Feature Selection using Multi-objective Genetic Algorithm: A Hybrid Approach

JYOTI AHUJA[1]
SAROJ DAHIYA RATNOO[1]

GJUST - Guru Jambheshwar University of Sciecne and Technology
DCSE - Department of Computer Science and Engineering
Hisar - Haryana - India - 125001
[1] (kwatra.jyoti, ratnoo.saroj)@gmail.com

Abstract. Feature selection is an important pre-processing task for building accurate and comprehensible classification models. Several researchers have applied filter, wrapper or hybrid approaches using genetic algorithms which are good candidates for optimization problems that involve large search spaces like in the case of feature selection. Moreover, feature selection is an inherently multi-objective problem with many competing objectives involving size, predictive power and redundancy of the feature subset under consideration. Hence, Multi-Objective Genetic Algorithms (MOGAs) are a natural choice for this problem. In this paper, we propose a hybrid approach (a wrapper guided by filter approach) for feature selection which employs a MOGA at filter phase and a simple GA at the wrapper phase. The MOGA at filter phase provides a non-dominated set of feature subsets optimized on several criteria as input to the wrapper phase. Now, Genetic Algorithm at wrapper phase does the classifier dependent optimization. We have used support vector machine (SVM) as the classification algorithm in the wrapper phase. The proposed hybrid approach has been validated on ten datasets from UCI Machine learning repository. A comparison is presented in terms of predictive accuracy, feature subset size and running time among the pure filter, pure wrapper, an earlier hybrid approach based on genetic algorithm and the proposed approach.

Keywords: Multi-objective Genetic Algorithm, hybrid, feature selection.

(Received November 28th, 2014 / Accepted May 12th, 2015)

1 Introduction

Feature selection is a process of selecting a subset of original features of a dataset based on certain criteria [13]. It is an important preprocessing task and plays an important role in building more accurate and comprehensible classifiers, by reducing the dimensionality of large datasets to contain relevant and non-redundant features. The selection of informative features with least redundancy is a challenging problem in domains such as Data Mining, Statistics and Pattern Recognition. Moreover, provided that the feature selection method is reasonably fast, the overall time to build the classifier model is reduced if only the selected features are used, rather than the entire set of features of a dataset.

Feature selection (FS) is a combinatorial optimization problem of selecting an optimal set of relevant features from amongst a large set of features. Exhaustive evaluation of all possible feature subsets 2^N is usually infeasible in practice due to the amount of computational effort required. As a result, many research studies focus on global search algorithms like Genetic Algorithms (GAs) to address the problem of feature selection. These algorithms start with a population of randomly initialized candidate solutions to the problem under consideration. These random candidate solutions evolve into an optimal/nearly optimal solution(s) by going through a process of reproduction, recombination and mutation over a large number of generations. Genetic algorithms exploit knowledge of the problem domain to guide the search towards global optimal solutions. Several GA based approaches proposed in the literature to solve the feature selection problem have shown promising results as compared to the traditional greedy algorithms developed for this domain [34] [6] [31]. These genetic algorithm based approaches converge to a single optimal or nearly optimal feature subset. However, often, discovering a single best subset of features is not of much interest in itself. In fact, finding several feature subsets reflecting a trade off among several objective criteria can be more beneficial. With many optimal feature subsets discovered with respect to multiple criteria, users can exercise a broad choice for good feature subsets according to their preferences. Thus, Multi-objective Genetic Algorithms (MOGAs) are more suitable to deal with the feature selection problem because of their ability to yield multiple non-dominated solutions (Pareto optimal solutions).

There are two most common approaches to feature selection, i.e. the filter and wrapper approaches [21]. The filter approaches assign a relevance score to each feature and select the features with high relevance score independent of any bias of classification algorithms. In contrast, the wrapper approaches take accuracy of the classifier as the measure to evaluate a feature subset [1] [11] [19]. The predictive accuracy of a classifier associated with the wrapper approach is mostly greater than the one associated with filter approach. However, the wrapper approaches are computationally expensive. Thus, both the techniques have their own advantages and disadvantages. In order to take benefit of both the methods, we propose a wrapper guided by filter, i.e. a hybrid approach for feature selection in this paper. The proposed approach employs the most referenced MOGA (Non- dominated Sorting Genetic Algorithm (NSGA-II)) [5] in the filter phase to discover diverse feature subsets optimized on multi-criteria without assigning a priority to any of the objectives. Our aim of applying multi-objective optimization is to find diverse feature subsets with small cardinality, high predictive power and least redundancy. At the end of the filter phase, users can use their domain knowledge or experience to select feature subset(s) reflecting the best trade-off between the conflicting objectives to suit their requirements. Subsequently, these subsets become input to the population of the wrapper phase genetic algorithm for further optimization/customization with respect to the classification method used in the wrapper phase. In this work, support vector machine, a well known classification technique, has been used in the wrapper phase.

The rest of the paper is organized as follows: Section 2 describes multi-objective optimization and Multi-objective Genetic Algorithm. A brief description of support vector machines is presented in Section 3. Section 4 contains the review of literature as a context to the work done in this paper. Section 5 details the proposed two phased framework for the hybrid approach to feature selection. Section 6 contains the experimental design and a discussion on results. Conclusions and future direction of the work are presented in Section 7.

2 Multi-objective optimization

In a multi-objective optimization problem (MOP), it is hard to find a single solution that is optimal with respect to all the objectives. This requires generating a set of solutions each of which is good enough to satisfy all the objectives to an extent without being dominated by any other solution in the solution space [17]. In reference to a multi-objective minimization problem with respect to k objectives, a feasible solution x is said to dominate another feasible solution y $(x > y)$ if and only if $f_i(x) \leq f_i(y)$ for $i = 1, 2...k$ and $f_j(x) < f_j(y)$ for at least one objective function j. A solution is said to be Pareto optimal if it is not dominated by any other solution in the solution space. The set of all possible non-dominated solutions in solution space is referred as the Pareto optimal set and the corresponding objective function values in the objective space are referred to as the Pareto front. The eventual task of multi-objective optimization is to work out solutions in the Pareto optimal set.

Multi-objective Genetic Algorithms are specifically suited to solve multi-objective optimization problems and do not require prioritizing the objectives a priori like simple GAs. The most referenced MOGA for the multi-objective optimization is NSGA II, which is an extension of the Non-dominated Sorting Genetic Algorithm (NSGA). It outperforms other Multi-objective Evolutionary Algorithms (MOEAs) like Pareto Archived Evolution Strategy(PAES) and Strength Pareto EA(SPEA), in terms of finding a diverse set of solutions while converging near the Pareto-optimal set [17]. The main features of NSGA II are Non-Domination Ranking Technique and diversity preservation using crowding distance. The Non-domination ranking approach explicitly utilizes the concept of Pareto dominance in evaluating fitness or assigning selection probabilities to solutions. [5] describes the detailed working of NSGA-II.

3 Support Vector Machines

A Support vector machine (SVM) is a machine learning algorithm based on the statistical learning theory [32]to classify the datasets with linear as well as non-linear decision boundary. The purpose of SVM is to search for the maximal marginal hyper-plane (MMH) which classifies all the training tuples corresponding to two classes and find out some essential training tuples, called as support vectors, which fall on the sides of the hyper-planes named as margins. SVM finds this hyper-plane by solving the following quadratic optimization problem:

$$max \frac{2}{\|w\|} \; subject \; to \; y_i \times (w^T x_i + b) \geq 1, i = 1, 2, ..., m \quad (1)$$

where,$w = w1, w2, ..., wn$, represents a weight vector associated with the attributes; m stands for the number of instances and b signifies a scalar referred to as the 'bias'. The tuples which satisfy Eq. 1 with equality are called support vectors. For non-linearly separable data, it uses a nonlinear mapping to map the data into higher dimensions so that the data can become linearly separable. This mapping is performed by a kernel function $k(x, y) = \Phi(x).\Phi(y)$, which is the dot product of two feature vectors in decision function. Some of the frequently used kernel functions are summed up below.

$$Linear \; kernel : x \times y \quad (2)$$

$$RBF \; kernel : exp(-g \times |x - y|^2) \quad (3)$$

$$Sigmoid \; kernel : tanh(g \times x^T y + r) \quad (4)$$

$$Polynomial \; kernel : (g \times x^T y + r)^d \quad (5)$$

In the above equations g is a predefined parameter called as gamma.

4 Related Work

Feature selection has long been a fertile field of research and a vast literature exists on the various techniques of feature selection. A comprehensive survey of existing feature selection techniques and a general framework for their unification can be found in [21]. Feature selection algorithms have been reviewed from a statistical learning point of view in [9]. There are two key issues in constructing a feature selection method: search strategies and evaluation measures. The feature selection methods has been categorized into two classes- classifier-specific and classifier independent- with respect to feature subset evaluation measures. Classifier independent (filter) methods use criteria based on statistics such as χ^2-statistics [20], T-statistics [18], F-statistics [26], Fisher criterion [7], information gain [22], mutual information [9] and entropy-based measures [4]. A well known algorithm that relies on relevance evaluation is the Relief algorithm [16]. Some existing evaluation measures that have also been shown to be effective in removing both irrelevant and redundant features include the consistency measures suggested in [1]. As feature selection involves combinatorial searches through the feature space, a number of GA based filter approaches [8][34][6]and wrapper approaches [28][10][24][15] have been proposed in the literature. GA based approaches have the disadvantage that they converge to a single best solution and hence the use of Multi-objective Genetic Algorithms is an obvious choice to discover diverse feature subsets based on simultaneous optimization of multiple and often conflicting criteria [17]. Most of the applications of MOGAs have been restricted to wrapper approaches [12] [25] and there have been comparatively less number of applications of MOGAs in filter methods [23][27][30].

Along with sole filter or wrapper approaches, many hybrid approaches have also been promulgated for features selection to take advantages of filter as well as wrapper approaches. A two phase feature selection method introduced in [35] begins by running a filter approach to remove the irrelevant features and then it runs the wrapper approach to remove redundant or useless features. Filter/wrapper approaches based on information theory and correlation measures in the filter phase have been suggested in [29] and [3]. We have already proposed a two phased genetic algorithm based hybrid approach for feature subset selection [14]. In this work, we employed genetic algorithms in the filter as well as in the wrapper phase. The filter stage ensured the selection of a feature subset consisting of highly predictive but non-redundant attributes whereas the wrapper stage further tuned the subset with respect to the SVM classifier used. Because feature subset selection is inherently a multi-objective problem, we have extended our earlier work to exploit the strength of MOGAs to discover diverse feature subsets in the filter phase itself.

5 The Proposed Hybrid Approach for Feature Selection

In this paper we propose a hybrid approach that benefits from the strengths of the filter approach, the wrapper approach and multi-objective optimization. The hybrid algorithm works in two phases- the filter phase and the wrapper phase. In the filter phase a MOGA (NSGA II) is employed to find Pareto optimal subsets of features. The MOGA minimizes the cardinality and redundancy, and maximizes the predictive power of the feature subsets with respect to the task of classification. A relevant but non-redundant feature subset must contain features that are highly correlated with the class attribute, yet are uncorrelated to each other. Keeping this in mind, we have devised a fitness function that uses correlation between a feature subset and class attribute to optimize the relevance/predictive power of features, intra-correlation among the features to reduce the redundancy and size of the feature subset to enhance the comprehensibility.

The MOGA based filter phase provides many non-dominated feature subsets varying in size and predictive power. With a number of solutions at hand, a user can exercise a broad choice in selecting feature subset(s) as the input to the wrapper phase for classifier specific tuning. As this work does not pertain to a specific application area and is applied to datasets from diverse domains, we have relied on an objective criterion for selecting features for the wrapper phase. For this purpose, we have sorted the feature subsets in descending order of first objective (predictive power) and ascending order of second objective (size). The features contained in the subsets whose relevance lye above the second quartile (median) of the relevance score, are passed to the wrapper phase. The wrapper phase employs a genetic algorithm to further optimize the features with respect to the accuracy of SVM classification. The wrapper phase GA converges very fast because its initial population contains highly relevant features and also because the redundant and noisy features have already been removed at the filter phase. Fig.1 shows the framework of the proposed hybrid approach.

We expect the following advantages in the proposed hybrid approach:

- The approach is successful in reducing the feature space for the wrapper phase significantly.
- The suggested two phased hybrid method is able to find a subset of highly relevant, non-redundant and well tuned features to the classifier under consideration. The accuracy of the classifier trained on such a subset of features is likely to be higher or comparable to the classifier trained with features discovered solely with filter or wrapper approach. The salient aspect of the envisaged approach is that the filter phase provides a set of non-dominated feature subsets consisting of relevant but non-redundant features. Thereafter, wrapper phase adds only the peculiar and specific features pertinent to the classifier used.
- The filter phase of the proposed hybrid approach provides the user choice from amongst a set of optimal solutions. A user is free to exercise his/her preference of feature subsets taking into account the trade-off between the objectives. This may be particularly useful in case the cost of attaining values for different features varies. In such cases the user can select a group of features with lower cost and still get a reasonably acceptable accuracy.
- The running time of the two phased approach is much less as compared to a purely GA based wrapper algorithm. This is because of faster convergence of wrapper phase of hybrid method, which would usually take a longer running time due to the computationally expensive fitness evaluations in successive iterations.

The flowchart describes the proposed algorithm in 2 and the implementation details of genetic algorithms for each phase are given below:

5.1 Chromosome Representation

Each chromosome represents a feature subset. A chromosome is represented as a sequence of 0's and 1's. The bit value 1 represents the presence of a feature whereas 0 denotes its absence.

In the filter MOGA phase the chromosomes are created completely randomly while in the wrapper phase the bits corresponding to the pre-discovered features, according to several criteria from the filter phase, are fixed as 1 and others are set as 0.

5.2 Fitness Functions

5.2.1 Fitness Function for filter MOGA Phase:

Since we are considering the optimality of feature subset with respect to two objectives, our fitness function comprises of two objective functions. In the filter phase the first objective is to maximize the inter-correlation (i.e. , the overall correlation between feature subset and the class attribute) and minimize intra-correlation (i.e. , mutual correlation among the predicting attributes). Thus the fitness function with respect to the first objective is taken as given in [33]

$$f_1 = \frac{RT(S,y)}{RI(S)} \tag{6}$$

where RT (S, y) is the overall correlation between the selected feature subset S and the corresponding class y, and is given by:

$$RT(S,y) = \frac{1}{|S|}\sum_{k=1}^{|S|} |corr(x,y)| \tag{7}$$

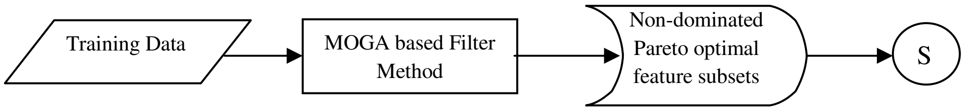

Fig. 1 a Filter phase

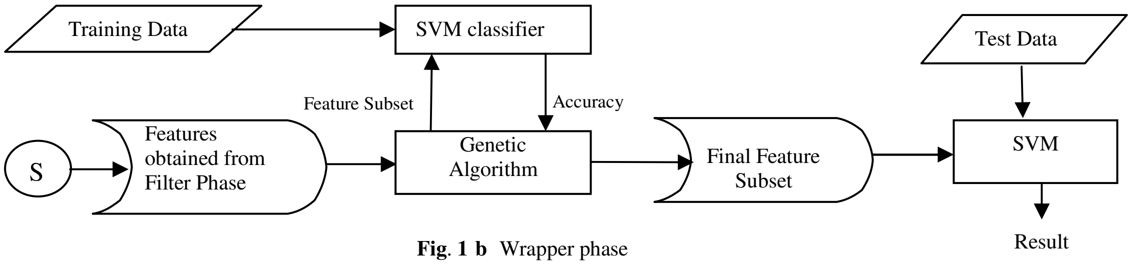

Fig. 1 b Wrapper phase

Figure 1: Framework of the proposed hybrid approach to Feature Selection

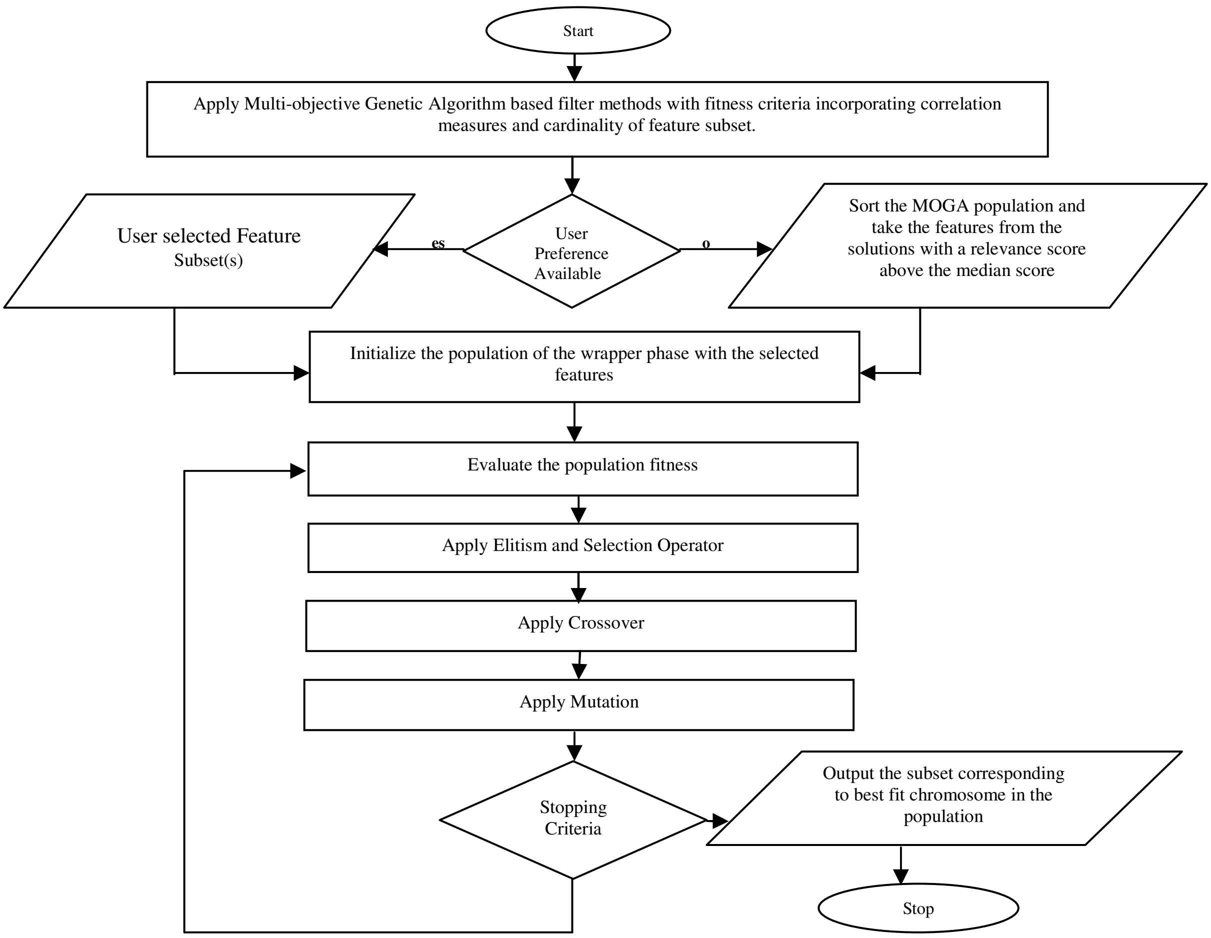

Figure 2: Flowchart of the proposed hybrid approach

Table 1: Contingency table

	v_1	v_2	v_3	$row\ sum$
y_1	$n_{11}(e_{11})$	$n_{12}(e_{12})$	$n_{13}(e_{13})$	$R_1 = \sum_1^3 n_{1i}$
y_2	$n_{21}(e_{21})$	$n_{22}(e_{22})$	$n_{23}(e_{23})$	$R_1 = \sum_2^3 n_{2i}$
$Column\ sum$	$C_1 = \sum_{j=1}^2 n_{j1}$	$C_2 = \sum_{j=1}^2 n_{j2}$	$C_3 = \sum_{j=1}^2 n_{j3}$	$Totalexamples = \lvert E\rvert = \sum_{i=1;j=1}^{3;2} n_{ij}$

RI(S) is the overall correlation within the selected feature subset (i.e. , intra-correlation) is defined as:

$$RI(S) = \frac{1}{C(S,2)} \sum_{k=1}^{|S|} \sum_{l=k+1}^{|S|} |corr(x_k, x_l)| \quad (8)$$

where C(|S|,2) is the number of $2-$combinations from the selected feature subset S. It is important to mention here that the fitness function as such is not restricted to nominal attributes but we have applied the chi-squared test of independence for calculating the inter-correlation and intra-correlation of a feature subset which is applicable to nominal features only. For computing inter and intra correlations between continuous and other types of variables the appropriate statistics needs to be used. For example we can use pearson's coefficient for continuous variables.

To find the chi-squared value between a feature and the class attribute, a contingency table is formed. Assuming a dataset with two classes (y_1 and y_2) and three values (v_1, v_2 and v_3) for the k^{th} feature (x_k), the contingency table is shown in 1:

In the above contingency table, n_{ij} and e_{ij} are the observed and expected counts of the i^{th} value of the k^{th} feature in the j^{th} class. The expected count e_{ij} is computed as $\frac{C_i \times R_j}{|E|}$. The chi-squared statistics for the k^{th} feature with respect to class is given by

$$\chi_k^2 = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (9)$$

$$corr(x_k, y) = critical\_value_{\chi^2}(df) - \chi_k^2 \quad (10)$$

$$where,\ df = (|Y| - 1) \times (|V| - 1) \quad (11)$$

In the above equations, the critical value for chi-square test is determined from the standard statistical tables for chi-square distribution at a significance level of 0.05, |Y| and |V| are the number of classes present in dataset and number of values of the k^{th} feature respectively. Intra-correlation between any two features ($corr(x_k, x_l)$) can be computed similarly. The second objective of the MOGA is to minimize the cardinality of a feature subset as given below:

$$f_2(X) = n - |X| \quad (12)$$

5.2.2 Fitness Function for the Wrapper Phase:

$$f_3(X) = Predictive\ Accuracy\ of\ SVM \quad (13)$$

5.3 Genetic Operators

- Selection- Tournament Selection has been used as selection operator in multi-objective genetic algorithm used in first phase and Roulette Wheel Selection has been used in genetic algorithm used in wrapper phase.
- Crossover- Single Point Crossover has been applied as the crossover operator.
- Mutation- We have used the simple bit flip mutation operator.

6 Experimental Design and Results

The proposed approach has been applied for feature selection on eleven benchmark datasets taken from UCI Machine Learning Repository. The methodology adopted for each dataset consists of the followings steps:

1. Setting Parameters for MOGA and GA used in filter and wrapper phases.
2. Tuning parameters for SVM classifier to obtain best predictive accuracy.
3. Measuring size of feature subsets selected using filter approach, wrapper approach and the proposed hybrid approach
4. Measuring predictive accuracy of the classifier using filter approach, wrapper approach and the proposed hybrid approach

Table 2: Dataset Description

Sr.No	*Dataset*	*#Attribute*	*#Instances*	*#classes*
1	Zoo	16	101	7
2	Mushroom	22	5644(8124)	2
3	Chess	36	3196	2
4	Splice	60	3190	3
5	Soyabean	35	562	19
6	German	24	1000	2
7	Vote	16	232(435)	2
8	Conect-4	42	67557	3
9	Solar Flare	11	1066	6
10	Tic-Tac-Toe	9	958	2
11	Lymphography	18	148	4

6.1 Description of the Datasets

The datasets considered for experimentation are very popular among the data mining researchers. These datasets comprise of diverse number of attributes as well as instances. The size of datasets varies from 101 to 5644 and number of features ranges from 9 to 60. We have also dealt with multi-class classification. Table 2 summarises all the 11 datasets.

6.2 Parameter setting for MOGA and GA for filter and wrapper phases

Initially some experimentation was done to tune the parameters for filter as well as wrapper phases. The parameter setting is given in Table 3.

In the second phase, the proposed approach calls classifier algorithm, i.e., SVM for each chromosome of the population in each of the iterations. Since fitness evaluations for this phase are computationally expensive, we have tuned the population size and number of generations to a minimal size. In addition, convergence is set as the stopping criteria. It is assumed that GA has converged if the best solution does not improve in the last five generations.

6.3 Parameter selection for SVM Classifier

As our focus is on using a hybrid method to select a non-redundant and highly informative set of small number of features, we have neither looked into the affect of different induction algorithms on the resulting feature subset nor have we used any parameter optimization techniques for SVM model selection. We have just compared the different kernel functions associated with SVM classifier and opted for the best one. In addition, appropriate values for the penalty parameter C are found through some experimentation for different datasets from the range given below:

Table 3: Genetic Algorithm Parameters

Parameters	Filter Phase	Wrapper Phase
Population Size	20	20
No. of Generations	400	50
Probability of Crossover	0.8	0.8
Probability of Mutation	0.01	0.01

Table 4: SVM Parameters

Sr.No	Dataset	Kernel Function	Value of C
1	Zoo	Linear kernel	10
2	Mushroom	Rbf kernel	10
3	Chess	Rbf kernel	100
4	Splice	Rbf kernel	100
5	Soyabean	Rbf kernel	10
6	German	Rbf kernel	1
7	Vote	Linear kernel	1
8	Conect-4	Rbf kernel	100
9	Solar Flare	Rbf kernel	100
10	Tic-Tac-Toe	Rbf kernel	1
11	Lymphography	Linear kernel	1

Table 5: Average number of features and classification accuracy of the feature subset resulted from Filter Wrapper and Hybrid Approaches

Dataset	Original #Features	Filter		Wrapper		Hybrid with GA		Hybrid with MOGA	
		# Features	#Accuracy	# Features	#Accuracy	# Features	#Accuracy	# Features	#Accuracy
Zoo	16	5.75	86.60	5.50	96.34	06.25	97.56	06.75	95.50
Mushroom	22	7.25	99.35	2.50	99.45	02.25	99.10	02.30	99.42
Chess	36	13.00	94.97	9.00	95.25	07.00	96.03	08.20	97.32
Splice	60	21.00	81.39	17.50	85.82	15.50	88.47	15.00	87.53
Soyabean	35	12.00	64.55	12.00	92.89	10.75	93.33	11.30	92.40
German	24	09.00	70.92	2.75	74.50	02.75	73.88	04.20	76.30
Vote	16	06.00	97.04	02.17	97.88	02.43	96.47	01.30	96.30
Conect-4	42	15.00	65.36	08.00	68.48	08.50	71.18	7.8	72.65
Solar Flare	11	04.00	58.08	02.25	76.17	02.00	74.12	02.30	73.45
Tic-Tac-Toe	09	04.00	73.01	02.60	75.11	05.75	77.44	04.00	77.24
Lymphography	18	07.00	82.50	04.50	82.92	06.25	86.67	06.00	86.56

C = { 0.1, 0.5, 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000}

A library for SVM called LIBSVM [2] for running the classifier is used. The SVM parameters for various datasets are shown in Table 4. All experiments are conducted on 3rd generation Intel core processor running at 3.30 GHz, 3 GB of RAM and Windows 7 Professional Operating System. We have used the MATLAB R2011a environment for implementing all the algorithms.

Every dataset is divided into training and test sets and 10 fold cross validation method is used for measuring the predictive performance of feature subsets. The results obtained for predictive accuracy and the feature subset size from various non-hybrid and hybrid approaches are given in Table 5. We have also measured the running time for five relatively larger data sets for all the approaches employed as shown in Table 6.

We have applied pairwise Wilcoxon signed rank test at the significance level of 0.05 to compare the performance of various approaches for feature selection across several datasets. It is found that wrapper and hybrid approaches are significantly better in terms of predictive accuracies and feature subset size as compared to the filter approach. However, the wrapper and hybrid approaches give comparable accuracies and subset sizes. Fig. 3 and Fig. 4 show the comparison of all the four approaches. It is obvious that pure filter approach takes the least time and pure wrapper approach takes the highest time across all the data sets. The proposed approach takes significantly less time as compared to the GA based hybrid approach. This is because MOGA at the filter stage is able to provide assorted features subsets optimized on multiple criteria and the wrapper phase GA has to do less work converging in lesser number of generations. In order to illustrate the added advantage of using MOGA at filter stage, non-dominated diverse solutions obtained for Chess and Mushroom datasets are shown in Fig. 5 Further, Table 7 gives some of the Pareto optimal solutions obtained for these datasets in descending order of the first objective (relevance and non-redundancy) and ascending order of the second objective (feature subset size). The fourth and seventh column of this table gives the predictive accuracies returned by the SVM corresponding to each of the feature subsets. The solutions given in the Table 7 are some of the alternate feature subsets that can serve as input to the wrapper phase for classifier specific tuning. The use of MOGA at filter stage provides the opportunity to select the feature subsets to be input to the wrapper phase according to subjective or objective criteria. Here, users can prefer one feature subset over the other based on their domain knowledge. In many applications, the cost of attaining different feature values may vary significantly. If these costs are known, the users can opt for cost sensitive feature subset selection. In other words users can compromise with accuracy, if required, in order to minimize the cost of feature subset. For example, if the cost associated with the 4^{th} feature subset is less than that of the 5th feature subset, a user can choose the 4^{th} solution instead of the 5^{th} with somewhat compromise on accuracy.

7 Conclusion and Future Work

This paper has proposed a novel hybrid approach for feature selection for the task of classification in data mining. The proposed approach worked in two phases-

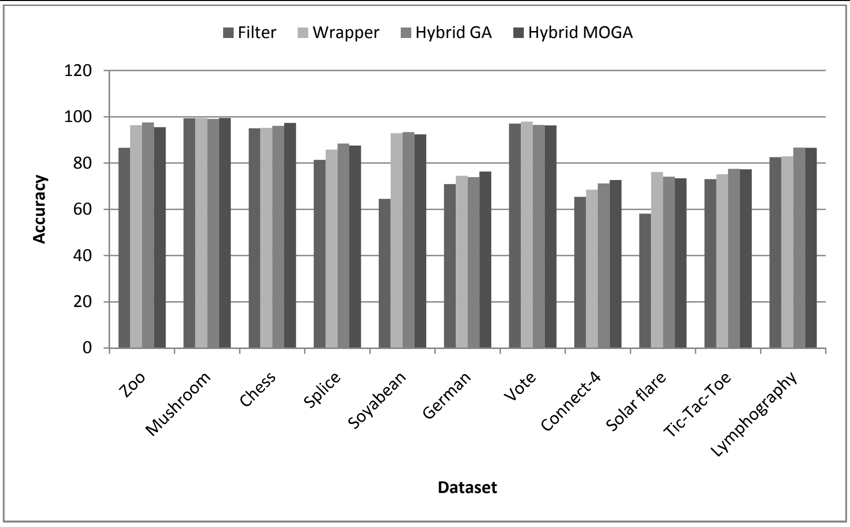

Figure 3: Variation of Number of Features amongst different approaches

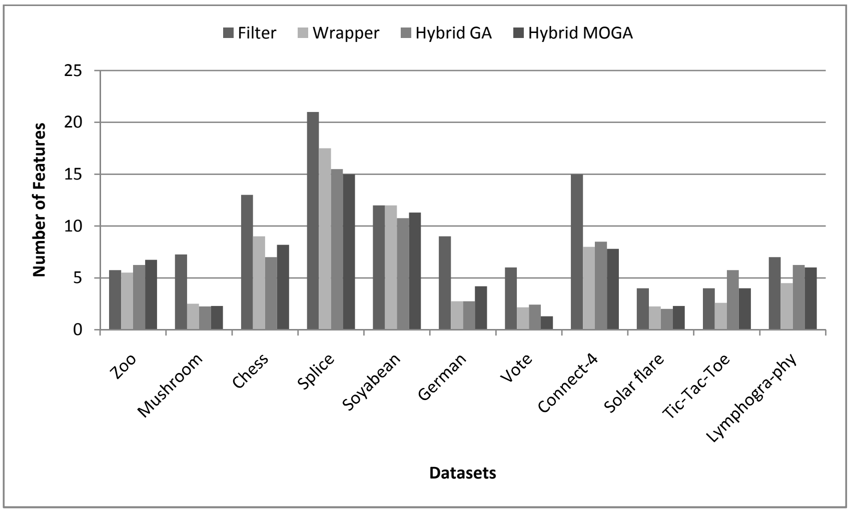

Figure 4: Variation of Predictive Accuracy amongst different approaches

a filter phase and a wrapper phase. A MOGA (NSGA II) was employed in the filter phase to provide a set of diverse non-dominated feature subsets. Subsequently, a GA based wrapper approach was implemented to perform classifier specific (SVM in this work) optimization by taking the features provided by the filter phase as the input population. The proposed approach was compared to pure filter approach, pure wrapper approach and the two phased GA approach suggested earlier in [14]. The proposed approach was found to be comparable and competitive with the pure wrapper and the two phased GA based approach and significantly better than the pure filter approach with respect to predictive accuracy and the feature subset size. Moreover, the MOGA based hybrid approach took significantly less running time as compared to other approaches except the pure filter approach. In addition, the novel approach provides scope for users to exercise their preferences in feature subset selection at the end of the filter stage.

The suggested approach currently works for the datasets containing nominal features. We intend to extend it to work with continuous features as well. Another promising research extension of this work is to take account of the cost of attaining feature values as

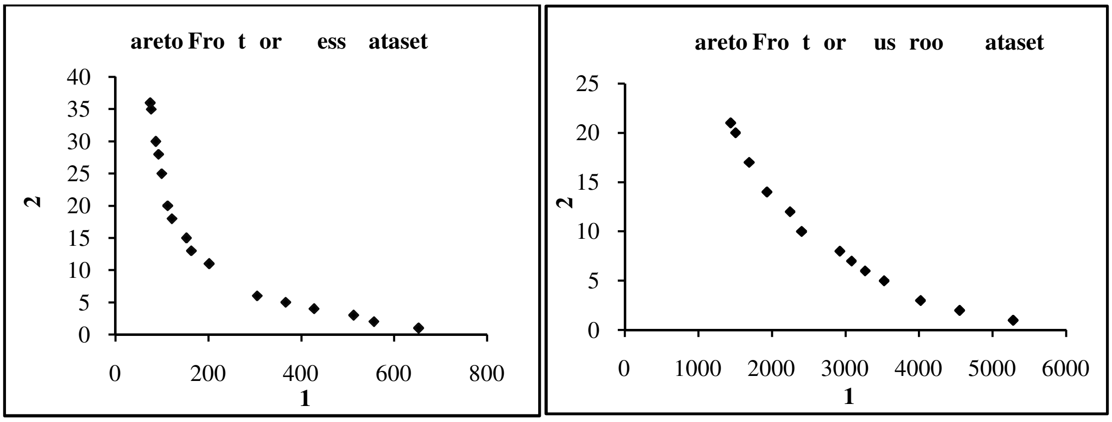

Figure 5: Non-dominated solutions obtained from filter phase of the

Table 7: Non-dominated Solutions in case of Chess Dataset

	Chess Dataset			Mushroom Dataset		
Sr.No	f1	f2	Accuracy(%)	f1	f2	Accurcy(%)
1	652.78	1	65.05	5277.1	1	98.7
2	556.76	2	76.30	4549.9	2	99.6
3	427.83	4	90.23	3697.2	4	99.6
4	512.98	3	90.23	3267.0	6	100
5	295.94	7	94.06	3057.8	7	100

one of the objectives of the MOGA at the filter stage.

Table 6: Average time taken (in seconds) to run Filter, Wrapper and Proposed Method

Dataset	Filter	Wrapper	Hybrid with GA	Hybrid with MOGA
Mushroom	3.60	13.64	49.69	21.33
Chess	4.64	476.97	96.46	44.32
Splice	8.39	538.48	349.92	282.97
German	1.64	23.25	21.92	16.32
Solar Flare	3.18	13.37	8.09	6.70

References

[1] Brönnimann, H., Chen, B., and Manoranjan Dash, P. S., Peter Haas. Efficient data reduction with EASE. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA*, pages 59–68, 2003.

[2] Chang, Chih-Chung, Lin, and Chih-Jen. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:1–27, 2011.

[3] Cheng-Huei Yang, C.-H. Y., Li-Yeh Chuang. Igga: a hybrid filter/wrapper method for feature selection of microarray data. *Journal Medical Biological. Engineering*, 30(1):23–28, 2010.

[4] Dash, M. and Liu, H. Handling large unsupervised data via dimensionality reduction. In *ACM Workshop on Research Issues in Data Mining and Knowledge DiscoverySIGMOD*, 1999.

[5] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, Apr. 2002.

[6] ElAlami, M. A filter model for feature subset selection based on genetic algorithm. *Knowledge-Based Systems*, 22(5):356–362, 2009.

[7] Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., and Haussler, D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906–914, 2000.

[8] Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[9] Guyon, I. and Elisseeff, A. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, Mar. 2003.

[10] Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422, 2002.

[11] Hall, M. A. Correlation-based feature selection for machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[12] Hamdani, T. M., Won, J.-M., Alimi, A., and Karray, F. Multi-objective feature selection with nsga ii. In Beliczynski, B., Dzielinski, A., Iwanowski, M., and Ribeiro, B., editors, *Adaptive and Natural Computing Algorithms*, volume 4431 of *Lecture Notes in Computer Science*, pages 240–247. Springer Berlin Heidelberg, 2007.

[13] Huan Liu, R. S., Hiroshi Motoda and Zhao, Z. Feature selection: An ever evolving frontier in data mining. In *4th Workshop on Feature Selection in Data Mining*, pages 4–13, Hyderabad, India, 2010.

[14] Jyoti and Saroj. Feature selection using genetic algorithm: A hybrid approach. In *7th International Conference on Advanced Computing and Communication Technologies (ICACCT-2013)*, pages 101–106, Panipat, India, 2013.

[15] Kabir, M. M., Shahjahan, M., and Murase, K. A new local search based hybrid genetic algorithm for feature selection. *Neurocomputing*, 74(17):2914–2928, 2011.

[16] Kira, K. and Rendell, L. A. The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, AAAI'92, pages 129–134, 1992.

[17] Konak, A., Coit, D. W., and Smith, A. E. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering System Safety*, 91(9):992–1007, 2006.

[18] Liu, H., Li, J., and Wong, L. A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. *Genome Informatics Series*, 13:51–60, 2002.

[19] Liu, H., Motoda, and Hiroshi. *Feature selection for knowledge discovery and data mining*. Springer, 1998.

[20] Liu, H. and Setiono, R. Chi2: Feature selection and discretization of numeric attributes. In *In Proceedings of the Seventh International Conference on Tools with Artificial Intelligence*, pages 388–391, 1995.

[21] Liu, H. and Yu, L. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering,*, 17(4):491–502, 2005.

[22] Liu, Y. A comparative study on feature selection methods for drug discovery. *Journal of Chemical Information and Computing Sciences*, 44(5):1823–1828, 2004.

[23] Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., and Coello, C. A. C. A survey of multiobjective evolutionary algorithms for data mining: Part i. *IEEE Transactions on Evolutionary Computation,*, 18(1):4–19, 2014.

[24] Oh, I.-S., Lee, J.-S., and Moon, B.-R. Hybrid genetic algorithms for feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 26(11):1424–1437, 2004.

[25] Pappa, G. L., Freitas, A. A., and Kaestner, C. Attribute selection with a multi-objective genetic algorithm. In Bittencourt, G. and Ramalho, G., editors, *Advances in Artificial Intelligence*, volume 2507 of *Lecture Notes in Computer Science*, pages 280–290. Springer Berlin Heidelberg, 2002.

[26] Peng, H., Long, F., and Ding, C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1226–1238, 2005.

[27] Saroj and Jyoti. Multi-objective genetic algorithm approach to feature subset optimization. In *IEEE International Conference on Advance Computing (IACC-2014)*, pages 544–548, 2014.

[28] Sebastián, M. and Richard, W. A wrapper method for feature selection using support vector machines. *Information Science*, 179(13):2208–2217, 2009.

[29] Sebban, M. and Nock, R. A hybrid filter/wrapper approach of feature selection using information theory. *Pattern Recognition*, 35(4):835 – 846, 2002.

[30] Spolaór, N., Lorena, A., and Lee, H. Multi-objective genetic algorithm evaluation in feature selection. In Takahashi, R., Deb, K., Wanner, E., and Greco, S., editors, *Evolutionary Multi-Criterion Optimization*, volume 6576 of *Lecture Notes in Computer Science*, pages 462–476. Springer Berlin Heidelberg, 2011.

[31] Tan, F., Fu, X., Zhang, Y., and Bourgeois, A. G. A genetic algorithm-based method for feature subset selection. *Soft Computing*, 12(2), 2007.

[32] Vapnik, V. N. *Statistical Learning Theory*. Wiley-Interscience, 1998.

[33] Wang, C.-M. and Huang, Y.-F. Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data. *Expert System with Applications*, 36(3), 2009.

[34] Yang, J. and Honavar, V. Feature subset selection using a genetic algorithm. *Intelligent Systems and their Applications, IEEE*, 13(2):44–49, 1998.

[35] Yuan, H., Tseng, S.-S., Gangshan, W., and Fuyan, Z. A two-phase feature selection method using both filter and wrapper. In *IEEE International Conference on Systems, Man, and Cybernetics, IEEE SMC '99*, volume 2, pages 132–136, 1999.