A HYBRID FILTER AND WRAPPER FEATURE SELECTION APPROACH FOR DETECTING CONTAMINATION IN DRINKING WATER MANAGEMENT SYSTEM

S. VISALAKSHI[1,*]; V. RADHA[2]

[1]Department of Information Technology, Sankara College of Science and Commerce, Sarvanampatti, Coimbatore, Tamil Nadu, India
[2]Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India
*Corresponding Author: visaraji@gmail.com

Abstract

Feature selection is an important task in predictive models which helps to identify the irrelevant features in the high - dimensional dataset. In this case of water contamination detection dataset, the standard wrapper algorithm alone cannot be applied because of the complexity. To overcome this computational complexity problem and making it lighter, filter-wrapper based algorithm has been proposed. In this work, reducing the feature space is a significant component of water contamination. The main findings are as follows: (1) The main goal is speeding up the feature selection process, so the proposed filter - based feature pre-selection is applied and guarantees that useful data are improbable to be detached in the initial stage which discussed briefly in this paper. (2) The resulting features are again filtered by using the Genetic Algorithm coded with Support Vector Machine method, where it facilitates to nutshell the subset of features with high accuracy and decreases the expense. Experimental results show that the proposed methods trim down redundant features effectively and achieved better classification accuracy.

Keywords: Mutual information, Fisher score, Markov blanket filter, Genetic algorithm, Mutation, crossover, SVM.

1. Introduction

Water supplied to the public must be potable. The water dataset contains several parameters such as physical (color, turbidity, total dissolved solids, etc.), chemical (pH, Alkalinity, Total Hardness etc.) and biological (total coliform, fecal coliform,

Nomenclatures

C	Class label
cov	Covariance
E_{ij}	Expected Frequency
F_i	Feature set
f_G	Projection off onto the variables in G
G	Subset of the overall feature
M_i	Markov Blanket
$n-$	Number of negative instances
$n+$	Number of positive instances
O_{ij}	Observed Frequency Features
$p(x)$	Probability density functions of variable X
$p(x,y)$	Joint probability of X and Y random Variables
$p(y)$	Probability density functions of variable Y
var	Variance
$\overline{X}_i^{(-)}$	Average of the ith feature of the negative
$\overline{X}_i^{(+)}$	Average of the ith feature of the positive
$\overline{X}_i$	Average of the whole dataset
$\overline{X}_{k,i}^{(-)}$	ith feature of the kth negative instance
$\overline{X}_{k,i}^{(+)}$	ith feature of the kth positive instance
(x_i, y)	Sample of the population

Abbreviations

2MGFS	2-Step process of Feature Selection
CS	Chi-Square
FS	Feature Selection
FS	Fisher Criterion Score
GA	Genetic Algorithm
M2FPS	Multiple 2-Step Filter-based Feature Pre-Selection
MB	Markov Blanket
MFMFPS	Multiple Filter Metrics-based Feature Pre-Selection
MI	Mutual Information
MIC	Mutual Information Criterion
NNs	Neural Networks
PC	Pearson Correlation
SVMs	Support Vector Machines

fecal strepto, etc.). The above parameters must be tested every day, the rise in massive and a high dimensional dataset which also challenges the traditional data mining task such as classification and clustering. Generally in data mining, the traditional and effective way of handling high dimensional data is feature selection, which aims to select relevant features and minimize the redundancy and maximize the relevance to the target class.

From the literature, the entire feature is reduced or removed using correlation coefficient or statistical test (f-test, t-test, mutual information, etc.) are used to filter the features. In this paper, the main objectives are to reduce the feature

space, solve the problem of scalability and improve the performance of classification algorithm. To meet out the objective both filter and wrapper based methods are used in this research. For selecting the relevant feature and to reduce the feature space, feature selection uses two different methods: Filter and Wrapper methods. Filter method uses statistical and ranking method for selecting relevant features [1]. The filters will extract features from the dataset without involving any learning process and wrapper uses learning techniques to evaluate each feature. The filter works independently without using classifier so it is computationally efficient. Filter method divided into multi-variate and uni-variant methods. Multivariate methods are able to find the relationship between the features, while uni-variate methods consider each feature separately. Wrapper uses a classifier for training and testing for each feature space, which leads to computationally expensive when the feature space grows [2].

Feature Selection (FS) is an important step in the water quality data analysis, as the concerned dataset exhibit high dimensionality (several tens of thousands of genes), which in turn, increases the computational challenges of the analysis algorithms [3]. A detailed review of various works related to feature selection with respect to microarray data are discussed briefly [4, 5]. The objective of this research is to select relevant features and improve the classification accuracy. So the proposed algorithm uses a filter-based method as a pre-subset feature selection and the result is then applied to wrapper-based to improve the classification performance.

This paper is organized as follows. In Section 2 the proposed 2-step Filter and Wrapper based Feature Selection is presented briefly. Section 3 discusses about the filter method metrics and explains about the wrapper-based method with a classifier. Section 4 investigates the experimental result and conclusion is given in Section 5.

2. A Proposed 2-step Filter and Wrapper based Feature Selection

The main motive of feature selection includes reducing dimensionality, removing irrelevant and redundant features, reducing the amount of data needed for the learning process and improving the predictive accuracy algorithm. The usage of feature selection in the design of contamination detection system is three-fold:-

- Improves performance of the classification algorithm,
- Handles the problem of scalability, and
- Reduces complexity of classification because of a reduced set of predictors.

The general steps involved in feature selection algorithm are shown in Fig. 1. [6, 7]. It includes generating the subset of features, evaluating the selected subset; stopping criterion and validation procedure are used to check the quality of a selected subset.

The method of filter and wrapper based feature selection methods used in this research is given below Fig. 2. A hybrid feature selection method consists of two stages which are presented in this research. In the first stage, different filter metrics are used and the Markov Blanket Filter effectively screens out irrelevant and redundant features. The proposed filter-based pre-selection method is Multiple Filter Metrics-based Feature Pre-Selection and is coded as (MFMFPS). In the second stage, uses wrapper-based to identify relevant features using

Genetic Algorithm (GA) + Support Vector Machine (SVM) which tries to determine the optimal feature subset [2].
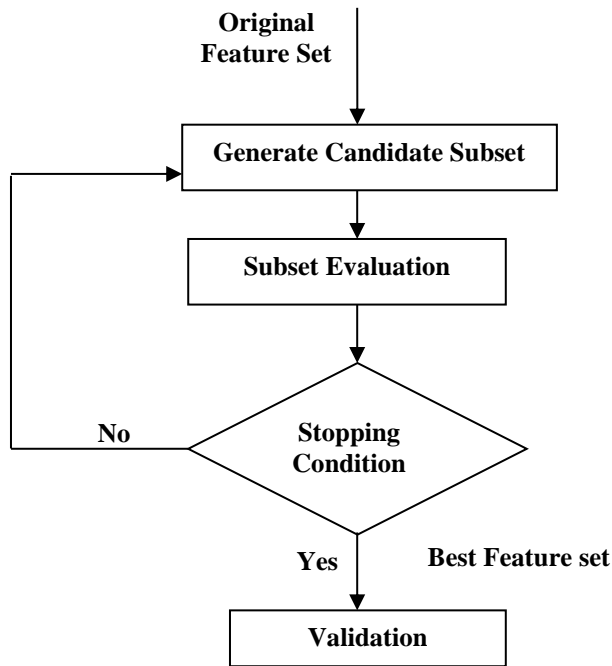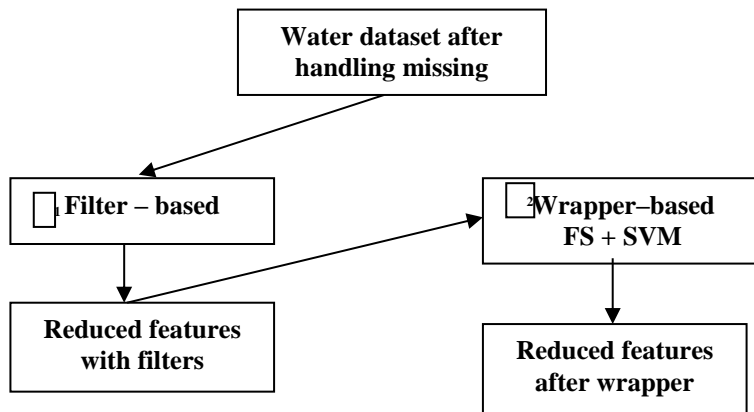

Fig. 1. Steps in feature selection.

Fig. 2. Flow of filter-wrapper process.

3. Materials and Methods

In this section describes the various feature selection methods in detail and genetic algorithm coded with support vector machine.

3.1. Filter-based feature selection

From the point of view the information theorem, the information of a set of features could be calculated by various statistical measures, and that is the core of the filter type of feature selection methods. Among the feature selection methods, filter-based method will separate the data before the classification process starts and then it calculates the feature weight values. In filter method, interactions among the features are very less [2]. Some of the methods in filter approach are correlation-based feature selection, t-test, information gain, mutual information and entropy based method. Because of the fast calculation, filters are often applied to feature selection in high-dimensional data.

3.1.1. Multiple 2 Step Filter-based Feature Pre-Selection (M2FPS) Algorithm

Filter-based feature selection techniques, as mentioned earlier, use statistical tests to determine the subset of features with the highest predictive power. Different filters yield different filtered subsets, which may leave out some relevant data. Generally, the feature selection algorithms often rely on information theory concepts like relevancy (measures of dependency between random features and features with high relevancy are selected) and redundancy (the repetitive nature of a feature and features with exhibit minimum redundancy are selected). The use of multiple filters, with different filter metrics, ensures that useful data are unlikely to be screened out in the initial pre-selection stage

This research work uses the four filter metrics, namely, Mutual Information (MI), Pearson Correlation (PC), Chi -Square (CS) and Fisher Criterion Score (FS). Markov blanket filtering is then used to generate optimal subsets of features, which are then combined, using union operation. The process of M2FPS is shown in Fig. 3. The first step of M2FPS algorithm uses four feature selection filters to select optimal features. The resulting features are used as input, to a more computationally intensive subset selection procedure, known as Markov blanket filtering which results with four sets of optimal features, namely, R-MI, R-PC, R-CS and R-FS (Steps 2). These four sets are combined using a simple Boolean Union operation, in order to select only common features that exist between the four subsets (Step 3).
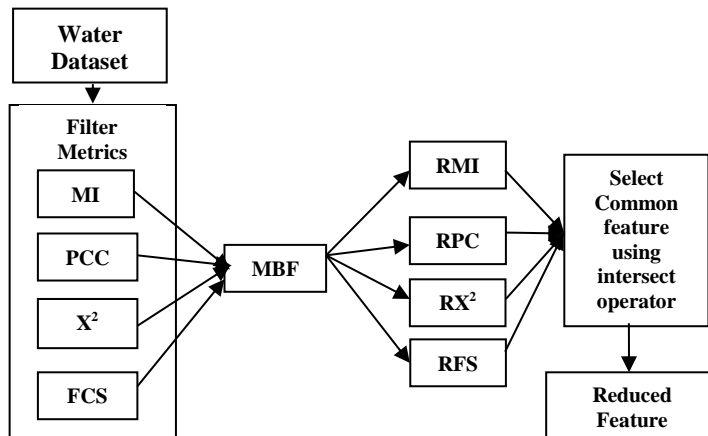

Fig. 3. Proposed Filter based feature selection with Markov blanket filter.

3.1.2. Selected Filter-based feature selection filters

Four feature selection filters, namely, Mutual Information, Pearson Co-efficient, Chi-Square and Fisher Criterion Score are used in the design of M2FPS algorithm. The details regarding these filters are presented below:-

Mutual information (MI)

Mutual Information Criterion (MIC) is a popular approach to analyze the correlation between features. It measures the contribution of a variable towards reducing uncertainty about the value of another variable. The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions [8]. For two features, X and Y, the MI is estimated using Eq. (1).

$$MI(X,Y) = -\sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{1}$$

where p(x, y) is the joint probability of X and Y random variables and p(x), p(y) are the probability density functions of variable X and Y respectively. A large value of MI signifies high correlation of two variables. Zero value indicates that two variables are not correlated.

Pearson correlation (PC)

Pearson's correlation returns a value that indicates the strength of the correlation and is also known in statistical models as the **r** value. For any two variables, it returns a value that indicates the strength of the correlation. Pearson's correlation co-efficient is computed by taking the covariance of two variables and dividing by the product of their standard deviations. The coefficient is not affected by changes of scale in the two variables. Let the sample of the population be (x_i, y), Eq. (2) then uses the mean of each feature and the target to calculate the Pearson coefficient.

$$R(i) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X_i \text{ var}(Y))}} \tag{2}$$

where *cov* and var are the covariance and variance, X and Y are any two features. The highest ranking values of features are considers and low rank will be discarded.

Chi-squared test (CS)

Chi-squared test is another popular statistical test of the divergence between the observed and expected distribution of a feature. In feature selection, it tests whether the distribution of a feature differs between groups. The chi-square score uses the summation of squared differences between observed and expected values divided by expected values. It is estimated using Eq. (3).

$$\chi^2 = \sum_{i=i}^{r} \sum_{j=i}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{3}$$

where O_{ij} is an observed frequency features and E_{ij} is an expected frequency, emphasize by the null hypothesis. The null hypothesis indicates that there is no correlation between two features.

Fisher criterion score (FS)

Fisher Criterion Score (F-Score) computes the importance of each feature independently of the other features by comparing that feature's correlation to the output labels. The score is computed by measuring the variance between the expected value of the information and the observed value. When variance is minimized, information is maximized.

F-score is a filter model which calculates the discriminative ability of each feature. That is to say, features with higher F-Score have better separation ability in classification problems. F-score is defined in the following Eq. (4).

$$F(i) = \frac{\left(\overline{X}_i^{(+)} - \overline{X}_i\right)^2 + \left(\overline{X}_i^{(-)} - \overline{X}_i\right)^2}{\frac{1}{n^+ - 1}\sum_{k-1}^{n_+}\left(\overline{X}_{k,i}^{(+)} - \overline{X}_i^{(+)}\right)^2 + \frac{1}{n_i - 1}\sum_{k-1}^{n_-}\left(\overline{X}_{k,i}^{(-)} - \overline{X}_i^{(-)}\right)^2} \tag{4}$$

where $\overline{X}_i^{(+)}$, $\overline{X}_i^{(-)}$ and $\overline{X}_i$ are the averages of the ith feature of the positive, negative and whole datasets; n+ and n- are the number of positive and negative instances, respectively; and $\overline{X}_{k,i}^{(+)}$ and $\overline{X}_{k,i}^{(-)}$ are the ith feature of the kth positive instance and the ith feature of the kth negative instance.

From Eq. (4), it can be seen that, the larger F(i) (F-Score) is, the stronger discriminative ability the feature has. The F-Score can only examine the discriminative ability of each individual feature. It cannot identify the discriminative ability of multiple features. Hence, features with low scores will be disregarded, even if they are complementary to the top features and might be very useful

3.1.3. Markov blanket filter

Let G be a subset of the overall feature set F. Let f$_G$ denote the projection off onto the variables in G. Markov blanket filtering aims to minimize the discrepancy between the conditional distributions P(C|F = f) and P(C|G = f$_G$), as measured by a conditional entropy Eq. (5).

$$\Delta_G = \sum_f P(f)D(P(C \mid F = f) \mid\mid P(C \mid G = f_G)) \tag{5}$$

where D(P||Q) = Σ_xP(x) log(P(x)/Q(x)) and is the Kullback-Leibler divergence. The goal is to find a small feature set G for which Δ_G is small. Intuitively, if a feature F$_i$ is conditionally independent of the class label given some small subset of the other features, then it should be possible to omit F$_i$ without compromising the accuracy of class prediction.

The Markov Blanket (MB) filter algorithm formalizes this idea using the notion of a Markov blanket. An MB is defined as follows. For a feature set G and class label C, the set M$_i$⊂ G (F$_i$∉ M$_i$) is a MB of F$_i$ (F$_i$∈ G) if F$_i$⊥ G−M$_i$ − {F$_i$},C | M$_i$. The following proposition establishes the relevance of the Markov blanket concept to the measure Δ_G.

For a complete feature set F, let G be a subset of F and G' = G − F$_i$. If ∃M$_i$⊆ G (where M$_i$ is a Markov blanket of F$_i$), then ΔG' = ΔG. This implies that once a MB of feature F$_i$ in a feature set G is found, it is safe to remove F$_i$ from G without

increasing the divergence to the desired distribution. The authors further proved that, in a sequential filtering process in which unnecessary features are removed one by one, a feature tagged as unnecessary based on the existence of a Markov blanket M_i remains unnecessary in later stages when more features have been removed [9]. In most cases, however, few if any, features will have a Markov blanket of limited size and therefore features that have an "approximate Markov blanket" should be searched. For this purpose, the following definition is used given in Eq. (6).

$$\Delta(F_i \mid M) = \sum_{f_M, f_i} P(M = f_M, F_i = f_i)D(P(C \mid M = f_M, F_i = f_i) \mid\mid P(C \mid M) = f_M)) \qquad (6)$$

If M is a Markov blanket for F_i then $\Delta(F_i|M) = 0$. Since an exact zero is unlikely to occur, the condition can be relaxed, and a set M can be found such that $\Delta(F_i|M)$ is small. If M is really a Markov blanket of F_i, then $P(C|M, F_i) = P(C|M)$. This suggests an easy heuristic way to search for a feature with an approximate Markov blanket.

Since the goal is to find a small non-redundant feature subset and those features that form an approximate Markov blanket of feature D_i are most likely to be more strongly correlated to F_i, a candidate MB for F_i is constructed by collecting the k features that have the highest correlations (defined by the Pearson correlations between the original non-quantized feature vectors) with F_i, where k is a small integer. The algorithm is given in Fig. 4. This heuristic sequential method is far more efficient than the methods that conduct an extensive combinatorial search over subsets of the feature set.

After using MBF, all reduced MI, PC, Chi-Square and Fisher Score features are combined to select common feature using the union operator. Hence by applying filter metrics, reduced and relevant features are selected. These features again will be given to wrapper based for training and testing the features.
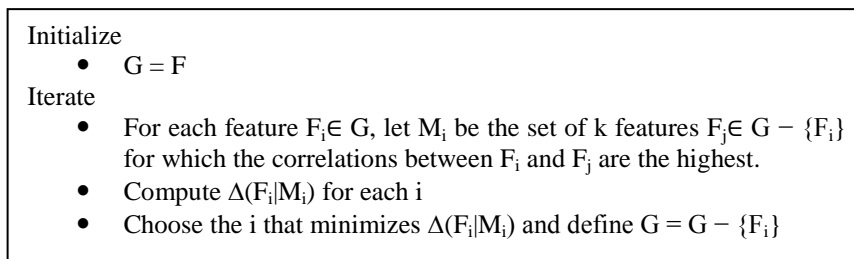

Initialize
- $G = F$

Iterate
- For each feature $F_i \in G$, let M_i be the set of k features $F_j \in G - \{F_i\}$ for which the correlations between F_i and F_j are the highest.
- Compute $\Delta(F_i|M_i)$ for each i
- Choose the i that minimizes $\Delta(F_i|M_i)$ and define $G = G - \{F_i\}$

Fig. 4. Markov blanket filter.

3.2. Wrapper-based feature selection

The wrapper-based FS method generally focuses on improving the classification accuracy and performs better compare to filter method. However, wrapper methods are computationally expensive than filter. Several methods are used to perform feature selection of training and testing data, such as branch and bound algorithm, sequential search algorithm, tabu search, binary particle swarm optimization and genetic algorithm [2, 10]. In this research, Genetic Algorithm (GA) with SVM is proposed and coded as GA-SVM and the flow is depicted in Fig. 5. SVM is based on statistical learning theory represented by Vapnik [1] and

it is used in most of prediction models. The main reason of using SVM is to incorporate the use of kernels (i.e. no need to concede the non-linear mapping function), and absence of local minima, the sparseness of result and generalization acquire by optimizing the margins. Out of all these reason, it gives more accuracy when compared to other classifiers [11]. Originally the SVM is built for binary classification later it extended to multi-class classification. In this research work, the cross validation is used to evaluate the classification performance. The cross-validation is used to test the full training dataset features by means of repeated resampling, thus maximize the total number of points used for testing which helps to protect against the overfitting. This hybrid method could significantly reduce the computational cost and also improves the classification accuracy.

Genetic Algorithm with SVM

According to Darwinian principle of 'survival of the fittest', GA obtains the optimal solution after a series of iterative computations [12]. In the evolutionary algorithm, one of the most significant algorithms is a Genetic Algorithm (GA). GAs has been applied to many searches, optimization and machine learning problems. GA comprise of search algorithm which guide to search the feature space based on a model of evolution. Based on the study, we combined GA with SVM classifier to provide better accuracy. The building blocks of the proposed system are shown in Fig. 5. Evolution means chromosome repetitively improves over generations, through selection, crossover and mutation. In this research work, GA-based feature selection algorithm is used and each feature is represented as a binary string encoding a feature subset. The genetic algorithm proceeds in an iterative manner by generating a new population of strings from old ones. Each string is encoded in the form of binary. The value of 0 indicates that the corresponding feature is not selected and the value of 1 means that the feature is selected. The initial population is generated. First, initially decide the coding for a solution which termed as a chromosome. It's usually described as a string of symbols from {0, 1}.
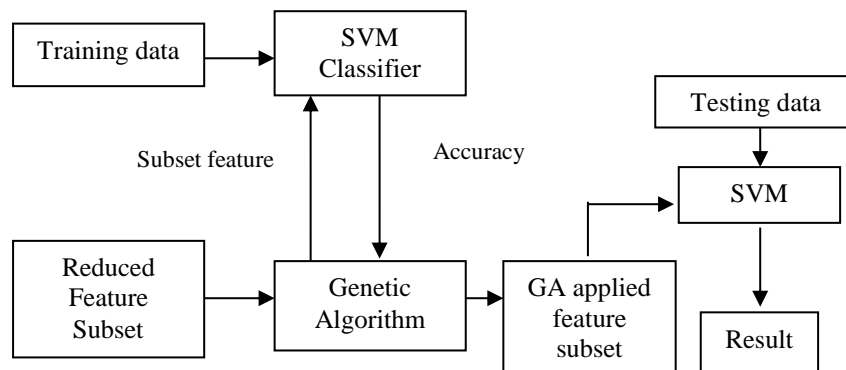

Fig. 5. Hybrid GA-SVM wrapper-based feature selection.

The genetic algorithm proceeds each feature is represented by binary digits with values 1 or 0, which describes whether that the particular feature is selected from a feature subset or not. This process is called solution (or chromosome) encoding. The general structure of chromosome is given in below Fig. 6.
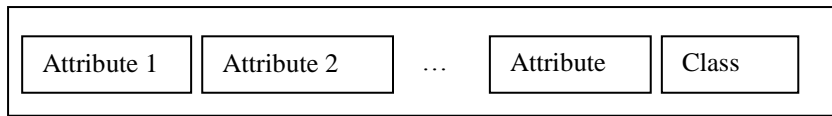
Attribute 1	Attribute 2	…	Attribute	Class

Fig. 6. General structure of a chromosome.

For example, in the 0010001101 string of ten binary digits, the features 3, 7 and 8 are selected in the corresponding feature subset. The final bit represents 0 or 1, '1' indicates whether the class in a dataset is present else '0' means otherwise. The proposed flow of GA- SVM wrapped-based feature selection is depicted in Fig. 7.

Fitness evaluation is done with the help of classification accuracy. Classification accuracy means the total number of selected features and the feature cost are the three criteria used to design a fitness function. High classification accuracy, small number of features and low total feature cost produce a higher fitness value. If the fitness value is satisfied, then terminate and produce the result, otherwise follow the next process of selection, crossover and mutation.
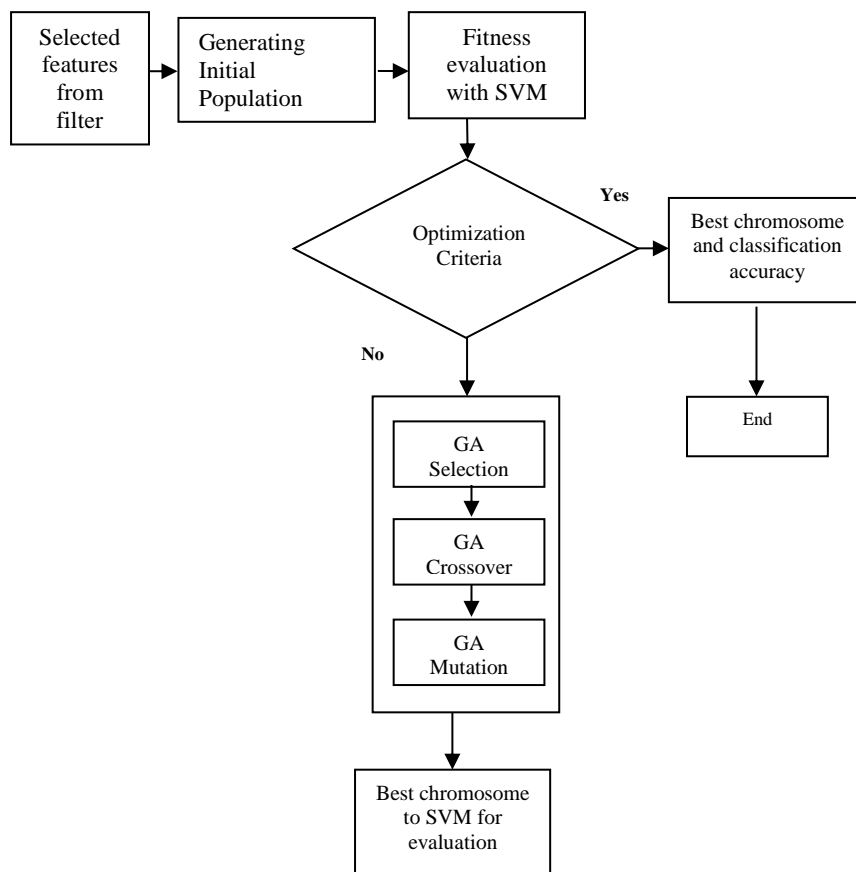

Fig. 7. Proposed GA with SVM for wrapper-based feature selection.

The fittest chromosome of each generation is chosen and given to the next population to ensure that good chromosome is passed to future generations. For crossover process, single point crossover is used in this research. Selected parent chromosomes are divided at a particular point which is randomly selected and their contiguous substrings are interchanged. For mutation process, turn over bits of a chromosome. In this work, implementation of mutation, a single bit is chosen randomly and inverted at a particular probability. The probability of the mutation and crossover are based on real life, evolution principle [2, 13]. The output features coming from genetic operators are checked again with cross validation to ensure the performance of classification.

The diagram of combined filter and wrapper feature selection methods is depicted in below Fig. 8. Hence, in the final process combine the proposed Multiple Filter Metrics-based Feature Pre-Selection (MFMFPS) and wrapper-based feature selection (GA- SVM) which is coded as a 2-Step process of feature selection (2MGFS). The metrics used to evaluate the proposed work are accuracy and normalized root mean square error rate. The experimental results are discussed in the below section.
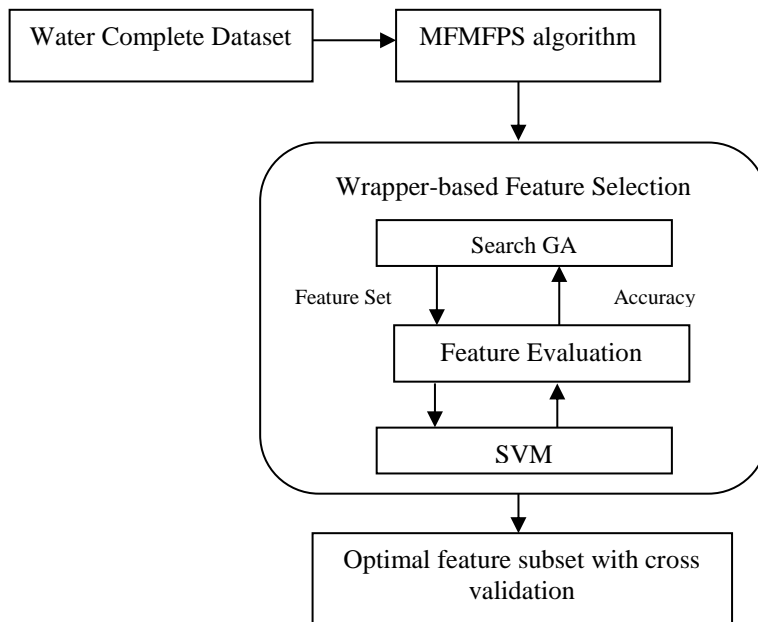

Fig. 8. Overview of proposed algorithm.

4. Experimental results

The proposed algorithms are evaluated and the experimental results are discussed in this section. The datasets are collected from Siruvani and Pillar treatment plant for summer and winter season. As we discussed earlier, the whole dataset is given for the feature selection process. In this work, the optimal features are selected by using both wrapper and filter methods. The widespread experiments were carried

out in Matlab. Finally the SVM classifier is used to evaluate the optimal features from the high dimensional dataset.

Figure 9 shows the results of feature selection with various techniques such as SVM, MI, PC, CS, FS, MFMFPS, GA with SVM and finally combined filter and wrapper are discussed. In Siruvani plant, when the season is summer, then the classification accuracy of SVM is 86.21%, MI is 88.60%, PC is 87.63%, CS is 88.66%, FS is 89.30%, proposed multi step filter feature selection is 90.90%, proposed wrapper based with GA-SVM is 91.07% and finally combined filter and wrapper the accuracy rate is increased to 91.91%.

Similarly, in Pillur, when the season is summer, then the classification accuracy of SVM is 85.15%, MI is 87.44%, PC is 87.24%, CS is 88.23%, FS is 89.04%, proposed multiple step filter feature selection is 91.04%, proposed wrapper based with GA-SVM is 91.20% and finally combined filter and wrapper the accuracy rate is increased to 92.06%. Similarly the winter season for both Siruvani and Pillur are evaluated which is given in Fig. 9. Hence, combined individual feature metrics and combined filter-wrapper gives a better accuracy rate.

The error rate of the dataset is evaluated in terms of Normalized Root Mean Square Error (NRMSE) which is shown in Fig. 10. The treatment plant of siruvani, when the season is summer then the error rate of SVM is 0.4227%, MI is 0.3636%, PC is 0.3796%, CS is 0.3464%, FS is 0.3287%, proposed system multiple step filter feature selection is 0.3050%, proposed wrapper based with GA-SVM is 0.2768% and finally combined filter and wrapper the error rate is decreased to 0.2490%. Similarly, in pillur, when the season is summer, then the error rate of SVM is 0.4308%, MI is 0.3639%, PC is 0.3793%, CS is 0.3440%, FS is 0.3308%, proposed system, multiple step filter feature selection is 0.3032%, proposed wrapper based with GA-SVM is 0.2791% and finally combined filter and wrapper the error rate is decreased to 0.2498%. Similarly the error rate of winter season for both Siruvani and Pillur are evaluated which is given in Fig. 10. Thus, the error rate is reduced compared with individual filter-wrapper methods.
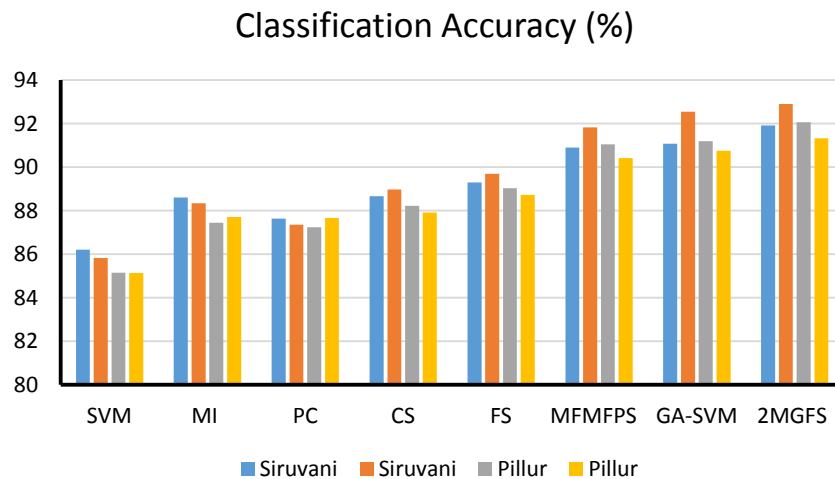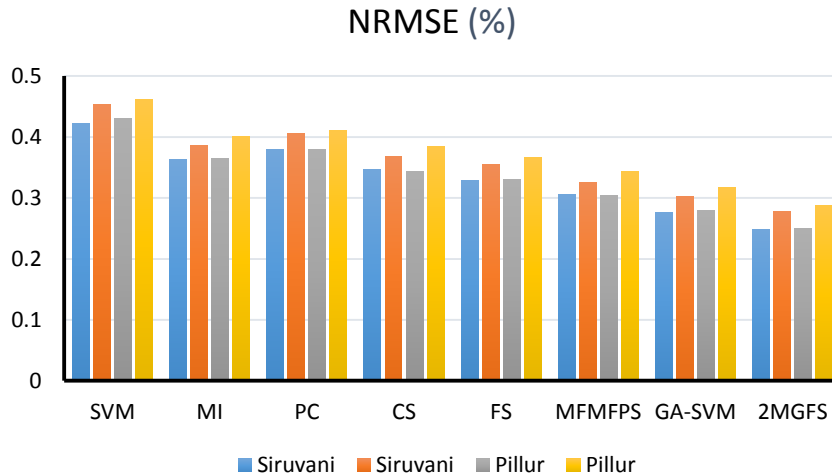

Fig. 9. Classification accuracy.

Fig. 10. NRMSE (%) Results

5. Conclusion

In recent years, feature selection methods in data mining achieved an excellent progress in prediction models. In this paper, optimal feature subset is selected using both filter and wrapper based methods. To improve the classification accuracy and remove relevant and redundant features, multiple filter and wrapper based evolutionary algorithm with SVM are proposed. Contamination detection in drinking water is very important to prevent the health hazards, so this research work proposes techniques to reduce the complexity of the wrapper based method using a 2-step method which combines filter and wrapper approaches. In filter based approach first pre-select optimal features, the resulting features are then applied to wrapper based approach to improve classification performance. The experiments are conducted with real time dataset collected from two treatment plants. The proposed approach uses an SVM classifier and performances are evaluated using classification accuracy and NRMSE. The classification accuracy is evaluated with existing Mutual Information, Pearson Coefficient, Chi-Square, Fisher Score and SVM technique and result shown in Fig. 9. For example, In Siruvani plant, when the season is winter, then the classification accuracy of SVM is 85.83%, MI is 88.34%, PC is 87.36%, CS is 88.97%, FS is 89.70%, proposed multi step filter feature selection is 91.82%, proposed wrapper based with GA-SVM is 92.55% and finally combined filter and wrapper the accuracy rate is increased to 92.90%. Likewise, for other datasets are evaluated and classification accuracy shows improved results compare with existing algorithms. Similarly the error rate is compared with the existing method which is shown in Fig. 10. For example, in treatment plant of Siruvani, when the season is summer then the error rate of SVM is 0.4534%, MI is 0.3859%, PC is 0.4051%, CS is 0.3676%, FS is 0.3552%, proposed system multiple step filter feature selection is 0.3259%, proposed wrapper based with GA-SVM is 0.3017% and finally combined filter and wrapper the error rate is decreased to 0.2775%. Similarly, error rate is also evaluated for other treatment plant and result shows, compare with existing, the error rate is reduced in proposed and hybrid algorithm of filter and wrapper -

based algorithms. From the results, it can be noted that all the proposed models are efficient when compared to its conventional counterparts and maximum efficiency can be obtained while combining the various proposed algorithms into a single framework. Thus, the proposed filter and wrapper based feature selection can be used by analyst to improve the performance of detecting contamination in water management systems. Finally the irrelevant and redundant features are pruned out successfully for contamination detection prediction model.

Reference

1. Chandrashekar, G.; and Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), 16-28.

2. Chuang, L.-Y.; Yang, C.-H.; Wu, K.-C.; and Yang, C.-H. (2011). A hybrid feature selection method for DNA microarray data. *Computers in Biology and Medicine,* 41(4), 228-237.

3. Somorjai, R.; Dolenko, B.; and Baumgartner, R. (2003). Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics*, 19, 1484-1491.

4. Saeys, Y.; Inza, I.; and Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507-2517.

5. Hira, Z.M.; and Gillies, D.F. (2015). A review of feature selection and feature extraction methods applied on microarray data. *Advances in Bioinformatics,* Volume 2015, Article ID 198363, 13 pages.

6. Koller, D.; and Sahami, M. (1996). Toward optimal feature selection. *Proceedings of the Thirteenth International Conference on Machine Learning*, 284-292.

7. Liu, H.; and Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4), 491-502.

8. Doquire, G.; and Verleysen, M. (2012). Feature selection with missing data using mutual information estimators. *Neurocomputing,* 90, 3-11.

9. Strobl, E.V.; and Visweswaranc, S. (2013). Markov blanket ranking using kernel-based conditional dependence measures. *NIPS 2013 Workshop on Causality*, 1-10.

10. Wang, H.; and Liu, S. (2016). An effective feature selection approach using the hybrid filter wrapper. *International Journal of Hybrid Information Technology*, 9(1), 119-128.

11. Uğuz, H. (2012). A hybrid system based on information gain and principal component analysis for the classification of transcranial Doppler signals. *Computer Methods and Programs in Biomedicine*, 107(3), 598-609.

12. Tsai, C.-F.; Lu, Y.-H.; and Yen, D.C. (2012). Determinants of intangible assets value: The data mining approach. *Knowledge-Based Systems*, 31, 67-77.

13. Huang, C.L.; and Wang, C.J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications*, 31(2), 231-240.