Feature Subset Selection Problem using Wrapper Approach in Supervised Learning

Asha Gowda Karegowda
Dept. of Master of Computer Applications
Siddaganga Institute of Technology
Tumkur, Karnataka,India

M.A.Jayaram
Dept. of Master of Computer Applications
Siddaganga Institute of Technology
Tumkur, Karnataka,India

A.S. Manjunath
Dept. of Computer Science & Engineering
Siddaganga Institute of Technology Tumkur, Karnataka, India

ABSTRACT
Feature subset selection is of immense importance in the field of data mining. The increased dimensionality of data makes testing and training of general classification method difficult. Mining on the reduced set of attributes reduces computation time and also helps to make the patterns easier to understand. In this paper a wrapper approach for feature selection is proposed. As a part of feature selection step we used wrapper approach with Genetic algorithm as random search technique for subset generation ,wrapped with different classifiers/ induction algorithm namely decision tree C4.5, NaïveBayes, Bayes networks and Radial basis function as subset evaluating mechanism on four standard datasets namely Pima Indians Diabetes Dataset, Breast Cancer, Heart Stat log and Wisconsin Breast Cancer. Further the relevant attributes identified by proposed wrapper are validated using classifiers. Experimental results illustrate, employing feature subset selection using proposed wrapper approach has enhanced classification accuracy.

General Terms
Experimentation, Verification.

Keywords
Feature Selection, filters, wrappers.

1.INTRODUCTION
Huge data repositories, especially in medical domains, contain enormous amounts of data. These data includes also currently unknown and potentially interesting patterns and relations, which can be uncovered using knowledge discovery and data mining methods. Medical data mining has enormous potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized for clinical diagnosis. Data preprocessing is a significant step in the knowledge discovery process, since quality decisions must be based on quality data. Data preprocessing includes data cleaning, data integration, data transformation and data reduction [4]. These data processing techniques, when applied prior to mining, can substantially improve the overall quality of the patterns mined and/or the time required for the actual mining. The goal of data reduction is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Mining on the reduced set of attributes has additional benefits. It reduces the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand. Further it enhances the classification accuracy and learning runtime. Section 2 briefs about the filter and wrapper approach for feature selection. The proposed wrapper approach with GA used for random search, wrapped with four different classifiers as evaluators is described in section 3 followed by results and conclusion in section 4 and 5 respectively.

2. FEATURE SELECTION
Feature selection is a process that selects a subset of original features. Feature selection is one of the important and frequently used techniques in data preprocessing for data mining. In real-world situations, relevant features are often unknown a priori. Hence feature selection is a must to identify and remove are irrelevant/redundant features. It can be applied in both unsupervised and supervised learning.

The goal of feature selection for unsupervised learning is to find the smallest feature subset that best uncovers clusters form data according to the preferred criterion [5].Feature selection in unsupervised learning is much harder problem, due to the absence of class labels. Feature election for clustering is the task of selecting important features for the underlying clusters [8]. Feature selection for unsupervised learning can be subdivided in filter methods and wrapper methods. Filter methods in unsupervised learning is defined as using some intrinsic property of the data to select feature without utilizing the clustering algorithm[5]. Entropy measure has been used as filter method for feature selection for clustering [9].Wrapper approaches in unsupervised learning apply unsupervised learning algorithm to each candidate feature subset and then evaluate the feature subset by criterion functions that utilize the clustering result[5]. Volker Roth and Tilman Lange proposes a wrapper method where Gaussian mixture model combines a clustering method with a Bayesian inference mechanism for automatically selecting relevant features[12].

In supervised learning, feature selection aims to maximize classification accuracy [10]. It is easier to select features for classification/supervised learning than for clustering, since the classification uses class label information. Though domain experts can eliminate few of the irreverent attributes, selecting the best subset of features usually requires a systematic approach. Feature selection method generally consists of four steps described below [9].

(a) Generate candidate subset: The original feature set contains n number of features, the total number of competing candidate subsets to be generated is 2^n , which is a huge number even for medium-sized n. Subset generation is a search procedure that produces candidate feature subsets for evaluation based on a certain search strategy. The search strategy is broadly classified as complete (eg. Breadth first search, Branch & bound, beam search, best first), heuristic (forward selection, backward selection, forward and backward selection), and random search (Las Vegas algorithm (LVW), genetic algorithm (GA), Random generation plus sequential selection (RGSS), simulated annealing (SA)).

(b) Subset evaluation function to evaluate the subset generated in the previous step (generate candidate subset) by using filter or wrapper approach. Filter and Wrapper approach differ only in the way in which they evaluate a subset of features. The filter approach is independent of the learning induction algorithm. Wrapper strategies for feature selection use an induction algorithm to estimate the merit of feature subsets. Wrappers often achieve better results than filters due to the fact that they are tuned to the specific interaction between an induction algorithm and its training data. (Filter and wrappers are described in section 2).

(c) Stopping Condition: Since the number of subsets can be enormous , some sort of stopping criterion is necessary. Stopping criteria may be based on a generation procedure/ evaluation function. Stopping criteria based on generation procedure include:

- Whether a predefined number of features are selected
- Whether a predefined number of iterations reached.

Stopping criteria based on an evaluation function can be:

- Whether addition (or deletion) of any feature does not produce a better subset
- Whether an optimal subset according to some evaluation function is obtained.

(d) Validation procedure to check whether the feature subset selected is valid. Usually the result of original feature set is compared with the feature selected by filters/wrappers as input to some induction algorithm using artificial/real-world datasets. Another approach for validation is to use different feature selection algorithm to obtain relevant features and then compare the results by using classifiers on each relevant attribute subset.

The above four steps are shown in the figure 1.

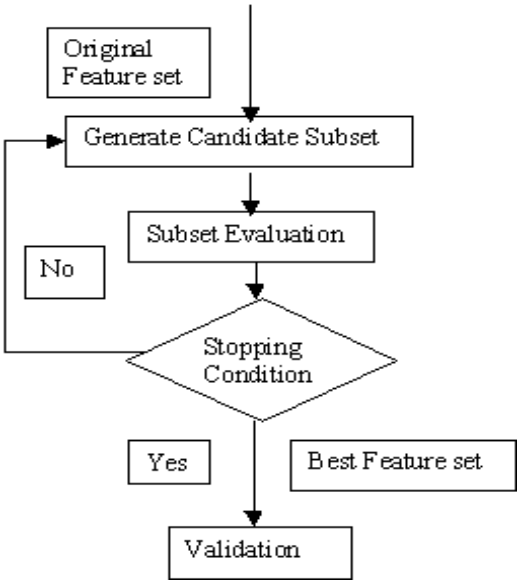

Figure 1. Steps for feature selection

2.1 The Filter Approach for Feature Selection

The filter approach actually precedes the actual classification process. The filter approach is independent of the learning induction algorithm [figure 2], computationally simple fast and scalable. Using filter method, feature selection is done once and then can be provided as input to different classifiers. Various feature ranking and feature selection techniques have been proposed such as Correlation-based Feature Selection (CFS), Principal Component Analysis (PCA), Gain Ratio (GR) attribute evaluation, Chi-square Feature Evaluation, Fast Correlation-based Feature selection (FCBF), Information gain, Euclidean distance, i-test, Markov blanket filter. Some of these filter methods do not perform feature selection but only feature ranking hence they are combined with search method when one needs to find out the appropriate number of attributes. Such filters are often used with forward selection, backward elimination, bi-directional search, best-first search, genetic search and other methods [7,11,13]. The authors have used decision tree as filter approach to provide the relevant features as input to neural network classifier [6]. Further Correlation based feature selection has been used in a cascaded fashion with GA as filter to provide relevant inputs to neural networks classifier [1].

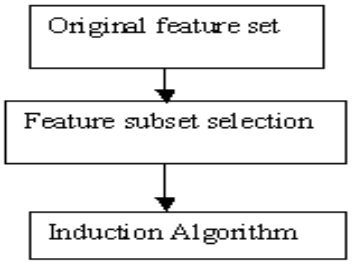

Figure 2. Filter approach for feature selection

2.2 The Wrapper Approach for Feature Selection

Wrapper model approach uses the method of classification itself to measure the importance of features set; hence the feature selected depends on the classifier model used. Wrapper methods generally result in better performance than filter methods because the feature selection process is optimized for the classification algorithm to be used. However, wrapper methods are too expensive for large dimensional database in terms of computational complexity and time since each feature set considered must be evaluated with the classifier algorithm used. [9,11,13]

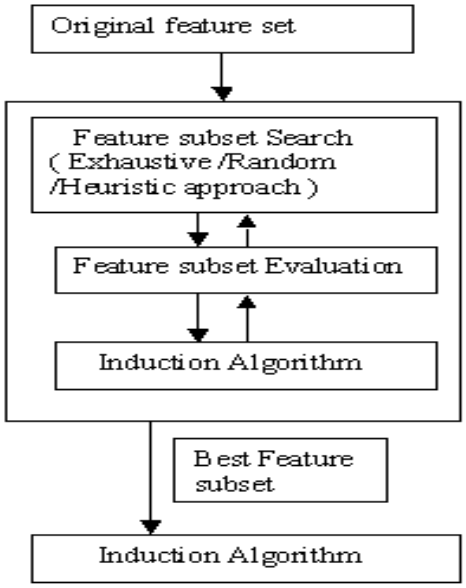

Figure 3. Wrapper approach for feature selection

3. PROPOSED WRAPPER METHOD

As a part of first step of feature selection, a random selection approach namely Genetic Algorithm (GA) has been used. GA[2] is a random search method, Capable of effectively exploring large search spaces, which is usually required in case of attribute selection. Further,

unlike many search algorithms, which perform a local, greedy search, GAs performs a global search. A genetic algorithm (GA) is a search algorithm inspired by the principle of natural selection. The basic idea is to evolve a population of individuals, where each individual is a candidate solution to a given problem.

A genetic algorithm mainly composed of three operators: reproduction, crossover, and mutation. Reproduction selects good string; crossover combines good strings to try to generate better offspring's; mutation alters a string locally to attempt to create a better string. In each generation, the population is evaluated and tested for termination of the algorithm. If the termination criterion is not satisfied, the population is operated upon by the three GA operators and then re-evaluated. This procedure is continued until the termination criterion is met. The working of proposed wrapper method is shown in figure 4. In this paper WEKA [3] GA is used as random search method with four different classifiers namely decision tree (DT) C4.5, NaïveBayes, Bayes networks and Radial basis function as induction method wrapped with GA. Further the relevant attributes identified by proposed wrapper is validated by different classifiers.

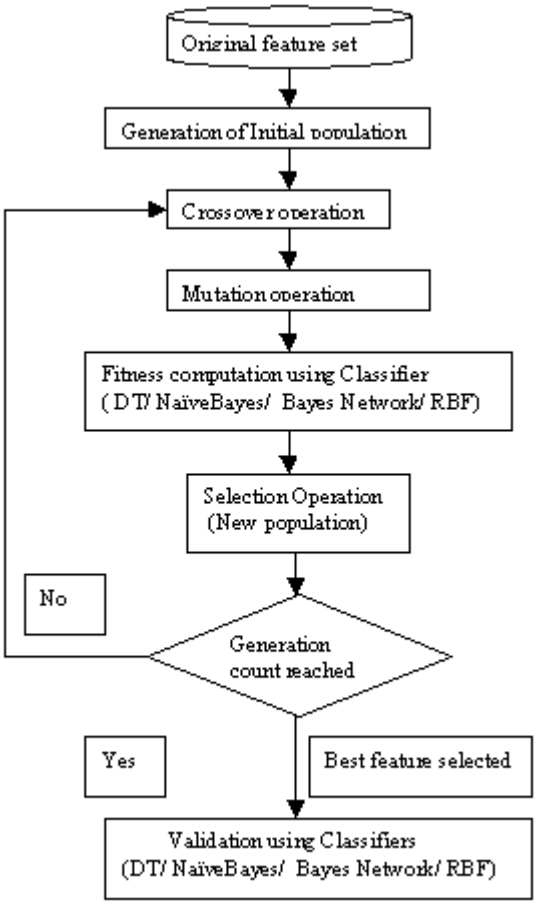

Figure 4. Proposed wrapper method

4.RESULTS

As a part of feature selection step we used wrapper approach with Genetic algorithm as random search technique wrapped with different classifiers/ induction algorithm namely decision tree C4.5, NaïveBayes, Bayes networks and Radial basis function as subset evaluating mechanism on four standard datasets namely Pima Indians Diabetes Dataset, Breast Cancer, Heart Statlog and Wisconsin Breast Cancer.

For GA, population size is 20, number of generation is 20 as terminating condition, crossover rate is 0.6 and mutation rate is 0.033. Table 1 to table 4 shows the reduced relevant attributes identified by different wrappers: GA+DT, GA+NaiveBayes, GA+ Bayesian and GA+RBF for four standard datasets and improved classification accuracy of different classifiers in validation step. Validation was done using four classifiers namely C4.5, RBF,

NaïveBayes& Bayesian classfiers (using 66 % percent training data and 36 % of test data)on four different dataset.

The results clears depicts the relevant attributes as identified by the various wrapper have indeed improved classification accuracy of the all the four classifiers used for validation when compared to classification accuracy with all the inputs. In few cases namely with (i) GA+NB as wrapper and DT as classifier in validations step for Breast cancer data set, (ii) GA+DT as wrapper and RBF as classifier for Heart statlog dataset, and (iii) GA+RBF as wrapper and NaïveBayes as classifier, the classification accuracy was reduced marginally. But most of the cases, experimental results show employing feature subset selection enhanced the classification accuracy.

The tabular results show that no one wrappers among the four wrappers experimented is best for all the datasets experimented. It was also found that the GA+DT wrapper always resulted in least number of relevant attributes for the all the datasets experimented except for Breast cancer dataset.

For PIMA dataset the relevant inputs identified by wrapper GA+NaiveBayes gave the best accuracy of 86.47% with NaïveBayes as classifier in validation step. Further it was observed that the relevant attributes identified by wrapper GA+RBF proved be best for RBF as classifier in validation step and not much improvement in accuracy for the remaining three classifiers in validation step. For Wisconsin Breast Cancer Database, the relevant inputs identified by wrapper GA+Naïve Bayes gave the best accuracy of 97.06 with Bayesian classifier in validation step. For Heart statlog dataset, the best accuracy of 85.86% using GA+RBF as wrapper and RBF / NaiveBayes as classifier in validation step. The same classification accuracy was also achieved using GA+NaiveBayes wrapper with Naïve Bayes as classifier in validation step. With Breast cancer dataset, it was found that the wrapper namely GA+DT and GA+RBF resulted in constant improved classification accuracy of 76.29% for the all the four classifiers in the validation step.

3.CONCLUTIONS

We have described the feature subset selection problem using wrapper approach in supervised learning. The experimented wrapper method used Genetic algorithm as random search technique wrapped with different classifiers/ induction algorithm namely decision tree C4.5, NaïveBayes, Bayes networks and Radial basis function as subset evaluating mechanism. Relevant attributes identified by different wrappers were compared using different classifiers in validation step. The results prove that there is no one standard wrapper approach, which is best for different datasets, however experiment results show that employing feature subset selection, surely enhances the classification accuracy.

4. REFERENCES

[1] Asha Gowda Karegowda and M.A. Jayaram, March 6-7, 2009. Cascading GA & CFS for Feature Subset Selection in Medical Data Mining. International Conference on IEEE International Advance Computing Conference (IACC'09), Thapar University, Patiala, Punjab India.

[2] D. Goldberg .1989. Genetic Algorithms in Search, Optimization, and Machine learning, Addison Wesley,

[3] I. H. Witten, E. Frank. 2005. Data Mining: Practical machine learning tools and techniques. 2nd Edition, Morgan Kaufmann, San Francisco.

[4] J. Han And M. Kamber. 2001. Data Mining: Concepts and Techniques. San Francisco, Morgan Kauffmann Publishers.

[5] Jennifer G. Dy. 2004. Feature Selection for Unsupervised Learning, Journal of Machine Learning, pp845-889.

[6] M.A.Jayaram, Asha Gowda Karegowda.2007. Integrating Decision Tree and ANN for Categorization of Diabetics Data. International Conference on Computer Aided Engineering, December 13-15, 2007, IIT Madras, Chennai, India.

[7] Mark A. Hall ,Correlation-based Feature Selection for Machine Learning, Dept of Computer science, University of Waikato .http://www.cs.waikato.ac.nz/ mhall/thesis.pdf

[8] Manoranjan Dash, Kiseiok Choi, Petr Scheuermann, Huan Liu. 2002. Feature Selection for Clustering – a Filter Solution. In Proceedings of the Second International Conference on Data Mining.

[9] M. Dash 1, H. Liu2. March 1997. Feature Selection for Classification, Intelligent Data Analysis 1 (131–156, www.elsevier.com/locate/ida]

[10] Ron Kohavi, George H. John.1997. Wrappers for feature subset Selection, *Artificial Intelligence*, Vol. 97, No. 1-2. pp. 273-324.

[11] Shyamala Doraisamy ,Shahram Golzari ,Noris Mohd. Norowi, Md. Nasir B Sulaiman , Nur Izura Udzir. 2008. A Study on Feature Selection and ClassificationTechniques for Automatic Genre Classification of Traditional Malay Music. ismir2008.ismir.net/papers/ISMIR2008 256.pdf(2008).

[12] Volfer Rotz, and Tilman Lange. 2003. Feature Selection in Clustering Problems", In Advances in Neural Information Processing Systems 16.

[13] Y.Saeys, I.Inza, and P. LarrANNaga,. 2007. A review of feature selection techniquesin bioinformatics, Bioinformatics, 23(19),, pp.2507-2517.

Table 1 Classification accuracy using wrapper feature selection approach for PIMA dataset

Wrapper Approach for Attribute selection Method	Number of Attributes	Classifiers Accuracy (%)			
		Decision Tree C4.5	Naïve Bayes	Bayesian classifier	RBF
GA+ Naïve Bayes	3	85.71	86.47	85.71	82.71
GA+ Bayesian	3	85.71	82.71	83.54	81.95
GA+ Decision Tree C4.5	2	85.71	84.21	85.71	81.95
GA+RBF	4	82.71	81.95	81.203	85.72
With all inputs	8	82.71	79.70	82.71	81.20

Table 2 Classification accuracy using wrapper feature selection approach for Heart Statlog dataset Breast Cancer dataset

Wrapper Approach for Attribute selection	Number of Attributes	Classifiers Accuracy (%)			
		DT	Naïve Bayes	Bayesian classifier	RBF
GA+ Naïve Bayes	11	76.09	85.87	82.61	83.70
GA+ Bayesian	5	84.78	83.70	84.78	83.70
GA+ Decision Tree C4.5	3	84.78	83.70	83.70	80.43
GA+RBF	11	76.09	85.869	82.61	85.86
With all inputs	13	76.09	83.70	82.61	82.61

Table 3 Classification accuracy using wrapper feature selection approach for Breast Cancer dataset

Wrapper Approach for Attribute selection	Number of Attributes	Classifiers Accuracy (%)			
		DT	Naïve Bayes	Bayesian classifier	RBF
GA+ Naïve Bayes	4	65.98	72.16	72.16	74.23
GA+ Bayesian	5	75.26	72.16	72.16	69.07
GA+ Decision Tree C4.5	3	76.29	76.29	76.29	76.29
GA+RBF	2	76.29	76.29	76.29	76.29
With all inputs	9	68.04	71.13	70.10	68.04

Table 4 Classification accuracy using wrapper feature selection approach for Wisconsin Breast Cancer dataset

Wrapper approach for Attribute selection	Number of Attributes	Classifiers Accuracy (%)			
		Decision Tree C4.5	Naïve Bayes	Bayesian classifier	RBF
GA+ Naïve Bayes	6	96.09	96.22	97.06	96.64
GA+ Bayesian	8	95.38	95.38	96.22	95.80
GA+ Decision Tree C4.5	3	95.38	94.96	95.38	94.54
GA+RBF	4	95.38	96.29	96.64	96.64
With all inputs	9	95.38	94.96	96.21	95.80