geq 1 - \xi_i\ ;\ i = 1, \dots \dots \dots, n \text{ and } \xi_i > 0 \tag{9}$$

The problem of finding the optimal separating Hyperplane now becomes

$$min\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \text{ s.t } y_i(w^T x_i + b) \geq 1 - \xi_i \text{ for } i = 1, \dots \dots \dots, n \text{ and } \xi_i > 0 \tag{10}$$

where C is a penalty parameter, trading off the margin size for the number of misclassified data points. The solution to this problem is also achieved by constructing a dual Lagrangian problem. The resulting dual problem is same as that of equation (6) except the positivity constraint on now α_i becomes $0 \leq \alpha_i \leq C$; $i = 1,\dots\dots\dots,n$

Non-Linear SVM

When a linear boundary is inappropriate, a nonlinear classification problem can be solved by mapping input vectors $\mathrm{x}_i \epsilon \Re^m$ into vectors $\phi(\mathrm{x}_i)\epsilon\Re^s$ of a high dimensional feature space S (where ϕ represents mapping:$\Re^m \to \Re^s$) and to solve a linear classification problem in this feature space

$$X\epsilon\Re^m \to \phi(x) = [\phi_1(x)\phi_2(x), \dots \dots., \phi_s(x)]^T \epsilon\Re^s \tag{11}$$

The dual Lagrangian optimization problem now becomes

$$maxL_d(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j}^{n} y_i\, y_j\, \alpha_i \alpha_j\, k(x_i, x_j), \text{ s.t } \alpha_i \geq 0 \text{ for } i = 1,\dots\dots\dots,n \text{ and } \sum_{i=1}^{n} \alpha_i\, y_i = 0 \tag{12}$$

where $k(x_i, x_j) = \phi_i^T \phi_j$ is the kernel function performing the nonlinear mapping into feature space. Kernel functions could be any symmetric function that satisfies the Mercer's conditions[6]. Commonly used kernel functions are summarized in table 1.

The nonlinear binary classification problem consists of assigning a label to each input vector x through

$$d(x) = sgn(\sum_{i=1}^{n} \alpha_i y_i\, k(x, x_i) + b) \tag{13}$$

where*n* is the number of support vectors.

Table 1: Standard Kernel Functions

Kernel Functions	Type of Classifier
$K(x, x_i) = (x^T x_i)$	Linear, dot product
$K(x, x_i) = [(x^T x_i) + 1]^d$	Polynomial of degree *d*
$K(x, x_i) = \exp(-[\lVert x - x_i \rVert^2]/2\sigma^2)$	Gaussian RBF
$K(x, x_i) = \tanh[(x^T x_i) + b]$	Multilayer Perceptron
$K(x, x_i) = \dfrac{1}{\sqrt{\lVert x - x_i \rVert^2 + \beta}}$	Inverse multiquadratic function

Multi-Class SVM

SVM was originally proposed as a binary classification problem. There are several methods of extending the binary SVM classification problem to a multi-class one. The multi-class problem is typically solved by decomposing it to several binary classification problem in which standard SVM can be used. Two commonly used algorithms for multi-class classification are one-against-all (1-a-a) proposed by Vapnik and one-against-one (1-a-1) method proposed by Knerr. In the first approach, *K* classifiers are constructed to solve a *K*-class problem. The n^{th} classifier constructs a hyperplane between class *n* and the *K*-1 other classes. The second approach involves constructing (*K*(*K*-1)/2) hyperplanes separating each class from each other class. Compared with 1-a-a method, 1-a-1 method gives better training results[7]. However, the number of binary classifiers used by 1-a-1 algorithm increases with number of classes, resulting in lower training speed.

FEATURE EXTRACTION

The identification and selection of features play a crucial role in classification. A wide range of features extracted from time domain or transformation domain such as spectral, cepstral and bispectral methods have been used for implementing various types of acoustic classifiers with varying success rates. Popular cepstral methods for feature extraction that have been successfully used for acoustic classification are Linear Predictive Cepstral Coefficients (LPCC) and Mel Frequency Cepstral Coefficients (MFCC). In this work we have used MFCC's for feature extraction, because it combines the advantages of cepstrum together with a frequency scale based on human auditory system[8].

MFCC is based on the human peripheral auditory system. The human perception of the frequency contents follows a logarithmic scale instead of a linear scale. MFCC coefficients are obtained by taking Discrete Cosine Transform (DCT) of the logarithm of the short-term energy spectrum obtained after mel-scale filtering which is expressed on a mel-frequency scale.

The mel-scale filter bank consists of a series of triangular band pass filters with spacing and bandwidth determined by a constant mel-frequency interval. On the frequency axis, such a filter bank corresponds to a set of non-uniformly spaced filters with more and narrow filters in the low frequency region and less and wide filters in the high frequency region. Such an arrangement of the filter bank is to simulate the subjective spectrum of humans which is better at discerning small changes in pitch at low frequencies that at high frequencies.

RESULTS AND DISCUSSIONS

SVM based classifier was used for underwater acoustic classification. The database consists of 114 real backscattered signals from four different types of targets. Two targets namely, humpback whale and sealion are of mammalian origin and the other two, ship and boat are of mechanical origin.

The classification involves two phases namely training and testing phase. In the training phase the acoustic features extracted from sounds of known origin are used to build a reference model. In the testing phase, acoustic features extracted from test signals are fed to an SVM based classification system for classification. The ratio of the acoustic files selected for training and testing are user decidable. But an intensive training phase will certainly improve the performance of the classifier. The performance of the classifier in the testing phase is commonly evaluated by the formula

$$Error\ rate\ (\%) = \frac{No:of\ misclassified\ \ data}{Total\ \ no:of\ data} \times 100 \tag{14}$$

In this work popular cross-validation technique is used for measuring the performance of the classifier. Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. In k-fold cross validation the data set is randomly partitioned into k-equal sized sub data sets. Of the k sub data sets, a single data set is retained as the validation data for testing the model, and the remaining k-1 subsets are used as training data. The cross-validation process is then repeated k-folds with each of sub data set used exactly once as validation data. The error rate at each of the k iterations are calculated and averaged to produce a single estimate. 5-fold, 10-fold and 15-fold cross-validation is attempted to measure the performance of the proposed SVM based classifier. The results are indicated in figure (2) and figure (3).

The classification system consists of a feature extractor based on MFCC. Before the feature extraction step, the digitized acoustic waveform is subjected to pre-emphasis to reduce noise effects and to boost the high frequency contents. Pre-emphasis is done using a first order high pass filter of the form

$$y[n] = \ x[n] - \ \alpha x[n-1], 0.9 \leq \alpha \leq 1 \tag{15}$$

In the filter implementation, α the pre-emphasis parameter is selected as 0.95.

For MFCC computation, each input acoustic signal is split into several frames. The length of the frame is chosen by the expression

$$frame\ length = 2^{(floor\ [log_2(0.03 \times fs)])} \tag{16}$$

An overlapping is also applied to the frames. The hop size for overlapping is chosen as half of the frame size. On each frame a windowing technique is applied which will get rid of some of the information at the beginning and end of each frame. Overlapping of frames is advantageous which helps to reincorporate this information back into the extracted features.

The next step in computing MFCC’s is windowing. Windowing is performed to avoid distortions in the underlying spectrum. A wide range of window functions may be used. Three window functions- Hamming, Hanning and Rectangular window functions are compared. Results shown in figure (2) and figure (3) indicate that Hamming window gives the best performance. The windowed signal is transformed to the frequency domain by taking n-point DFT. The choice of n depends on N, the total number of samples in a frame. For even number of samples, n is chosen as (N+2)/2 and for odd number of samples n is chosen as (N+1)/2. Next step towards the computation of MFCC is mel-frequency scaling followed by mel-scale filtering. For Hz to mel transformation popular O’Shangnessy’s formula as in equation (17) is used.

$$m = \frac{\ln\left(1+\frac{f}{700}\right)\times 1000}{\ln\left(1+\frac{1000}{7000}\right)} \tag{17}$$

The mel-scale filter bank employed consists of a series of 12 triangular band pass filters with constant bandwidth and spacing on a mel-frequency scale. However these filters are non-uniformly spaced on the frequency axis with more filters in the low frequency regions and less filters in the high frequency regions. The logarithm of the filter bank energies is then taken. The log filterbank energies are then converted back to the time domain using Discrete Cosine Transform (DCT). The output of the DCT stage gives the MFCC’s. 12 coefficients are computed for each frame of an acoustic input. Thus for a single acoustic input the feature vector size would be no: of frames × 12.

The feature vectors so obtained are clustered using the *k*-means clustering algorithm (with *k*=1) to form the codewords. Each codeword will be a row vector with 12 columns. The codebook generated by arranging the set of codewords will thus be a vector of dimension no: of acoustic inputs×12. Each row of the codebook is a codeword which corresponds to the acoustic feature vector of an acoustic input. These are then fed to a nonlinear multiclass SVM classifier which employs one-against-all (1-a-a) approach for classification. Two commonly used kernel functions, namely Gaussian RBF and multilayer perceptron (MLP) have been used for classification. The results are indicated in figure (2) and figure(3). MLP showed lowest error rate when evaluated with cross-validation technique.

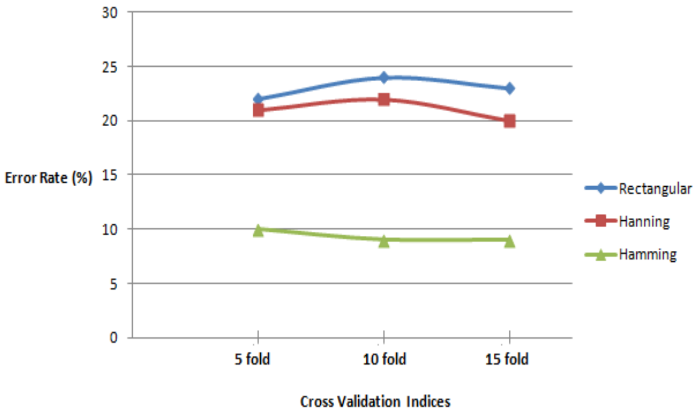

Figure 2: Error Rate with MLP Kernel

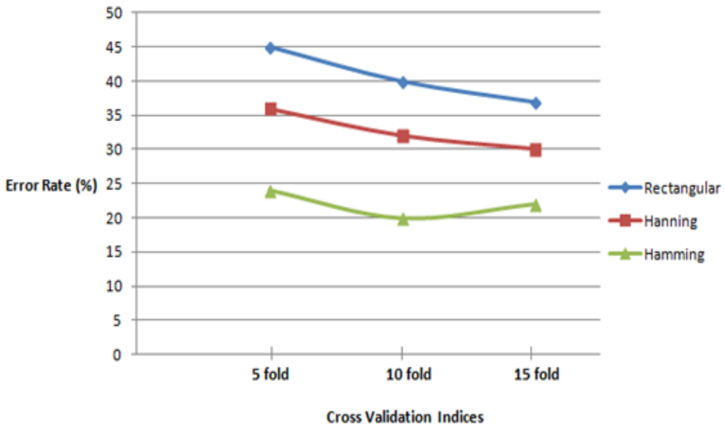

Figure 3: Error Rate with RBF Kernel

The performance of the classifier in the oceanic environment was simulated by calculating the error rate when the signals are corrupted in the ocean wave noise. The corrupted signals were obtained by forming a composite signal combining the target signal and the wave noise. The error rate according to equation (14) with the signals corrupted in wave noise was calculated to be 6.14%. MLP kernel which showed the best performance when evaluated with cross-validation was employed.

CONCLUSIONS

The proposed classifier showed best results with MLP kernel. SVM classifier offers many advantages over other classification methods and has been used successfully in many fields of machine learning. In this paper an underwater target classifier based on SVM is attempted. The results showed the superiority of MLP kernel over other standard kernel functions. The proposed classifier showed good generalization ability with an error rate of only 9% when evaluated with 10-fold cross-validation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Department of Electronics, Cochin University of Science and Technology, Kochi, India for providing the necessary facilities for carrying out this work.

REFERENCES

1. Donald Michie, David Spiegelhalter and Charles Taylor, *Machine Learning: Neural and Statistical Classification*, 2009 Edition, Overseas Press, ISBN: 978-8-188-68973-6.
2. Huang, Kecman and Kopriva, *Kernel Based Algorithms for Mining Huge Data Sets*, Studies in Computational Intelligence, Vol 17, Springer Publications, ISBN 3-540-26892-8.
3. VladmirVapnik, *The Nature of Statistical Learning Theory*, Springer, 2nd Edition 2000, ISBN: 978-0-387-98780-4.
4. Andrey Temko and Climent Nadeu, *Classification of Acoustic Events using SVM-based Clustering schemes*, The Journal of Pattern Recognition Society, 2005, pp.682-694..
5. Nello Cristianin and John Shawe-Taylor, *An Introduction to Support Vector machines and other Kernel based Learning methods*, Cambridge University Press, 2000, ISBN: 978-0-521-78019-3.
6. Bernhard Schölkopf and Alexander J. Smola, *Learning with Kernels*, ISBN:978-0-262-19475-4.
7. Chih-Wei Hsu and Chih-Jen Lin, *A comparison of Methods for Multiclass Support Vector Machines*, IEEE Transactions on Neural Networks, Vol.13, No.2, March 2002, pp.415-425.
8. C. J. Van der Merwe and J. A. Du Preez, *Calculation of LPC-based Cepstrum Coefficients using Mel-Scale Frequency Warping*, Proceedings of South African Symposium on Communications and Signal Processing, 1991, pp 17-21.