Journal of Engineering and Applied Sciences 12 (Special Issue 10): 9021-9027, 2017
ISSN: 1816-949X

A Novel Feature Selection Framework for Improving Detection Performance of Supervised Classifiers

[1]Sivakumar Venkataraman, [2]Rajalakshmi Selvaraj and [2]Venu Madhav Kuthadi
[1]Department of Computer Science and Information Systems, Faculty of Computing, Botho University, Botswana International University of Science and Technology (BIUST), Gaborne, Botswana
[2]Department of Computer Science and Information Systems, Botswana International University of Science and Technology (BIUST), Gaborne, Botswana

Abstract: This research aims to develop a novel feature selector for improving the detection performance of supervised classifiers. Handling large number of features is a tedious process. One solution is to select only the relevant features and eliminate both irrelevant, redundant features from the original set. A new feature selection method based on Class Conditional Probability (CCP) is proposed in this research. The CCP for every attribute is calculated using Naive Bayes approach. The related attributes which has the CCP value greater than the threshold value is selected as relevant features. Then, the reduced feature set is applied to different classifiers such as C4.5, Naive Bayes (NB), Support Vector Machine (SVM), Nearest Neighbour (NN) and K-Nearest Neighbour (K-NN). Different datasets from UCI repository are considered to prove the efficacy of the proposed feature selector based on the number of selected features, time taken to build the model and classification accuracy.

Key words: Class conditional probability, C4.5, Naive Bayes, support vector machine, nearest neighbour, K-nearest neighbour

INTRODUCTION

The Feature Selection (FS) method is mostly employed to advance the detection performance of classifiers and for diminish the time taken to build the model. The elimination of irrelevant and redundant features is the first objective of any feature selection algorithm followed by selection of relevant features through various selection measures. Filter, wrapper, embedded and hybrid methods are four major category of FS approach which is constructed on the supervised learning procedure is working. Figure 1 shows the process on how filter approach acquires the features by not including any supervised learning algorithm, forward this sets for all classifications. This filter approach compared with wrapper method, attains less computational complexity by more generality. This is more fit for high-dimensional feature space. The validation of generated features by incorporating the supervised learning algorithms is the main idea of wrapper based feature selector as shown in Fig. 2. It is not ensuring to produce higher classification accuracy for every one of the classifiers, since, it doesn't have a high generality and the computational intricacy is relatively higher than filter method. The embedded approach employs supervised

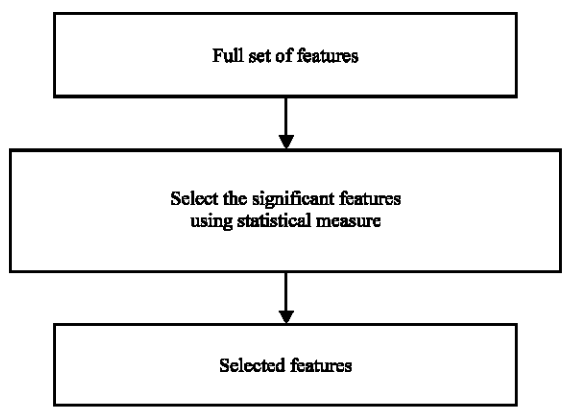

Fig. 1: Filter approach for feature selection

learning algorithm for feature selection process to obtain better accuracy. Hence, the embedded method does not have a high generality and involves high computational cost than the filter and cheaper than the wrapper method. In simple, the combination of wrapper and filter approach is known as embedded method (Saeys *et al.*, 2007).

In practical, the FS is ordered into two classes specific feature subset based method and feature ranking methods. The feature subset based method produces the

Corresponding Author: Sivakumar Venkataraman, Department of Computer Science and Information Systems, Faculty of Computing, Botho University, Botswana International University of Science and Technology (BIUST), Gaborne, Botswana

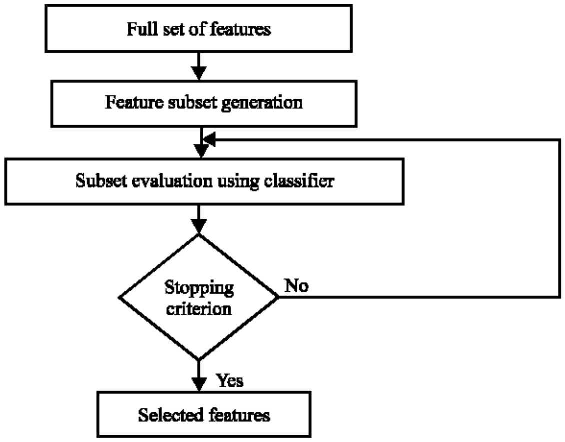

Fig. 2: Wrapper approach for feature selection

conceivable number of blends of the feature subsets utilizing any of the penetrating systems. By using the feature ranking methods, FS are independently calculated by any of the selection criteria. This way involves low space and calculational intricacy than feature subset selection method.

By using the feature ranker based method all independent features are ranked using a selection method like top rank feature method, information-gain method and gain-ratio method are suitable features having prerequisite threshold value. Hence, this method is more c computationally inexpensive and involves minimum space intricacy.

Towards, reducing computational complexity and to achieve high generality and also to overcome the problems related to the memory space, generality and higher classification accuracy, the proposed feature selection frame research is designed with two different mechanisms for relevancy and redundancy analysis in order to select the significantly relevant features by identifying and removing the redundant and irrelevant features from the high dimensional space.

Literature review: Reviews on various FS methods are provided and the procedure for on the features collectively evaluated to select the substantial features with the available dataset is analyzed.

The feature subset is evaluated by using some of the statistical measures under the filter method or by using the supervised learning algorithms under wrapper methods. Then the most significant subset is selected will have the more prominent features which helps in improving the classifier accuracy for the given datasets.

Generally, feature subset selection algorithm results in various possible groupings of the FS through one of searching approaches. By applying the statistical measures or with the help of learners the best one among the feature subset selected is found and the accuracy is predicted for those selected features. These selected features are treated as the most significant features and their corresponding subset is the most significant subset. The same procedure is followed for any kind of the datasets driven from their domain sources.

As the part of literature survey Correlation-based Feature Subset Selection (CRFS) is the finest sample for Feature Subset Based Method (FSB) developed by Hall (1999) is considered in our study. In this method, the correlation is predicted in two ways, feature-feature correlation and feature-class correlation.

Towards in determine the likely features happening in N number of features from FS the heuristic-based best-first search is applied. The subset which has maximum feature-class correlation is selected as the suitable FS to be used in the classification. The Consistency Based Feature Subset Selection (CBFS) method proposed by Liu and Setiono (1995) as the suitable feature subset selection approach. This technique utilizes the class consistency by calculation metric with a specific end goal to classify the suitable the FS from the data set. These stand alone approaches derive under the filter based techniques as not using the supervised learning algorithm in evaluating the FS.

On comparing with simple searching methods, exhaustive strategy is computationally quite expensive since to produce 2n number of feature subset. Some of the in-depth approaches are the Simulated Annealing (SA) (Lin *et al*., 2008a), Tabu Searching (TS), Ant Colony Optimization (ACO), Genetic Algorithm (GA) (Ghosh and Bagchi, 2009), Particle Swarm Optimization (PSO) (Lin *et al*., 2008b), etc. Number of investigators has suggested the same approach to achieve the minimum feature from minimum feature subset for this task by Lisnianski *et al*. (2010).

Some of the feature subset method uses heuristic searching in which the heuristic function is applied to obtain the preceding knowledge in directing the examination procedure for obtaining feature subsets. Then the feature subsets is intended through Supervised Machine Learning Algorithm (SMLA) and cost vise high than the wrapper approach.

Many researchers have implemented Tabu search based Feature Selection to obtain the FS for experiments. According to Lin *et al*. (2008a, b) suggested the SA and GA used to obtain the FS to calculated using SMLA such as Back Propagation Network (BPN) for selecting the best feature subset. In marketing, based applications simulated annealing-based feature selection is proposed by Meiri and Zahavi (2006). According to Zhang and Sun (2002) have proposed a Tabu search based feature selection as a method to obtain the feature subset and evaluated on the selected feature subset. By using Tabu search based feature selection (Tahir *et al*., 2007) generated the feature subsets for selecting the feature subsets and calculated by K-nearest neighbor classifier and it produces the classification errors to show the accuracy of the significant feature subset.

According to Kanan and Faez (2008) have applied the ant colony optimization as feature selection method as face recognition classification. By using this method, the adjacent neighbor classifier is accepted to test the obtained subset. Similarly, Sivagaminathan and Ramakrishnan (2007) for the medical diagnosis have applied ant colony optimization for selecting the features and accuracy is predicted with the help of Artificial Neural Networks (ANN).

GA in combination with artificial neural network to obtain the Electro Encephalogram (EEG) signal classification proposed by Erguzel *et al*. (2015). For mining the medical dataset genetic algorithm with Support Vector Machine (SVM) is employed by Welikala *et al.* (2015) to selecting the features. According to Li *et al*. (2011) proposed an GA combined by Support Vector Machine (SVM) method for hyper-spectral image classification. Credit risk assessment is analyzed by the genetic based feature selection algorithm with the neural network classifier by Oreski and Oreski (2014). According to Das *et al*. (2012) expressed an GA combined by SVM for hand-written digit recognition. For data classification applications (Wang *et al*., 2011) used the GA to obtain the subset through SVM for evaluating the selected features.

According to Xue *et al*. (2013) intended PSO (Particle Swarm Optimization) as a feature selection for classification and evaluated through SVM. In land cover classification application, PSO-based feature selection remains active (Yang *et al*., 2012) for feature subset selection. FS approaches using PSO for sleep disorder diagnosis system proposed by Chen *et al*. (2012).

The high dimensional datasets includes thousands of features for analysis. Handling huge dataset by using the common methods not appropriate. This methods trail an wrapping method which is cost vise high, similarly, produces best classification accurateness. But letdowns to attain the maximum generalization.

In the feature-ranker based approach, statistical or information-theoretic measures are used to select each feature of the dataset and the feature ranking method are used to rank the features based on weightage. From the ranking, the highly ranking features are used as substantial feature. Chi-square-based Feature Selection (CQFS) is the best example for this approach. As Liu and Setiono (1995) suggested Chi-Square Statistic Measure (CSSM) to estimate the feature weightages in ranking the features. Same way the Gainratio, information gain, SU-Symmetric Uncertainty used in finding the weightage for each feature which helps in ranking during selection process.

Additionally, the ranking methods uses the statistic measures or information-theoretic measures finds the weightage for each feature seeing the relation among them and the final group. As exposed by Bolon-Canedo *et al*. (2013), the ranking takes the minimum execution time, though does not eliminates the unrelated features.

In some conditions, ranking approaches implements an filter method due to the property of not involving the SLMA to test the suitable features. Accordingly, the above approaches are self-determining from SLMA, so, attains additionally over-all and minimum computational complexity. In identifying the best and appropriate features from huge datasets for redundancy analysis mechanism, the above methods are the best methods.

MATERIALS AND METHODS

Proposed Class Conditional Probability based Feature Selector (CCPFS): The proposed method is developed using Naive Bayes approach. For example, data set $D = \{X_1, X_2, X_3, ..., X_n\}$ where the sample occurrence are presented as $X_i = \{X_{i1}, X_{i2}, X_{i3}, ..., X_{ih}\}$. Dataset D has attributes like $\{A_1, A_2, A_3, ..., A_n\}$. In the respectively attributes, the values for the attributes in A_i are $\{A_{i1}, A_{i2}, A_{i3}, ..., A_{ih}\}$. The data belong to a dataset of class $C = \{C_1, C_2, C_3, ..., C_m\}$. The Posteriori hypothesis $P(C_i|X)$ is calculated using the bayes theorem:

$$P(C_i|X) = \frac{P(X|C_i)\,P(C_i)}{P(X)}$$

$P(C_i)$ denotes the Class Preceding probability. This calculation is done $P(C_i) = |C_iD = |C_iD|/|D|$, $|C_i, D|$ denotes the amount of occurrences in Class C_i. This above equation can be written as:

$$\text{Posterior} = \frac{\text{Prior}\times\text{Likelihood}}{\text{Evidence}}$$

The attributes in the dataset are conditionally independent of one another. The following equations are used to compute $P(X|C_i)$:

$$P(X|C_i) = \prod_{k=1}^{n} P(xk|C_i)$$

$$P(X|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times, ..., \times P(x_n|C_i)$$

The Probability $P(A_1|C_i)$, $P(A_2|C_i)$, ..., $P(A_n|C_i)$ are estimated for every attribute in the dataset. If the values of the attribute are continuous then it has to be discretized before applying the algorithm. The features that are above threshold value are considered as relavant features. Remaining features are eliminated by considering them as irrelevant and redundant features. The selected features are class conditional independence. Thus, the reduced set of features is obtained. And the dataset with reduced features are classified using classification algorithms.

Algorithm:

Input: $D = \{A_1, A_2, ..., A_n\}$
Output: Reduced features
Method:

Step 1. for {individual Class C_i of dataset D}do
Step 2. Identify the preceding probabilities, P (C_i)
Step 3. Close for loop
Step 4. for {individual attribute value Aij of Dataset D} do
Step 5. Identify the class conditional Probabilities P(Aij|C_i)
Step 6. Close for loop
Step 7. If {class conditional probabilities is greater then threshold value} do
Step 8. have Aji in dataset
Step 9. Else
Step 10. eliminate Aji from dataset
Step 11. Do until {each attributes are covered}
Step 12. Apply various classifiers on original and reduced features
Step 13. Evaluate the performance of different classifiers on both original and reduced features

Inductive probability describing of an arrangement of conceivable foundations for a given watched occasion can be figured from learning of the probability of each cause and the conditional probability of the result of each cause.

The proposed algorithm is used to examine the relations of each paired features. Conditional probability is used to measure the dependency between pairs of attributes. Any attribute is defined as relevant if the class conditional probability satisfies a predefined threshold otherwise considered as irrelevant when the class conditional probability not satisfies a predefined threshold.

RESULTS AND DISCUSSION

The accuracy of the proposed algorithm are examined through the datasets available in UCI machine learning repository and the UCI KDD archive. The execution time and features selected by several traditional feature selectors including SU, GA, SU-GA and the proposed CCPFS were recorded. The overall performance of NB, C4.5, NN, SVM and K-NN on both original features and reduced features by each feature selection algorithm was recorded.

Table 1: Features selected through different FS algorithm

Dataset	All	SU	GA	SU-GA	Proposed CCPFS
Ionosphere	34	17	14	13	11
Soybean	35	18	22	16	15
Diabetes	8	4	4	3	3
Segment	19	10	8	4	4
Vote	16	8	4	2	2
Dermatology	35	18	22	11	10
Lung cancer	57	28	21	13	11
Wine	14	7	12	6	6
Hepatitis	20	10	11	8	7
Vehicle	19	10	11	4	4
Average	25.7	13.0	12.9	8.0	7.3

Table 2: Accuracy obtained through Naive-Bayes by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GAA	CCPFS
Ionosphere	82.6	87.2	90.3	90.6	91.20
Soybean	93.0	90.6	92.1	89.3	93.00
Diabetes	76.3	75.4	77.5	76.4	78.20
Segment	81.1	76.9	82.9	82.8	82.90
Vote	90.1	91.3	96.1	95.6	96.40
Dermatology	97.3	91.8	98.1	92.9	97.80
Lung cancer	50.0	68.8	75.0	78.1	78.10
Wine	97.2	96.1	97.2	97.8	97.80
Hepatitis	84.5	84.5	84.5	85.8	85.80
Vehicle	44.8	43.4	48.5	44.1	51.20
Average	79.7	80.6	84.2	83.3	85.24

Several FS algorithms such as SU, GA and SU-GA was used to select the features which drives more features that are selected by the proposed method using dependency analysis. Proposed method removes the dependent features and classifies only independent features that are adequate for effective data classification. This shows the FS by class conditional probability approach attains the maximum of dimensionality reduction dataset through selecting of the minimum number of relevant features.

It is inferred from the results that for the many datasets, the feature reduction through inductive probability confirms the best results with the datasets with the original number of features there by the proposed system are adequate for data classification Table 1.

Table 2 shows the accuracy of Naive-Bayes data classification on both original features. It also shows the reduced selected features by using the FS algorithm. It is observed that features selected through conditional probability improves the performance of Naive-Bayes algorithm. For the datasets such as ionosphere, diabets, segment, vote, lung cancer, wine, hepatitis, soybean and vehicle the proposed feature selector shows superior results over the conventional FS algorithms including SU, GA and SU-GA.

The average detection performance of various classification algorithms such as C4.5, NB, SVM, K-NN and NN are shown from Table 2-6. The proposed class conditional probability based feature selector followed by GA feature selector gives higher accuracy compared to other methods such as SU and SU-GA when the average accuracy value of each feature selection method calculated. The classification accuracy is significantly improved after the application of attribute reduction by class conditional probability to the original dataset. It is found that the proposed CCPFS is sufficient for effective data classification through selection of only the most relevant features rather than irrelevant and redundant features.

The running time over different datasets with the conventional and proposed feature selector is shown from Table 7-11. It is found that the average time taken for

Table 3: Accuracy obtained through C4.5 by the selected features using FS algorithm

Data set	Full set	SU	GA	SU-GAA	CCPFS
Ionosphere	91.5	91.7	92.6	92.0	92.00
Soybean	91.5	90.6	90.8	90.5	90.80
Diabetes	73.8	74.3	74.9	74.6	74.60
Segment	95.7	94.9	95.4	94.9	95.70
Vote	96.3	95.2	96.1	95.6	96.30
Dermatology	94.0	88.0	94.3	90.4	92.60
Lung cancer	50.0	62.5	62.5	62.5	62.50
Wine	93.8	94.4	93.8	94.4	94.40
Hepatitis	83.9	82.6	83.2	83.9	83.90
Vehicle	72.5	69.5	68.3	66.1	67.80
Average	84.3	84.4	85.2	84.5	85.06

Table 4: Accuracy obtained through SVM by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GA	CCPFS
Ionosphere	88.6	88.0	88.0	88.3	88.0
Soybean	93.7	93.4	93.5	93.4	93.5
Diabetes	77.3	76.0	76.8	75.9	76.8
Segment	91.9	90.2	89.4	82.1	89.4
Vote	96.1	95.6	95.6	95.6	95.6
Dermatology	95.4	84.1	97.3	98.2	97.3
Lung cancer	40.6	62.5	65.6	68.8	65.6
Wine	98.3	96.6	98.3	96.6	98.3
Hepatitis	85.2	83.2	83.2	83.2	83.2
Vehicle	74.3	66.7	58.0	46.7	58.0
Average	84.1	83.6	84.6	82.9	84.6

Table 5: Accuracy obtained through NN by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GA	CCPFS
Ionosphere	91.2	92.0	92.6	92.9	93.40
Soybean	93.4	93.0	94.0	93.3	93.40
Diabetes	75.4	77.2	75.5	76.4	76.40
Segment	96.7	96.3	94.5	91.5	94.50
Vote	94.7	95.2	95.9	95.4	95.90
Dermatology	96.2	92.1	96.4	93.7	96.40
Lung cancer	37.5	53.1	65.6	71.9	71.90
Wine	97.2	98.3	99.4	98.3	99.40
Hepatitis	80.0	83.9	83.9	82.6	83.20
Vehicle	81.7	70.1	71.1	64.2	81.70
Average	84.4	85.1	86.9	86.0	88.62

Table 6: Accuracy obtained through KNN by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GA	CCPFS
Ionosphere	84.6	86.3	86.6	84.0	86.60
Soybean	88.0	90.9	89.2	90.5	90.50
Diabetes	69.1	71.6	69.9	72.4	72.40
Segment	96.6	97.1	97.4	96.7	97.40
Vote	93.3	92.9	96.1	95.6	96.10
Dermatology	94.3	90.7	96.7	93.2	96.70
Lung cancer	40.6	53.1	62.5	65.6	62.50
Wine	98.9	96.6	97.8	96.1	97.80
Hepatitis	81.9	83.2	83.9	81.9	83.90
Vehicle	71.4	72.5	66.2	67.3	69.70
Average	81.9	83.5	84.6	84.3	85.36

Table 7: Running time on C4.5 by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GA	CCPFS
Ionosphere	0.08	0.01	0.010	0.010	0.010
Soybean	0.03	0.01	0.010	0.010	0.010
Diabetes	0.01	0.00	0.010	0.000	0.000
Segment	0.04	0.03	0.020	0.020	0.020
Vote	0.00	0.00	0.000	0.000	0.000
Dermatology	0.03	0.02	0.030	0.000	0.000
Lung cancer	0.00	0.00	0.000	0.000	0.000
Wine	0.01	0.01	0.010	0.010	0.010
Hepatitis	0.05	0.00	0.000	0.010	0.010
Vehicle	0.04	0.02	0.020	0.000	0.000
Average	0.029	0.01	0.011	0.006	0.006

Table 8: Running time on SVM by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GA	CCP
Ionosphere	0.000	0.080	0.080	0.010	0.010
Soybean	1.490	1.240	1.120	1.080	1.070
Diabetes	0.090	0.070	0.060	0.010	0.010
Segment	0.400	0.130	0.280	0.110	0.110
Vote	0.090	0.010	0.010	0.010	0.010
Dermatology	0.280	0.260	0.210	0.080	0.070
Lung cancer	0.020	0.010	0.020	0.010	0.010
Wine	0.020	0.020	0.020	0.020	0.020
Hepatitis	0.050	0.010	0.010	0.010	0.010
Vehicle	0.070	0.050	0.050	0.050	0.050
Average	0.251	0.188	0.186	0.139	0.137

Table 9: Running time on NB by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GA	CCPFS
Ionosphere	0.010	0.000	0.000	0.00	0.00
Soybean	0.000	0.000	0.000	0.00	0.00
Diabetes	0.010	0.000	0.000	0.00	0.00
Segment	0.020	0.010	0.000	0.00	0.00
Vote	0.000	0.000	0.000	0.00	0.00
Dermatology	0.010	0.000	0.010	0.00	0.00
Lung cancer	0.000	0.000	0.000	0.00	0.00
Wine	0.000	0.000	0.010	0.00	0.00
Hepatitis	0.000	0.000	0.000	0.00	0.00
Vehicle	0.010	0.000	0.000	0.00	0.00
Average	0.006	0.001	0.002	0.00	0.00

Table 10: Running time on NN by the selected features using FS algorithm

Data set	Full set	SU	GA	SU-GA	CCPFS
Ionosphere	2.270	0.650	0.540	0.430	0.390
Soybean	31.780	15.670	15.940	13.470	13.020
Diabetes	0.660	0.450	0.460	0.320	0.320
Segment	5.310	3.000	2.530	1.880	1.880
Vote	0.840	0.420	0.200	0.140	0.140
Dermatology	23.660	7.390	10.680	4.770	4.500
Lung cancer	5.750	1.430	0.780	0.360	0.270
Wine	0.290	0.150	0.230	0.130	0.130
Hepatitis	0.360	0.150	0.150	0.130	0.110
Vehicle	2.090	0.950	1.130	0.630	0.630
Average	7.301	3.026	3.264	2.226	2.139

Table 11: Running time on K-NN by the selected features using FS algorithm

Dataset	Full set	SU	GA	SU-GA	CCPFS
Ionosphere	0.00	0.00	0.00	0.00	0.00
Soybean	0.00	0.00	0.00	0.00	0.00
Diabetes	0.00	0.00	0.00	0.00	0.00
Segment	0.00	0.00	0.00	0.00	0.00
Vote	0.00	0.00	0.00	0.00	0.00
Dermatology	0.00	0.00	0.00	0.00	0.00
Lung cancer	0.00	0.00	0.00	0.00	0.00
Wine	0.00	0.00	0.00	0.00	0.00
Hepatitis	0.00	0.00	0.00	0.00	0.00
Vehicle	0.00	0.00	0.00	0.00	0.00
Average	0.00	0.00	0.00	0.00	0.00

Table 12: Average performance of different classification algorithms

Classification algorithms	All	SU	GA	SU-GA	CCPFS
NB	79.7	80.6	84.2	83.3	85.24
C4.5	84.3	84.4	85.2	84.5	85.06
SVM	84.1	83.6	84.6	82.9	84.60
NN	84.4	85.1	86.9	86.0	88.62
K-NN	81.9	83.5	84.6	84.3	85.36
Average	82.8	83.2	84.3	83.7	85.77

Table 13: Running time on selected features for each classification algorithm

Running time (msec)	DT	NB	SVM	NN	K-NN
Full set	0.029	0.006	0.251	7.301	0.00
SU	0.010	0.001	0.188	3.026	0.00
GA	0.011	0.002	0.186	3.264	0.00
SU-GA	0.006	0.000	0.139	2.226	0.00
CCPFS	0.006	0.000	0.137	2.139	0.00

attribute reduction with class conditional probability approach is much lower than the other traditional algorithms average. It can also be observed that for the proposed feature selector, the running times over different datasets are consistent which verifies its superior computational efficiency.

The detection performance of various classifiers happening in various sets is shown in Table 12. It is observed from the average accuracy that the feature reduction using class conditional probability increases the performance of Naive Bayes (NB) classifier, Nearest Neighbour (NN) classifier and K-Nearest Neighbors (K-NN) classifier. It is also observed from the individual accuracy values that feature reduction using the class conditional probability could sustain or improve the performance. This can highly reduce the features in dataset, also improve classification performance through with predominant attributes. In overall, the proposed feature selector is used to identify the dependent features and to eliminate the redundant features amongst them. An advanced feature reduction algorithm combined with class conditional probability is executed and the results are examined evaluated using various experiments through the comparison of the correlated feature selection algorithms as in literature. The execution period with the selected attributes on various classification algorithm is shown in Table 13.

CONCLUSION

In this research, a novel framework for feature selection is proposed using class conditional probability and their effectiveness is evaluated on the datasets taken from the UCI machine learning repository. Feature Selection process seeks to reduce the dataset dimension by analyzing and understanding the impact of its features and their accuracy is predicted with the help of the classification model. The objective of the proposed feature selection is to eliminate the redundant and irrelevant features. Usually, high dimensional data contains high degree of irrelevant and redundant information which may greatly degrade the performance of learning algorithms. Hence, it is always advisable to develop the induction algorithm for performing the feature selection to increase the classifiers accuracy. A new feature selector using class conditional probability is implemented and the performance of a number of supervised classification algorithms on the UCI machine learning repository databases was compared and it was found that the proposed feature selector performs better than the popular and computationally more expensive traditional feature selection algorithms in the literature.

REFERENCES

Bolon Canedo, V., N. Sanchez-Marono and A. Alonso-Betanzos, 2013. A review of feature selection methods on synthetic data. Knowl. Inf. Syst., 34: 483-519.

Chen, L.F., C.T. Su, K.H. Chen and P.C. Wang, 2012. Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis. Neural Comput. Appl., 21: 2087-2096.

Das, N., R. Sarkar, S. Basu, M. Kundu, M. Nasipuri and D.K. Basu, 2012. A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application. Applied Soft Comput., 12: 1592-1606.

Erguzel, T.T., S. Ozekes, O. Tan and S. Gultekin, 2015. Feature selection and classification of electroencephalographic signals: An artificial neural network and genetic algorithm based approach. Clin. EEG. Neurosci., 46: 321-326.

Ghosh, P. and M.C. Bagchi, 2009. QSAR modeling for quinoxaline derivatives using genetic algorithm and simulated annealing based feature selection. Curr. Med. Chem., 16: 4032-4048.

Hall, M.A., 1999. Correlation-based feature selection for machine learning. Ph.D. Thesis, University of Waikato, Hamilton, New Zealand.

Kanan, H.R. and K. Faez, 2008. An improved feature selection method based on Ant Colony Optimization (ACO) evaluated on face recognition system. Appl. Math. Comput., 205: 716-725.

Li, S., H. Wu, D. Wan and J. Zhu, 2011. An effective feature selection method for hyper spectral image classification based on genetic algorithm and support vector machine. Knowl. Based Syst., 24: 40-48.

Lin, S.W., K.C. Ying, S.C. Chen and Z.J. Lee, 2008b. Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Syst. Appl., 35: 1817-1824.

Lin, S.W., Z.J. Lee, S.C. Chen and T.Y. Tseng, 2008a. Parameter determination of support vector machine and feature selection using simulated annealing approach. Applied Soft Comput., 8: 1505-1512.

Lisnianski, A., I. Frenkel and Y. Ding, 2010. Multi-State System Reliability Analysis and Optimization for Engineers and Industrial Managers. Springer, New York, USA., ISBN:978-1-84996-319-0, Pages: 392.

Liu, H. and R. Setiono, 1995. Chi_2: Feature selection and discretization of numeric attributes. Proceedings of the 7th International Conference on Tools with Artificial Intelligence, November 5-8, 1995, IEEE, Herndon, Virginia, pp: 388-391.

Meiri, R. and J. Zahavi, 2006. Using simulated annealing to optimize the feature selection problem in marketing applications. Eur. J. Oper. Res., 171: 842-858.

Oreski, S. and G. Oreski, 2014. Genetic algorithm-based heuristic for feature selection in credit risk assessment. Exp. Syst. Appl., 41: 2052-2064.

Saeys, Y., I. Inza and P. Larranaga, 2007. A review of feature selection techniques in bioinformatics. Bioinformatics, 23: 2507-2517.

Sivagaminathan, R.K. and S. Ramakrishnan, 2007. A hybrid approach for feature subset selection using neural networks and ant colony optimization. Exp. Syst. Applic., 33: 49-60.

Tahir, M.A., P. Bouridane and F. Kurugollu, 2007. Simultaneous feature selection and feature weighting using Hybrid Tabu Search/k-nearest neighbor classifier. Pattern Recognation Lett., 28: 438-446.

Wang, Y., X. Chen, W. Jiang, L. Li and W. Li *et al*., 2011. Predicting human microRNA precursors based on an optimized feature subset generated by GA-SVM. Genomics, 98: 73-78.

Welikala, R.A., M.M. Fraz, J. Dehmeshki, A. Hoppe and V. Tah *et al*., 2015. Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. Comput. Med. Imaging Graphics, 43: 64-77.

Xue, B., M. Zhang and N. Browne, 2013. Particle swarm optimization for feature selection in classification: A multi-objective approach. IEEE. Trans. Cybern., 43: 1656-1671.

Yang, H., Q. Du and G. Chen, 2012. Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. IEEE. J. Sel. Top. Appl. Earth Obs. Remote Sens., 5: 544-554.

Zhang, H. and G. Sun, 2002. Feature selection using TABU search method. Pattern Recogn., 35: 701-711.