Optimal Feature Selection for Cluster Based Ensemble Classifier using Meta Heuristic Function for Medical Disease Data Classification for Symptom Prediction

Rahul Deo Sah*

The diversity and applicability of data mining are increase day to day in the field of medical science for the predication of symptom of disease. The data mining provide lots of technique for mine data in several field, the technique of mining as association rule mining, clustering technique, classification technique and emerging technique such as called ensemble classification technique. The process of ensemble classifier increases the classification rate and improved the majority voting of classification technique for individual classification algorithm such as KNN, Decision tree and support vector machine. The new paradigms of ensemble classifier are cluster oriented ensemble technique for classification of data.

Keyword:- Data Mining Technique (Ensemble Classifier), Medical Diseases Symptoms, Meta HeuristicFunction.

Introduction

The diversity and applicability of data mining are increase day to day in the field of medical science for the predication of symptom of disease.This dissertation apply classification proceed based on classifier selection to medical disease data and propose a clustering-based classifier selection method. In the method, many clusters are selected for a ensemble process. Then, the standard presentation of each classifier on selected clusters is calculated and the classifier with the best average performance is chosen to classify the given data. In the computation of normal act, weighted average is technique is used. Weight values are calculated according to the distances between the given data and each selected cluster. There are generally two types of multiple classifiers combination: multiple classifiers selection and multiple classifiers fusion. Multiple classifiers selection assumes that each classifier has expertise in some local regions of the feature space and attempts to find which classifier has the highest local accuracy in the vicinity of an unknown test sample. Then, this classifier is nominated to make the final decision of the system.

Performance of a classifier is frequently the most important aspect of its value and is measured using a variety of well known method and matrix is used. On the other hand knowledge of a classifier is often treated as less important or even neglected. However it is vital for the users of the classifier as they belief it more if they can realize how the classifier works and because additional knowledge about the relations in observed data can be extracted by involved classifier. Consequently some of the old methods focus on knowledge of learned classifiers or transforming non-knowledge classifiers into human- knowledge structure. There is lack of algorithms that treat accuracy and knowledge of classifiers as uniformly significant, converting the domain of constructing a classifier intoheuristic optimization crisis. Such algorithms are especially important in domains where there are parts of attribute space that can be classified with high accuracy using knowledgeable classifier and parts that require non- knowledge classifiers to achieve required classification accuracy.

**Lecturer, Department of CA & IT, Ranchi College, Ranchi, Jharkhand*

Technique

The process of combining different clustering output (cluster ensemble or clustering Aggregation) emerged as an alternative approach for improving the quality of the Results of clustering algorithms. It is based on the success of the combination of supervised classifiers. Given a set of objects, a cluster ensemble method consists of two principal steps: Generation, which is about the creation of a set of partitions a of these objects, and Consensus Function, where a new partition, which is the integration of all partitions obtained in the generation step, is computed. Over the past years, many clustering ensemble techniques have been proposed, resulting in new ways to face the problem together with new fields of application for these techniques. Besides the presentation of the main methods, the introduction of taxonomy of the different tendencies and critical comparisons among the methods is really important in order to give a practical application to a survey. Thus, due to the importance that clustering ensembles have gained facing cluster analysis, we have made a critical study of the different approaches and the existing methods.

Feature selection technique is used for selecting subset of relevant features from the data set to build robust classification models. Classification accuracy is improved by removing most irrelevant and redundant features from the dataset. Ensemble model is proposed for improving classification accuracy by combining the prediction of multiple classifiers. In this dissertation used cluster based ensemble classifier. The performance of each classifier and ensemble model is evaluated by using statistical measures like accuracy, specificity and sensitivity.

Classification of medical data is an important task in the prediction of any disease. It even helps doctors in their diagnosis decisions. Cluster oriented Ensemble classifier is to generate a set of classifiers instead of one classifier for the classification of a new object, hoping that the combination of answers of multiple classification results in better performance.

Literature Review

We demonstrate the algorithmic use of the classification technique by extending SVM the most popular binary classification algorithms. From the studies above, the key to improve cluster oriented classifier is to improve binary classification. In the final part of the thesis, we include empirical evaluation that aim at understanding binary classification better in the context of ensemble learning.

Cluster and classification play an important role in data mining and machine learning paradigm. The evaluation of ensemble classifier is great advantage over binary and conventional classifier. The process of prototype classification is combined two or more method with same nature. The prototype classification of data brings come in form of cluster oriented ensemble classifier. The need and requirement of online transaction of data is stream classification, due to stream classification save time of computation and storage area of network. For the purpose of stream data classification various machine learning algorithm are used, such as clustering, classification, and neural network. In the classification process of a growing data stream, also the temporary or long-standing activities of the stream may be more significant, or it often cannot be known a priori as to which one is more important. We decide the window or horizon of the training data to use so as to obtain the best classification accuracy. For the proper selection of window and horizon used ensemble classifier with support of clustering technique. In ensemble methods, the main strategy is to maintain a dynamic set of classifiers. When a decrease in performance is practical, new classifier are fused into the ensemble while aged and horrific the stage fusion process are disinterested. For classification, decisions of fusion in the ensemble are combined, usually with a cluster scheme [10]. The advantage of cluster ensembles over single classifiers in the data stream classification problem has been proved empirically and theoretically [1, 3]. However, few ensemble methods have been designed to take into consideration the problem of recurring contexts [6, 7]. Specifically, in problems where concepts re-occur, models of the ensemble should be maintained in memory even if they do not perform well in the latest batch of data. Moreover, every classifier should be specialized in a unique concept, meaning that it should be trained from data belonging to this concept and used for classifying similar data. In [9, 12], a methodology that identifies concepts by grouping classifiers of similar performance on specific time intervals is described. Clusters are then assigned to classifiers according to performance on the latest batch of data. Predictions are made by using weighted averaging. Although this strategy fits very well with the recurring contexts problem, it has an offline step for the discovery of concepts that is not suitable for data streams. In particular, this framework will probably be inaccurate with concepts that did not appear in the training set. To classified real-world data set with overlapping features from different classes. The training of class borders between overlap class features in such cases is a hard crisis. Extreme preparation of the base classifiers will lead to accurate training of the decision border but resulting in over fitting thus miss-classifying instances of test data. On the other hand, learning generalized boundaries will avoid over fitting but at the cost of always miss-classifying some overlapping features. This problem on learning the class boundaries of overlapping features remains inherent in all the base classifiers and is propagated to the decision fusion stage as well even though the base classifier errors are uncorrelated. We come bring in clustering at this point. Clustering is the process of partitioning a data set into multiple groups where each group contains data points that are very close in Euclidean space. The clusters have well defined and easy to learn boundaries. Let's assume that the features are labeled with their cluster number. Now if the base classifiers are trained on the modified data set they will learn the cluster boundaries. As the clusters have well defined easy to learn boundaries the base classifiers can learn them with high accuracy. Clusters can contain overlapping features from multiple classes. The cluster oriented ensemble classifier performs a great role in classification technique, but the selection process of cluster is not well defined. Now these overcome are removed by ant colony optimization. Ant colony optimization well knows multi-objective function used in feature optimization. This chapter gives an extensive literature survey on the existing ensemble classification technique of data mining classification using genetic algorithm and other process of optimization. We study various research paper and journal and know about optimization technique of association rule mining. All methodology and process are not described here. But some related work in the field of ensemble classification by the name of authors and their respective title.

Brijesh Verma and Ashfaqur Rahman entitled "Cluster-Oriented Ensemble Classifier: Impact of Multi-cluster Characterization on Ensemble Classifier Learning" a novel cluster-oriented ensemble classifier is presented [1]. This cluster oriented ensemble classifier is based on original concepts where cluster boundaries are learned by the base classifier and cluster confidences are mapped with the help of fusion classifier to the class decision. According to this paper an ensemble classifier is constructed using a set of base classifier which learns the class boundaries separately over the pattern. It is a difficult problem to learn class boundaries between overlapping class because this problem remains inherent in all the base classifier. So concept of clustering comes into existence. Clustering is the process of separating an item set into multiple item sets group.

Anne-Laure Bianne-Bernard, Fares Menasri, Rami Al-Hajj Mohamad, ChaficMokbel, Christopher Kermorvant, and Laurence Likforman-Sulem a study of building an efficient word recognition system resulting from the combination of three handwriting recognizers as building an efficient word recognition system resulting from the combination of three handwriting recognizers [11]. The main component of this combined system is an HMM-based recognizer which considers dynamic and contextual information for a better modeling of writing units. Among popular applications of handwriting recognition are bank check processing, mailed envelopes reading, and handwritten text recognition in documents and videos, for which different systems have been successfully developed.

NayerM.Wanas, Rozita A. Dara and Mohamed S. Kamel entitled an investigation of Adaptive fusion and co-operative training for classifier ensembles [2] as ensembles are designed in such a way that each classifier is trained independently and the decision In pattern classification, multiple classifier systems are often considered a practical and effective solution for difficult recognition problems fusion is performed as a post-process module. In some cases, the empirical observations of the performance of specialized classifiers justify the use of multiple classifiers. In other cases, the adoption of multiple classifiers stems from the problem decomposition such as the need to employ a variety of sensor types, or the need to avoid making commitments to arbitrary initial conditions and parameters. There are many ways to utilize more than one classifier in a recognition problem. A divide-and-conquer approach isolates the types of input from which a specific classifier performs well, and directs that input accordingly. While in sequential approach, one classifier is applied first, and invokes others only if it fails to yield a decision with sufficient confidence. The objective of this research is to develop an architecture that makes decision fusion a more adaptive process, by introducing learning to the aggregation phase. In this paper, we empirically investigated various aggregation architectures and algorithms for the multiple classifiers. We also proposed and developed a new architecture. In this architecture, a set of classifiers, called detectors, was applied to make the aggregation scheme a more adaptive process. These classifiers were responsible for using distinctive features to help the aggregation module.

Leo Breiman entitled a Bagging predictor [3] as a method for generating multiple version of a predictor and using these to get an aggregated predictor. The aggregation averages over the version when predicting a numerical outcome and does a plurality vote when predicting a class. The multiple versions are formed by making bootstrap replication of the learning set and using these as new learning sets. Tests on real and simulated data sets using classification and regression trees and subset selection in linear regression show that bagging can give substantial gains in accuracy.

Yoshua Bengio entitled Learning Deep Architectures for AI [4] as Theoretical results suggest that in order to learn the kind of complicated functions that can represent high-level abstractions (e.g., in vision, language, and other AI-level tasks), one may need deep architectures. Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. Automatically learning features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from data, without depending completely on human-crafted features. Since a deep architecture can be seen as the composition of a series of processing stages, the immediate question that deep architectures raise is: what kind of representation of the data should be found as the output of each stage (i.e., the input of another)? What kind of interface should there be between these stages? This monograph started with a number of motivations: first to use learning to approach AI, then on the intuitive plausibility of decomposing a problem into multiple levels of computation and representation, followed by theoretical results showing that a computational architecture that does not have enough of these levels can require a huge number of computational elements, and the observation that a learning algorithm that relies only on local generalization is unlikely to generalize well when trying to learn highly varying functions.

Giorgio Fumera, Fabio Roli and Alessandra Serrau entitled Analysis of Bagging [5] as a Linear Combination of Classifiers as applying an analytical framework for the analysis of linearly combined classifiers to ensembles generated by bagging. This provides an analytical model of bagging misclassification probability as a function of the ensemble size, which is a novel result in the literature. Experimental results on real data sets confirm the theoretical predictions. This allows us to derive a novel and theoretically grounded guideline for choosing bagging ensemble size. Several methods for the construction of classifier ensembles, like bagging, the random subspace method, tree randomization and random forests, are based on introducing some kind of randomness into the design process of individual classifiers. Bagging is perhaps the most popular method, and its effectiveness has been empirically shown in many real pattern recognition problems. Author applied an analytical framework for linear combiners developed in, and to the particular case of linearly combined classifiers generated by bagging. This provided two main contributions. First, an analytical model of bagging expected added error as a function of the ensemble size. Second, based on such model, a practical guideline on the choice of bagging ensemble size which is an advance with respect to empirical guidelines proposed in the literature. We also showed that our theoretical results support the optimality of the simple average combining rule for classifier ensembles generated by bagging.

Albert Hung-RenKo and Robert Sabourin entitled an Implication of Data Diversity for a Classifier-free Ensemble Selection in Random Subspaces [6] as Ensemble of Classifiers (EOC) has been shown effective in improving the performance of single classifiers by combining their outputs. The goal of pattern recognition systems is to achieve the best possible classification performance. The two key issues that are crucial to the success of an EOC routine are the following: first, we need diversity for ensemble creation, because an EOC will not perform well without it; and second, we need to select classifiers once they have been created, because not all the classifiers created are useful. For this First, we need to evaluate the hypothesis that the clustering diversity of different feature subsets can be used as an objective function for ensemble selection in Random Subspaces. Even though the clustering diversities might only be able to represent data diversities in Random Subspaces, for Bagging, which only use a part of the samples, there is still no adequate measure for their data diversities. It will be of great interest to figure out how to measure the data diversities in Bagging. Finally, we have to mention that, due to its special ensemble generating mechanism, the scheme is not likely to be applicable in Boosting.

Zhihui Lai, Zhong Jin, Jian Yang and W.K Wong entitled major disadvantages of the linear dimensionality reduction algorithms [7] as Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA), are that the projections are linear combination of all the original features or variables and all weights in the linear combination known as loadings are typically non-zero. Low-dimensional representation of high- dimensional data is an important problem in many application fields. There are many classical approaches for dimensionality reduction such as principle component analysis (PCA, linear discriminant analysis (LDA). Recently, locality based method was also proposed for feature extraction. However, one of the major disadvantages of these linear methods mentioned above is that the learned projective axes are linear combinations of all the original features or variables, hence they are often hard to give a reasonable interpretation on which features or variables play an important role in real world applications. Author develops a supervised learning technique called sparse local discriminant projections (SLDP) for linear dimensionality reduction of high-dimensional data. By characterizing the local between-class severability and within-class neighborhood geometry relationship, SLDP aims at maximizing the between-class severability and simultaneously preserving within-class geometry with sparsity constraint.

Juan J. Rodrıguez and Jesu s Maudes entitled a method for the construction of classifier ensembles called boosting [8] as Boosting is a set of methods for the construction of classifier ensembles. The differential feature of these methods is that they allow obtaining a strong classifier from the combination of weak classifiers. Therefore, it is possible to use boosting methods with very simple base classifiers. One of the simplest classifiers is decision stumps, decision trees with only one decision node. This work proposes a variant of the most well-known boosting method, AdaBoost. It is based on considering, as the base classifiers for boosting, not only the last weak classifier, but a classifier formed by the last r selected weak classifiers (r is a parameter of the method). If the weak classifiers are decision stumps, the combination of r weak classifiers is a decision tree. Given one or more classification methods, one of the most natural ways of obtaining more accurate classifiers is the use of ensembles. There are methods that combine classifiers obtained with different methods. There are some ensemble methods that have been designed specifically for combining classifiers obtained with methods from a certain kind, normally decision trees one of the most successful ensembles methods is Boosting. There are several variants; AdaBoost is the most well-known. These methods assign a weight to each example. Initially, all the examples have the same weight. This paper presents an approach for improving the results obtained with boosting and decision stumps. The idea is to combine several decision stumps in a not so weak tree. In order to improve the accuracy of a classifier obtained with AdaBoost and a learning method, there are two direct approaches: To add more weak classifiers and to use more complex base classifiers. Although it is possible to use boosting with strong base classifiers (e.g., decision trees, neural networks, etc.) the name of the method comes from its ability to obtain strong classifiers from weak methods. A weak method commonly used with boosting are decision stumps, decision trees with only one decision node and two leaves.

Oriol Pujol and David Masip entitled ensembles Toward a Structural Characterization of the Classification Boundary [9] as a novel binary discriminative learning technique based on the approximation of the nonlinear decision boundary by a piecewise linear smooth additive model. The decision border is geometrically defined by means of the characterizing boundary points that belong to the optimal boundary under a certain notion of robustness. The intuitive geometric reasoning behind the well-known support vector machines technique comes from the maximization of the margin, defined as the minimum distance from the closest data points to the boundary. Although it is simple to understand this concept when the optimal separation is a hyper plane, it becomes much more complex in front of nonlinear boundaries. The most well-known strategy to deal with this issue is the kernel approach that represents a change in the metric space when computing the margin. An approach to combine the outputs of multiple classifiers to support the decision-making process in classification tasks as ensemble learning methods have attracted growing attention from both academia and industry recently, it is critical to understand the fundamental issues of the combining rule. Motivated by the signal strength concept, our proposed SSC algorithm can effectively integrate the individual vote from different classifiers in an ensemble learning system. Classifier combination has been an important research topic in ensemble learning. In fact, no matter what kinds of mechanisms are used to obtain the multiple classifiers, a classifier combination method is needed to combine all the individual votes for the final decision. Classifier combination has been an important research topic in ensemble learning. In fact, no matter what kinds of mechanisms are used to obtain the multiple classifiers, a classifier combination method is needed to combine all the individual votes for the final decision. Theoretical analysis followed by a voting algorithm, namely, SSC, was presented in detail. Simulation results of this method compared with those of nine major voting strategies were used to show the effectiveness of this method.

Nandita Tripathi, Stefan Wermter, Chihli Hung, and Michael Oakes entitled a technique using the maximum significance value to detect a semantic subspace [10] as Subspace detection and processing is receiving more attention nowadays as a method to speed up search and reduce processing overload. Subspace Learning algorithms try to detect low dimensional subspaces in the data which minimize the intra-class separation while maximizing the inter-class separation. Subspace learning methods are therefore nowadays being increasingly researched and applied to web document classification, image recognition and data clustering. This work is an effort to explore semantic subspace learning with the overall objective of improving document retrieval in a vast document space.

Terry Windeatt entitled a describe a measure of MLP Classifier Design [20] as it iscapable of predicting the number of classifier training epochs for achieving optimal performance in an ensemble of MLP classifiers. The measure is computed between pairs of patterns on the training data, and is based on a spectral representation of a Boolean function. This representation characterizes the mapping from classifier decisions to target label, and allows accuracy and diversity to be incorporated within a single measure Multi-layer perceptrons (MLP) make powerful classifiers that may provide superior performance compared with other classifiers, but are often criticized for the number of free parameters. Ensemble classifiers, also called committees or Multiple Classifier Systems (MCS) offer a way of solving some of these problems. The idea of combining multiple classifiers is based on the observation that achieving optimal performance in combination is not necessarily consistent with obtaining the best performance for an individual (base) classifier. It is shown experimentally that, over a range of k-class datasets, k^3 2, a pair-wise measure computed over training patterns is well correlated with base classifier test error when number of training epochs of MLP base classifiers are systematically varied. Bootstrapping significantly improves the estimate of this measure, while making little difference to the ensemble test error. It is also demonstrated that correlation of the spectral measure with ensemble test error is not as strong. These can be thought of as two separate problems, the first being concerned with the prediction of over-fitting of the base classifier and which is the main focus of this paper. The second problem is to determine the relationship between ensemble and base classifier test error.

Dacheng, Tang, Xiaoou, Li, Xuelong, Wu and Xindong entitled "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval" a new asymmetric bagging and random subspace mechanism is designed [23]. Relevance feedback schemes based on support vector machines (SVM) have been widely used in content-based image retrieval (CBIR). However, the performance of SVM-based relevance feedback is often poor when the number of labeled positive feedback samples is small. This is mainly due to three reasons: 1) an SVM classifier is unstable on a small-sized training set; 2) SVM's optimal hyper plane may be biased when the positive feedback samples are much less than the negative feedback samples, and 3) over fitting happens because the number of feature dimensions is much higher than the size of the training set. The proposed method addressed all these three problems. In a relevance feedback process, the user first labels a number of relevant retrieval results as positive feedback samples and some irrelevant retrieval results as negative feedback samples. Then, a CBIR system refines all retrieval results based on these feedback samples. These two steps are carried out iteratively to improve the performance of the image retrieval system by gradually learning the user's preferences. Many relevance feedback methods like discriminant learning, heuristic method the density estimation method have been developed in recent years. They either adjust the weights of various features to adapt to the user's preferences or estimate the density of the positive feedback examples.

Rich Caruana, AlexandruNiculescu-Mizil, Geoff Crew and Alex Ksikes entitled "Ensemble Selection from Libraries of Models" a method for constructing ensembles from libraries of thousands of models is presented [24]. Using distinct learning algorithms and parameter settings, model libraries are generated. To maximize the performance of the ensemble models a forward stepwise selection is added. An ensemble is a collection of models whose predictions are combined by weighted averaging or voting. According to Dietterich "A necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is if the classifiers are accurate and diverse." The simple forward model selection procedure is fast and effective, but sometimes over fits to the hill climbing set, reducing ensemble performance. To reduce the over fitting selection with replacement, stored ensemble initialization and bagged ensemble selection methods are added.

Sandrine Dudoit and Jane Fridlyand entitled "Bagging to improve the accuracy of a clustering procedure" an application of bagging to cluster analysis is proposed [47]. Bagging can substantially improve clustering accuracy and yields information on the accuracy of cluster assignments for individual observations. In addition, bagged clustering procedures are more robust to the variable selection scheme, i.e. their ac-curacy is less sensitive to the number and type of variables used in the clustering. Improving and assessing the accuracy of a given clustering procedure using a resampling method is known as bagging. In supervised learning bagging is used to generate and aggregate multiple clustering's. In this paper two new sampling methods BagClust1 and BagClust2 are proposed to improve and assess the accuracy of a given clustering procedure. In BagClust1 the clustering procedure is repeatedly applied to each bootstrap sample and the final partition is obtained by plurality voting. The BagClust2 method forms a new dissimilarity matrix by recording for each pair of observations the proportion of time they were clustered together in the bootstrap clusters.

Nikunj C. Oza and KaganTumer entitled “Classifier Ensembles: Select Real-World Applications” classifier ensembles and ensemble applications are presented [26]. Ensuring that the particular classification algorithm matches the properties of the data is crucial in providing results that meet the needs of the particular application domain. One way in which the impact of this algorithm/application match can be alleviated is by using ensembles of classifiers, where a variety of classifiers are pooled before a final classification decision is made. Classifier ensembles provide an extra degree of freedom in the classical bias/variance tradeoff, allowing solutions that would be difficult to reach with only a single classifier. Many learning algorithms generate a single classifier that can be used to make predictions for new examples. The ways in which multiple classifiers are combined are simple averaging, weighed averaging, stacking, bagging and boosting.

Robert E. Banfield, Lawrence O. Hall, Kevin W. Bowyer and W.P. Kegelmeyer entitled “A Comparison of Decision Tree Ensemble Creation Techniques” Randomization-Based technique for creating an ensemble of classifiers is proposed [27]. BAGGING is one of the older, simpler, and better known techniques for creating an ensemble of classifiers. Bagging creates an ensemble of classifiers by sampling with replacement from the set of training data to create new training sets called “bags”. A number of other randomization-based ensemble techniques boosting, random subspaces, random forests, and randomized C4.5 have been introduced. In bagging, only a subset of examples typically appears in the bag which will be used in training the classifier. Out-of-bag error provides an estimate of the true error by testing on those examples which did not appear in the training set. Authors have developed an algorithm which appears to provide a reasonable solution to the problem of deciding when enough classifiers have been created for an ensemble. It works by first smoothing the out-of-bag error graph with a sliding window in order to reduce the variance.

Inferences

The data mining play an important role in the field of medical science for the diagnosis of critical disease such as cancer, brain tumor, liver damage prediction and diabetic condition prediction. Automated prediction of critical disease symptoms is big issue, now a day’s various researchers and medical scientists used the application of data mining for the prediction of symptoms. For the prediction of symptoms various data mining tools are available such as clustering, classification, rule mining and regression. The lacking of feature selection process these tools not predicts accurate result. So increase the prediction rate of classification and clustering used feature optimization and feature selection method for the medical disease symptoms prediction.

Optimal cluster selection is important part in cluster oriented classifier. COEC on classification of data set is efficient process, but this method generates some better result in compression of clustering method on the consideration of loss of data. When the scale of data set is increase the complexity of classification is also increase, it is difficult to select optimal number of cluster for individual classifier of data set. We introduce a new feature sub set selection method for finding similarity matrix for clustering without alteration of cluster oriented classifier. The proposed features sub set selection method based on ant colony optimization, ant colony optimization is very famous meta-heuristic function for searching for finding similarity of data. In this method we introduced continuity of ants for similar features and dissimilar features collect into next node. In that process ACO find optimal selection of features sub set. Suppose ants find features of similarity in continuous root. Every ant of features compares their property value according to initial features set.

Conclusion

During the tenure of the research work following work done.

1. Design a cluster oriented classification model for dynamic selection of feature for the prediction of disease symptoms
2. Derived a meta-heuristic function for the feature optimization of medical disease.
3. Increase the classification and prediction ratio of classification technique of medical data.
4. Reduces the complexity of method for fast execution of process.
5. Explode the large amount of data for the further researcher in data mining for new comer scholar student

Future Proposed Work Focus on medical Disease symptoms Prediction

This proposed work focus on medical disease symptoms predication based on cluster oriented ensemble classifier. The proposed work tested on various medical disease dataset such as cancer dataset, brain tumor dataset, diabetic dataset and much more critical disease.

1. Improved the prediction ratio.
2. Minimized the value of false prediction.
3. Minimized the time complexity of algorithm.
4. Improve the performance of classifier
5. Provide automated algorithm for medical disease prediction.

Reference

1. *BrijeshVerma and AshfaqurRahman "Cluster-Oriented Ensemble Classifier: Impact of Multi cluster Characterization on Ensemble Classifier Learning" in IEEE Transactions on knowledge and data engineering, 2012.*
2. *NayerM.Wanas, Rozita A. Dara and Mohamed S. Kamel "Adaptive fusion and co-operative training for classifier ensembles" in Pattern Analysis and Machine Intelligence Lab, University of Waterloo, 2006.*
3. *LEO BBEIMAN "Bagging Predictors" in Kluwer Academic Publishers, 1996.*
4. *YoshuaBengio "Learning Deep Architectures for AI" in Foundations and Trends in Machine Learning, 2009.*
5. *Giorgio Fumera, Fabio Roli and Alessandra Serrau "A Theoretical Analysis of Bagging as a Linear Combination of Classifiers" in IEEE Transactions.*
6. *Albert Hung-RenKo and Robert Sabourin "The Implication of Data Diversity for a Classifier-free Ensemble Selection in Random Subspaces" in IEEE Transactions.*
7. *Zhihui Lai, Zhong Jin, Jian Yang and W.K Wong "Sparse Local Discriminant Projections for Face Feature Extractio" International Conference on Pattern Recognition, 2010.*
8. *Juan J. Rodrıguez and Jesus Maudes "Boosting recombined weak classifiers" in ScienceDirect, 2007.*
9. *OriolPujol and David Masip "Geometry-Based Ensembles: Toward a Structural Characterization of the Classification Boundary" in IEEE Transactions, 2009.*
10. *NanditaTripathi, Stefan Wermter, Chihli Hung and Michael Oakes "Semantic Subspace Learning with Conditional Significance Vectors" in IEEE Transactions.*
11. *Anne-Laure Bianne-Bernard, Fare`s Menasri, Rami Al-Hajj Mohamad, ChaficMokbel, Christopher Kermorvant and Laurence Likforman-Sulem "Dynamic and Contextual Information in HMM Modeling for Handwritten Word Recognition" in IEEE transactions on pattern analysis and machine intelligence, 2011.*
12. *NayerM.Wanas, Rozita A. Dara and Mohamed S. Kamel "Adaptive fusion and co-operative training for classifier ensembles" in Pattern Analysis and Machine Intelligence Lab, University of Waterloo, 2006.*
13. *P. Kraipeerapun and C. C. Fung, "Binary classification using ensemble neural networks and interval neutrosophic sets," Neurocomput., vol. 72, pp. 2845-2856, 2009.*
14. *YoshuaBengio "Learning Deep Architectures for AI" in Foundations and Trends in Machine Learning, 2009.*
15. *Giorgio Fumera, Fabio Roli and Alessandra Serrau "A Theoretical Analysis of Bagging as a Linear Combination of Classifiers" in IEEE Transactions.*
16. *Albert Hung-RenKo and Robert Sabourin "The Implication of Data Diversity for a Classifier-free Ensemble Selection in Random Subspaces" in IEEE Transactions.*
17. *Zhihui Lai, Zhong Jin, Jian Yang and W.K Wong "Sparse Local Discriminant Projections for Face Feature Extractio" International Conference on Pattern Recognition, 2010.*
18. *Juan J. Rodrı´guez and Jesu ´s Maudes "Boosting recombined weak classifiers" in ScienceDirect, 2007.*
19. *Haibo He and Yuan Cao "SSC: A Classifier Combination Method Based on Signal Strength" in IEEE Transactions on neural networks and learning systems, 2012.*
20. *Terry Windeatt "Accuracy/Diversity and Ensemble MLP Classifier Design" in IEEE Transactions.*
21. *Xueyi Wang "A New Model for Measuring the Accuracies of" in IEEE World Congress on Computational Intelligence, 2012.*
22. *Gavin Brown, and Ludmila I. Kuncheva ""Good" and "Bad" Diversity in Majority Vote Ensembles" in IEEE Transaction.*
23. *Tao, Dacheng, Tang, Xiaoou, Li, Xuelong, Wu and Xindong "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval" in IEEE Transactions, 2006.*

24. *Rich Caruana, AlexandruNiculescu-Mizil, Geoff Crew and Alex Ksikes "Ensemble Selection from Libraries of Models" 21st International Confer-ence on Machine Learning, 2004.*

25. *Valerio Grossi, Alessandro Sperduti "Kernel-Based Selective Ensemble Learning for Streams of Trees" in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence 2010.*

26. *Nikunj C. Oza and KaganTumer "Classifier Ensembles: Select Real-World Applications" in Elsevier, 2007.*

27. *Robert E. Banfield, Lawrence O. Hall, Kevin W. Bowyer and W.P. Kegelmeyer "A Comparison of Decision Tree Ensemble Creation Techniques" in IEEE TRANSACTIONS, 2007.*

28. *Leo Breiman "Bagging Predictors" in Kluwer Academic Publishers, 2006.*

29. *Thomas G. Dietterich "Ensemble Methods in Machine Learning" in IEEE TRANSACTIONS.*

30. *S. B. Kotsiantis "Supervised Machine Learning: A Review of Classification Techniques" in Informatica 30, 2007.*

31. *Thomas G. Dietterich "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization" in Kluwer Academic Publishers, 1999.*

32. *Guoqiang Peter Zhang entitled "Neural Networks for Classification: A Survey" in IEEE TRANSACTIONS, 2000.*

33. *Xue-wen Chen1, Byron Gerlach, and David Casasent "Pruning Support Vectors for Imbalanced Data Classification"in Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4, 2005*

34. *XiujuFu , Lip0 Wang , KokSeng Chua and Feng Chu "Training RBF neural networks on unbalanced data" in Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'OZ) , Vol. 2,2006*

35. *L. Bruzzone and S.B. Serpico "Classification of imbalanced remote-sensing data by neural networks" in Elsevier Science B.V. PII S0167-8655,2009*

36. *Yetian Chen" Learning Classifiers from Imbalanced, Only Positive and Unlabeled Data Sets" in ijcse-2010*

37. *Suzan Koknar-Tezel and Longin Jan Latecki "Improving SVM Classification on Imbalanced Data Sets in Distance Spaces" Ninth IEEE International Conference on Data Mining,2009*

38. *Haibo He, Member, IEEE, and Edwardo A. Garcia "Learning from Imbalanced Data" in IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 9, SEPTEMBER 2009."*

39. *Hong Guo and Yi L. Murphey "Neural Learning From Unbalanced Data Using Noise Modeling" in IJCNN , pp.309-314,June 2009.*

40. *Shuiping Gou, Member, IEEE, Hui Yang, Licheng Jiao , Senior Member, IEEE and XiongZhuang "Algorithm of Partition based Network Boosting for Imbalanced Data Classification" in IEEE 2010.*

41. *Xue-wen Chen and Michael Wasikowski "FAST: A ROC-based Feature Selection Metric for Small Samples and Imbalanced Data Classification Problems" in ijser conference 2010.*

42. *VaishaliGanganwar "An overview of classification algorithms for imbalanced datasets" in International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 4, April 2012.*

43. *S, eydaErtekin, Jian Huang, Leon Bottou and C. Lee Giles "Learning on the Border:*

44. *Active Learning in Imbalanced Data Classification" in Journal of Machine Learning Research 6 1579–1619,2005.*

45. *Yuchun Tang, Yan-Qing Zhang, Nitesh V. Chawla, and Sven Krasser "svmsModeling for Highly Imbalanced Classification" journal of latex class files, vol. 1, no. 11, november 2002.*

46. *Salvador Garcia, Jose Ramon Cano, Alberto Fernandez and Francisco Herrera "Prototype Selection for Class Imbalance Problems" in Eighth International Symposium on Natural Language Processing, 2009.*

47. *Sandrine Dudoit and Jane Fridlyand "Bagging to improve the accuracy of a clustering procedure" in IEEE Transactions, 2002.*