International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 23 (2015) pp 43426-43429

A Comparative Study on Feature Selection Algorithms for Improving Classification Performance

K. Sutha

Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India suthas@rediffmail. com

Dr. J. Jebamalar Tamilselvi

Director, Department of MCA, Jaya Engineering College, Chennai, Tamil Nadu, India jjebamalar@gmail. com

Abstract
Database with high dimensionality, affects the effectiveness of mining algorithms. Feature Selection is one of the data pre-processing steps in data mining process, reduces dimensionality and improves the performance of learning algorithms. Feature Selection selects the most relevant features, removes irrelevant and redundant attributes. Researchers developed various feature selection algorithms, based on different method for reducing the original feature space. But those algorithms suffer due to drastically increasing dimensionality. In this article, we perform experimentation on five data sets with some popular feature selection algorithms SFS, SBE, FCBF, GA, mRmR, Relief and FCBF+SFS. The result shows that SFS outperforms the other algorithms in improving classification accuracy.

Keywords-feature selection, Data mining, classification; filter, wrapper, hybrid.

Introduction
Data sets containing hundreds and thousands of features (variables or attributes) are said to be high dimensional. High dimensional data contain irrelevant, noise and weakly informative features along with the most useful (relevant) features for learning process. Data pre-processing is the most essential step to be followed, before applying any data mining tasks. Data cleaning, Data Integration, Data Transformation and Data Reduction are the categories of data pre-processing techniques. Feature Selection or subset selection is a Data Reduction technique. It determines the optimal feature subset with minimal number of most relevant features.
Classification [1] is one of the most useful data mining tasks, used for describing the data classes from the dataset. It involves in building a model or classifier by analyzing the training data, Model accuracy is then estimated using the testing data. Presence of irrelevant and redundant features, affects the accuracy of classification algorithms [2]. Feature Selection, as a pre-processing step plays an important role in improving the classification accuracy and efficiency [3]. Feature Selection method has four basic steps, subset generation, subset evaluation, stopping criterion and validation [4].
On the basis of evaluation approach, used in subset evaluation step, the feature selection methods are broadly classified into filter, wrapper and hybrid methods [5]. Filter method evaluates the goodness of feature subset, by exploiting the inherent characteristics of the training data, independent of any learning algorithm [6]. Advantages of Filter methods are its high computational efficiency and generality. But it does not always produce satisfactory result. Wrapper method uses the predictive accuracy of a learning algorithm, as an evaluation measure. This method guarantee better result, but it is more computationally expensive when it is applied to large data set [7]. Due to this reason, wrapper method is not usually preferred. Hybrid method combines the filter and wrapper methods[8][17][18][19][20]. It takes the advantages of both filter and wrapper methods. In this method, filter method is first applied to the original data set to get the reduced feature subset. Wrapper method is then applied to the reduced feature subset to obtain the optimal feature subset [9].
This paper compares the following feature selection algorithms: Sequential Forward Selection (SFS), Sequential Backward Elimination (SBE), Genetic Algorithm, Relief, Fast Correlation-Based Feature Selection (FCBF), and minimum Redundancy maximum Relevance (mRmR).
In the remainder of this paper, Section 1 discusses the existing feature selection algorithms, Section 2 presents the description about dataset chosen for experimentation, Experimental results are discussed in the Section 3, and section 4 provides conclusion of this work.

A. Feature Selection Algorithms
i. Relief
Relief [10] is a well known feature selection algorithm which depends on Relevance Evaluation. A feature is selected if its weight of relevance is greater than a threshold value, where the relevance of feature is estimated on the basis of how well their values distinguish between the instances of the same and different classes that are nearer to each other.
Relief Algorithm is very scalable to data set with increasing dimensionality. It cannot eliminate the redundant features. It selects the features if they are relevant to the class concept even though they are highly correlated to each other. Redundant features affect the speed and accuracy of learning algorithm (Hall 2000, Kohavi & John 1997).

ii. Fast Correlation Based Filter (FCBF)
FSBF[11] uses predominant correlation as a goodness measure, which is based on symmetric uncertainty(SU).

$$SU(X, Y) = 2\left[\frac{IG(X|Y)}{H(X)+H(Y)}\right]$$

Information Gain, IG(X|Y) = H(X) – H(X|Y)
Information Gain[12] measures the dependence between the feature and the target class. If X represents a feature and Y represents the class labels, IG is calculated as,
IG(X|Y) = H(X) – H(X|Y)
H(X) is the entropy of X and H(X|Y) is the entropy of X after observing Y. Entropy(H) measures the uncertainty associated with a random variable.
Entropy, H(X) = $-\sum_{x\in X} p(x)log_2(p(x))$
H(X|Y) = $-\sum_{y\in Y} p(y)\sum_{x\in X} p(x|y)log_2(p(x|y))$
Features with highest IG (maximum value is 1), is relevant to the target class. According to FCBF, a feature F_i is selected if it is predominant in predicting the target class(C). A feature F_i is predominant iff $SU_{i,c} \geq$ □ and $\forall F_j \in S'(j \neq i)$, there exists no F_j such that the $SU_{j,i} \geq SU_{i,c}$
It drastically reduces the dimensionality of high dimensional data set and improves the classification accuracy.

iii. Genetic Algorithm

Genetic Algorithm is a global search technique [12][16], based on natural evolutionary process. Selection, Crossover and mutation are the three operators of GA. The idea behind the GA is the Darwin's theory of survival of the fittest [13]. A population consists of a number of individuals or candidates. Each individual is characterized by its chromosome (a string of symbols). Fitness function is applied to evaluate the fitness of the individuals. The selection operator uses the fitness function to select a good string (an individual with high fitness), for breeding a new generation or offspring. The Crossover operator generates better offspring by combining the good strings. The Mutation operator alters one or more components of a selected string to retain genetic diversity from one generation to the next. The population is then evaluated in each generation for termination of the algorithm. The three operators of GA are applied repeatedly and re-evaluated, until the termination condition is met.

iv. Minimum Redundancy Maximum Relevance

Minimum Redundancy Maximum Relevance [14] uses mutual information as a measure of relevance. The maximum mutually dissimilar features are selected to achieve minimum redundancy. The minimum redundancy condition is given by **Min W_I, $W_I = \frac{1}{|S|^2}\sum_{i,j\in S} I(i,j)$.** The Maximum relevance condition is given by **Max V_I, $V_I = \frac{1}{|S|}\sum_{i\in S} I(h,i)$,** where I(x, y) is the mutual information of variables x and y, **I(x, y)=$\sum_{i,j} p(x_i,y_j)\log\frac{p(x_{i,}y_j)}{p(x_{i)}p(y_{j)}}$**
MRMR improves the class prediction accuracy.

v. Sequential Forward Selection

Sequential Forward Selection [15] is the greedy search algorithm. It starts with an empty set, features which increases the value of objective function are sequentially added to the feature set. SFS produces better results when there are a small number of features in the optimal feature subset. It uses the following steps to select the relevant features.
Step 1: Start with an empty set S={ }
Step 2: Select the relevant feature x, which maximizes the function f(S+x), x ∉ S
Step 3: Add the feature x to S.
Step 4: Go to step 2.

vi. Sequential Backward Elimination

Sequential Backward Elimination [15] is also referred as Sequential Backward Selection. It is the reverse process of sequential forward selection. It starts with the complete set of features; the features that decrease the value of objective function (irrelevant features) are removed one by one sequentially. The removal of an irrelevant feature increases the value of objective function. SBE performs best when there are a larger number of features in the feature subset. The following steps are involved in the SBE algorithm.
Step 1: Start with the entire set FS.
Step 2: Remove the irrelevant feature x from feature set FS.
Step 3: Update FS= {FS-x}
Step 4: Go to Step 2.

COMPARISON OF FEATURE SELECTION ALGORITHMS

This section compares the algorithms SFS, SBE, GA, FCBF, Relief, MRMR and one hybrid algorithm (FCBF + SFS). The efficiency of algorithms is analyzed in terms of its classification accuracy and the number of features selected to produce the optimal feature subset. The experiment was conducted on five datasets which are downloaded from Keel dataset. The details of dataset are listed in Table 1.

TABLE 1: Datasets used in the experiment

Dataset	Features	Instances	Classes
Wdbc	30	569	2
Ionosphere	33	351	2
Spectfheart	44	267	2
Sonar	60	208	2
Heart	13	270	2

EXPERIMENTAL RESULTS

Naïve Bayesian classifier is used in this experimentation to test the classification accuracy of optimal feature subset selected by different feature selection algorithms. The experiment was conducted using Rapidminer tool. The results are listed in Table 2. It shows the classification accuracies and the number of features selected by SFS, SBE, GA, FCBF, Relief, mRmR and a hybrid algorithm FCBF + SFS. The results are also compared with the classification accuracy of entire feature set.
The experimental results show that the optimal feature subset selected by SFS contains 4 average numbers of features and has the highest average classification accuracy 85. 91%. The next highest accuracy is achieved by SBE, but it has failed to meet the main goal of feature selection**:** the best feature subset should be as small as possible.
The average classification accuracy of GA is 82. 42%, which outperforms the mRmR algorithm its average accuracy is 78. 88%. The algorithms FCBF, Relief and mRmR select the least number of features when compared with SBE and GA.

SFS is a wrapper algorithm. The demerit of wrapper algorithm is its time complexity when applied to large datasets, but it guarantees good result. FCBF is a Filter algorithm. The filter algorithms are highly computational efficient and does not produce satisfactory result all the time. The hybrid algorithm takes the advantages of both filter and wrapper approaches. FCBF is first applied to the dataset. Then SFS is applied to the reduced feature subset to obtain the optimal feature subset. The results reveal that the hybrid algorithm-the combination of FCBF and SFS, selects the least average number of features when compared with the other algorithms and its average classification accuracy is 79. 12%.

TABLE 2 : Accuracy and No. of features selected with different algorithm tested by Naïve bayes classifier

Dataset	All Features Accuracy	SFS		SBE		FCBF		GA	
		Accuracy	Features	Accuracy	Features	Accuracy	Features	Accuracy	Features
Wdbc	94.02%	97.01%	4	94.90%	25	93.15%	3	94.73%	9
Ionosphere	90.60%	91.74%	4	91.17%	31	74.07%	2	90.88%	14
Spectfheart	69.29%	79.40%	1	70.79%	42	79.40%	1	71.91%	19
Sonar	73.08%	75.48%	2	71.15%	58	62.02%	3	71.63%	35
Heart	86.30%	85.93%	7	85.93%	11	83.70%	6	82.96%	6
Average	82.66%	**85.91%**	**4**	82.79%	33	78.47%	3	82.42%	17

TABLE. 2. Accuracy and No. of features selected with different algorithm tested by Naïve bayes (CONTINUED)

Dataset	All Features Accuracy	MRMR		Relief		FCBF+SFS	
		Accuracy	Features	Accuracy	Features	Accuracy	Features
Wdbc	94.02%	91.04%	5	90.86%	3	93.35%	2
Ionosphere	90.60%	92.59%	7	90.03%	7	72.08%	1
Spectfheart	69.29%	79.40%	1	71.16%	3	79.40%	1
Sonar	73.08%	57.69%	1	74.04%	4	67.79%	2
Heart	86.30%	73.70%	6	76.30%	3	82.96%	4
Average	82.66%	78.88%	4	80.48%	4	**79.12%**	**2**

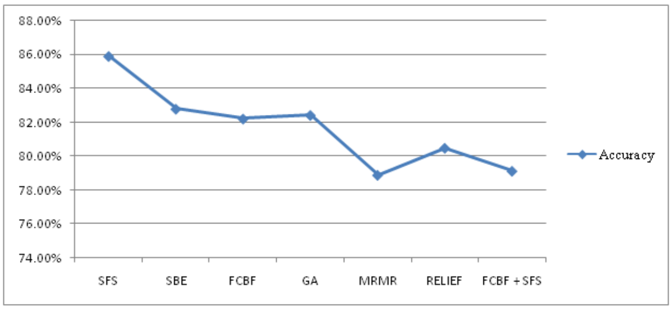

Fig. 1 Classification Accuracy

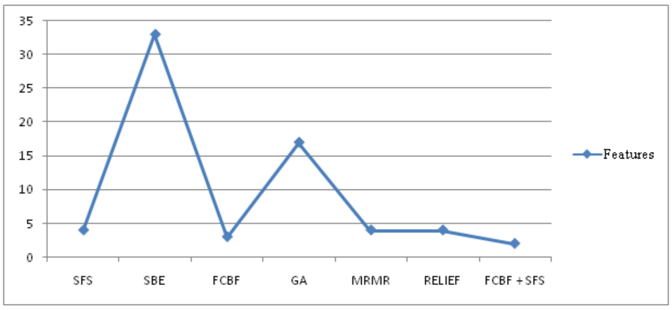

Fig. 2. Number of Features selected by different algorithms

Conclusion

Feature selection plays an important role in Data Mining process. It is one of the essential data pre-processing tasks. The main objective of feature selection is to determine an optimal feature subset as small as possible. And it should also achieve the highest prediction accuracy. Experimental result on five datasets shows that the hybrid approach can improve classification accuracy reasonably and it also selects minimum number of features for finding the optimal feature subset, when compared with the filter and wrapper methods. The study shows that FCBF+SFS select the minimal number of features when compared to the other algorithms. Our work concludes that the best combination of filter and wrapper method can improve the prediction accuracy with minimum number of features.

References

[1] Jiwaei Han, Kamber Micheline, Jian Pei " Data Mining: Concepts and Techniques", Secon Edition, Morgam Kaufmann Publishers, March 2006.

[2] N. N. R Ranga suri, M. N. Murty, G. Athitahan, "Unsupervised Feature Selection for Outlier Detection in Categorical Data using Mutual Information", IEEE, 978-1-4673-5116-4, 2012

[3] Kajal Naidu, Aparna Dhenge, Kapil Wankhade "Feature Selection Algorithm for Improving the Performance of Classification : A Survey", Fourth International Conference on Communication Systems and Network Technologies, 2014

[4] M. Dash and H. Liu "Feature Selection for Classification" Intelligent Data Analysis, Vol. 1, no 1-4, pp. 131-156, 1997.

[5] Lei Yu, Huan Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", Department of Computer Science & Engineering, Arizone State University, Tempe, AZ 85287-5406, USA, 2003

[6] Huan Liu and Lei Yu, " Towards Integrating Feature Selection Algorithms for Classification and Clustering", IEEE Transactions on Knowledge and Data Engineering, Vol. 17 No. 4 2005

[7] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning", Artificial Intelligence, vol 97, pp 245-271, 1997

[8] S. Das, "Filters, Wrappers and a Boosting bassed Hybrid for Feature Selection", Proc. 18th International Conference on Machine learning", pp74-81, 2001

[9] Qinbao Song, Jingjie Ni and Guangtao Wang, "A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data", IEEE Transactions on Knowledge and Data Engineering, Vol 25, No. 1, 2013

[10] K. Kira and L. A Rendell, "The Feature Selection Problem: Traditional methods and A New Algorithm, " *Proc. 10th National Conference Artificial Intelligence, pp. 129-134, 1992.*

[11] Lei Yu and Huan Liu., " Feature Selection for high-dimensional data: A fast correlation-based filter solution", *Proceedings of the 21th International Joint Conference on Machine Learning(ICML-03), August 21-24, 2003*, pages 856-863.

[12] Rajdev Tiwari and Manu Pratap Singh, "Correlation-based Attribute Selection using Genetic Algorithm:, International Journal of Computer Applications, Vol. 4-No. 8 pp28-34, Aug 2010

[13] Holland, J. H., "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor, MI., 1975

[14] C. H. Q. Ding and H. Peng. "Minimum redundancy feature selection from microarray gene expression data".. In CSB, pages 523-529. IEEE Computer Society, 2003.

[15] A. Veeraswamy, S. Appavu, "A Survey on Feature Selection Algorithms in Data Mining", International Journal of Advanced Research In Technology, Vol. 1 Issue 2, pages 108-117, 2011

[16] Jihoon Yang and Vasant Honavar, "Feature Subset Selection Using A Genetic Algorithm", Lowa State University, Ames, IA 50011, 1997

[17] A. Y Ng, "On Feature Selection, Learning with Exponentially Many Irrelavant Features as Training Examples", Proc. 15th International Conference on Data mining, pp. 306-313, 2002

[18] E. Xing, M. Jordan, and R Karp, "Feature Selection for High dimensional Genomic Microarray Data"' Proc. 18th International conference on Machine Learning, pp 601-608, 2001

[19] J. Souza, "Feature Selection with a General Hybrid Algorithm", Ph. D Dissertation, University of Ottawa, 2004

[20] J. Yu, S. S. R Abidi, and P. H. Artes, "A Hybrid Feature Selection Strategy Image Defiing Features, Towards Interpretation of Optic Nerve Images", Proc International Conference on Machine Learning and cybernetics, Vol 8, pp. 5127-5132, 2005