A Hybrid Framework using RBF and SVM for Direct Marketing

M.Govidarajan
Assistant Professor
Department of Computer Science and Engineering
Annamalai University
Annamalai Nagar-608002
Tamil Nadu, India

***Abstract*—one of the major developments in machine learning in the past decade is the ensemble method, which finds highly accurate classifier by combining many moderately accurate component classifiers. This paper addresses using an ensemble of classification methods for direct marketing. Direct marketing has become an important application field for data mining. In direct marketing, companies or organizations try to establish and maintain a direct relationship with their customers in order to target them individually for specific product offers or for fund raising. A variety of techniques have been employed for analysis ranging from traditional statistical methods to data mining approaches. In this research work, new hybrid classification method is proposed by combining classifiers in a heterogeneous environment using arcing classifier and their performances are analyzed in terms of accuracy. A Classifier ensemble is designed using Radial Basis Function (RBF) and Support Vector Machine (SVM) as base classifiers. Here, modified training sets are formed by resampling from original training set; classifiers constructed using these training sets and then combined by voting. Empirical results illustrate that the proposed hybrid systems provide more accurate direct marketing system.**

Keywords—Direct Marketing; Ensemble; Radial Basis Function; Support Vector Machine; Classification Accuracy.

I. INTRODUCTION

Data mining methods may be distinguished by either supervised or unsupervised learning methods. In supervised methods, there is a particular pre-specified target variable, and they require a training data set, which is a set of past examples in which the values of the target variable are provided.

Direct marketing [20] has become an important application field for data mining. In direct marketing [2] companies or organizations try to establish and maintain a direct relationship with their customers in order to target them individually for specific product offers or for fund raising. Large databases of customer and market data are maintained for this purpose. The customers or clients to be targeted in a specific campaign are selected from the database, given different types of information such as demographic information and information on the customer's personal characteristics like profession, age and purchase history.

The customers of a company are regarded as valuable business resources in competitive markets, leading to efforts to systematically prolong and exploit existing customer relations. Consequently, the strategies and techniques of customer relationship management (CRM) have received increasing attention in management science. CRM features data mining as a technique to gain knowledge about customer behaviour and preferences.

Data mining problems in the CRM domain, such as response optimization to distinguish between customers who will react to a mailing campaign or not, chum prediction, in the form of classifying customers for churn probability, cross-selling, or up-selling are routinely modeled as classification tasks, predicting a discrete, of- ten binary feature using empirical, customer centered data of past sales, amount of purchases, demographic or psychographic information etc. Customer retention has a significant impact on firm profitability. Gupta et al find that a 1% improvement in retention can increase firm value by 5%. [9]. Churn refers to the tendency for customers to defect or cease business with a company. Marketers interested in maximizing lifetime value realize that customer retention is a key to increasing long-run firm profitability. A focus on customer retention implies that firms need to understand the determinants of customer defection (churn) and are able to predict those customers who are at risk of defection at a particular point in time.

Response modeling is usually formulated as a binary classification problem. The customers are divided into two classes, respondents and non-respondents. A classifier is constructed to predict whether a given customer will respond or not. From a modeling point of view, however, several difficulties arise [22] [28]. One of the most noticeable is a severe class imbalance resulting from a low response rate: typically less than 5% of customers are respondents [7]. A typical binary classifier will result in lopsided outputs to the non-respondent class [14].

In other words, the classifier will predict most or even all customers not to respond. Although the classification accuracy may be very high since a majority of customers are in fact non-respondents. In this work, a model which identifies a subset of customers is constructed that includes as many respondents and as few non-respondents as possible Various classification methods have been used for response modeling such as statistical and machine learning methods. Recently, SVMs have drawn much attention and a few researchers have implemented them for response modeling [22] [27].

Classification is a very common data mining task. In the process of handling classification tasks, an important issue usually encountered is determining the best performing method for a specific problem [13]. Several studies address the issue. For example, Michie, Spiegelhalter, and Taylor try to find the relationship between the best performing method and data types of input/output variables.[18] Hybrid models have been suggested to overcome the defects of using a single supervised learning method, such as radial basis function and support vector machine techniques. Hybrid models combine different methods to improve classification accuracy. The term combined model is usually used to refer to a concept similar to a hybrid model. Combined models apply the same algorithm repeatedly through partitioning and weighting of a training data set. Combined models also have been called Ensembles. Ensemble improves classification performance by the combined use of two effects: reduction of errors due to bias and variance [11]. Recently, hybrid data mining approaches have gained much popularity; however, a few studies have been proposed to examine the performance of hybrid data mining techniques for response modeling [17].

This paper proposes a new hybrid classification method to improve the Classification accuracy. The primary objective of this paper is to construct ensemble of radial basis function and Support Vector Machine is to predict whether a given customer will respond or not for direct marketing in terms of classification accuracy.

The rest of this paper is organized as follows: Section 2 describes the related work. Section 3 presents hybrid direct marketing system and Section 4 explains the performance evaluation measures. Section 5 focuses on the experimental results and discussion. Finally, results are summarized and concluded in section 6.

II. Related Work

Direct marketing aims at obtaining and maintaining direct relations between suppliers and buyers within one or more product/market combinations. In marketing, there are two main different approaches to communication: mass marketing and direct marketing [16]. Mass marketing uses mass media such as print, radio and television to the public without discrimination. While direct marketing involves the identification of customers having potential market value by studying the customers' characteristics and the needs (the past or the future) and selects certain customers to promote. Direct marketing becomes increasingly popular because of the increased competition and the cost problem. It is an important area of applications for data mining, data warehousing, statistical pattern recognition, and artificial intelligence. In direct marketing, models (profiles) are generated to select potential customers (from the client database) for a given product by analyzing data from similar campaigns, or by organizing test mail campaigns [21]. Various classifiers have been employed such as logistic regression, neural networks and support vector machine.

Aristides Gionis et al have shown that the numbers of clusters discovered by their algorithms seem to be very reasonable choices: for the Votes dataset most people vote according to the official position of their political parties, so having two clusters is natural; for the Mushrooms dataset, notice that both ROCK and LIMBO achieve much better. [1]. Many aspects of churn have been modeled in the literature. First, whether churn is hidden or observable influence the overall approach to modeling. In some industries, customer defection is not directly observed, as customers do not explicitly terminate a relationship, but can become inactive. In other industries, however, the defection decision is observable as customers cease their relationship via actively terminating their contract with the firm [4].

The modeling approach could also depend critically on the relative importance placed on explanation/interpretation *vis a vis* prediction. Models that are better at explanation may not necessarily be better at prediction. The empirical literature in marketing has traditionally favored parametric models (such as logistic or probit regression or parametric hazard specifications and zero-inflated poisson models) that are easy to interpret. Similar to the previous discussion on acquisition, churn is a rare event that may require new approaches from data mining, machine learning and non-parametric statistics that emphasize predictive ability [10]. These include projection-pursuit models, jump diffusion models, neural network models, tree structured models, spline-based models such as Generalized Additive Models (GAM), and Multivariate Adaptive Regression Splines (MARS), and more recently approaches such as support vector machines and boosting [15].

Tang applied feed forward neural network to maximize performance at desired mailing depth in direct marketing in cellular phone industry. He showed that neural networks show more balance outcome than statistical models such as logistic regression and least squares regression, in terms of potential revenue and churn likelihood of a customer [23].

Xu et al proposed four combining classifier approaches according to the levels of information available from the various classifiers. The experimental results showed that the performance of individual classifiers could be improved significantly. [26]

Freund and Schapire proposed an algorithm the basis of which is to **a**daptively **r**esample and **c**ombine (hence the acronym--arcing) so that the weights in the resampling are increased for those cases most often misclassified and the combining is done by weighted voting. [5] [6]. Previous work has demonstrated that arcing classifiers is very effective for RBF-SVM hybrid system. [8]. It is surprising there are only few papers that seek to assess the state of research in this area, or outline the challenges unique to this area. This paper seeks to address this void.

In this paper, a direct marketing system is proposed using radial basis function and support vector machine and the effectiveness of the proposed RBF-SVM hybrid system is evaluated by conducting several experiments on voting database. The performance of the RBF-SVM hybrid classifier is examined in comparison with standalone RBF and standalone SVM classifier. This work focuses to understand the relative merits of the base and proposed hybrid approaches for CRM applications.

III. HYBRID DIRECT MARKETING SYSTEM

This section shows the proposed RBF-SVM hybrid system which involves Radial Basis Function (RBF) and Support Vector Machine (SVM) as base classifiers.

A. RBF-SVM Hybrid System

The proposed hybrid direct marketing system is composed of three main phases; preprocessing phase, classification phase and combining Phase.

1) Voting Dataset Preprocessing: First the data is collected from the United States Congressional Voting Records Database. Before performing any classification method the data has to be preprocessed. In the data preprocessing stage it has been observed that the datasets consist of many missing value attributes. By eliminating the missing attribute records may lead to misclassification because the dropped records may contain some useful pattern for Classification. The dataset is preprocessed by removing missing values using supervised filters.

2) Existing Classification Methods

*a) Radial Basis Function Neural Network:*The RBF [19] design involves deciding on their centers and the sharpness (standard deviation) of their Gaussians. Generally, the centres and SD (standard deviations) are decided first by examining the vectors in the training data. RBF networks are trained in a similar way as MLP. The output layer weights are trained using the delta rule. The RBF networks used here may be defined as follows.

- RBF networks have three layers of nodes: input layer, hidden layer, and output layer.
- Feed-forward connections exist between input and hidden layers, between input and output layers (shortcut connections), and between hidden and output layers. Additionally, there are connections between a bias node and each output node. A scalar weight is associated with the connection between nodes.
- The activation of each input node (fanout) is equal to its external input where is the th element of the external input vector (pattern) of the network (denotes the number of the pattern).
- Each hidden node (neuron) determines the Euclidean distance between "its own" weight vector and the activations of the input nodes, i.e., the external input vector the distance is used as an input of a radial basis function in order to determine the activation of node. Here, Gaussian functions are employed. The parameter of node is the radius of the basis function; the vector is its center.
- Each output node (neuron) computes its activation as a weighted sum The external output vector of the network, consists of the activations of output nodes, i.e., The activation of a hidden node is high if the current input vector of the network is "similar" (depending on the value of the radius) to the center of its basis function. The center of a basis function can, therefore, be regarded as a prototype of a hyper spherical cluster in the input space of the network. The radius of the cluster is given by the value of the radius parameter.

b) Support Vector Machine:

The support vector machine (SVM) is a recently developed technique for multi dimensional function approximation. The objective of support vector machines is to determine a classifier or regression function which minimizes the empirical risk (that is the training set error) and the confidence interval (which corresponds to the generalization or test set error) [24].

Given a set of N linearly separable training examples $\mathbf{S} = \{\mathbf{x_i} \in \mathbf{R^n} | \mathbf{i} = 1,2,...,\mathbf{N}\}$, where each example belongs to one of the two classes, represented by $y_i \in \{\div 1, -1\}$, the SVM learning method seeks the optimal hyperplane **w . x** +b = 0, as the decision surface, which separates the positive and negative examples with the largest margins. The decision function for classifying linearly separable data is:

$$f(\mathbf{X}) = sign(W.X + b) \quad (1)$$

Where **w** and b are found from the training set by solving a constrained quadratic optimization problem. The final decision function is

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} a_i y_i (x_i..x) + b\right) \quad (2)$$

The function depends on the training examples for which a_i s is non-zero. These examples are called support vectors. Often the number of support vectors is only a small fraction of the original data set. The basic SVM formulation can be extended to the non linear case by using the nonlinear kernels that maps the input space to a high dimensional feature space. In this high dimensional feature space, linear classification can be performed. The SVM classifier has become very popular due to its high performances in practical applications such as text classification and pattern recognition. The support vector regression differs from SVM used in classification problem by introducing an alternative loss function that is modified to include a distance measure. Moreover, the parameters that control the regression quality are the cost of error C, the width of tube ε and the mapping function ϕ.

In this research work, the values for polynomial degree will be in the range of 0 to 5. In this work, best kernel to make the prediction is polynomial kernel with epsilon = 1.0E-12, parameter d=4 and parameter c=1.0. A hybrid scheme based on coupling two base classifiers using arcing classifier adapted to data mining problem is defined in order to get better results.

3) Proposed RBF-SVM Hybrid System

Given a set D, of d tuples, arcing works as follows; For iteration i (i =1, 2,…..k), a training set, D_i, of d tuples is sampled with replacement from the original set of tuples, D. some of the examples from the dataset *D* will occur more than

once in the training dataset D_i. The examples that did not make it into the training dataset end up forming the test dataset. Then a classifier model, M_i, is learned for each training examples *d* from training dataset D_i. A classifier model, M_i, is learned for each training set, D_i. To classify an unknown tuple, X, each classifier, M_i, returns its class prediction, which counts as one vote. The hybrid classifier (RBF-SVM), M^*, counts the votes and assigns the class with the most votes to X.

Algorithm: Hybrid RBF-SVM using Arcing Classifier
Input:

- D, a set of d tuples.
- $k = 2$, the number of models in the ensemble.
- Base Classifiers (Radial Basis Function, Support Vector Machine)

Output: Hybrid RBF-SVM model, M^*.
Procedure:

1. For i = 1 to k do // Create k models
2. Create a new training dataset, D_i, by sampling D with replacement. Same example from given dataset *D* may occur more than once in the training dataset D_i.
3. Use D_i to derive a model, M_i
4. Classify each example d in training data D_i and initialized the weight, W_i for the model, M_i, based on the accuracies of percentage of correctly classified example in training data D_i.
5. endfor

To use the hybrid model on a tuple, X:

1. if classification then
2. let each of the k models classify X and return the majority vote;
3. if prediction then
4. let each of the k models predict a value for X and return the average predicted value;

The basic idea in Arcing [3] is like bagging, but some of the original tuples of D may not be included in Di, where as others may occur more than once.

IV. Performance Evaluation Measures

A. Cross Validation technique

Cross-validation (Jiawei Han and Micheline Kamber**,** 2003) sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. 10-fold cross validation is commonly used. In stratified K-fold cross-validation, the folds are selected so that the mean response value is approximately equal in all the folds.

B. Criteria for Evaluation

The primary metric for evaluating classifier performance is classification Accuracy: the percentage of test samples that the ability of a given classifier to correctly predict the label of new or previously unseen data (i.e. tuples without class label information). Similarly, the accuracy of a predictor refers to how well a given predictor can guess the value of the predicted attribute for new or previously unseen data.

V. Experimental Results and Discussion

A. Dataset Description

This data set includes votes for each of the U.S. House of Representatives Congressmen on the 16 key votes identified by the CQA. The CQA lists nine different types of votes: voted for, paired for, and announced for (these three simplified to yea), voted against, paired against, and announced against (these three simplified to nay), voted present, voted present to avoid conflict of interest, and did not vote or otherwise make a position known (these three simplified to an unknown disposition).

B. Experiments and Analysis

An experimental evaluation of the competing methods is conducted in the domain of CRM, striving to exemplify the adequacy and performance of RBF versus SVM versus proposed hybrid RBF-SVM for the task of response optimization in terms of accuracy based upon an numerical experiment. The voting dataset are taken to evaluate the proposed RBF-SVM direct marketing system. All experiments have been performed using Intel Core 2 Duo 2.26 GHz processor with 2 GB of RAM and weka software [25].

TABLE I. The Performance of base and hybrid Classifiers

Dataset	Classifiers	Classification Accuracy
Voting dataset	RBF	94.48 %
	SVM	96.09 %
	Proposed Hybrid RBF-SVM	99.31 %

The data set described in section 5 is being used to test the performance of base classifiers and hybrid classifier. Classification accuracy was evaluated using 10-fold cross validation. In the proposed approach, first the base classifiers RBF and SVM are constructed individually to obtain a very good generalization performance. Secondly, the ensemble of RBF and SVM is designed. In the ensemble approach, the final output is decided as follows: base classifier's output is given a weight (0–1 scale) depending on the generalization performance as given in Table 1.

The results of the computational experiments are presented in Table 1, comparing the performance of RBF, SVM and proposed hybrid RBF-SVM models on the generalization set. For the case of response optimization the accuracy is of primarily importance, as it measures the amount of correctly classified respondents. The accuracy of SVM was found to be higher than RBF classifier and the proposed hybrid RBF-SVM exhibits higher percentage accuracy than the individual classifiers. Thus the proposed hybrid RBF-SVM model can be regarded as very good for the application domain.

According to Table 1, the proposed hybrid model shows significantly larger improvement of classification accuracy than the base classifiers and the results are found to be

statistically significant. The x^2 statistic x2 is determined for all the above approaches and their critical value is found to be less than 0.455. Hence corresponding probability is $p < 0.5$. This is smaller than the conventionally accepted significance level of 0.05 or 5%. Thus examining a x^2 significance table, it is found that this value is significant with a degree of freedom of 1. In general, the result of x^2 statistic analysis shows that the proposed classifiers are significant at $p < 0.05$ than the existing classifiers.

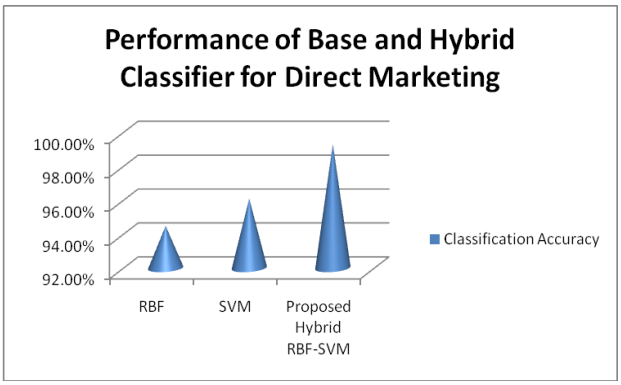

Fig. 1. Classification Accuracy

The experimental results show that proposed ensemble of RBF and SVM is superior to individual approaches for direct marketing problem in terms of Classification accuracy.

VI. Conclusion

In this research, some new techniques have been investigated for direct marketing and their performance is evaluated based on the Voting dataset. Recently, various architectures from computational intelligence and machine learning, such as artificial neural networks (ANN) and support vector machines (SVM) have found increasing consideration in practice, promising effective and efficient solutions for classification problems in real-world applications through robust generalization in linear and non-linear classification problems, deriving relationships directly from the presented sample data without prior modeling assumptions. Hence RBF and SVM are explored as direct marketing models. Next a hybrid RBF-SVM model is designed using RBF and SVM models as base classifiers. Thus, a hybrid intelligent direct marketing system is proposed to make optimum use of the best performances delivered by the individual base classifiers and the hybrid approach. The hybrid RBF-SVM shows higher percentage of classification accuracy than the base classifiers. This paper provides some insights on the relative performance of base and hybrid approaches to predictive modeling for churn based on classification accuracy for modeling defections.

The numerical results show, that RBF and SVM are both suitable for the task of response optimization, leading to classification accuracy that can be considered as very good for practical problems. This robustness makes SVM best suited for users who are less experienced in data mining and model building, which is not untypical in business environments. Consequently, the hybrid RBF-SVM is recommended in standard data mining software packages like WEKA as the proposed technique is easy to manage and provides competitive results. Finally, early detection and prevention of customer attrition can also enhance the total lifetime of the customer base, if efforts are focused on the retention of valuable customers. The future research will focus on investigate other ways to combine basic models in order to create more accurate models in response modeling and therefore, minimize marketing expenses and bring in more profits to the company.

Acknowledgment

Author gratefully acknowledges the authorities of Annamalai University for the facilities offered and encouragement to carry out this work.

References

[1] Aristides Gionis and Heikki Mannila and Panayiotis Tsaparas. (2005), Clustering Aggregation. ICDE.

[2] C. L. Bauer (1998). A direct mail customer purchase model, Journal of Direct Marketing, 2:16–24.

[3] Breiman. L, (1996), "Bias, Variance, and Arcing Classifiers", Technical Report 460, Department of Statistics, University of California, Berkeley, CA.

[4] Fader, P. S., B. G. S. Hardie, and K. L. Lee. (2004). "Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model," Working Paper, Wharton Marketing Department.

[5] Freund, Y. and Schapire, R. (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In proceedings of the Second European Conference on Computational Learning Theory, pp 23-37.

[6] Freund, Y. and Schapire, R. (1996) Experiments with a new boosting algorithm, In Proceedings of the Thirteenth International Conference on Machine Learning, 148-156 Bari, Italy.

[7] Gonul, F. F., Kim, B. D., & Shi, M. (2000). Mailing smarter to catalog customers. Journal of Interactive Marketing, 14(2), 2–16.

[8] M.Govindarajan, RM.Chandrasekaran, (2012), "Intrusion Detection using an Ensemble of Classification Methods", In Proceedings of International Conference on Machine Learning and Data Analysis, pages 459-464.

[9] Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart. (2004). "Valuing Customers," Journal of Marketing Research 41(1), 7–18.

[10] Hastie, T., R. Tibshirani, and J. Friedman. (2001). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer-Verlag.

[11] Haykin, S. (1999). Neural networks: a comprehensive foundation (second ed.). New Jersey: Prentice Hall.

[12] Jiawei Han , Micheline Kamber**,** (2003), " Data Mining – Concepts and Techniques" Elsevier Publications.

[13] Joon Hur, Jong Woo Kim, (2008), "A hybrid classification method using error pattern modeling", Expert Systems with Applications, 34, 231–241.

[14] Kubat, M., Holte, R., & Matwin, S. (1997). Learning when negative examples abound. Lecture Notes in Artificial Intelligence (LNAI 1224), 146–153, Prague, The Czech Republic

[15] Lemmens, Aur´elie and Christophe Croux. (2003). "Bagging and Boosting Classification Trees to Predict Churn", Working Paper, Teradata center.

[16] Ling, C. X., & Li, C. (1998). Data mining for direct marketing: Problems and solutions Proceedings of the KDD98 pp. 73–79.

[17] Maryam Daneshmandi, Marzieh Ahmadzadeh (2013), "A Hybrid Data Mining Model to Improve Customer Response Modeling in Direct Marketing, Indian Journal of Computer Science and Engineering, Vol. 3 No.6, 844-855.

[18] Michie, D., Spiegelhalter, D. J., & Taylor, C. (1994). Machine learning. Neural and statistical classification. Ellis Horwood.

[19] Oliver Buchtala, Manuel Klimek, and Bernhard Sick, Member, IEEE**,** (2005) **"**Evolutionary Optimization of Radial Basis Function Classifiers for Data Mining Applications", IEEE Transactions on systems, man, and cybernetics—part b: cybernetics, vol. 35, no. 5.

[20] Sara Madeira Joao M.Sousa (2000), "Comparison of target selection methods in direct Marketing" Technical University of Lisbon, Institution Superior Technician, Dept. Mechanical Eng./IDMEC, 1049-001 Lisbon, Portugal.

[21] Setnes, M., & Kaymak, U. (2001). Fuzzy modeling of client preference from large data sets: an application to target selection in direct marketing. IEEE Transactions on Fuzzy Systems, 9(1), 153–163.

[22] Shin, H. J., & Cho, S. (2006). Response modeling with support vector machines. Expert Systems with Applications, 30(4), 746–760.

[23] Tang, Z. (2011). "Improving Direct Marketing Profitability with Neural Networks." International Journal of Computer Applications 29(5): 13-18.

[24] Vapnik, V. (1998). Statistical learning theory, New York, John Wiley & Sons.

[25] Weka: Data Mining Software in java http://www.cs.waikato.ac.nz/ml/weka/

[26] L. Xu, A. Krzyzak, and C. Y. Suen, (1992), "Methods of Combining Multiple Classifiers and Their Applications to Handwritten Recognition", IEEE Transactions on Systems, Man, Cybernetics, Vol. 22, No. 3, pp. 418-435.

[27] Yu, E., & Cho, S. (2006). Constructing response model using ensemble based on feature subset selection. Expert Systems with Applications, 30(2), 352–360.

[28] Zahavi, J., & Levin, N. (1997). Issues and problems in applying neural computing to target marketing. Journal of Direct Marketing, 11(4), 63–75.

AUTHOR PROFILE

M.Govindarajan received the B.E and M.E and Ph.D Degree in Computer Science and Engineering from Annamalai University, Tamil Nadu, India in 2001 and 2005 and 2010 respectively. He did his post-doctoral research in the Department of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, United Kingdom in 2011. He is currently an Assistant Professor at the Department of Computer Science and Engineering, Annamalai University, Tamil Nadu, India. He has presented and published more than 70 papers in Conferences and Journals. His current research interests include Data Mining and its applications, Web Mining, Text Mining, and Sentiment Mining. He was the recipient of the Achievement Award for the field and to the Conference Bio-Engineering, Computer science, Knowledge Mining (2006), Prague, Czech Republic and All India Council for Technical Education "Career Award for Young Teachers (2006), New Delhi, India. He is active Member of various professional bodies.