Feature Selection Based On Ant Colony

Kritika, Ritika Mehra
[1]Research scholar, [2]Assistant Professor,
[1]Department of Computer Science Engineering
[1]RPIIT Technical Campus , Karnal , India

Abstract - **Feature selection involves identifying a subset of the most useful features that produces compatible results as the original entire set of features. A feature selection algorithm may be evaluated from both the efficiency and effectiveness points of view. While the efficiency concerns the time required to find a subset of features, the effectiveness is related to the quality of the subset of features. Based on these criteria, a clustering-based feature selection algorithm is proposed and experimentally evaluated in their work. Features are divided into clusters by using graph-theoretic clustering methods Most representative feature that is strongly related to target classes is selected from each cluster to form a subset of features.**

Index Terms - **CFS,SVM, RFE ,RMR**

I. 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. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing?"
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 a distinctive approach or attitude to data analysis. The emphasis is not so much on extracting facts, but on generating hypotheses. The aim is more to yield questions rather than answers. Insights gained by data mining can then be verified by conventional analysis.

II.FEATURE SELECTION TECHNIQUES

With regard to the relationship between a feature selection algorithm and the inductive learning method used to infer a model, three major approaches can be distinguished: –

- Filters, which rely on the general characteristics of training data and carry out the feature selection process as a pre-processing step with independence of the induction algorithm.
- Wrappers, which involve optimizing a predictor as a part of the selection process.
- Embedded methods, which perform feature selection in the process of training and are usually specific to given learning machines.

Uni-variate methods (such as Info Gain) are fast and scalable, but ignore feature dependencies. On the other hand, multivariate filters (such as CFS, INTERACT, etc.) model feature dependencies, but at the cost of being slower and less scalable than uni-variate techniques.
Besides this classification, feature selection methods can also be divided according to two approaches: individual evaluation and subset evaluation. Individual evaluation is also known as feature ranking and assesses individual features by assigning those weights according to their degrees of relevance. On the other hand, subset evaluation produces candidate feature subsets based on a certain search strategy. Each candidate subset is evaluated by a certain evaluation measure and compared with the previous best one with respect to this measure. While the individual evaluation is incapable of removing redundant features because redundant features are likely to have similar rankings, the subset evaluation approach can handle feature redundancy with feature relevance. However, methods in this framework can suffer from an inevitable problem caused by searching through feature subsets required in the subset generation step, and thus, both approaches will be studied in this research.
The feature selection methods included in this work are subsequently described according to how they combine the feature selection search with the construction of the classification model: filter methods. All of them are available in the Weka tool environment or implemented in Matlab. These feature selection methods belong to different families of techniques and conform an heterogeneous suite of methods to carry out a broad and complete study.

1.1 Filter methods

- *Correlation-based Feature Selection (CFS)* is a simple multivariate filter algorithm that ranks feature subsets according to a correlation-based heuristic evaluation function [9]. The bias of the evaluation function is toward subsets that contain

features that are highly correlated with the class and uncorrelated with each other. Irrelevant features should be ignored because they will have low correlation with the class. Redundant features should be screened out as they will be highly correlated with one or more of the remaining features. The acceptance of a feature will depend on the extent to which it predicts classes in areas of the instance space not already predicted by other features.

- *The Consistency-based Filter* evaluates the worth of a subset of features by the level of consistency in the class values when the training instances are projected onto the subset of attributes.
- *The INTERACT Algorithm* is a subset filter based on symmetrical uncertainty (SU) and the consistency contribution, which is an indicator about how significantly the elimination of a feature will affect consistency. The algorithm consists of two major parts. In the first part, the features are ranked in descending order based on their SU values. In the second part, features are evaluated one by one starting from the end of the ranked feature list. If the consistency contribution of a feature is less than an established threshold, the feature is removed, otherwise it is selected. The authors stated that this method can handle feature interaction, and efficiently selects relevant features.
- *Information Gain* is one of the most common attribute evaluation methods. This uni-variate filter provides an ordered ranking of all the features, and then a threshold is required. In this work the threshold will be set up selecting the features which obtain a positive information gain value.
- *ReliefF* is an extension of the original Relief algorithm. The original Relief works by randomly sampling an instance from the data and then locating its nearest neighbor from the same and opposite class. The values of the attributes of the nearest neighbors are compared to the sampled instance and used to update relevance scores for each attribute. The rationale is that an useful attribute should differentiate between instances from different classes and have the same value for instances from the same class. ReliefF adds the ability of dealing with multiclass problems and is also more robust and capable of dealing with incomplete and noisy data. This method may be applied in all situations, has low bias, includes interaction among features and may capture local dependencies which other methods miss.
- *The mRMR* (minimum Redundancy Maximum Relevance) method selects features that have the highest relevance with the target class and are also minimally redundant, i.e., selects features that are maximally dissimilar to each other. Both optimization criteria (Maximum Relevance and Minimum Redundancy) are based on mutual information.
- *The Mdfilter* is an extension of mRMR which uses a measure of monotone dependence (instead of mutual information) to assess relevance and irrelevance. One of its contributions is the inclusion of a free parameter (λ) that controls the relative emphasis given on relevance and redundancy. In this work, two values of lambda will be tested: 0 and 1. When(λ) is equal to zero, the effect of the redundancy disappears and the measure is based only on maximizing the relevance. On the other hand, when (λ) is equal to one, it is more important to minimize the redundancy among variables. These two values of (λ) were chosen because we are interested in checking the performance of the method when the effect of the redundancy disappears. Also, the authors state that λ =1 performs better than other λ values.

1.2 Embedded Methods

- *SVM-RFE* (Recursive Feature Elimination for Support Vector Machines) method performs feature selection by iteratively training a SVM classifier with the current set of features and removing the least important feature indicated by the SVM. Two versions of these methods will be tested: the original one, using a linear kernel and an extension using a nonlinear kernel in order to solve more complex problems.
- *FS-P* (Feature Selection- Perceptron) is an embedded method based on a perceptron. A perceptron is a type of artificial neural network that can be seen as the simplest kind of feed forward neural network: a linear classifier. The basic idea of this method consists on training a perceptron in the context of supervised learning. The interconnection weights are used as indicators of which features could be the most relevant and provide a ranking.

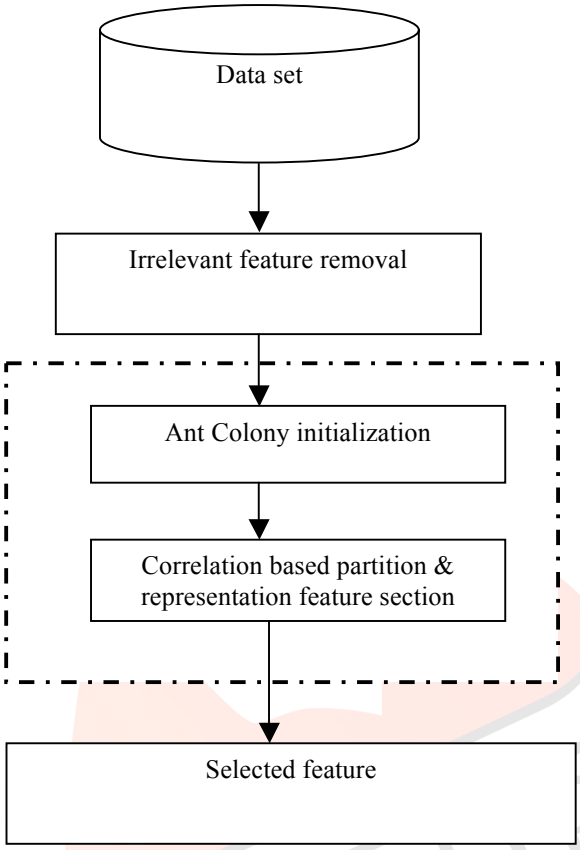
III. Related Work

Zheng Zhao et al., 2009 [5] 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. Feature (attribute) interaction presents a challenge to feature subset selection for classification. This is because a feature by itself might have little correlation with the target concept, but when it is combined with some other features; they can be strongly correlated with the target concept. Thus, the unintentional removal of these features may result in poor classification performance. It is computationally intractable to handle feature interactions in general. Daniela M. Witten et al., 2010 [6] 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. They propose a novel framework for sparse clustering, in which one clusters the observations using an adaptively chosen subset of the features. The method uses a lasso-type penalty to select the features. They use this framework to develop simple methods for sparse K-means and sparse hierarchical clustering. A single criterion governs both the selection of the features and the resulting clusters. These approaches are demonstrated on simulated

data and on genomic data sets. Iffat A. Deng Cai et al., 2010 [8] 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. In this paper, they consider the feature selection problem in unsupervised learning scenario, which is particularly difficult due to the absence of class labels that would guide the search for relevant information

IV.PROPOSED SYSTEM

Feature selection or variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features.



Redundant features are those which provide no more information than the currently selected features, and irrelevant features provide no useful information in any context. Feature selection techniques are a subset of the more general field of feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features.,

- Selection of Multi-dimensional Data for study, Open Source Multi-dimensional data sets will be used for the implementation.
- Data Pre-Processing and Noise and Irrelevant feature removal from the data sets.
- Implementation of Local Search based Ant Colony Optimization (LS-ACO) for feature Selection with ability to data partition and feature selection.

Extensive experiments will be carried out to compare the proposed algorithm with existing representative feature selection algorithms, including, *Fast Correlation-Based Filter* (FCBF), Relief, *CFS*. (Correlation based *Feature Selection*), with respect to various types of well-known classifiers, namely, the probability-based Naive Bayes, the tree-based, the instance-based, and the rule-based Classifiers before and after feature selection application.

V.RESULT ANALYSIS

5.1 Selected Databases

Name	Num Features	Num Samples
Iris	4	150
Cancer	9	699
Bodyfat	13	252

House	13	506

Table 5.1 Selected Dataset for the Evaluation of Proposed Scheme

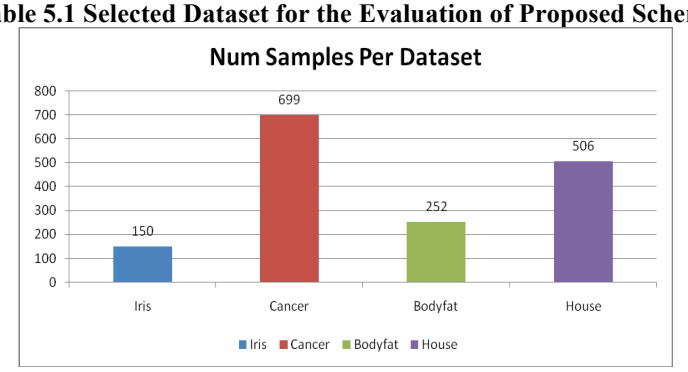


Figure. 5.1 Number of Samples Per Dataset.

5.1.1 *Iris Dataset*

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

5.1.2 *Cancer Dataset*

This is one of three domains provided by the Oncology Institute that has repeatedly appeared in the machine learning literature. (See also lymphography and primary-tumor.)

This data set includes 201 instances of one class and 85 instances of another class. The instances are described by 9 attributes, some of which are linear and some are nominal. Attribute Information:

1. Class: no-recurrence-events, recurrence-events
2. age: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99.
3. menopause: lt40, ge40, premeno.
4. tumor-size: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59.
5. inv-nodes: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39.
6. node-caps: yes, no.
7. deg-malig: 1, 2, 3.
8. breast: left, right.
9. breast-quad: left-up, left-low, right-up, right-low, central.
10. irradiat: yes, no.

5.1.3 *Bupa Dataset*

The first 5 variables are all blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption. Each line in the dataset constitutes the record of a single male individual.

Important note: The 7th field (selector) has been widely misinterpreted in the past as a dependent variable representing presence or absence of a liver disorder. This is incorrect [1]. The 7th field was created by BUPA researchers as a train/test selector. It is not suitable as a dependent variable for classification. The dataset does not contain any variable representing presence or absence of a liver disorder. Researchers who wish to use this dataset as a classification benchmark should follow the method used in experiments by the donor (Forsyth & Rada, 1986, Machine learning: applications in expert systems and information retrieval) and others (e.g. Turney, 1995, Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm), who used the 6th field (drinks), after dichotomising, as a dependent variable for classification. Because of widespread misinterpretation in the past, researchers should take care to state their method clearly. Attribute Information:

- mcv mean corpuscular volume
- alkphos alkaline phosphotase
- sgpt alanine aminotransferase
- sgot aspartate aminotransferase
- gammagt gamma-glutamyl transpeptidase

- drinks number of half-pint equivalents of alcoholic beverages drunk per day
- selector field created by the BUPA researchers to split the data into train/test sets

5.1.4 *Body Fat Data Set*

Lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. This data set can be used to illustrate multiple regression techniques. Accurate measurement of body fat is inconvenient/costly and it is desirable to have easy methods of estimating body fat that are not inconvenient/costly.

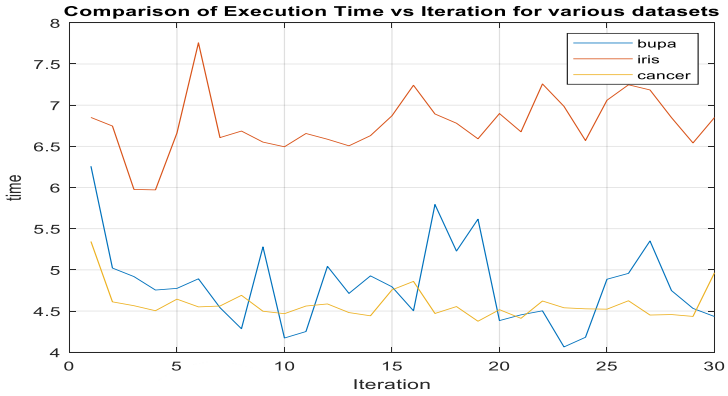


Figure 5.2 Comparison of Execution Time vs Iteration on Bupa, Cancer and Iris Datasets.

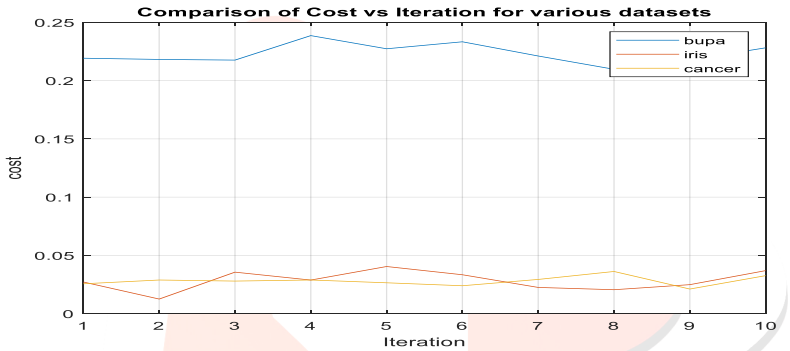


Figure. 5.3 Comparison of Cost vs Iteration on Bupa, Cancer and Iris Datasets

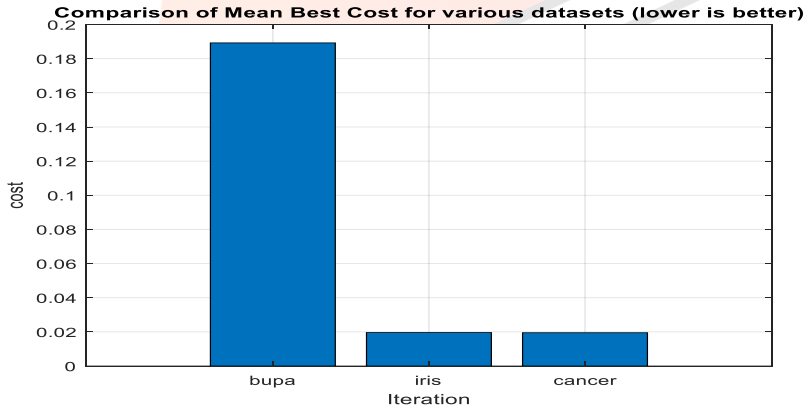


Figure 5.4 Comparison of Mean Best Cost for various datasets (lower is better)

Datasets	Number of features	Reduced Features	Existing Work	Proposed ACO
Bupa	6	3	84.9664	99.045
Bodyfat	13	7	97.03703	98.041
Iris	4	2	97.33	99.979467
House	6	3	93.56	99.97
Cancer	10	5	94.70	98.65

Table 5.2 Comparison of Proposed work with existing optimization With Accuracy and Reduced No of Features

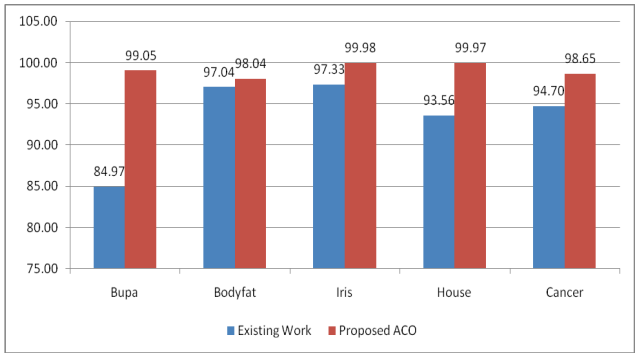


Figure 5.5 Comparison of Accuracy achieved after Application of ACO FSS

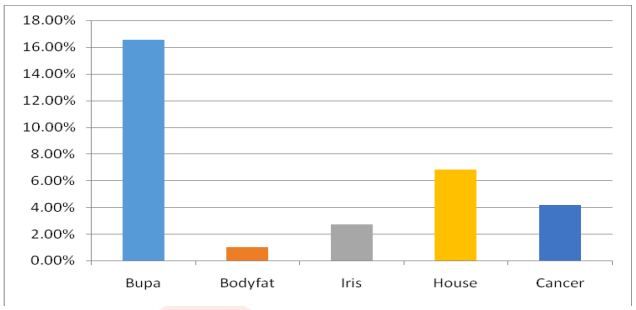


Figure 5.6 Improvement in Accuracy achieved after Application of ACO FSS

VI. Conclusion and Future Scope

Feature subset selection is the process of selecting a subset of relevant feature for the construction of a model or better classification and description of the data. The core concept of the feature subset selection technique is that the raw data we are using has many unappropriate, redundant and irrelevant features. Redundant features are those features those provide no information as compared to already selected feature and irrelevant feature can be considered as a feature of no use in context of the information. Feature selection techniques are the subset of the Feature extraction field. Feature extraction makes new features from the attribute set or the original features, whereas feature selection returns a subset of the features. In this work, we have address the problem of feature selection using Ant Colony optimization approach, however variable selection in high-dimensional feature space is not yet tackled. The problem of reliable variable selection in high-dimensional is important in many scientific areas where simple models are needed to provide insights into complex systems. Existing research has focused primarily on establishing results for prediction consistency, ignoring feature selection. In future we will we bridge this gap, by analyzing variable selection properties of the using thr ACO procedure and establishing sufficient conditions required for successful recovery of the set of relevant variables. This analysis can be complemented by analyzing the information theoretic limits, which provide necessary conditions for variable selection in discriminated analysis.

References

[1] Brand, Matthew. "Charting a manifold." In *Advances in neural information processing systems*, pp. 961-968. 2002.

[2] DeMers, David, and Garrison Cottrell. "Non-linear dimensionality reduction."*Advances in neural information processing systems* (1993): 580-580.

[3] 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.

[4] 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.

[5] Zheng Zhao and Huan Liu. "Searching for interacting features in subset selection." Intelligent Data Analysis 13, no. 2 (2009): 207-228.

[6] Daniela M. Witten and Robert Tibshirani. "A framework for feature selection in clustering." Journal of the American Statistical Association 105, no. 490 (2010).

[7] Iffat A. Gheyas and Leslie S. Smith. "Feature subset selection in large dimensionality domains." Pattern recognition 43, no. 1 (2010): 5-13.

[8] 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.

[9] Sheng Ding. "Feature selection based F-score and ACO algorithm in support vector machine." In Knowledge Acquisition and Modeling, 2009. KAM'09. Second International Symposium on, vol. 1, pp. 19-23. IEEE, 2009.

[10]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 mapreduce." In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 690-698. ACM, 2011.