International Journal of Scientific Research and Management (IJSRM)
||Volume||5||Issue||08||Pages||6764-6770||2017||
Website: www.ijsrm.in ISSN (e): 2321-3418
Index Copernicus value (2015): 57.47 DOI: 10.18535/ijsrm/v5i8.25

Improved Feature Subset Selection using Hybrid Ant Colony and Perceptron Network

Nidhi Grover[1], Rakesh Chawla[2]

*M.Tech Scholar, CSE Department, DITM , Sonipat
**Assistant Professor, CSE Department, DITM, Sonipat

Abstract
As classification accuracy is strongly dependent on the set of features used as input variables. For automatic feature acquisition, the literature provides us with numerous strategies aiming to find a “best” set of features. Thus we need to find a “best” set of features given a constraint on the computational complexity or cost of the feature acquisition, which may dominate the cost of the classifier. In our work we improve on the accuracy of ACO FSS algorithm by incorporating the Perceptron Classifier as fitness function of ACO whose classification error is taken as the Cost of Classifier. This method improves on accuracy of selecting minimum number of features with maximum accuracy achieved.

Keywords: ACO; CSO; ICSO; MSE.

Introduction

Data mining is used to mine useful data from a large amount of data in the same way as extraction of minerals from mine fields. Clustering is the classification of similar objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset share some common trait. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. Data mining is more than just conventional data analysis. It uses traditional analysis tools like statistics and graphics plus those associated with artificial intelligence such as rule induction and neural nets. It is all of these, but different. It is a distinctive approach or attitude to data analysis. The emphasis is not so much on extracting facts, but on generating hypotheses.

Feature selection problem deals with selection of an optimum relevant set of features or attributes that are necessary for the recognition process (classification or clustering). It helps reduce the dimensionality of the measurement space. The goal of feature selection is mainly threefold. First, it is practically and computationally difficult to work with all the features if the number of features is too large. Second, many of the given features may be noisy, redundant, and irrelevant to the classification or clustering task at hand. Finally, it is a problem when the number of features becomes much larger than the number of input data points. For such cases, reduction in dimensionality is required to permit meaningful data analysis. Feature selection facilitates the use of easily computable algorithms for efficient classification or clustering. In general, the feature selection problem (Ω, P) can formally be defined as an optimization problem: determine the feature set F^* for which

$$P(F^*) = \min \ P(F,X) \ ; \quad F \in \Omega$$

Where Ω is the set of possible feature subsets, F refers to a feature subset, and P: $\Omega \times \psi \rightarrow (R)$ denotes a criterion to measure the quality of a feature subset with respect to its utility in classifying/clustering the set of points $X \in \psi$. The elements of X, which are vectors in d-dimensional space, are projected into the subspace of dimension dF=|F|=d defined by F. P is used to judge the quality of this subspace.

Feature selection can be either supervised or unsupervised. For the supervised case, the actual class labels of the data points are known. In filter approaches for supervised feature selection, features are selected based on their discriminatory power with regard to the target classes. In wrapper approaches for supervised feature selection, the utility of F is usually measured in terms of the performance of a classifier by comparing the class labels predicted by the classifier for feature space F with the actual class labels.

The organization of present paper is as follow. Section II presents the literature survey which highlights the facts of various researchers. Section III describes the methodology used for proposed work as in this paper ant colony optimization is used. Result analysis is presented in section IV following the concluding remarks in section V.

I. Literature review

This section will provide the brief description and highlights the contribution, remarks and factors of the work done by the researchers. Many attempts have been made in the past to achieve minimization of mean square error & execution time.

Hong Zeng et al., 2008 [13] This paper addresses the problem of feature selection for the high dimensional data clustering. In this paper, they propose a novel feature weighting scheme for a kernel based clustering criterion, in which the weight for each feature is a measure of its contribution to the clustering task.

Zheng Zhao et al., 2009 [14] In this research paper the evolving and adapting capabilities of robust intelligence are best manifested in its ability to learn. Machine learning enables computer systems to learn, and improve performance. Feature selection facilitates machine learning (e.g., classification) by aiming to remove irrelevant features.

Daniela M. Witten et al., 2010 [15] In this paper they consider the problem of clustering observations using a potentially large set of features. One might expect that the true underlying clusters present in the data differ only with respect to a small fraction of the features, and will be missed if one clusters the observations using the full set of features.

Iffat A. Gheyas et al., 2010 [16] This paper Searching for an optimal feature subset from a high dimensional feature space is known to bean NP complete problem. They present a hybrid algorithm, SAGA, for this task. They compare the performance overtime of SAGA and well-known algorithms on synthetic and real data sets.

Deng Cai et al., 2010 [17] In this research paper in many data analysis tasks, one is often confronted with very high dimensional data. Feature selection techniques are designed to find the relevant feature subset of the original features which can facilitate clustering, classification and retrieval.

John Q.Gan et al., 2011 [18] This paper in brain-computer interface (BCI) development, temporal/spectral/ spatial/statistical features can be extracted from multiple electroencephalography (EEG) signals and the number of features available could be up to thousands.

Cordeiro Ferreira et al., 2013 [19] In this paper given a very large moderate-to-high dimensionality dataset, how could one cluster its points. For datasets that don’t fit even on a single disk, parallelism is a first class option. In this paper they explore Map Reduce for clustering this kind of data.

DOI: 10.18535/ijsrm/v5i8.25

Manuel Ignacio Lopez et al., 2012 [20] This paper proposes a classification via clustering approach to predict the final marks in a university course on the basis of forum data. The results show that the Expectation-Maximisation (EM) clustering algorithm yields results similar to those of the best classification algorithms, especially when using only a group of selected attributes.

Houtao Deng et al., 2012 [21] In this research paper they propose a tree regularization framework, which enables many tree models to perform feature selection efficiently. The key idea of the regularization framework is to penalize selecting a new feature for splitting when its gain (e.g. information gain) is similar to the features used in previous splits.

Quanquan Gu et al., 2012 [22] In this paper Fisher score is one of the most widely used supervised feature selection methods. In this paper, they present a generalized Fisher score to jointly select features. It aims at finding a subset of features, which maximize the lower bound of traditional Fisher score.

Kuan-Cheng Lin et al., 2016 [18] This paper presented an improved version of the cat swarm optimization (CSO) has been proposed recently and the authors demonstrated that their version of algorithm outperformed earlier PSO based models and initial version of CSO algorithm. Since CSO is restricted due to long computation time taken thus, the authors modified and improved CSO to present ICSO algorithm. The algorithm was experimented to selects features for text classification in big data. Results showed that ICSO outperforms CSO in terms of classification accuracy when applied to UCI datasets. This section has provided the brief review of the work done in past. It also highlighted the factors, contribution and remarks on the achievement.

II. Frame Work for Implementation

In this research work following objectives are being analysed in the presented research work:-

1. Selection of Multi-dimensional Data for study, Open Source Multi-dimensional data sets will be used for the implementation.
2. Data Pre-Processing and Noise and Irrelevant feature removal from the data sets.
3. Implementation of Correlation based Ant Colony Optimization (CB-ACO) while hybridizing it with Perceptron network for feature Selection with ability to data partition and feature selection.

While fulfilling the above objectives Extensive experiments will be carried out to compare the proposed algorithm with existing representative feature selection algorithms, including, Cat Swarm Optimization (CSO) and its improved version ICSO [18] in order to show the performance improvement during feature subset selection based on prediction analysis, when implemented with the same datasets. The steps of CB-ACO can be described by the flowchart shown in Fig. below, which are described in more details as follows:-

Step 1: Initialize the parameters of ACO, including the number of ants m, the maximum number of iteration Tmax, the tunable parameters α , β and ρ , the initial pheromone level τ_0, the pheromone trail interval [τ_{min} , τ_{max}] and the heuristic information, η of all edges using one of the methods described in the previous sub-section. In proposed ACO based Tmax is stalled using a gradient. Stalling is defined as “If gradient is does not changed for specified number of iterations and a above a threshold” the Tmax is updated.

Step 2: Construct candidate solutions from a randomly selected node and select the rest 2(n-1) roads ending to (n-1) features, follow the proposed probabilistic transition rule. Choosing sub-node 1 of a feature means selecting, and sub-node 0 means deselecting that feature. Stay in this step while all features (nodes) have been visited by each ant. Size of candidates is defined using an Heuristic function utilizing number features present in the database. The proposed flow chart of the process explained above is as shown in Figure 1.

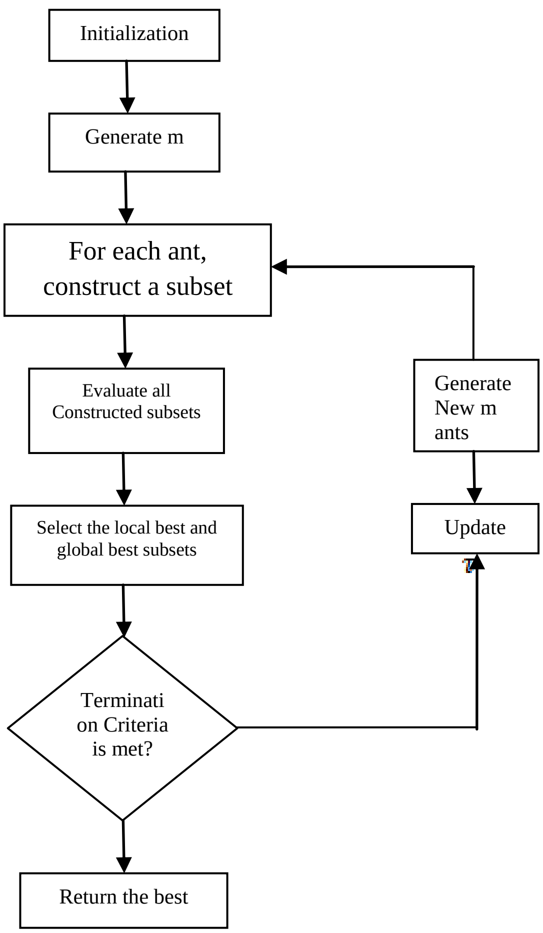

Fig 1 Flow Chart of the Proposed Process

Step3: Evaluate the candidate feature subsets using the trained classifier by testing the classification accuracy on the training set.

Step4: After evaporation of the laid pheromone, find the ant with the best feature subset. It is the subset with the best classification accuracy. Only permit this ant to deposit pheromone according to the scheme explained in the previous section. Keep the pheromone values of edges in the interval [τ_{min}, τ_{max}]. If the maximum number of iteration, Tmax is not reached go back to step 2; otherwise, go to the next step.

Step5: search for the global best subset which produces the highest classification accuracy among all local best solutions.

III. Result Analysis

MATLAB platform has been used to evaluate the results. Some assumptions are made to simulate the results as discussed The different parameters used in proposed work are given in table 1. The performance parameters are analyzed using MATLAB 2016a.

Table 1 Names Number of Features and total number of samples of the selected databases

Name	Num Features	Num Samples
Iris	4	150
Cancer	9	699
Body fat	13	252
House	13	506

Table 1 gives database attributes

No

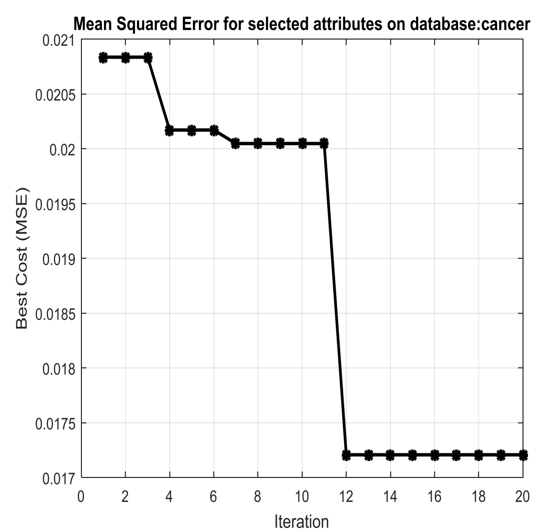

Fig 2 MSE for Selected attributes (num = 6) on cancer

DOI: 10.18535/ijsrm/v5i8.25

database

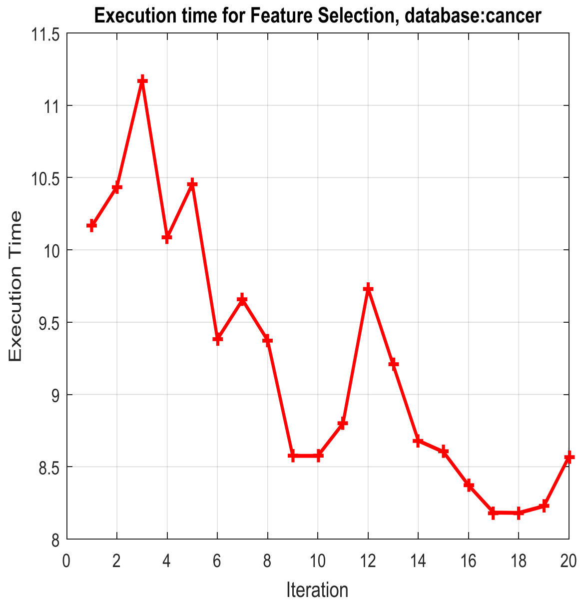

Fig 3 Execution time for Selected attributes (num = 6) on cancer database

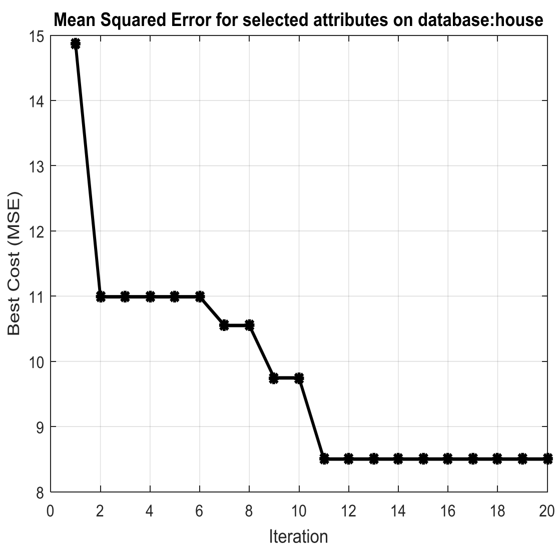

Fig 4 MSE for Selected attributes (num = 6) on house database

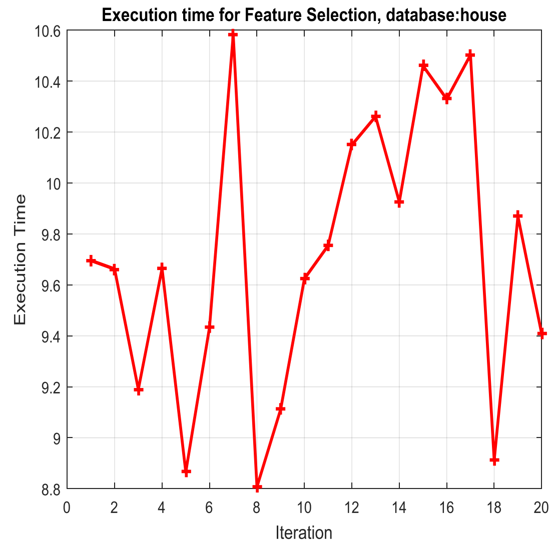

Fig 5 Execution time for Selected attributes (num = 6) on house database

Datasets	**No. of features**	**CSO Avg Accuracy**	**ICSO Avg Accuracy**	**Proposed ACO Avg Accuracy**
Bupa	6	82.6302	84.9664	99.045
Pima	8	81.3807	82.4248	99.097
Glass	9	89.7403	92.05633	99.073
Vowel	10	99.899	100	98.095
Wine	13	100	100	99.050
Heart	13	95.5556	97.03703	98.041
Australian	14	91.7391	92.9188	99.026
Vehicle	18	88.2927	90.89643	99.052
German	24	80.9	82.80003	98.099
Ionosphere	34	99.4286	100	99.009

DOI: 10.18535/ijsrm/v5i8.25

Table 2 Comparison of Average Accuracy of Proposed Technique with Cat Swarm optimization and its variants

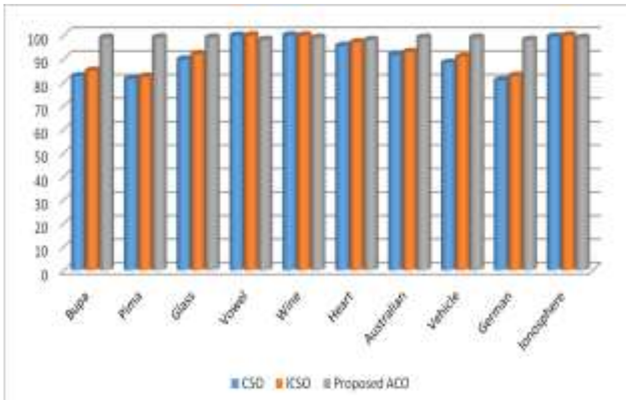

Fig 6 Comparison of proposed ACO with Cat Swarm optimization and its variants

IV. Conclusion

In machine learning and statistics, feature selection plays a very important role as a process for the selection of the relevant features from the pool of features for use in model construction and for the better classification and understanding of the data. There are many irrelevant features or attributes in the data that does not give us any information at all or gives very less information. So it is very much compulsory to have a subset of features that can classify describe the data at its best. In this work Feature subset selection algorithm is done. Later, we hybridized ACO and Neural Network Classifier to obtain their excellent features by synthesizing them. The Back propagation Neural Network is used as a classifier which is trained for minimum errors the Task of ACO is to Select Best Features for ANN such that the error is minimum and the number of features is also minimum. The results as compared in Table 2 show that proposed algorithm provides better performance in terms of Average Accuracy Rate as compared to CSO and ICSO. In future it is intended to analyze variable selection properties using the ACO procedure and establishing sufficient conditions required for successful recovery of the set of relevant variables. Work must also be focused towards reducing time complexity taken by FSS process in high-dimensional data using meta variables.

References:

1. Gert Van Dijck and Marc M. Van Hulle. "Speeding up the wrapper feature subset selection in regression by mutual information relevance and redundancy analysis." In Artificial Neural Networks–ICANN 2006, pp. 31-40. Springer Berlin Heidelberg, 2006.
2. Hong Zeng and Yiu-ming Cheung. "Feature selection for clustering on high dimensional data." In PRICAI 2008: Trends in Artificial Intelligence, pp. 913-922. Springer Berlin Heidelberg, 2008.
3. Zheng Zhao and Huan Liu. "Searching for interacting features in subset selection." Intelligent Data Analysis 13, no. 2 (2009): 207-228.
4. Daniela M. Witten and Robert Tibshirani. "A framework for feature selection in clustering." Journal of the American Statistical Association 105, no. 490 (2010).
5. Iffat A. Gheyas and Leslie S. Smith. "Feature subset selection in large dimensionality domains." Pattern recognition 43, no. 1 (2010): 5-13.
6. Deng Cai, Chiyuan Zhang, and Xiaofei He. "Unsupervised feature selection for multi-cluster data." In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 333-342. ACM, 2010.
7. John Q.Gan,Bashar Awwad Shiekh Hasan, and Chun Sing Louis Tsui. "A hybrid

approach to feature subset selection for brain-computer interface design." In Intelligent Data Engineering and Automated Learning-IDEAL 2011, pp. 279-286. Springer Berlin Heidelberg, 2011.

8. Cordeiro Ferreira, Robson Leonardo, Caetano Traina Junior, Agma Juci Machado Traina, Julio López, U. Kang, and Christos Faloutsos. "Clustering very large multi-dimensional datasets with map reduce." In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 690-698. ACM, 2011.
9. Manuel Ignacio Lopez, J. M. Luna, C. Romero, and S. Ventura. "Classification via Clustering for Predicting Final Marks Based on Student Participation in Forums." International Educational Data Mining Society (2012).
10. Houtao Deng and George Runger. "Feature selection via regularized trees." In Neural Networks (IJCNN), The 2012 International Joint Conference on, pp. 1-8. IEEE, 2012.
11. Quanquan Gu, Zhenhui Li, and Jiawei Han. "Generalized fisher score for feature selection." arXiv preprint arXiv:1202.3725 (2012).
12. Roland Schwarzkopf, Matthias Schmidt, Christian Strack, Simon Martin, and Bernd Freisleben. "Increasing virtual machine security in cloud environments." Journal of Cloud Computing 1, no. 1 (2012): 1-12.
13. Tao Chen, Nevin L. Zhang, Tengfei Liu, Kin Man Poon, and Yi Wang. "Model-based multidimensional clustering of categorical data." Artificial Intelligence 176, no. 1 (2012): 2246-2269.
14. Guorong Li, Qingming Huang, Junbiao Pang, Shuqiang Jiang, and Lei Qin. "Online selection of the best k-feature subset for object tracking." Journal of Visual Communication and Image Representation 23, no. 2 (2012): 254-263.
15. Qinbao Song, Jingjie Ni, and Guangtao Wang. "A fast clustering-based feature subset selection algorithm for high-dimensional data." Knowledge and Data Engineering, IEEE Transactions on 25, no. 1 (2013): 1-14.
16. A. Srikrishna, B. Eswara Reddy, and V. Sesha Srinivas. "Automatic Feature Subset Selection using Genetic Algorithm for Clustering." International Journal on Recent Trends in Engineering & Technology 9, no. 1 (2013).
17. Mladen Kolar and Han Liu. "Feature selection in high-dimensional classification." In Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 329-337. 2013.
18. Lin, Kuan-Cheng, et al. "Feature selection based on an improved cat swarm optimization algorithm for big data classification." The Journal of Supercomputing 72.8 (2016): 3210-3221.