Improved the Performance of Cluster Oriented Classifier using Ant Colony Optimization (AECC)

Shivam Mishra[1,] Anurag Jain[2]

shivammishrajhansi@gmail.com[1], anurag.akjain@gmail.com[2]

Department of Computer Science & Engineering[1, 2]

RITS, Bhopal, India

Abstract

The generation and selection of optimal cluster for cluster based ensemble classifier is important parameter. The selection of optimal cluster impact the performance of ensemble classifier, in conventional cluster based ensemble classifier used cluster diversity such as hard clustering with agglomerative cluster and some other cluster technique. The diversity of clustering technique used fixed number of cluster selection due to fixed number of generated cluster. For the improvement of optimal cluster selection used ANT colony optimizations technique for generation of multiple cluster index and cluster confidence value. The multiple cluster confidence value gives the better selection of optimal cluster selection process. In this paper proposed ant selection based cluster ensemble classifier for data classification. Empirical evaluation shows better result in compression of COEC.

Key words: - cluster, ensemble and ANT

INTRODUCTION

The increasing rate multiple diversity of pattern classification and data mining need some process of prototype classification, such as cluster ensemble classification and fusion of classifier. The ensemble and fusion provide great advantage over conventional technique of clustering and classification. The cluster oriented classifier is result of multi-pattern classification technique. The diversity of cluster generation techniques gives multiple cluster ensemble process as variable cluster and improved the performance of clustering technique. 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 merging of clusters, models of the ensemble should be maintained in process even if they do not perform well in the latest real time of data. Furthermore, every classifier should be dedicated in a distinctive concept, meaning that it should be trained from data belonging to this concept and used for classifying similar data. Clustering means the act of partitioning an unlabeled dataset into groups of similar objects. The goal of clustering is to group sets of objects into classes such that similar objects are placed in the same cluster while dissimilar objects are in separate clusters. Clustering is used as a data processing technique in many different areas, including artificial intelligence, bioinformatics, biology, computer vision, city planning, data mining, data compression, earth quake studies, image study, image portion, query retrieval, machine learning, marketing, medicine, object recognition, pattern recognition, spatial database analysis, statistics and web mining. ACO approach was proposed by Marco Dorigo to solve several discrete optimization problems. ACO deals with artificial systems that are inspired from the foraging behavior of biological ants, which are gives, the contribution of optimization of discrete problem [26]. The main idea is the indirect communication between the ants by means of liquid of ant trials, which enables them to find short paths between their nest and food. Studies of ant colonies have contributed in abundance to the set of smart algorithms. The process of pheromone drop by ants in their search for the shortest paths to food sources resulted in the development of shortest path optimization algorithms. Ant base ensemble cluster classifier (AECC) is a novel algorithm proposed in this paper through optimization framework to find optimal cluster selection subset. AECC optimizes the cluster selection objective function in the solution space of the cluster selection algorithms which makes AECC feasible to analyze large scale data. One of the interesting properties of AECC is that some state-of-the-art data classification algorithm algorithms have been proven to be special cases of AECC and it is applicable to all machines learning technique. AECC basically computes confidence of original cluster and selects cluster with the highest confidence, that is, according to AECC; they are the optimal cluster set for data analysis. AECC has been proven experimentally to be more effective than other previous techniques. The remainder of this paper is structured as follows. In section 2, related concept of ensemble classifier is described. In section 3, ANT ensemble cluster classifier.

Section 4 experimental result and cluster analysis. Section 5 Finally, conclusion and future research directions.

II.RELATED WORK

In this section describe method for ensemble classifier for data classification using clustering technique and other method for ensemble classifier. The method of cluster ensemble classifier and fusion of ensemble classifier reduces the bottleneck problem of individual classifier. Clustering and other data grouping technique provide flexibility for classification fusion in different domain of data.

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. Clustering is the method of separating an item set into multiple item sets group. Clustering assumed that if the patterns are labeled with their cluster number and the base classifiers are trained on the modified data set then base classifier will learn the cluster boundaries [1]. To gain improved and better accuracy of the ensemble classifier clusters are classified into multiple clusters and cluster decisions produced by the base classifier are combined into class decision by a fusion classifier.

Ensembles are designed in such a way that each classifier is trained independently and the decision in pattern classification, multiple classifier systems are often use a practical and effective solution for difficult recognition problems fusion is performed as a post-process module. In some cases, the experimental observations of the performance of specialized classifiers justify the use of multiple classifiers [2]. In other cases, the implementation 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 methods to use more than one classifier in a recognition problem.

A method for generating multiple version of a predictor and using these to get an aggregated predictor [3].The aggregation averages over the description when predicting a numerical outcome and does a plurality vote when predicting a class. Number of constraints is formed by making bootstrap replication of the learning set and using these as new learning groups. Tests on experimental datasets using classification and regression technique and feature selection in linear regression show that bagging gives better result predication. Bagging is one of the oldest, simpler, and better known methodology for creating an ensemble of classifiers. A number of other randomization-based ensemble techniques have been introduced. Some of the include boosting random sub spaces, random forests [20].

Analysis of Bagging as a Linear Combination of Classifiers as applying an analytical framework for the analysis of linearly combined classifiers to ensembles generated by bagging [5]. This provides an analytical model of bagging misclassification probability as a function of the ensemble size, which is a novel outcome. This permits us to derive a novel and theoretically grounded guideline for choosing bagging ensemble size. The technique of ensemble classifier are bagging, boosting and random forest tree, are based on introducing some kind of randomness into the design process of single classification technique. The ensemble process of linear technique is going on boosting process. Author applied an analytical framework for linear combiners created in specific case of linearly combined classifiers generated by bagging. Several methods for the construction of classifier ensembles, like bagging, random subspace technique, tree randomization and random forests technique, these methods are based on introducing some kind of randomness into the design process of individual classifiers. Bagging is perhaps the most admired method, and its efficiency has been empirically shown in number of real pattern recognition problems. Author applied a systematic framework for linear combiners created in specific case of linearly combined classifiers generated by bagging [16,24].

Ensemble of Classifiers (EoC) has been shown effective in improving the performance of single classifiers by combining their outputs [6]. Even though the clustering diversities might only be able to represent data diversities in random Subspaces, for Bagging method, which only use a part of the samples, there is still no adequate measure for their data diversities. It will be big interest to *figure* out how to calculate the data diversities in Bagging. Finally, we have to point out that, due to its special ensemble generating methods, which are not likely to be related in Boosting.

A bagging can push a good but unstable procedure a major step towards optimality. On the other hand, it can slightly degrade the performance of established steps. There has been latest work in the literature with some of the flavor of bagging [4].

The goal of ensemble learning methods is to construct a collection (an ensemble) of individual classifiers that are diverse and yet accurate. the highly accurate classification decisions can be obtained by voting the decisions of the individual classifiers in the ensemble.. Two of the most popular techniques for constructing ensembles are bootstrap aggregation and the Adaboost family of algorithms. Both of these methods operate by taking a base learning algorithm and invoking it many times with different training sets. In bagging, each training set is constructed by forming a bootstrap replicate of the original training set[22].Ensemble

ISSN 2229-5518

learning methods have become an active research topic within the computational intelligence community. Over the past decade, many theoretical analyses, practical algorithms, and empirical studies have been proposed in this field. Ensemble training techniques also have been widely applied in many real-world applications, including Web mining, financial engineering, geosciences and remote Sensing, biomedical data analysis. Bootstrap aggregating (bagging) is an ensemble learning method based on the idea of developing multiple hypotheses by bootstrap sampling (with replacement) of the available training instances. In the bagging method, the probability sampling function is uniformly distributed across all the training instances. In order to dynamically adjust the weights for different data instances according to their distributions, various boosting algorithms have been developed [10,11].

Ensemble methods make predictions by combining the predictions from a set of individual classifiers. To achieve high prediction accuracy, traditionally it is believed that ensemble methods should have accurate and diverse individual classifiers. "Accurate classifiers" means the prediction accuracy of each classifier should be better than random, that is, larger than 0.5 for a binary classifier. "Diverse classifiers" means each classifier should make prediction independently, so that a combination of these predictions will result in high prediction accuracy for ensemble methods [14].

Boosting is a set of methods for the construction of classifier ensembles [8]. The differential feature of these methods is that they allow obtaining a strong classifier from the combination of minor classifiers. Therefore, it is possible to use boosting methods with very simple base classifiers. [9,10] The simple classifier as decision tress is one decision node. This method is a alternative of the boosting method. It is based on considering, as the base classifiers for boosting, but a classifier formed by the last selected weak classifiers. If the weak classifiers are decision tree, the combination of weak classifiers is a decision tree.

Author describe a process of stream data classification by Kernel-Based Selective Ensemble Learning as Kernel methods enable the modeling of structured data in learning algorithms. Kernel methods provide a dominant technique for modeling structured objects in instant based technique. they require a high computational complexity to be used in streaming environments. This method is the first that demonstrates how kernel methods can be employed to define an ensemble approach able to quickly react to concept drifting and guarantees an efficient kernel computation [12,17].

[21,23]There are several applications for Machine Learning (ML), the most significant of which is data mining. People are often prone to creating errors during analyses or probably, when trying to creating relationships between various dataset. This makes it difficult for them to find solutions to some certain problems. Machine learning can be successfully applied to these problems. Every instance in any feature set used by machine learning algorithms is represented using the same data sets. The features may be continuous, categorical or binary. particular, this work is concerned with classification problems in which the output of instances admits only discrete, unordered values.[18]The library of machine learning algorithm used tools and technique of kernel function. To maximize the recital of the ensemble models a forward process selection is joined. An ensemble is a group of models who's voting are combined by weighted averaging value of classifier. A necessary and ample condition for an ensemble of classifiers to be more accurate than any of its individual members is if the classifiers are good and bad.

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. An essential and sufficient condition for an ensemble of classifiers to be more specific than any of its individual members is if the classifiers are precise and diverse. The area of 'diversity' has been a favorite's buzzword in the multiple classifier systems community for long time. Various diversity measures have been proposed, measured and maximized and all with the goal to increase ensemble performance by balancing "individual accuracy" against "diversity". It is therefore ironic that after so much time and attempt, we still have no distinctively agreed definition for "diversity" [13,15].

[25]The simple forward model selection procedure is fast and effective, but sometimes over fits to the hill climbing set, reducing ensemble execution process. To decrease the over fitting selection with alternate, stored ensemble initialization and bagged ensemble selection methods are used.

III ANT ENSEMBLE CLUESTER CLASSIFIER

Proposed clustering ensemble method based on ant colony optimization where the compromise clustering is selecting optimal cluster criterion function using a ant colony algorithm. This method uses a metric between clustering's based on the ant between partitions. It also uses class level method to solve the label correspondence problem. The search capabilities of ant colony algorithms are used in these methods. It allows exploring partitions that are not easy to be found by other techniques. However, a negative aspect of these algorithms is that a solution is better only in comparison to another; such an algorithm actually has no concept of an optimal solution or any way to test whether a solution is optimal or not. This selection function combines partitions obtained by using locally adaptive clustering AECC algorithms. When a AECC algorithm is applied to a set of objects X, it gives as an output a partition P = { C1,C2,….,Cq}, which can be also identified by two sets {c1,…..cq}and {w1,…..wq}, where ci and wi are the centroid and the confidence associated to the cluster Ci respectively. The AECC algorithms are designed to work with numerical data, i.e. this method assumes that the object representation in the dataset is made up of numerical features: X={x1,…,xn}, with xj € R, j =1, …..,n ; n. Also, ci € Rand wi €2 R, i = 1,….., k. The set of partitions P = {p1,p2,…….,pm} is generated by applying AECC algorithms m times with deferent parameters initialization. The process of ant colony algorithm is to use the ensemble to choose, the selection of pheromone update (increment and decrement of constant deposit of interval value of phenomenon) parameters of ant colony optimization to control the sensitive value. For a given cluster assembling the problem of optimal cluster selection can be stated as follows: given the variable set of cluster index, F ,of n data point, find optimal S ,which consists of m classifier($m < n, S c F$),such that the classification accuracy is maximized. The cluester index selection representation exploited by artificial ants includes the following:

1. n Data point that constitutes the cluster index set,$F = \{f_1 \ldots\ldots f_2\}$.
2. Distribution of the cluster index space (na ants).
3. τ_i , the intensity of pheromone trail associated with cluster index f_i ,which reflects the previous knowledge about the importance off_i.
4. For each ant j , a list that contains the selected optimal cluster subset, $R_j = \{R_1 \ldots\ldots R_m\}$.

AECC evaluation measure that is able to estimate the overall performance of subset as well as the local importance of cluster. A classification algorithm is used to estimate the performance of optimal cluster selection .On the other hand, the local importance of a given cluster measured using the correlation based evaluation function, which is a filter evaluation function. In the first iteration, each ant will randomly choose a cluster index of m classifier. Only the best k subsets, $k < na,$ be used to update the pheromone trail and influence the optimal subset of the next move. In the second and following moves, each ant will start with $m - p$ cluster that are randomly chosen from the previously selected $k -$ best subsets, where p is a number that limit between 1 and $m - 1$. In this way, the ensemble that constitutes the best k subsets will have more chance to be present in the subsets of the next iteration. However, it will still be possible for each ant to consider other cluster index as well. For a given ant j, those cluster index are the once that achieve the best compromise between pheromone trails and local importance with respect to S_j , where S_j is the subset that consists of the cluster that have already been selected by ant j. The Updated selection Measure (USM) is used for this purpose the defined as:

$$USM_i^{S_j} = \begin{cases} \dfrac{(\tau_i)^\alpha \left(LI_i^{S_j}\right)^\beta}{\sum_{g \notin S_j} (\tau_g)^\alpha \left(LI_g^{S_j}\right)^\beta} & if \quad i \notin S_j \\ 0 & otherwise \end{cases} \qquad (1)$$

Where $LI_i^{S_j}$ is the local importance of cluster index f_i given the subset S_j.The parameters α and β control the effect of pheromone trail intensity and local cluster importance respectively. $LI_i^{S_j}$ is measured using the correlation measure and defined as:

$$LI_i^{S_j} = \frac{|C_{iR}|}{\sum_{f_s \notin S_j} |C_{is}|} \qquad (2)$$

Where $|C_{iR}|$ is the absolute value of the correlation between cluster index $i(f_i)$ and the response (class) variable , and $|C_{is}|$ is the absolute value of the inter- correlation between cluster index $i(f_i)$ and opimal $S(f_s)$ that belongs to S_j

Below are the steps of the algorithm:

1. Initialization:
   - Set $\tau_i = cc$ and $\Delta T_i = 0, (i = 1, \ldots., n)$,where cc is a constant and $\Delta\tau_i$ is the amount of change of pheromone trail quantity for variable cluster index f_i.
   - Assign the maximum number of moves.
   - Assign , where the $k - best$ subsets will influence the subsets of the next iteration.
   - Assign , where $m - p$ is the number of cluster indexes that each ant will start with in the second and following moves.
2. If the first iteration,
   - For$j = 1 \; to \; na,$
     - Randomly assign a subset of m classifier to S_j.
   - Goto step 4.
3. Select the remaining p cluster index for each ant:
   - For $mm = m - p + 1 \; to \; m,$
     - For $j = 1 \; to \; na$,

- Given subset S_j, choose cluster index f_i that maximizes $USM_i^{S_j}$
- $S_j = S_j \cup \{f_i\}$.

- Merge the duplicated subsets, if any, with randomly chosen index.

4. estimated the selected index of each cluster ant using a chosen classification algorithm:

- For $j = 1\ to\ na$,
  - Estimate the Error (E_j) of the classification results obtained by classifying the optimal cluster of S_j.

- Sort the subsets according to their E. Update the minimum E (if achieved by any ant in this iteration), and store the corresponding subset of cluster.

5. Using the ensemble subsets of the best k ants, update the pheromone trail intensity:
   - For $j = 1\ to\ k$,

$$\Delta\tau_i = \begin{cases} \dfrac{\max\limits_{g=1:k}(E_g) - E_j}{\max\limits_{h=1:k}\left(\max\limits_{g=1:k}(E_g) - E_h\right)} & if\ \ f_i \in S_j \quad (3) \\ 0 & otherwise \end{cases}$$

$$\tau_i = \rho.\tau_i + \Delta\tau_i \qquad (4)$$

Where ρ is a constant such that $(1-\rho)$ represents the evaporation of pheromone trails.

6. If the number of moves is less than the maximum number of moves, or the desired E has not been achived, initialize the subsets for next iteration and goto step3:

- For $j = 1\ to\ na$,
  - From the selected cluster of the best k ants, randomly produce $m-p$ classifeir subset for ant j, to be used in the next iteration, and store it in S_j.
- Goto step 3.

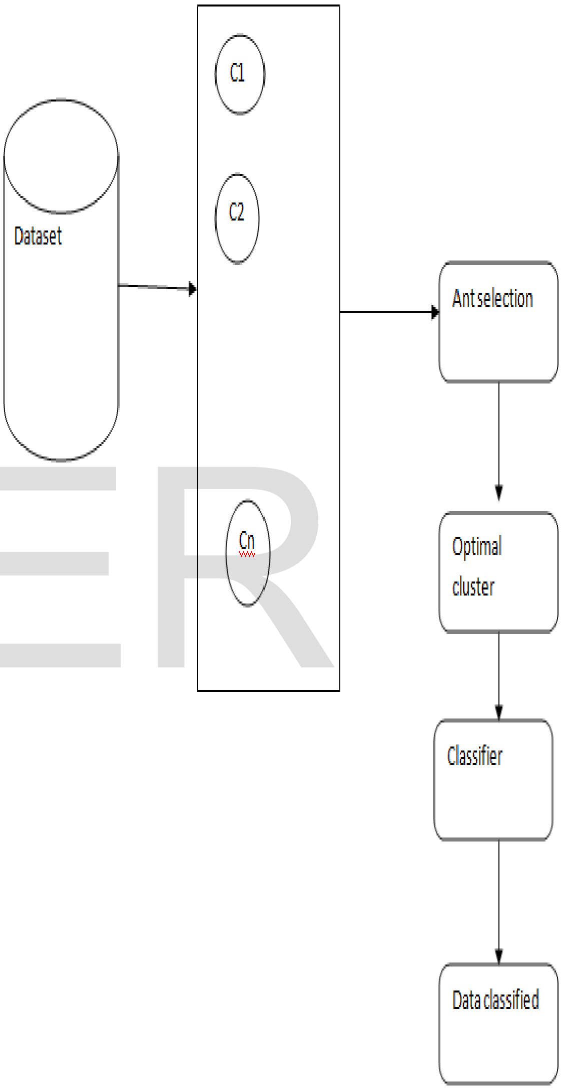

Figure 1 shows proposed model of AECC

IV EXPERIMENTAL RESULT

In this section describe the optimal selection of cluster for improvement of cluster based classifier. The basic classifier used as support vector machine and KNN. The kernel of support vector machine is replaced with RBF kernel function. For the performance evaluation we used six dataset forms UCI machine learning repository. These datasets are cancer; glass, iris, page, cancer and finally wine dataset are used. Our modified classifier implements in MATLAB 7.8.0 software package and used library function of support vector machine. Here we show some classified data region using cluster ensemble classifier.

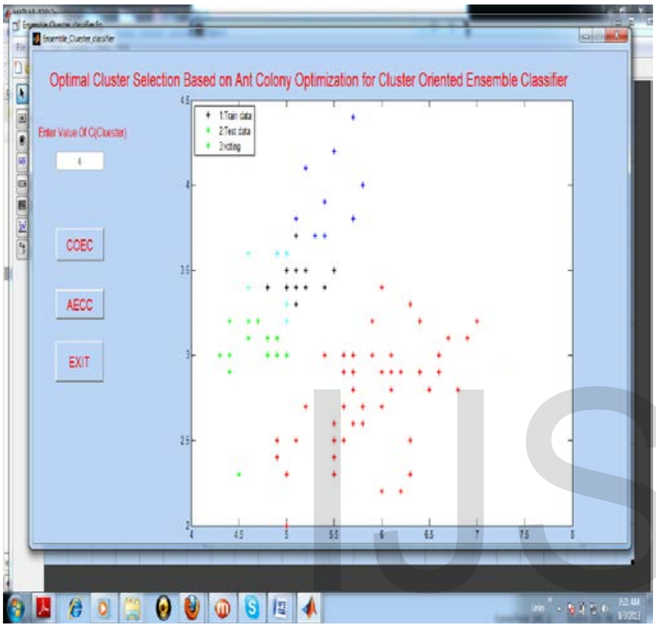

Figure 2 shows that classification of classified data of cluster ensemble classifier with fixed number of cluster value

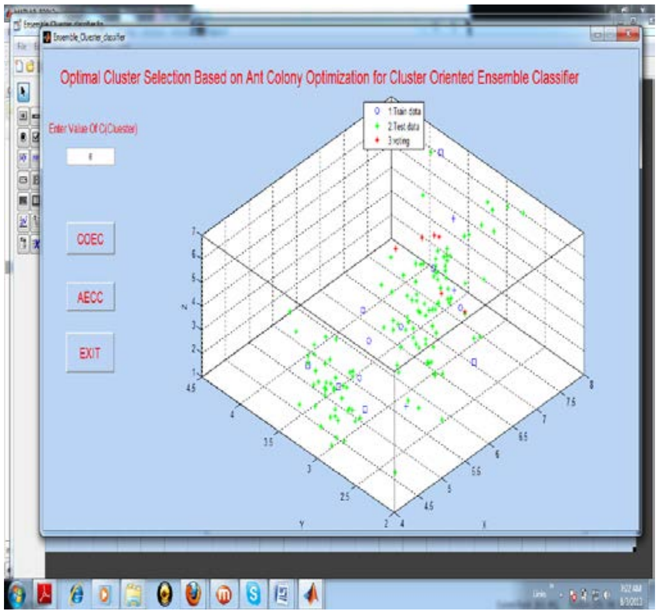

Figure 3 shows that classification of classified data of cluster ensemble classifier with optimal number of cluster of cluster value

Performances evaluation matrix of all data

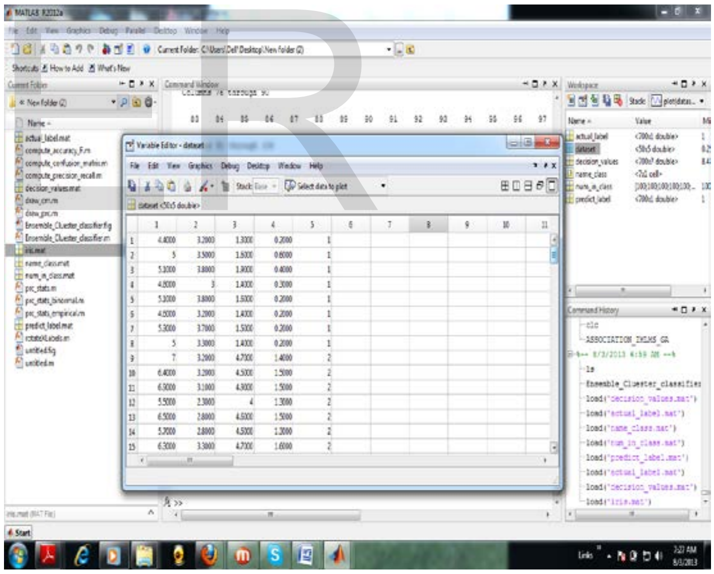

Figure 4 shows that confusion matrix of iris data for measuring accuracy and classification of iris data in three classes with optimal cluster selection technique.

Performance evaluation parameter

Classification Accuracy

The classification accuracy of all six dataset is increased significantly by the proposed method.

Accuracy rate = Total no. of correctly classified instance * 100 ...... (a)
Total no. of instance

Mean absolute Error Rate

In classification the mean absolute error (MAE) is a quantity used to measure how close real or predictions are to the eventual outcomes. The mean absolute error is given by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$
.... (b)

As the name suggests, the mean absolute error is an average of the absolute errors $e_i = |f_i - y_i|$, where f_i is the prediction and y_i the true value. Note that alternative formulations may include relative frequencies as weight factors.

Table 1 gives the information about result analysis of all data such as iris, cancer, glass and wine the number of cluster value is used 8 and base classifier is support vector machine.

METHOD	MAE	ACCURACY
COEC	3.45	96.00
ACEE	2.7	97.85
COEC	28.86	71.08
AECC	25.0	73.46
COEC	6.5	92.95
AECC	4.5	93.95
COEC	8.68	90.95
AECC	7.67	91.59

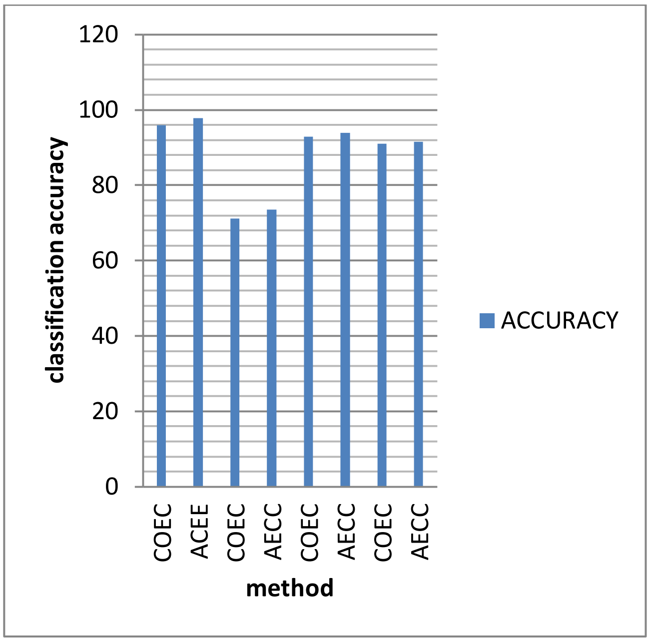

Figure 4 shows that comparative result analysis of all data set used in cluster ensemble technique and ant technique the classification result show that our process of method is better than COEC.

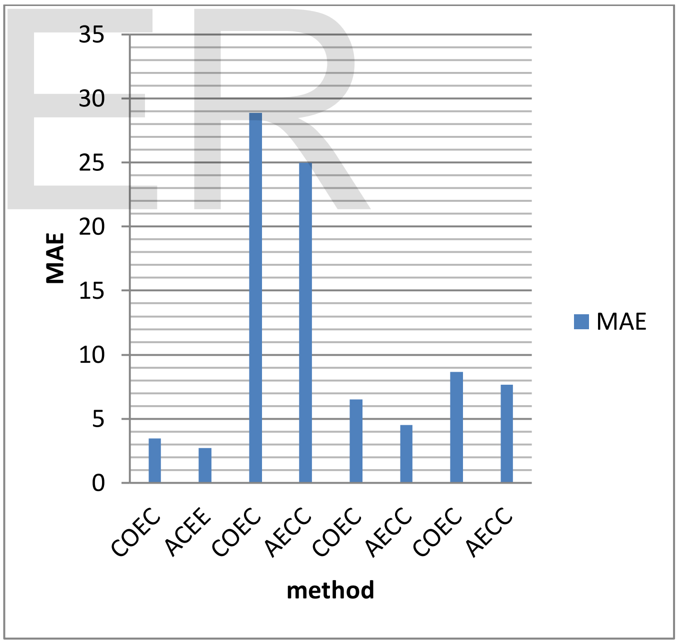

Figure 6 shows that comparative result analysis of all data set used in cluster ensemble technique and ant technique the classification result show that our process of method is better than COEC.

V CONCLUSION AND FUTURE WORK

In this paper proposed a ant ensemble cluster classifier (AECC) which is based on variable of cluster index by the base classifiers leading to better classification capability and cluster-to-class mapping by a classifier leading to better classification accuracy. The proposed AECC has been evaluated on benchmark data sets from UCI machine learning repository. The detailed experimental results and their significance give details in section IV. In the future, we will further study and analyze the utilization of the proposed approach for other text mining tasks such as text retrieval and document clustering.

REFERENCES:-

[1] Brijesh Verma and Ashfaqur Rahman "Cluster-Oriented Ensemble Classifier: Impact of Multicluster Characterization on Ensemble Classifier Learning" in IEEE Transactions on knowledge and data engineering, 2012.

[2] Anne-Laure Bianne-Bernard, Fare`s Menasri, Rami Al-Hajj Mohamad, Chafic Mokbel, 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.

[3] Nayer M.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.

[4] LEO BBEIMAN "Bagging Predictors" in Kluwer Academic Publishers, 1996.

[5] Yoshua Bengio "Learning Deep Architectures for AI" in Foundations and Trends in Machine Learning, 2009.

[6] Giorgio Fumera, Fabio Roli and Alessandra Serrau "A Theoretical Analysis of Bagging as a Linear Combination of Classifiers" in IEEE Transactions.

[7] Albert Hung-Ren Ko and Robert Sabourin "The Implication of Data Diversity for a Classifier-free Ensemble Selection in Random Subspaces" in IEEE Transactions.

[8] Zhihui Lai, Zhong Jin, Jian Yang and W.K Wong "Sparse Local Discriminant Projections for Face Feature Extractio" International Conference on Pattern Recognition, 2010.

[9] Juan J. Rodrı´guez and Jesu ´s Maudes "Boosting recombined weak classifiers" in ScienceDirect, 2007.

[10] Oriol Pujol and David Masip "Geometry-Based Ensembles: Toward a Structural Characterization of the Classification Boundary" in IEEE Transactions, 2009.

[11] Haibo He and Yuan Cao "SSC: A Classifier Combination Method Based on Signal Strength" in IEEE Transactions on neural networks and learning systems, 2012.

[12] Nandita Tripathi, Stefan Wermter, Chihli Hung and Michael Oakes "Semantic Subspace Learning with Conditional Significance Vectors" in IEEE Transactions.

[13] Terry Windeatt "Accuracy/Diversity and Ensemble MLP Classifier Design" in IEEE Transactions.

[14]Xueyi Wang "A New Model for Measuring the Accuracies of" in IEEE World Congress on Computational Intelligence, 2012.

[15]Gavin Brown, and Ludmila I. Kuncheva ""Good" and "Bad" Diversity in Majority Vote Ensembles" in IEEE Transaction.

[16]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.

[17]Rich Caruana, Alexandru Niculescu-Mizil, Geoff Crew and Alex Ksikes "Ensemble Selection from Libraries of Models" 21st International Confer-ence on Machine Learning, 2004.

[18]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.

[19]Nikunj C. Oza and Kagan Tumer "Classifier Ensembles: Select Real-World Applications" in Elsevier, 2007.

[20]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.

[21]Leo Breiman "Bagging Predictors" in Kluwer Academic Publishers, 2006.

[22]Thomas G. Dietterich "Ensemble Methods in Machine Learning" in IEEE TRANSACTIONS.

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

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

[25]Guoqiang Peter Zhang entitled "Neural Networks for Classification: A Survey" in IEEE TRANSACTIONS, 2000.

[26]Shivam Mishra and Anurag Jain " A Review of Cluster Based Classification Technique" in International Journal of Computer Technology & Applications,2013,Vol 4 issue-4 PP673-677