IFSS – An Improved Filter-Wrapper Algorithm for Feature Subset Selection

Saurabh Soni
M.E. Scholar
Computer Science Engineering
Gujarat Technological University

Pratik Patel
Asst. Professor
C.S.E Department
Parul Institute of Technology

ABSTRACT
The ever increasing growth of databases in the real time application is a major issue for the handling of large data. The data mining of the same is also a tedious task. The feature subset selection is a process for finding the irrelevant and redundant data and handling them. The proposed algorithm IFSS- Improved Feature Subset Selection works in 2 major steps: 1. Find the irrelevant features and 2. Evaluate its fitness with Ant Colony Optimization (ACO). The Computation time taken to derive the results is taken to compare with different FSS algorithms.

Keywords
FSS, filter, wrapper, ACO, IFSS

1. INTRODUCTION
We live in the era of information. We have lots of data but raw data is useless unless we process this data and gain some knowledgeable information from it. This knowledge can be very useful in some decision making for business that is why it is also called business intelligence. Feature Subset Selection (FSS) is a technique to identify the most co-related quality set of features which helps to predict the class label very precisely.

A feature selection algorithm can be evaluated by the efficiency and effectiveness points of view. Where the efficiency means the time needed to search a subset of features, the effectiveness is concerned to the subset of features' quality. Among various feature subset selection algorithms, some can effectively remove the irrelevant features but fail to tackle redundant features [1], [2], [3] still some of the others can remove the irrelevant while taking care of the features which are redundant [4], [5].

In this proposed method IFSS, first the relevancy of the features is calculated then the parameters of ACO are initiated. Then each feature subset is evaluated and the fitness is calculated. The process is repeated until the stopping criteria are met. Thus the best feature subset is generated.

The remaining paper is organized as follows: section 2 illustrates various FSS techniques in brief, the section 3 contains the proposed method details, and section 4 provides result analysis and comparisons. Finally, in the section 5 some conclusions are drawn.

2. VARIOUS FSS TECHNIQUES
The existence of irrelevant features can increase the size of space of search and time acquired by the algorithm. But in contradiction, if neural networks have less input neurons than needed, the algorithm will be useless to search expected classification function and if it has more inputs than needed, the result will lead to poor generalization. The filter methods are not depended on an induction algorithm and the drawback is that it ignores the biased of the algorithm and do effect the performance of the algorithm. The wrapper method is dependent on the algorithm and uses it as a part of evaluation technique. But the drawback of wrapper method is its time consuming and expensive. The 2-phase subset selection method for neural networks is designed. The algorithm is effecting in eliminating useless redundant and irrelevant features [6]

Table 1: Comparison of Filter and Wrapper Approach

Sr. No.	Measures	Wrapper	Filter
1	**Flow**	Predetermined algorithm	Independent of learning algorithm
2	**Efficiency**	High	Not guaranteed
3	**Computation**	Expensive	Low
4	**Performance**	Better	Lesser than Wrapper
5	**Usage**	Used when no. of features are less	Used when no. of features are more

MRANNIGMA- Maximum Relevance Artificial Neural Network Input Gain Measurement Approximation. The filter methods are computationally cheap because they are applied to the dataset as a pre-processing step. Their experiments show that gathered heuristic score in the MR-ANNIGMA ranks the features in a way that the internal proceedings of the wrapper step gives better subsets of features than both filter and wrapper approaches. [7]

The FAST algorithm consists of these steps: (i) eliminating irrelevant features, (ii) design a MST from relative features, and (iii) partition the MST and select a representative from the given features. [8]

Not only the performance and computational efficiency but stability referred to robustness of datasets. The problem of stability, its importance, and many measures of stability used to evaluate feature subsets in context of bio-informatics is discussed. Based on the lack of research in the field of bioinformatics, more work is needed in this area of research. [9]

3. PROPOSED METHOD

As referred various for FSS algorithms, it can be observed that the time and accuracy should be maintained while using any algorithm. With the help of T-relevance [8], the irrelevant features will be removed; as the irrelevant features increase the overhead in computation time.

Redundant features should be eliminated during the execution, failing to which it will consume more time to select the feature subset. And that will affect the output in terms of accuracy. The former generates the features relevant to the target concept by removing irrelevant features, and the latter eliminates redundant features from relevant ones via picking up the representatives from different feature clusters, and then provides the final subset.

3.1 Some Definitions:

T-Relevance: the relevance between the features F and target concept C is referred to as T-relevance of F and C and denoted by SU (F, C). If SU (F, C) is greater than predetermined threshold, we can say that F is a strong T-relevance feature. [8]

The symmetric uncertainty (*SU*) [8] is concluded from the mutual information by normalizing it to the entropies of feature values or feature values and target classes, and had been used to compute the fineness of features for classification by various researchers (e.g. Hall and Smith [10], Hall [4], Yu and Liu [11], [12], Zhao and Liu [13], [14]). Therefore the symmetric uncertainty is chosen as the measure of correlation between either a feature & the target concept or between two features.

Symmetric Uncertainty is defined as follows

$$SU(X,Y) = \frac{2 \times Gain\,(X|Y)}{H(X)+H(Y)}$$

Where,

1) $H(X)$ is the entropy of a discrete random variable X. Suppose $p(x)$ is the prior probabilities for all values of X, $H(X)$ is defined by

$$H\,(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

2) Gain $(X|Y)$ is the amount by which the entropy of Y decreases. It reflects the additional information about Y provided by X and is called the information gain [15] which is given by

$$Gain\,(X|Y) = H(X) - H(X|Y)$$
$$= H(Y) - H\,(Y|X)$$

Where $H(X|Y)$ is the conditional entropy which quantifies the remaining entropy (i.e. uncertainty) of a random variable X given that the value of another random variable Y is known. Suppose $p(x)$ is the prior probabilities for all values of X and $p(x|y)$ is the posterior probabilities of X given the values of , $H(X|Y$) is defined by

$$H(X|Y) = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_2 p(x|y)$$

Information gain is a symmetrical measure. That is the amount of information gained about X after observing Y is equal to the amount of information gained about Y after observing X. This ensures that the order of two variables (e.g. (X, Y) or (Y, X)) will not affect the value of the measure. [8]

3.2 The IFSS algorithm works as follows:

Step 1: Load the dataset.

Step 2: Find Relevancy of the features using T-Relevance Symmetric Uncertainty value for each feature.

Step 3: Initialize the parameters of ACO.

Step 4: Evaluate each feature subset and fitness (accuracy).

Step 5: Repeat the process until the stopping criteria do not meet.

Step 6: Report the best feature subset as final more appropriate set.

Step 7: End.

4. RESULT ANALYSIS

The other FSS algorithms like CFS, FCBF and Focus-F [4], [11] are compared with the proposed IFSS algorithm. The IFSS algorithm is evaluated with other algorithm on basis of runtime in milliseconds. Different datasets like chess, mfeat Fourier, elephant and arrhythmia are taken for calculation. The results show that the runtime comparison among these algorithms. The comparison is shown as a table and also as a graph below.

Table 2: Runtime (in ms) for 4 feature subset selection algorithms

	Runtime in (ms)			
Data Set	**CFS**	**FCBF**	**Focus-SF**	**IFSS**
chess	355	60	665	575
mfeat fourier	938	716	659	699
elephant	905	312	1098	618
arrhythmia	826	115	2945	567

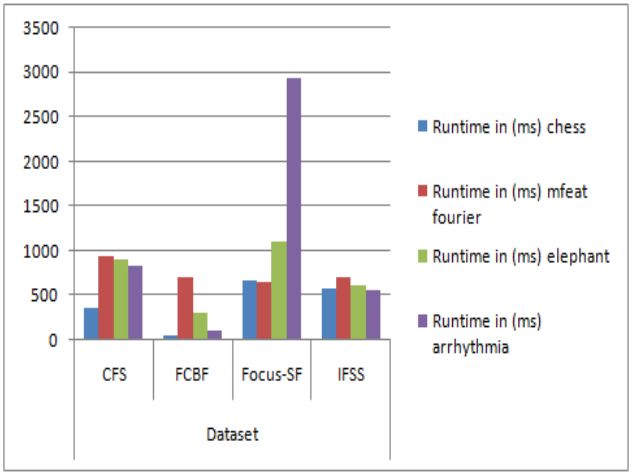

Fig.1 Runtime (in ms) for 4 different algorithms in a Graph

5. CONCLUSION

FSS - Feature Subset Selection algorithms which provide irrelevant and redundant features removal but the computation time of generating results is also concerned. In the proposed work, a hybrid algorithm IFSS is designed and it is compared with the other Feature Subset Selection algorithms. It is found that IFSS consumes less running time than other FSS algorithms.

In the future work, further more enhancements can be done in the proposed method IFSS and it can be compared in terms of the accuracy with other FSS algorithms.

6. ACKNOWLEDGEMENT

I would like to thank Assistant Professor Mr. Pratik Patel of Parul Institute of Technology, Baroda, India for his constant guidance and support in this research.

7. REFERENCES

[1] Forman G., An extensive empirical study of feature selection metrics for text classification, Journal of Machine Learning Research, 3, pp 1289-1305, 2003.

[2] Hall M.A., Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning, In Proceedings of 17th International Conference on Machine Learning, pp 359-366, 2000.

[3] Scherf M. and Brauer W., Feature Selection By Means of a Feature Weighting Approach, Technical Report FKI-221-97, Institut fur Informatik, Technische Universitat Munchen, 1997.

[4] Hall M.A., Correlation-Based Feature Subset Selection for Machine Learning, Ph.D. dissertation Waikato, New Zealand: Univ. Waikato, 1999.

[5] Yu L. and Liu H., Feature selection for high-dimensional data: a fast correlation-based filter solution, in Proceedings of 20th International Conference on Machine Leaning, 20(2), pp 856-863, 2003.

[6] A Two-phase Feature Selection Method using both Filter and Wrapper. By Huang Yuan, Shian-Shyong Tseng, Wu Gangshan, Zhang Fuya, 1999, IEEE.

[7] Hybrid wrapper-filter approaches for input feature selection using Maximum Relevance and Artificial Neural Network Input Gain Measurement Approximation, by Shamsul Huda, John Yearwood, Andrew Strainieri, 2011 IEEE.

[8] A Fast Clustering-Based Feature Subset Selection Algorithm for High Dimensional Data, by Qinbao Song, Jingjie Ni and Guangtao Wang, 2013, IEEE.

[9] A Survey of Stability Analysis of Feature Subset Selection Techniques. By Taghi M. Khoshgoftaar, Alireza Fazelpour, Huanjing Wang and Randall Wald, 2013, IEEE.

[10] Hall M.A. and Smith L.A., Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper, In Proceedings of the Twelfth international Florida Artificial intelligence Research Society Conference, pp 235-239, 1999.

[11] Yu L. and Liu H., Feature selection for high-dimensional data: a fast correlation-based filter solution, in Proceedings of 20th International Conference on Machine Leaning, 20(2), pp 856-863, 2003.

[12] Yu L. and Liu H., Efficient feature selection via analysis of relevance and redundancy, Journal of Machine Learning Research, 10(5), pp 1205-1224, 2004.

[13] Zhao Z. and Liu H., Searching for interacting features, In Proceedings of the 20th International Joint Conference on AI, 2007.

[14] Zhao Z. and Liu H., Searching for Interacting Features in Subset Selection, Journal Intelligent Data Analysis, 13(2), pp 207-228, 2009.

[15] Quinlan J.R., C4.5: Programs for Machine Learning. San Mateo, Calif: Morgan Kaufman, 1993.