International Journal of Electrical, Electronics and Computer Engineering **3**(1): 206-211(2014)

ISSN No. (Online): 2277-2626

Selecting Feature Using Ant Colony Optimization Algorithm

Shruti Tiwari* and Prof. Anurag Jain**

**Research Scholar, Department of Computer Science & Engineering.*
RITS, Bhopal, (MP) India
***Professor, Department of Computer Science & Engineering.*
RITS, Bhopal, (MP) India

(Corresponding author: Shruti Tiwari)
(Received 05 March, 2014 Accepted 25 June, 2014)

ABSTRACT: The paper on Multiple Feature Subset Selection Using Meta-heuristic Function is being presented. Classification problems needs selection of a subset of attributes or features from a more enormous dataset to represents the patterns to be classified. Various splendid different feature selection technique such as Hill Climbing (HC), Genetic Algorithms (GAs), Simulated Annealing (SA), Tabu Search (TS) has been frequent amongst research group. Although, these things still confront when the various feature of dataset is accessible and we need to choose those attribute which is best amongst the accessible features. So considering the issue of multiple features is analyzed and implemented and tested on different dataset. The experiment is being conducted on and abalone dataset. We may conclude that the algorithm shows highest accuracy amongst all other method which is being used in this thesis, Therefore we are focusing on the classification of multiple features.

Keywords: Meta-heuristic Function, Ant colony optimization, k-NN, Tabu Search, Genetic Algorithms, multiple features,

I. INTRODUCTION

The increasing computerization with in the world around us has meant that the existence of database containing vast quantities of data is now a fact of everyday life. The enormous quantities of data which are now stored in databases create a problem in that it becomes very difficult to make meaningful sense of such a large quantity of data. For human investigators, the [1] process of extracting meaningful information from such a large amount of data becomes the classic problem of information overload. The field of data mining seeks to address [3] this problem by the use of computer modeling techniques to derive useful knowledge in a concise form from these large databases. Using the ability of computers to sort ,analyze and categorize large volumes of data extremely quickly, data mining methods seek to redress the problem of information overload by allowing fast and reliable methods of data modeling and representation .A typical database or dataset to which data mining methods can be applied will consist of a number of data elements or examples, which are termed tuples in [4] the field of relational databases.

Each elements or examples of data are made up of a set of attributes, each of which encodes a value relevant to the type of the given attribute. Generally, all data elements making up a dataset will consist of the same attributes, giving each data element a consistent form. This allows data mining methods to look for patterns and commonality between the data elements. The general field of data mining encompasses a range of tasks, each specific to a different set of target problems. Major areas of data mining are as follows:

(i) Summarization: the process of summarization is that of finding a simple model or description for a given set of data. Some examples of this might be finding summary rules to describe common properties for a subset of data, or visualization methods for displaying relationships within the data.

(ii) Deviation detection: this process is used to detect the significant and consistent changes in a set of data over a series of time steps. The process derives a model which defines which attributes change and how they change over time.

(iii) Classification: this is the task of finding some model or function which maps each element of data to one class, out of a discrete set of possible classes. Each element of data must belong to [5] only one class, and the classification task is to accurately assign each element of data to the appropriate class.

(iv) Regression/ Numeric Prediction: similar to classification the task of numeric prediction is to derive some model or function which maps each element of data to a value. In this case however the prediction value is a continuous value. [8] A model must predict an output value to the appropriate value as possible for each element of dat.

(v) Association Finding / Dependency Modeling: this problem involves the derivation of a model which describes dependencies or associations between variables of a problem. The model will indicate that on variable, or a set of variables, influences the value of another variable, and will often indicate the strength of this influence.

(vi) Clustering: the task of clustering is to find some model which distinguishes a subset of data elements from the main body of data elements based upon some ver the space of the attributes of the problem. characteristics of the subset. The elements which form the cluster should be closely related in some way, in which data elements outside the cluster are not related. The measure of how closely related elements of the cluster are and how greatly this differs from the wider body of data elements indicates the strength of the cluster.For a good discussion of general data mining issues and methods see work such as [1, 2].

II. CLUSTERING AND CLASSIFICATION

Two common data mining techniques [3] for detecting hidden patterns in data are classification and clustering analysis. Although classification & clustering are frequently declared in the similar breath, these are different analytical approaches. Imaging a database of customer records, where each record derived customer's attributes. [10] These includes identifiers like name & address, demographic information such as gender and age and financial attributes such as income & revenue spend. Clustering is a method to classify equivalent records together. And related records are classified together on the starting of having same values for attributes. These method of separating the database via clustering analysis is usually used an exploratory approach because it is not required for the end-users to specify ahead of time how records should be related together. Actually, the objective of analysis is usually to find segments or clusters & then testing the attributes and values which explain the clusters or segments. As a interesting and surprising way of grouped customers together become apparent and this can be used to drive promotion and marketing criteria to target specific types of customers There are many types of algorithms used for clustering, but they all are share the property of iteratively assigning records to the cluster, determining a estimate(generally equivalence and distinctiveness) and re-assigning data to clusters till the calculated estimate don't change more specifying that the method has converge to stable segments. Records within the cluster are much equal to each other and more differ from data that are in the other clusters. Depends upon the particular implementation and there are variety of estimates of equivalence that are used (for e.g. found on statistical variability [6] or even adaptations, based on spatial distance of Condorcet assess used in voting schemes) but the altogether aim is for the method to converge to the set of similar data. There are various algorithms for clustering but they all are part of characteristics of iteratively mentioning data to the cluster, determining a measure (generally similarity, and distinctiveness) and re-assigning data to clusters till the evaluated measures don't change more indicating that the method has converge to stable segments. Records within the cluster are much similar to every other and more different from records that are in the other clusters [9]. Depends upon the particular implementation and there are variety of parameter of equality that are used (e.g. depends on spatial distance, statistical variability or even adaptations of Condorcet values used in voting schemes) but the whole aim is for the approach to converge to groups of related records. Classification is a individual technique than clustering. [11] Classification is equal to the clustering, that it also part of customer data records into different parts called classes. But not like clustering, the classification analysis requires that the end-user know ahead of time how classes are defined. For eg: classes are defined to show the possibility that a customer defaults on a loan whether (Yes/No). It is requires that every data in the dataset used to built the classifier already contains a value for the attribute used to explain classes. Because every data has value for the attribute which used to define the classes and because the end-users decide to use on the attribute, classification is less exploratory than clustering. The objective of classifier is not to explore the data to discover interesting segment, but decide how new records should be classified, i.e. is this new customer likely to default on the loan?

Classification procedure in data mining also used many types of algorithms and the given algorithm used may affect the data are classified. A common approach for the [14] classifiers is to use the decision trees to partition & segment the records. Fresh data may be classified by traversing the tree from the root via branches and nodes to a leaf which represents a class. The path of record takes by a decision tree then can be constitute. For example: Income < $30,000 and age<25 & debt=High, then Default Class =Yes).

But due to the continuous nature of the way a decision tree break the records (most discriminative attribute-values) can result in a decision tree being over sensitive to previous break. Therefore, calculating the goodness of fit a tree, it is important to check the error rate for every leaf node (proportion of records incorrectly classified). A characteristics of decision tree classifiers are because paths may be conveyed as a rule then it becomes possibly to use measure for calculating the useness of rules like Support, Confidence and Lift to evaluate the usefulness of the tree. To conclude, although clustering & classification are frequently used for uses of segmenting data records, they have such variety of objectives and achieve their segmentation through various ways. Knowing which approach is use to important for decision-making.

III. PROPOSED WORK

In the process of classification we find that the normal classification algorithm is very slow in term of processing time as well as the classification rate is very low. Now we are applying some existing methodology for the improvement of classification of K.N.N algorithm such as MFS, MDF and FC-MMNC.But all these methods of classification not up to the mark for the different- different dataset to increase accuracy and decrease errors such as abalone, iris, cancer, aids and internet advertisement. [15] Because classification required a search space for better classification of multiple feature dataset. Now we proposed a new technique for multiple feature classification of data. We adopt meta-heuristic algorithm ant colony optimization with multiple feature subset classification in order to improve accuracy of classification.

Ant Colony Optimization (ACO) it is the ability of real ants to discover shortest path is mostly due to their rest of pheromone as they travel each ant probabilistically favour to follow a supervision rich in this chemical. [13]. The pheromone decompose over time proceed in more less pheromone on less favoured routes. Stated that over time the shortest path will have the higher rate of ant traversal, this route will be strengthen and the others diminished until all ants follow the same shortest route. The overall process of ACO feature selection can be seen in Fig. 1.

For a given classification function, the feature selection problem can be stated as follows: given the original set F of n features find subset S which consists of m features (m < n, S Ì F) such that the classification accuracy is maximized.

The feature selection description utilize by artificial ants includes the following: • n features that constitute the authentic set, F = [f1 ….. fn]

•A number of artificial ants to find through the feature space (na ants). is the intensity of pheromone trail associated with feature fi which reflects the previous knowledge about the importance of fi.

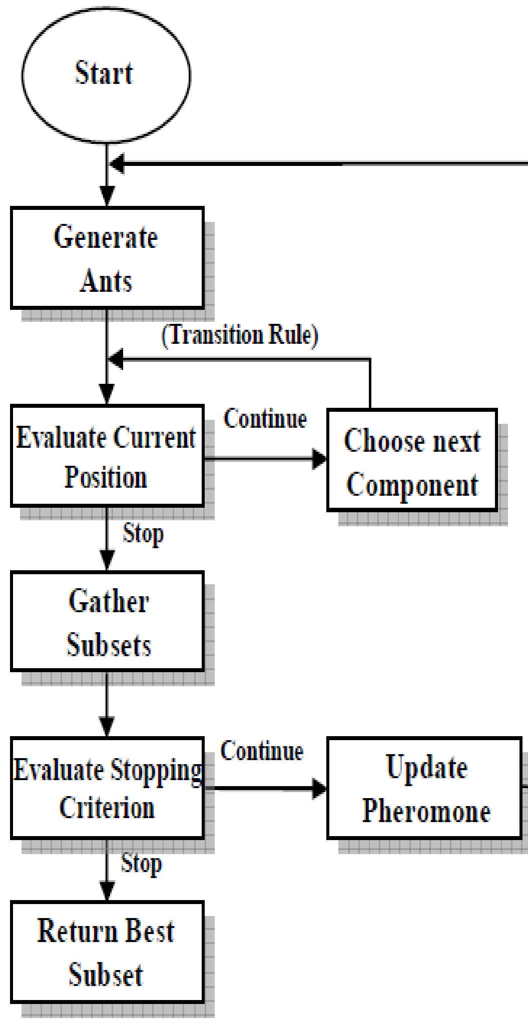

Fig.1. Over all ACO Feature Selection Process.

•For every ant j a list that having the selected feature subset

Sj = {S1… Sn}

We propose to use a hybrid evaluation measure that is able to estimate the overall performance of subsets as well as the local importance of features. A classification algorithm is used to estimate the performance of subsets (i.e., wrapper evaluation function).

On the other hand, the local value of a given feature is measured using the correlation based evaluation function which is a filter evaluation function. In the first step each ant will randomly choose a feature subset of m features. Only the best k subsets k < na, will be used to update the pheromone trail and influence the feature subsets of the next step. In the second and following step, each ant will start with m - p features that are randomly chosen from the previously selected k -best subsets where p is an integer that ranges between 1 and m -1. In this way the features that constitute the best k subsets will have more chance to be present in the subsets of the [7] next step. However it will still be possible for each ant to consider other features. For a stated ant j those features are the ones that achieve the best compromise between pheromone trails and local importance with respect to Sj where Sj is the subset that consists of the features that have already been selected by ant j. The Updated Selection Measure (USM) is used for this purpose and defined as: Where is the local importance of feature specify the subset [18] the parameters α and β manage the result of pheromone stream strength and limited feature consequence respectively. Is measured using the correlation measure and defined asWhere is the entire value of the correlation between feature i and the response (class) variable R and is the absolute value of the inter-correlation between feature i(fi) and feature s(fs) that belongs to Sj .Below are the steps of the algorithm:

1. Initialization:
    - Set ti = CC and ΔTi = 0, (i= 1………n) where cc is a constant and Δti is the amount of change of pheromone trail quantity for feature fi.
    - Define the extreme number of steps.
    - Define k, where the k -best subsets will influence the subsets of the next steps.
    - Define p, where m - p is the number of features that every ant will begin with in the second and following steps.
2. If in the first step,
    - For j =1 to na,

Randomly assigns a subset of m features to Sj.

- Goto step 4.

3. Select the remaining p features for each ant:

For mm = m - p +1 to m ,

For j =1 to na ,

Given subset Sj, choose feature fi that decreases.

- Replace the replicate subsets if any with randomly chosen subsets.

4. Assess the selected subset of every ant using a chosen classification algorithm:

- For j =1 to na.

Estimate the Error (EJ) of the classification results obtained by classifying the features of Sj.

- 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 features.

5. Using the feature subsets of the best k ants update the pheromone trail intensity:

- For j =1 to k. /* update the pheromone trails */

=

6. If the number of iterations is less than the maximum number of iterations, or the desired E has not been achieved initialize the subsets for next iteration and go to step 3

- For j =1 to na,

O From the features of the perfect k ants, randomly create m - p feature subset for ant j to be used in the next iteration and store it in SJ.

- Goto step 3.

IV. RESULT

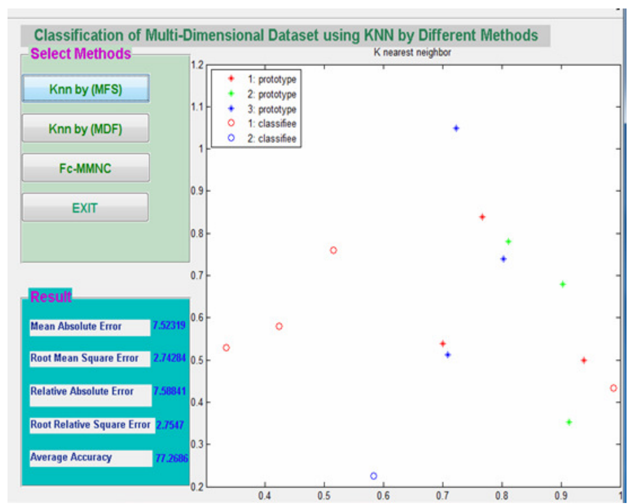

Fig. 2. Generating function value α = 0.25 K-NN by (MFS) METHOD.

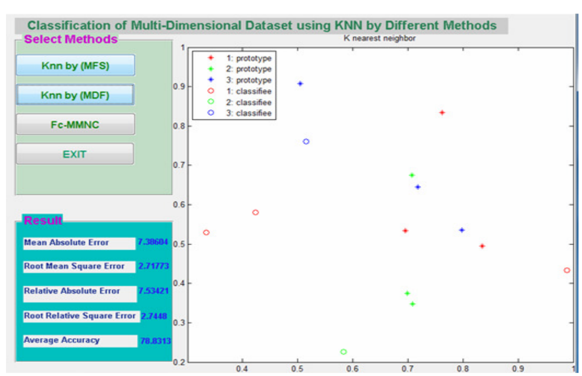

Fig. 3. Generating function value α = 0.25 K-NN by (MDF) METHOD.

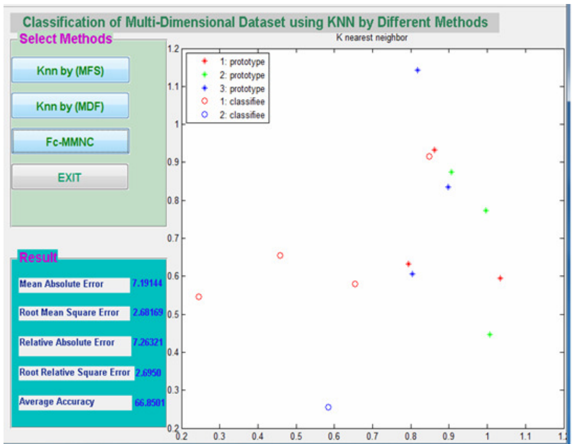

Fig. 4. Generating function value α = 0.25 Fc-MMNC METHOD.

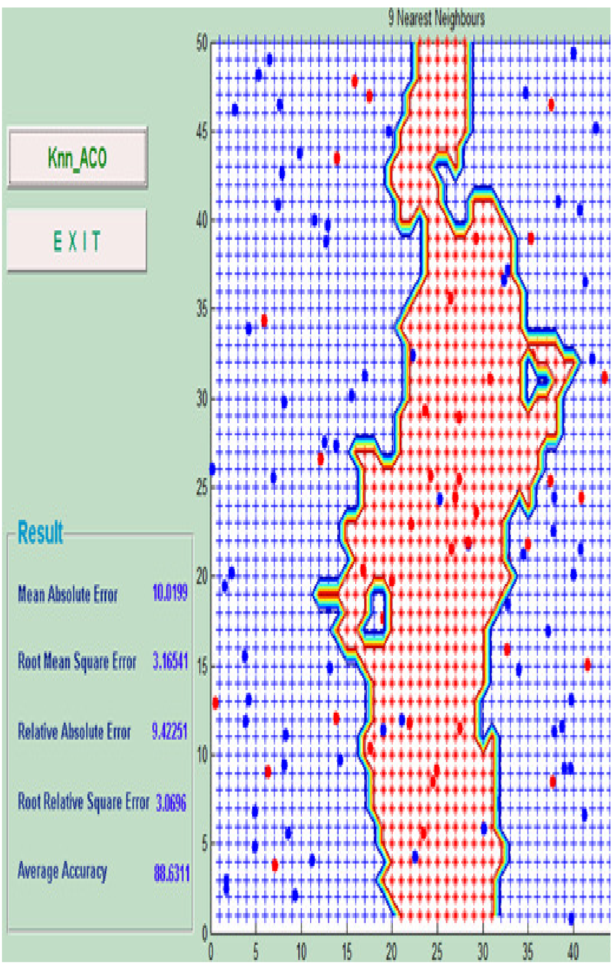

Fig. 5. Generating function value α = 0.25 K-NN by ACO METHOD.

V. CONCLUSION

In this thesis, I presented a feature selection search procedure based on the Ant Colony Optimization. The proposed algorithm uses both local consequence of features and overall presentation of subsets to find through the feature margin for optimal solutions. When used to select features in presented datasets the proposed algorithm outperformed other feature selection methods (MFS, MDF and FC-MMNC). Results show this algorithm selects a good subset of features that are useful to common machine learning algorithms by improving their accuracy and making their results easier to understand especially in data sets with irrelevant or redundant features. It is evident that in feature selection, enhancement in accurate classification rate based on the link between features and hence based on dataset. Therefore in datasets with uncorrelated features and without peripheral features and the feature selection may be result reducing of correct classification rate. Other advantage of algorithm is that it scales only with the number of features. Therefore does not require extra computation cost if the number of the data points in a dataset increases.

VI. FUTUREWORK

Now the problem with algorithm is more computation time so we can enhance the given algorithm by using different methods With the help of new method we can improve feature subset selection problem.

REFERENCES

[1]. U. Fayyad, G.Piatetsky-Shapiro and P.Smyth. From data mining to knowledge discovery in databases. *Ai Magazine*, Volume **17**, pages 37-54, 1996.

[2]. Ian H.Written and Eibe Frank. Data *Mining*. Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, 2000

[3]. Ed *Colet* "Clustering and classification*:* Data mining approaches*"* July 4, 2000 / Vol. **4** NO.27.

[4]. J.R. Quinlan. *C4.5:* programs for machine learning. Morgan Kaufmann, 1992.

[5]. J.R. Quinlan. Improved use of continuous attributes in C4.5. Journal of artificial intelligence, Volume **4**, pages 77-9-, 1996.

[6]. Sreerama K.Murthy, Simon Kasif, Steven Salzberg and Richard Beigel. Ocl: A randomized induction of oblique decision trees. In National Conference on Artificial Intelligence, pages 322-327, 1993.

[7]. Sreerama K. Murthy. Automatic construction of decision trees from data: A multidisciplinary survey. *Data mining and Knowledge Discovery*, Volume **2**, Number 4, pages 345-389, 1998.

[8]. Brian R. Gaines and Paul Compton. Induction of ripple-down rules applied to modeling large databases. *Journal of Intelligent Information Systems*, Volume **5**, Number 3, pages 211-228, 1995

[9]. David W. Aha. Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man Machine studies*, Volume **36,** Number 2, pages 267-287, 1992

[10]. Leo Breiman. Bagging predictors. Machine Learning, Volume **24**(2), pages 123-140, 1996.

[11]. Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. *In proc. 13th International Conference on Machine Learning*, pages 148-146 Morgan Kaufmann, 1996.

[12]. David H. Wolpert. Stacked generalization. Neural Networks, Volume **5**, Number 2, pages 241-259, 1992.

[13]. H. Liu and L. Yu. Towards integrating feature selection algorithms for classification and clustering, *IEEE Transactions on Knowledge and Data Engineering,* **17**(3), pages. 1–12, 2005.

[14]. T.G. Dietterich. Ensemble methods in machine learning. In J. Kittler and F.Roli (editors), First International Workshop on Multiple Classifier systems, pages 1-15 Springer-Verlag, 2000

[15]. Stephen D. Bay "Combining nearest neighbor classifiers through multiple feature subsets" Proceeding 17th Intl. Conf. on Machine Learning: 191-209,1998.

[16]. Li-Juan Wang, Xiao-Long Wang, Qing-Cai Chen "GA-based Feature Subset Clustering for Combination of Multiple Nearest Neighbor Classifiers" Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005.

[17]. L. Louisa Lam and C Y. Suen "Application of majority voting to pattern recognition: An analysis of its behavior and performance*", IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, Vol. **27**, no. 5, pp.553-567, 1997.

[18]. Takahiro Yamada, Kyohei Yamashita, Naohiro Ishii "Text Classification by Combining Different Distance Functions with Weights" Proceedings of the Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing- SNPD, 2006.