Indian Journal of Science and Technology, *Vol 9(4), DOI: 10.17485/ijst/2016/v9i4/87038, January 2016*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A Novel Clustering based Feature Subset Selection Framework for Effective Data Classification

Sivakumar Venkataraman[1*] , Subitha Sivakumar[1] and Rajalakshmi Selvaraj[2]

[1]Faculty of Computing, Botho University, Botswana and Department of Information Systems, BIUST, Botswana; Sivakumar.venkataraman@bothouniversity.ac.bw, subitha.sivakumar@bothouniversity.ac.bw

[2]Department of Information Systems, BIUST, Botswana; selvarajr@biust.ac.bw

Abstract

Background/Objectives: A novel feature selection framework using minimum variance method is proposed. The purpose of the proposed method is to reduce the computational complexity, reduce the number of initial features and increase the classification accuracy of the selected feature subsets. **Methods/Statistical Analysis:** The clusters are formed using minimum variance method. The process must be repeated for different pairs of records and voting is done on the different sets of cluster pairs. The cluster pair which has the maximum number of votes is chosen and the highest priority member is chosen from each cluster using information gain and removing the remaining attributes, thus reducing the number of attributes. **Findings:** The proposed feature selector is evaluated by comparing it with existing feature selection algorithms over 9 datasets from UCI and WebKb Datasets. The proposed method shows better results in terms of number of selected features, classification accuracy, and running time than most existing algorithms. **Improvements/Applications:** A new feature selector using minimum variance method is implemented and found that it performs better than the popular and computationally expensive traditional algorithms.

Keywords: Classification, Data Mining, Dimensionality Reduction, Feature Selection, Information Gain, Minimum Variance Method

1. Introduction

Data mining is the process of finding interesting patterns in data. Data mining often involves datasets with a large number of attributes. Many of the attributes in most real world data are redundant and/or simply irrelevant to the purposes of discovering interesting patterns. Attribute reduction selects relevant attributes in the dataset prior to performing data mining. This is important for the accuracy of further analysis as well as for performance. Because the redundant and irrelevant attributes could mislead the analysis, including all of the attributes in the data mining procedures not only increases the complexity of the analysis, but also degrades the accuracy of the result. For instance, clustering techniques, which partition entities into groups with a maximum level of homogeneity within a cluster, may produce inaccurate results. In particular, because the clusters might not be strong when the population is spread over the irrelevant dimensions, the clustering techniques may produce results with data in a higher dimensional space including irrelevant attributes. The usage of PC, laptop and database terms are grown quickly, information's are gathered in an unmatchable speed by human ability of records handling.

For data pre-processing, the Feature Selection (FS) practice in data mining is the best and repeatedly used techniques. FS decreases the annoying features, eliminates the unrelated features, and decreases the noisy features and the redundant features from the huge dataset, transports the direct possessions for the system, speedup the algorithm to perform better, through improving the predictive in accurateness and outcome. FS is a practice used to extract a subset of features from original set of features. The finest feature of a subset is defined by an evaluation principles or measures.

**Author for correspondence*

All individual feature subset are evaluated, compared along through the preceding best feature until a certain evaluation standard. This feature subset generation procedure and the evaluation procedure are repeated until the defined criterion is fulfilled. Then, the selected feature subset should be assessed through the previous facts or various investigations via artificial and/or actual dataset. FS are used in various fields in data mining such as classification, clustering, association rules, and regression. This paper focuses on feature selection algorithms for classification. Initially research was mainly focused on the FS for classification with categorized or branded data (supervised FS) for the available classes. After the modern changes in research, it has been proved the above process can be implemented through clustering in FS for uncategorized or unbranded data (unsupervised FS) for the classes.

Feature selection algorithms designed with different evaluation criteria broadly fall into three categories. The filter model, the wrapper model, and the hybrid model Wrappers utilize the learning machine of interest as a black box to score subsets of variable according to their predictive power. Filters select subsets of variables as a pre-processing step, independently of the chosen predictor. Embedded methods perform variable selection in the process of training and are usually specific to given learning machines. In this paper; we give an overview of the popularly used feature selection algorithms under a unified framework. Furthermore a novel FS algorithm has proposed based on the minimum variance technique for defining the dependent attributes and in eliminating independent attributes to reduce the number of attributes and to increase the predication accuracy and to reduce the time. Experiments and research on the original datasets illustrates the planned approach is encouraging and positive in relations of its effectiveness, efficiency compared with other state-of-art algorithms.

For research in Data Mining, Feature Selection is the growth collection to achieve. Abdul Razak et al. proposed data mining framework for Credit Card Fraud detection. Asma Feki, Anis Ben Ishak and Saber Feki[2] proposed a feature selection using Bayesian and multiclass Support Vector Machines approaches for bank risk prediction. Blum and Langley[3] classified the feature selection techniques into three basic approaches. In the first approach, known as the embedded approach, a basic induction method is used to add or remove features from the concept description in response to prediction errors on new instances. The second approach is known as the filtering approach, in which, various subsets of features are explored to find an optimal subset, which preserves the classification. The third approach is known as wrapper methods which evaluate alternative feature sets by running some induction algorithm on the training data and using the estimated accuracy of the resulting classifier as its metric. A number of feature selection techniques based on the evolutionary approaches have also been proposed. Casillas et al.[4] presented a genetic feature selection technique which is integrated into a multi-stage genetic learning process to obtain a Fuzzy Rule Based Classification System (FRBCS). In the first phase of this method, a filtering approach is used to determine an optimal feature subset for a specific classification problem using class-separability measures. This feature subset along with expert opinion is used to obtain the adequate feature subset cardinality in the second phase which is used as the chromosome length.

Hu Huang et al.[5] proposed an ant colony optimization-based feature selection method for surface electromyography signals classification.

Kohari, John et al.[6] proposed another feature selection framework known as the wrapper technique. The wrapper methods evaluate alternative feature sets by running some induction algorithm on the training data and using the estimated accuracy of the resulting classifier as its metric. The major disadvantage of the wrapper approach is that it requires much computation time.

YouShyang Chen[7] classified credit ratings for Asian Banks using integrating feature selection and the CDPA-based rough sets approach. Jung Hwan Cho and Pradeep U. Kurup[8] proposed a dimensionality reduction method on electronic nose data.

Sombut Foithong, Ouen Pinngern and Boonwat Attachoo[9] proposed a feature subset selection wrapper based on mutual information and rough sets.

A number of feature selection techniques based on the evolutionary approaches have also been proposed. Kira and Rendell[10] proposed a different approach to feature selection and the filter based feature ranking algorithm (RELIEF) also proposed by them assigns a weight to each feature based on the ability of the feature to distinguish among the classes and then selects those features whose weights exceed a user defined threshold as relevant features. The weight computation is based on the probability of the nearest neighbors from two different classes having different values for an attribute and the probability of two nearest neighbors of the same class having the same value of the attribute. The higher the difference between these two probabilities, the more significant is the attribute. Inherently, the measure is defined for a two-class problem which can be extended to handle multiple classes, by splitting the problem into a series of two-class problems. Kononenko[11] suggested to use k-nearest neighbors to increase the reliability of the probability approximation. It also suggested how RELIEF can be extended to work with multiple sets more efficiently. Weighting schemes are easier to implement and are preferred for their efficiency.

Learning to classify objects is an inherently difficult problem for which several approaches like instance-based learning or nearest neighbor-based algorithms are used. However, the nearest neighbor algorithms need some kind of distance measure. Cost and Salzberg[12] emphasized the need to select appropriate metrics for symbolic values. Stanfill and Waltz[13] proposed the Value Difference Metric (VDM) which measures the distance between values of symbolic features. It takes into account the overall similarity of classification of all instances for each possible value of each feature. Based on this, Cost and Salzberg proposed the Modified Value Distance Metric (MVDM) which is symmetric, and satisfies all the metric properties. They showed that nearest neighbor algorithms perform well even for symbolic data using this metric. It is observed that distance-values are similar if the pairs occur with the same relative frequency for all classes. Zhao and Tsang[14] proposed an attribute reduction with fuzzy approximation operators. Sharma and Paliwal[15] proposed a rotational linear discrimination analysis technique for dimensionality reduction which is a supervised learning technique that finds a linear transformation such that the overlap between the classes is minimum for the projected feature vectors in the reduced feature space.

2. Proposed work

This paper introduces a novel method for Feature Selection in a huge collection of dataset using a minimum variance method. In the dataset, the dependent and unrelated attributes are taken-out by expending the novel techniques. Dependency between attributes are calculated by first grouping them into clusters using minimum variance method and then using information gain to find the highest ranked attribute among the cluster members. The suggested method provides the good outcome in selecting the number of designated attributes, accuracy in classification method, and provides the shortest execution time compared to most other algorithms.

The variance method fails to calculate the distance among the groups or clusters, somewhat, this method groups by mounting the clusters within the homogeneity. The sum of squares in the mounted group or within-clusters helps to measure the amount of homogeneity. That is, the variance method tries to minimize the total within-group or within cluster sum of squares. The clusters remains to generated at every stage to produce the resulting cluster where to have least within-cluster sums of squares, also known as the Error Sums of Squares (ESS) and generalized as follows. The dataset with 5 features is shown in Table 1. The computation steps of ward's minimum variance method is shown in Figure 1.

First need to compute E for each of the ten possible mergers. Take the first one: (12),3,4,5 Calculate the cluster mean for (12) = mean(12) = [mean(10,20), mean(5,20)] = [15,12.5]

Table 1. General structure of dataset

X	1	2	3	4	5
X1	10	20	30	30	5
X2	5	20	10	15	10

Step	Possible paritions	E
1	(12) 3 4 5	162.5
	(13) 2 4 5	212.5
	(14) 2 3 5	250.0
	(15) 2 3 4	25.0
	(23) 1 4 5	100.0
	(24) 1 3 5	62.5
	(25) 1 3 4	162.5
	(34) 1 2 5	12.5
	(35) 1 2 4	312.5
	(45) 1 2 3	325.0
2	(34) (12) 5	175.0
	(34) (15) 2	37.5
	(34) (25) 1	175.0
	(134) 2 5	316.7
	(234) 1 5	116.7
	(345) 1 2	433.3
3	(234) (15)	141.7
	(125)(34)	245.9
	(1345) 2	568.8

Figure 1. Procedure Minimam variance method.

For the first possible merger the value of E is

E = (10 – 15)2 + (5 – 12.5)2 + (20 – 15)2 + (20 – 12.5)2
(30 – 30)2 + (10 – 10)2 + (30 – 30)2 + (15 – 15)2
(5 – 5)2 + (10 – 10)2 = 162.5

The general structure of a training set is shown in Table 1. The predictor attribute a1 can take values {a11, a12, · · ·, a1n}, a2 can take values {a21, a22, · · ·, a2n},· · ·, an can take values {an1, an2, · · ·, ann} and the class attribute c can take the values {c1, c2, · · ·, cn}

The main steps of the proposed algorithm are given below.

- Let A = {a1, a2, a3, · · ·, an} be the initial set of attributes and a1 = {a11, a12, · · ·, a1n}, · · ·, an = {an1, an2, · · ·, ann}.
- Group the similar attributes using wards method.

$$e_X^2 = \overset{nk}{\underset{i=1}{å}} \overset{p}{\underset{j=1}{å}} [y_{ij}^{(kt)} - m_j^{(kt)}]$$

- Apply voting on the combination of clusters and the combination which has the maximum votes are considered relevant.
- Calculate information gain for the cluster members and select the higher ranked cluster members as most relevant attribute.

The proposed algorithm is enumerated as follows:

Step 1:
Let the Initial set of attributes be A = {a1, a2, a3, · · ·, an}, where a1 = {a11, a12, · · ·, a1n }, a2 = {a21 , a22 , · · ·, a2n}, · · ·, an = {an1, an2, · · · , ann} and class attribute c = {c1, c2, · · ·, cn}

Step 2:
Mean11 = {a11+a12}/2 Mean12 = {a12+a13}/2 … Mean lm={alm+a1m+1}/2

Mean21 = {a21+a22}/2 Mean12 = {a22+a23}/2 … Mean 2m={a2m+a2m+1}/2

..

Meann1 = {an1+an2}/2 Meann2 = {an2+an3}/2 … Mean nm={anm+anm+1}/2

E11 = (a11-Mean11)2 + (a12 – Mean11)2 +(a21-Mean21)2 + (a22-Mean21)2

E12 = (a12-Mean12)2 + (a13 – Mean12)2 +(a22-Mean22)2 + (a23-Mean22)2

… E1m = (a1m-1-Mean1m)2 + (a1m – Mean1m)2 +(a2m-1-Mean2m)2 + (a2m-1-Mean2m)2

..

En1 = (an,1-Meann,1)2 + (an,2 – Meann,1)2 +(an+1,1-Meann+1,1)2 + (an+1,2-Meann+1,1)2

… Enm = (an, m-1-Meann, m-1)2 + (an, m–Meann, m-1)2 + (an+1, m-1-Meann+1, m-1)2 + (an+1, m - Meann+1, m-1)2

Calculate the minimum Eij value,

i ->1 to n
j ->1 to m

minimum Eij value fixed and compare Eij with remaining attributes

The dataset with 5 attributes has the minimum ESS value as E12. The resulting cluster groups are as follows

E12 minimum value m = 5

- (E12)(E3)(E45)
- (E12)(E34)(E5)
- (E123)(E4)(E5)

Step 3:
Apply voting on the combinations and the combination which gets the maximum votes are considered as relevant attributes.

(E12)(E3)(E45) -> X
(E12)(E34)(E5) -> Y
(E123)(E4)(E5) -> Z

Where, X, Y, Z are integers.

If X>Y>Z

The final cluster is (E12)(E3)(E45)

Step 4:
Calculate the information gain for the clusters obtained with minimum variance method. Choose the attribute with highest priority as relevant attribute.

(E12) 1 >2
(E3)
(E45) 5>4

Selected attribute: (1, 5)
Removed attribute: (2, 3, 4)

3. System Implementation

The proposed algorithm is implemented using Java. The stepwise approach is as follows.

- The input to the system is given as a text file format. The results are the clusters formed.
- The implemented java file also produces the number of votes produced in each combination of cluster.

- The combination which has got the maximum votes are given as input to WEKA (Weikato Environment For Knowledge Analysis).
- The highest ranked cluster member is chosen using information gain available in WEKA.
- The classifier accuracies from various classifiers for each feature selection method are recorded in tables from the results got from WEKA.

4. Experimental results and discussion

The feature selection using minimum variance method is applied to many datasets, and the performance evaluation is done. We presented the performance evaluation on both UCI and WebKB datasets. The general structure of dataset's used in the experiment is shown in Table 2.

We applied wards minimum method to each dataset and ran all traditional feature-selection algorithms including wrapper sub set evaluation, consistency subset evaluation, Info Gain attribute evaluation, Gain Ratio attribute evaluation, One R attribute evaluation, principal components, classifier subset evaluation, respectively, and recorded the number of selected features by each feature selection algorithm. Number of Selected and Removed features by each conventional featute selection method is shown in Table 3 and Table 4 respectively. Attribute Clustering with proposed variance method is shown in Table 5. Number of selected and removed attributes by proposed varaiance and information gain baesd feature selector is shown in Table 6.

Accuracy of classifiers with different feature selectors such as proposed variance and Information Gain, Principal Component, One R attribute Evaluation, Relief, Gain Ratio, Symmetric Uncertain, Information Gain are shown in Table 7 to Table 13 respectively.

Table 2. Data set description

Dataset	No.of aattributes	No.of Instances
Webkb	6	56
Webkb2	7	92
Webkb3	14	291
Webkb4	10	298s
Webkb5	15	414
Webkb6	14	391
Webkb7	16	432
Webkb8	18	557
Webkb9	18	585
Glass	10	214
Iris	5	150
Diabetes	9	768

Table 3. No. of selected features by each feature selection method

Dataset	Initial No. Of Attributes	Cfs	Princ. Compts.	Info Gain	Relief	Gain Ratio	One R Attribute Eval	Wrapper Subset	Symmertic Uncert Attribute Eval	Proposed Variance Method
WEB KB	5	5	5	5	5	5	5	5	5	4
WEB KB 2	6	3	6	6	6	6	6	5	6	3
GLASS	10	7	6	9	9	9	9	5	9	7
IRIS	5	2	2	4	4	4	4	4	4	3
DIABETES	9	4	8	8	8	8	8	5	8	7
WEB KB3	14	12	12	13	13	13	13	13	13	12
WEB KB4	10	9	8	9	9	9	9	9	9	9
WEB KB5	15	13	13	14	14	14	14	13	14	13
WEB KB6	14	12	12	13	13	13	13	13	13	12
WEB KB7	16	13	14	15	15	15	15	15	15	14
WEB KB8	18	15	16	17	17	17	17	17	16	16
WEB KB9	18	16	17	17	16	17	17	17	16	16
AVERAGE	11.66	9.25	9.91	10.83	10.75	10.83	10.83	10.08	10.66	9.66

Table 4. Number of removed attributes by each feature selection algorithms

Dataset	No of Attribute	Cfs	Princ. Compts.	Info Gain	Relief	Gain Ratio	One R Attribute Eval	Wrapper Subset	Symmertic Uncert Attribute Eval	Proposed variance method
WEB KB	5	-	-	-	-	-	-	-	-	1
WEB KB 2	6	2,3,6	-	-	-	-	-	5	-	2,5
GLASS	10	2,5,10	7,8,9,10	10	10	10	10	10	10	8,10
IRIS	5	1,2,5	3,4,5	5	5	5	5	5	5	2,5
DIABETES	9	4	8	8	8	8	8	5	8	13,14,17
WEB KB3	14	4,14	13,14	14	14	14	14	14	14	3,14
WEB KB4	10	10	9,10	10	10	10	10	10	10	2
WEB KB5	15	4,15	14,15	15	15	15	15	13,15	15	13,15
WEB KB6	14	6,14	13,14	14	14	14	14	14	14	14
WEB KB7	16	5,11,16	15,16	16	16	16	16	16	16	10,11,16
WEB KB8	18	1,6,18	17,18	18	18	18	18	18	18	17,18
WEB KB9	18	18	18	18	18	18	18	18	18	12,18

Table 5. Attribute clustering with minimum variance method

Dataset	No. of Attributes (Without Class)	Cluster 1	Cluster 2	Cluster N
WEBKB	5	(1)	(2,3,4,5)	----
WEBKB2	6	(2,4,5,6)	Remaining Single Clusters	Remaining Single Clusters
WEBKB3	13	(11,13)	(2,3,12)	Remaining Single Clusters
WEBKB4	9	(4,5,7,9)	Remaining Single Clusters	Remaining Single Clusters
WEBKB5	14	(3,4,5,14)	Remaining Single Clusters	Remaining Single Clusters
WEBKB6	13	(2,3,4,13)	Remaining Single Clusters	Remaining Single Clusters
WEBKB7	15	(2,4,5,15)	Remaining Single Clusters	Remaining Single Clusters
WEBKB8	17	(1,4,5,17)	Remaining Single Clusters	Remaining Single Clusters
WEBKB9	17	(1,4,5,17)	Remaining Single Clusters	Remaining Single Clusters
DIABETES	8	(4,6,8)	(1,7)	Remaining Single Clusters
GLASS	9	(7)	(1,4,8,9)	Remaining Single Clusters
IRIS	4	(2,4)	(1,3)	----

Table 6. Number of selected and removed attributes by proposed variance and information gain feature selector

Dataset	No. of Attributes (without class)	Removed Attributes	Selected Attributes
WEBKB	5	(2,3,4)	(1,5)
WEBKB2	6	(2,4,6)	(1,3,5)
WEBKB3	13	(2,3,11)	(1,4,5,6,7,8,9,10,12,13)
WEBKB4	9	(5,7,9)	(1,2,3,4,6,8)
WEBKB5	14	(3,4,5)	(1,2,6,7,8,9,10,11,12,13,14)
WEBKB6	13	(2,3,13)	(1,4,5,6,7,8,9,10,11,12)
WEBKB7	15	(2,4,5)	(1,3,6,7,8,9,10,11,12,13,14,15)
WEBKB8	17	(1,4,5)	(2,3,6,7,8,9,10,11,12,13,14,15, 16,17)
WEBKB9	17	(1,4,5)	(2,3,6,7,8,9,10,11,12,13,14,15, 16,17)
DIABETES	8	(4,6,7)	(1,2,3,5,8)
GLASS	9	(1,4,8)	(2,3,5,6,7,9)
IRIS	4	(1,2)	(3,4)

Our findings are summarized as follows:

- Minimum variance method performs well than most traditional feature selection methods such as Principal Components, Information gain, relief, gain ratio, One R Attribute Evaluation, Wrapper Subset Evaluation, Symmetric uncertain attribute evaluation in terms of classifier accuracy and the no. of attributes reduced. The reduced no of attributes using minimum variance and information gain is about 17.15% which is more compared to leading attribute selection methods like principal components which is 15%.

Table 7. Accuracy of classifiers (variance + information gain as feature selector)

DATASETS	Naive Baysien	K star	Multilayer Perceptron	J 48
WEBKB	100	98.2143	100	100
WEBKB 2	88.0435	96.7391	96.7391	96.7391
GLASS	41.5888	66.3551	65.4206	61.215
IRIS	96	95.3333	95.3333	96
CLDIABETES	73.0469	71.224	74.4792	72.1354
WEBKB3	91.7526	90.0344	89.6907	89.0034
WEBKB4	90.2685	91.9463	91.2752	91.9463
WEBKB5	95.6522	95.1691	95.6522	91.5459
WEBKB6	94.6292	96.6752	95.9079	94.1176
WEBKB7	94.6759	95.1389	95.1389	92.1296
WEBKB8	95.5117	95.1526	95.3321	91.921
WEBKB9	94.7009	94.7009	95.5556	92.9915
AVERAGE	87.9891	90.5569	90.8770	89.1454

Table 8. Accuracy of classifiers (principal component as feature selector)

Data Sets	Naive Baysien	K star	Multilayer Perceptron	J 48
WEBKB	73.00	87.50	87.50	87.50
WEBKB2	83.6	88.05	84.78	89.13
GLASS	41.12	75.23	65.34	69.67
IRIS	96.00	94.66	95.33	94.76
DIABETES	74.56	69.14	73.97	73.92
WEBKB3	93.42	91.75	92.43	91.75
WEBKB4	90.26	91.99	91.27	91.94
WEBKB5	95.16	96.61	94.92	92.02
WEBKB6	95.45	94.88	96.16	94.88
WEBKB7	93.28	96.03	96.06	93.28
WEBKB8	94.56	94.97	95.69	93.53
WEBKB9	94.01	94.07	94.70	92.64
AVERAGE	85.36	89.57	89.01	88.75

Table 9. Accuracy of classifiers (one r attribute evaluation as feature selection method)

Data Sets	Naive Baysien	K star	Multilayer Perceptron	J 48
WEBKB	100	98.2	100	100
WEBKB2	88.02	97.73	96.73	96.73
GLASS	49.08	68.54	65.34	67.43
IRIS	92.00	92.44	94.00	94.00
DIABETES	73.23	71.67	74.21	73.92
WEBKB3	91.65	90.75	89.06	89.02
WEBKB4	90.26	91.99	91.27	91.94
WEBKB5	95.16	95.67	95.93	91.34
WEBKB6	94.56	96.57	95.98	94.88
WEBKB7	93.28	96.03	96.06	93.28
WEBKB8	95.67	96.23	96.34	91.53
WEBKB9	94.01	94.07	94.70	92.64
AVERAGE	88.07	90.82	90.8	89.72

Table 10. Accuracy of classifiers (relief as feature selection method)

Data sets	Naive Baysien	K star	Multilayer Perceptron	J 48
Webkb	100	96.4286	100	100
Webkb2	93.4783	94.5652	96.7391	96.7391
Glass	48.5987	75.2336	68.2243	66.8224
Iris	96	94.6667	92.6667	96
Diabetes	93.8462	95.2137	95.3864	92.6496
Webkb3	94.1581	96.2199	89.6907	92.0962
Webkb4	90.9396	91.6107	92.6174	90.604
Webkb5	95.8937	95.6522	91.0625	94.4444
Webkb6	96.4194	97.1857	91.5601	93.3504
Webkb7	93.15	96.0648	96.2963	92.1296
Webkb8	95.1526	95.3321	91.0233	92.9982
Webkb9	94.7009	96.0684	92.1368	93.5043
Average	91.023	93.675	91.445	91.773

- In feature selection approach, we have shown that the minimum variance method is a promising approach for automatic feature selection. The classification accuracy by minimum variance as a feature selector is 89.6420675% which is greater than the traditional feature selection methods like principal components which showed 88.2225% and One R Attribute evaluation method which showed 89.5525%.
- Thus, have implemented a new feature selector using minimum variance method and found that it performs better than the popular and computationally expensive traditional algorithms.

Table 11. Accuracy of classifiers (gain ratio as feature selection method)

Data sets	Naive Baysien	K star	Multilayer Perceptron	J 48
Webkb	100	96.4286	100	100
Webkb2	93.4783	94.5652	96.7391	96.7391
Glass	48.5987	75.2336	68.2243	66.8224
Iris	96	94.6667	92.6667	96
Diabetes	76.8299	75.9115	73.8291	71.875
Webkb3	94.1581	96.2199	89.6907	92.0962
Webkb4	96.6443	92.7835	92.4399	91.4089
Webkb5	95.8937	95.6522	91.0625	94.4444
Webkb6	97.3734	95.1407	94.8849	92.4399
Webkb7	93.9815	96.7593	92.8291	94.4444
Webkb8	95.1526	95.3321	91.0233	92.9982
Webkb9	94.7009	96.0684	92.1368	93.5043
Average	90.23	92.058	89.621	90.225

Table 12. Accuracy of classifiers (symmetric uncertain attribute evaluation as feature selection method)

Data sets	Naive Baysien	K star	Multilayer Perceptron	J 48
Webkb	100	96.4286	100	100
Webkb2	93.4783	94.5652	96.7391	96.7391
Glass	48.5987	75.2336	68.2243	66.8224
Iris	96	94.6667	92.6667	96
Diabetes	76.3021	69.1406	71.229	73.8281
Webkb3	94.1581	96.2199	89.6907	92.0962
Webkb4	96.6443	94.2953	94.2913	95.6376
Webkb5	95.8937	95.6522	91.0625	94.4444
Webkb6	96.4194	97.1857	91.5601	93.3504
Webkb7	93.9815	96.7593	96.7593	94.4444
Webkb8	95.1526	95.3321	91.0233	92.9982
Webkb9	94.7009	96.0684	92.1368	93.5043
Average	90.106	91.790	89.610	91.150

Table 13. Accuracy of classifiers (information gain as feature selection method)

Data sets	Naive Baysien	K star	Multilayer Perceptron	J 48
Webkb	100	96.4286	100	100
Webkb2	93.4783	94.5652	96.7391	96.7391
Glass	48.5987	75.2336	68.2243	66.8224
Iris	96	94.6667	92.6667	96
Diabetes	76.3021	69.1406	71.229	73.8281
Webkb3	94.1581	96.2199	89.6907	92.0962
Webkb4	92.4399	94.2953	94.2913	95.6376
Webkb5	95.8937	95.6522	91.0625	94.4444
Webkb6	96.4194	97.1857	91.5601	93.3504
Webkb7	93.9815	96.7593	92.8291	94.4444
Webkb8	95.1526	95.3321	91.0233	92.9982
Webkb9	94.7009	96.0684	92.1368	93.5043
Average	89.755	91.785	89.289	90.817

5. Conclusion

This paper proposes a novel feature selection algorithm using minimum variance method. The algorithm can remove redundancy from the original dataset. The main idea provided is to find the dependent attributes from a cluster and remove the other members in the cluster. The technology to obtain the clusters is based on minimum variance method. A new attribute reduction algorithm of using minimum variance method is implemented and evaluated through extensive experiments via comparison with related attribute reduction algorithms.

6. References

1. Abdul RT, Najeeb AG. Detecting credit card fraud using data mining techniques-meta learning. Indian Journal of Science and Technology. 2015; 8(28):1–9.
2. Feki A, Ben IA, Feki S. feature selection using bayesian and multiclass support vector machines approaches: Application to bank risk prediction. Expert Systems with Applications. 2011; 39(3):3087–99.
3. Blum AL, Langley P. Selection of relevant features and examples in machine learning. Artificial Intelligence. 1997; 97(1-2):245–71.

4. Casillas J, Gordon O, Del Jesus MJ, Herrera F. Genetic feature selction in a fuzzy rule-based classification system learning process for high dimensional problems. Information Sciences. 2001; 36(1):135–57.
5. Huang H, Xie HB, Guo J-Y, Chen H-J. Ant colony optimization–based feature selection for surface electromyography signals classification. Computers in Biology and Medicine. 2011; 42(1):30–8.
6. Kohavi R, John GH. Wrappers for feature subset selection. Artificial Intelligence. 1997; 97(1-2):273–324.
7. Chen Y-S. Classifying credit ratings for asian banks using integrating feature selection and CPDA-based rough sets approach. Knowledge-Based Systems. 2011; 26:259–70.
8. Cho JH, Kurup PU. Decision tree approach for classification and dimensionality reduction of electronic nose data. Sensors and Acutators B: Chemical. 2011; 160(1): 542–8.
9. Foithong S, Pingern O, Atachoo B. Feature subset selection wrapper based on mutual information and rough sets. Expert Systems with Applications. 2011; 39(1):574–84.
10. Kira K, Rendell LA. The feature selection problem: Traditional methods and a new algorithm. Proceedings of the 10th National Conference on Artificial Intelligence. San Jose, California, USA: MIT Press; 1992. p. 129–34.
11. Kononenko L. Estimating attributes: Analysis and extensions of relief. Proceedings of Europe International Conference on Machine Learning. New York, USA: Springer-Verlag; 1994. p. 171–82.
12. Cost S, Salzberg S. A weighted nearest algorithm with symbolic features. Machine Learning. 1993; 10(1):57–8.
13. Stan C, Waltz D. Towards memory based reasoning. Communications of the ACM. 1986; 29(12):1213–28.
14. Zhao S, Tsang ECC. On fuzzy approximation operators in attribute reduction with fuzzy rough sets. Information Sciences. 2008; 178(16):3163–76.
15. Sharma A, Paliwal KK. Rotational linear discriminate analysis technique for dimensionality reduction. IEEE Transaction on Knowledge and Data Engineering. 2008; 20(10):1336–47.