

Analysis of Feature Selection Algorithms on Classification: A Survey

S.Vanaja
Research Scholar
Bharathiar University, Coimbatore
Tamil Nadu, India

K.Ramesh kumar
Research Supervisor (Part-Time – External)
Bharathiar University, Coimbatore
Tamil Nadu, India

ABSTRACT

The aim of this paper is to discuss about various feature selection algorithms applied on different datasets to select the relevant features to classify data into binary and multi class in order to improve the accuracy of the classifier. Recent researches in medical diagnose uses the different kind of classification algorithms to diagnose the disease. For predicting the disease, the classification algorithm produces the result as binary class. When there is a multiclass dataset, the classification algorithm reduces the dataset into a binary class for simplification purpose by using any one of the data reduction methods and the algorithm is applied for prediction. When data reduction on original dataset is carried out, the quality of the data may degrade and the accuracy of an algorithm will get affected. To maintain the effectiveness of the data, the multiclass data must be treated with its original form without maximum reduction, and the algorithm can be applied on the dataset for producing maximum accuracy. Dataset with maximum number of attributes like thousands must incorporate the best feature selection algorithm for selecting the relevant features to reduce the space and time complexity. The performance of Classification algorithm is estimated by how accurately it predicts the individual class on particular dataset. The accuracy constrain mainly depends on the selection of appropriate features from the original dataset. The feature selection algorithms play an important role in classification for better performance. The feature selection is one of the preprocessing techniques in the classification. This research paper deals with different feature selection algorithms and their performance on different dataset.

General Terms

Feature selection algorithm, Medical dataset, Accuracy, Filter method, Wrapper method, embedded method.

Keywords

Classification, Binary class, Multiclass, feature selection algorithm, Medical dataset, high dimensional dataset.

1. INTRODUCTION

Data mining is a process of knowledge discovery. The KDD is an automated process of knowledge discovery from the original data. The KDD consists of many steps like data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation. Among the steps the data selection is very much important to select the relevant feature and remove the irrelevant attributes. Classification is one of the datamining techniques used to discover the unknown class.

The different classification methods in data mining are Bayesian classification (Statistical classifier), Decision tree

induction, and Rule based classification (IF THEN Rule), Classification using Back propagation (Neural network algorithm), Support vector machine, Classification using Association Rule, k-nearest neighbor classifiers, case-based reasoning classifiers, Rough set approach, Genetic algorithm, Fuzzy set approach.

2. PROBLEM STATEMENT

Complex data analysis and mining of huge amount of data can take a long time and is infeasible. So the Data reduction techniques such as Data cube aggregation, Attribute subset selection, Dimensionality reduction, Numerosity reduction, Data discretization and Concept hierarchy generations are used. When the user applies any one of the data reduction methodologies it must reduce the dataset without missing any valuable data.

Another problem in classification is an accuracy of the classifier. The accuracy of the classifier depends not only with the classification algorithm but also on the feature selection method. Selection of irrelevant and inappropriate features may lead to increase the complexity. The feature selection method plays the major role which increases the efficiency of classification. Though different kinds of feature selection methods are available, for selecting an appropriate features, the best algorithm should be chosen to maximize the accuracy of the classification and also the feature selection algorithm should consumes less space and time for it better performance. The greatest requirement of the feature selection algorithm is to support the high dimensional multiclass dataset.

The recent researches on medical datamining are carried out in the following categories. By using the finite number of attributes, the presence of disease is found out which results as binary class. The real world data can also be used for prediction of disease as well as to find out their type which leads to multiclass like kinds of cancer, kinds of tumor.

When real world datasets are used, the researches concentrate on preprocessing of the data, the best feature selection methodology to be used to filter out the irrelevant and redundant features and the classification accuracies are estimated. The best feature selection algorithm is one which supports for both binary and multiclass dataset as well as it should improves the accuracy of the classifier and also reduce the time and space requirement.

3. FEATURE SELECTION ALGORITHMS

The feature selection algorithm removes the irrelevant and redundant features from the original dataset to improve the classification accuracy. The feature selections also reduce the dimensionality of the dataset; increase the learning accuracy, improving result comprehensibility.

The feature selection avoid over fitting of data. The feature selection also known as attributes selection which is used for best partitioning the data into individual class. The feature selection method also includes the selection of subsets, evaluation of subset and evaluation of selected feature.

The two search algorithms *forward selection* and *backward eliminations* are used to select and eliminate the appropriate feature. The feature selection is a three step process namely search, evaluate and stop.

Different kinds of feature selection algorithms have been proposed. The feature selection techniques are categorized into three *Filter method*, *Wrapper method*, and *Embedded method*. Every feature selection algorithm uses any one of the three feature selection techniques.

According to the class label present or not the feature selection can be further classified into two categories. Supervised and unsupervised feature selections. In the supervised method the relevance between the feature and the class is evaluated by calculating the correlation between the class and the feature. The relevance is evaluated by checking some property of the data in an unsupervised method.

When we consider the microarray datasets it has thousands of features and also has high dimensional dataset. The feature selection algorithm plays an important role to maximizing the performance of such high dimensional datasets. To maximize the accuracy of high dimensional dataset we must follow the steps:

- step1. Select the appropriate feature selection method
- step2. Select the suitable classification algorithm

Originally we have two categories in datasets, Binary and multiclass datasets. The selected feature selection method and the classification algorithm must support to classify the data into both binary and multi classes and maximize it accuracy.

In Most of the applications the multiclass dataset can be divided into number of binary classes and can be merged to get multiclass. Not all the algorithms produce better accuracy on binary as well as multiclass datasets. Some of the algorithms perform well on binary dataset and the other performs well on multiclass datasets.

If the attribute A is a discrete value then binary tree must be produced. For the Binary class the Dataset D is split into two classes say C1, C2. Class C1 represents the attributes related to the constraint1 and Class C2 represents the attributes related to the constraint2. The mid-point or split-point is used to split the given attribute A into binary class values. Finally we can get D1 is set of tuples which satisfies the condition1 and tuple D2 which satisfies the condition2.

For multiclass, the dataset is split into multiple classes Ci where i=1, 2...n. The class C1 corresponds to the tuples which satisfies the constraint1, C2 corresponds to the constraint2 and so on. In multi class dataset we get Di dataset which corresponds to Ci classes.

4. FILTER METHOD

The filter attribute selection method is independent of the classification algorithm. Filter method is further categorized into two types

1. Attribute evaluation algorithms
2. Subset evaluation algorithms

The algorithms are categorized based on whether they rate the relevance of individual features or feature subsets. Attribute evaluation algorithms rank the features individually and assign a weight to each feature according to each feature's degree of relevance to the target feature. The attribute evaluation methods are likely to yield subsets with redundant features since these methods do not measure the correlation between features. Subset evaluation methods, in contrast, select feature subsets and rank them based on certain evaluation criteria and hence are more efficient in removing redundant features [1].

The main disadvantage of the filter method is it ignores the dependencies among the features and treats the features individually. The basic feature selection algorithm is given by [2] is

4.1. Basic feature selection algorithm

Input:

- S - Data sample f with features X, |X| = n
- J - Evaluation measure to be maximized
- GS – successor generation operator

Output:

- Solution – (weighted) feature subset
- L: = Start Point(X);
- Solution: = {best of L according to J };

Repeat

- L: = Search Strategy (L, GS (J), X);
- X':= {best of L according to J};
- If J (X') =J (Solution) or (J (X') =J (Solution) and |X'|

< |Solution|) then

Solution: =X';

Until Stop (J, L).

The filter method uses the discriminating criteria for feature selection. The correlation coefficient or statistical test like t-test or f-test is used to filter the features in the filter feature selection method.

The FSDD, RFS, CFS are some of the feature selection algorithm which uses the Filter methodology. The relevance score is calculated for the features to check the correlation between the features. The calculated score is high with some threshold value then the particular feature is selected for further classification. When the ranking is low those feature are removed. This method is very simple, fast and also independent of classification algorithm. The following are the basic filter feature selection algorithm

1. χ^2 test
2. Euclidian distance
3. T-test
4. Information gain
5. CFS-correlation based feature selection method
6. MBF- Markov blanket filter
7. FCBF-fast correlation based feature selection

4.1. χ^2 Test

The chi-squared test is one of the feature selection methodology used in filter method. The chi squared statistical test checks the independence between the two events. If X, Y are two events then the statistical independence is denoted by the following equations

$$P(XY) = P(X) P(Y) \text{ or}$$

$$P(X/Y) = P(X) \text{ and } P(Y/X) = P(Y)$$

The null hypothesis indicates that there is no correlation between the events. The events in the classification denote the class.

4.2. Euclidian Distance

The Euclidian distance is a feature selection technique used in filter method. In this method the correlation between the features is calculated interns of Euclidian distances. If there are n number of features in a sample feature say 'a' is compared with other n-1 features by calculating the distance between them using the following equation.

$$d(a,b) = \{\sum_i (a_i - b_i)^2\}^{1/2}$$

The addition of new feature will not affect the distance between any two samples.

4.3. T-Test

The filter method uses the t-test for calculating the relationship between the two samples by comparing its mean value. The t-test uses the following formula to compare the mean value.

$$T = \frac{X' - Y'}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}}$$

The result of the formula is a ratio which indicates the difference between the two mean values.

4.4. Information Gain

The entropy and the information gain is an attribute measure which indicates how much percentage the given attribute separate the training dataset according to their final classification. The Entropy for a set S is calculate as

$$Entropy(s) = \sum_{i=1}^n -P_i \log_2 P_i$$

Where 'n' is the number of classes, and the P_i is the probability of S belongs to class i.

The gain of A and S is calculated as

$$Gain(A) = Entropy(S) - \sum_{k=1}^m \frac{|S_k|}{|S|} \times Entropy(s_k)$$

S_k is the subset of s

4.5. CFS

CFS is a Correlation-based Feature Selection algorithm which uses the filter method for selecting the attributes. It is described by Hall Correlation. The CFS algorithm uses a heuristic which measures the usefulness of individual features for predicting the class label along with the level of inter-correlation among them. The highly correlated and irrelevant features are avoided. The equation used to filter out the

irrelevant, redundant feature which leads the poor prediction of the class is calculate using the equation

$$Fs = \frac{N \overline{r_{ci}}}{N + N(N-1) \overline{r_{ij}}}$$

N is the number of features in the subset, r_{ci} is the mean feature correlation with the class and r_{ij} is the average feature inter-correlation. For computing the correlations necessary for equation a number of information based measures of association were proposed such as: the uncertainty coefficient, the gain ratio or the minimum description length principle. The best results however were achieved with the gain ratio used for feature-class correlations and symmetrical uncertainty coefficients used for feature inter correlations.

4.6. MBF

The Markov Blanket of a feature is calculated to remove the irrelevant features from the feature set.

Let G is the subset of the feature set S. F_i is the feature in G. M is some other subset independent of F_i . M is a Markov Blanket of F_i , if F_i is conditionally independent of G-M- F_i . If M is the Markov Blanket of F_i then F_i can be remove from the set G. The remaining features are denoted by

$$G' = G - F_i$$

The algorithm steps for MBF filter method [1]

- step1. Initialize $G = F$
 - step2. Iterate
 - step3. For each feature $F_i \in G$, let M_i be the set of k feature $F_j \in G - \{F_i\}$ for which the correlations between F_i and F_j are the highest.
 - step4. Compute $\Delta(F_i | M_i)$ for each i
 - step5. Choose the i that minimizes $\Delta(F_i | M_i)$ and define $G = G - \{F_i\}$
- The Markov blanket filtering minimizes the discrepancy between the conditional distributions using conditional entropy.

4.7. FCBC

The FCBC algorithm for selecting the features using the filter method is as follows[1]

- Step1. S is the set of candidate predictors,
 $M = \emptyset$ is the set of selected Predictors
- Step2. Searching X^* (among S) which Maximizes its correlation with $Y \rightarrow \rho$
- Step3. If $\rho_{y, X^*} \geq \delta$ add X^* into M and Remove X^* from S
- Step4. Remove also from S all the variables X such $\rho_{X, X^*} \geq \rho_{y, X^*}$
- Step5. If $S \neq \emptyset$ then GOTO (2), else
- Step6. Stop

The algorithm supports a dataset of very large number of candidate predictors.

4.8. Sequential Forward Selection (SFS)

The SFS is a greedy search algorithm, which start with an empty set and add the relevant features one by one. The

SFS algorithm uses the following steps to select the relevant features.

Step1: start with the Empty set $S_k = \{ \}$

Step2: Select the relevant feature x to maximize the function $F(S_k+x); x \notin S_k$

Step3: Update S_k as $S_{k+1} = S_k+x$

Step4: Go to step2

5. WRAPPER METHOD

The wrapper method is slower and expensive than the filter method. The interaction between the feature subsets and the maintaining the dependencies between the features are the main advantageous of wrapper method. Wrapper methods are better in defining optimal features rather than simply relevant features and that they do this by allowing for the specific biases and heuristics of the learning algorithm and the training set. The wrapper method uses the backward elimination to remove the insignificant features from the subset. The SVM-RFE is one of the feature selection algorithms which use the Wrapper method. The Wrapper method need some predefine learning algorithm to identify the relevant feature. It has interaction with classification algorithm. The over fitting of feature is avoided using the cross validation. But it takes much time comparing to the filter method. The following are some of the wrapper feature selection methods

1. SFS-sequential forward selection
2. SBE-sequential backward elimination
3. plus L minus R
4. Beam search
5. Randomized hill climbing
6. Genetic algorithms
7. Estimation of distribution algorithms
8. Simulated annealing

5.1. SFS-sequential forward selection

The algorithm discussed in section 5.8 is also support for wrapper method to select the features.

5.2. SBE - sequential backward elimination

The sequential backward elimination is a reverse of SFS algorithm. Initially the algorithm start with the entire set of features and it removes the irrelevant features from the set in backward direction. The SBE algorithm is executed using the following steps

Step1: Start with the Full set S

Step2: Remove the Irrelevant feature f_i from the set S

Step3: Update the set S as $S_{k+1} = S - f_i$

Step4: Go to Step 2

5.3. Plus L minus R (LRS)

The Plus L is a generalization of SFS and the Minus R is the generalization of SBE algorithms. If $L > R$ then The algorithm start with the empty set and add the necessary features to the resultant set, else the algorithm start with the entire set and start eliminating the irrelevant features and produce the resultant set.

5.4. Beam search

The beam search is a more generic feature selection algorithm. The algorithm steps are

Step1. Compute the classifier performance using each of the n features individually (n I-tuples).

Step2. Select the best K (beam-width) features based on a pre-defined selection criterion among these I-tuples

Step3. Add a new feature to each of these K features, forming $K(n-1)$ 2-tuples of features. The tuple-size t is equal to 2 at this stage.

Step4. Evaluate the performance of each of these t -tuples. Of these, select the best K , based on classification performance.

Step5. Form all possible $(t + 1)$ tuples by appending these K t -tuples with other features (not already in that t -tuple).

step6. Repeat steps 4 to 5 until the stopping criterion is met; the tuple size at this stage is m .

step7. The best K m -tuples are the result of beam search.

5.5. Randomized hill climbing

The hill climbing is a local search algorithm.

Step1.procedure hill-climbing (Max_Flips)

Step2.restart: $s \leftarrow$ random valuation of variables;

Step3.for $j: =1$ to Max_Flips do

Step4.if eval(s) =0 then return s endif;

Step5.if s is a strict local minimum then

Step6.goto restart

Step7.else

Step 8. $s \leftarrow$ neighborhood with smallest evaluation value

Step9.endif

Step10.endfor

Step11.goto restart

Step12.end

5.6. Genetic algorithms

Genetic algorithm based on the Darwinian survival of the fittest theory. It is an optimization algorithm. The genetic algorithm starts from a group of points coded as finite length alphabet instead of one real parameter set. The genetic algorithm does not require the derivative information and step size calculation.

The three operators of genetic algorithms are:

1. Selection
2. Crossover
3. Mutation.

It selects some individuals with stronger adaptability from population according to the fitness, and then decides to copy the number of individual according to the selection methods such as Backer stochastic universal sampling. It exchanges and rearranges a pair of chromosome through crossover. Mutation is done to change certain point. The suitable crossover and mutation probability is to be chosen through real problems.

5.7. Estimation of distribution algorithms

The EDA algorithm incorporates the advanced machine learning into genetic an evolutionary algorithms. The main advantage of the EDA is it scalability. The algorithm steps are

Step 1.Initialize the population with $l=0$;

Step .generate initial population $P(0)$;

Step 3.while (not done) {

Step 4.select population of promising solutions $S(l)$;

Step 5.generate probabilistic model $P(l)$ for the solution $S(l)$;

Step 6.sample $P(l)$ to generate $O(l)$;

Step 7.incorporate $O(l)$ into $P(l)$;

Step 8. $l := l+1$; }

5.8. Simulated annealing

The simulated annealing is a method which is used for optimization problems. Initially the solution S and the related cost function $F(c)$ is given as input values, and the algorithm is executed to find out the resultant solution S' and the minimum cost function $F'(c)$. Then algorithm start with the input values S' and $F'(c)$ to find out the new solution and the minimum cost function. The algorithm proceeds till the optimum result.

6. EMBEDDED METHOD

The embedded method has the interaction with the classification algorithm

The KP-SVM is the example for embedded method. It consumes less time than the wrapper method. The embedded method uses the support vector mechanism to select the feature.

The following are some of the embedded feature selection algorithms.

1. Decision trees
2. Weighted naïve bayes
3. Feature selection using the weighted vector of SVM

Based on the output the feature selection algorithm, it is categorized into two minimum subset algorithm and feature ranking algorithm. In the minimum subset algorithm the feature are not ranked. The features in the subsets have the relevance.

7. EXISTING WORK

7.1. Feature selection algorithms on medical data set

The medical dataset are complex to handle. There are number of researches carried out on the medical dataset. Most research uses the synthetic data. The preprocessing is important for real time medical data. The medical data with large attribute need efficient feature selection algorithm for predicting it into binary and multiclass data with improve accuracy.

Three feature selection methods FCBF, Multi thread FCBF, and decision dependent and decision independent which are applied on c4.5 decision tree on lung cancer, breast cancer an insurance company data sets. The multi thread FCBC approach takes less computational time. The DDC-DIC took, not only the less time but also gives the best accuracy. When the number of features are reduced, both the feature selection algorithms gives the better accuracy and is executed in less time which concludes the dimensionality reduction will improve the accuracy and reduce the time complexity[3].

Three filter based feature selection algorithms Quick reduct, Chi-squared and CFS were applied on medical data set [4]. The decision rules were generated using the MLEM2 (Modified learning from example model version2) algorithm and the classification result were test using the J48 decision tree and the WEKA tool. The CFS and the χ^2 algorithms give the better performance for selecting the features on medical dataset.

In [5] the author use two feature selection algorithms FSSMC (Feature Selection via Supervise Model Construction) and ReliefF on diabetic dataset. For prediction the naïve Bayes, Ib1 and c4.5 classification algorithms were used and the accuracies are estimated. FSSMC feature selection algorithm

produce the better performance on the classification of different medical dataset than the ReliefF.

In [6] the author proposed the filter based MK-SVM (Multiple Kernel Support Vector Machine scheme) feature selection method on the colon tumor and ALL-AML leukemia datasets to evaluate the performance of the classification. The MK-SVM feature selection method produces the better performance than the other methods

The decision tree based filter method for setting attribute weights for use with naïve Bayes which perform well on different medical datasets [7]. T-static based filter feature selection method on SVM-RFE is SVt-RFE which is applied for selecting the gene of various medical dataset for classification and it produce the better performance [8].

In [9], [10] SNB (Selective Naïve Bayes) is used for filtering and ranking the attribute of medical dataset which gives the better performance. The prior of single attribute is measured by adding the new attribute with the existing set and comparing the accuracy of the current and previous sets. The SNB support only for the feature of discrete values and is not support for the continuous valued attributes.

Finite dataset is use to predict the heart disease prediction which result as binary class on the classification algorithms decision tree, Naïve Bayes an neural networks[11].

The presence of stroke disease is predicted using the classification methods naïve Bayes, neural networks and decision trees by reducing the dimensionality using the principal component analysis which gives the result as binary class. The relevance analysis is used to select the feature subset [12].

Breast cancer patient are classified into “*carcinoma in situ*” and “*malignant potential*” which is the binary class on SEER cancer dataset using c4.5 classification algorithm is determined and the accuracy of the Classifier is estimated [13].

In [14] the author use PCA data reduction method to reduce the features of Parkinson’s disease and the FKNN classifier is used for classification. The wrapper based improved naïve Bayes algorithm called RSNB is use to select the features randomly and is applied on large UCI datasets which gives the better classification accuracy [15].

In [16] the author proposed a MLRM (modified Local Range Modification) method of filtering the noise on mammogram images to predict the breast cancer which gives the 97.9% than the existing method LRM which is 93%.

The NPPC (Nonparallel plan proximal classifier) ensemble is used for both the binary and multiclass classification of cancer gene analysis. Wrapper method is used for feature subset and filter method is used for selection of genes which saves the computational cost [17].

In [18] the author presents a filter based feature selection using selective sampling. The Kd-Tree is used for partitioning the data. The performance of the active feature selection is evaluated and proved that it save significant amount of time. The filter based FSDD feature selection algorithm is compared with the ReliefF and mrmrMid with different datasets and on the classification algorithm like KNN, Naïve Bayes, Decision Tree and SVM. The algorithm predicts both binary as well as the multiclass

data. The FSDD outperforms well than the other two when the number of features are increased and it reduces the computational cost [19].

In [20] the author proposed filter based feature selection algorithm IT-IN which use backward elimination method. The algorithm is compared with FCBF, Relief and CFS. The classification accuracy is measured using the Naïve Bayes, SVM and ELM classifier. The algorithm provides minimum computation time than the other three and the algorithm produce better classification accuracy with ELM classifier.

In [21] the author presents the embedded based feature selection method KP-SVM and is applied on four bench mark datasets. The KP-SVM automatically obtains the optimal feature subset. It employs the filter method to filter out the irrelevant feature.

In [22] the author presents a feature selection algorithm MRMS using rough sets to select the relevant feature from high dimensional microarray dataset. The algorithm is applied on SVM and KNN for predictive method and proved that the performance is better than the existing algorithms.

In [23] the author proposed a roughest based RS-SVM feature selection algorithm to remove the redundant features which gives the high classification accuracy on Wisconsin breast cancer diagnosis.

Author in [24] proposed a new feature selection algorithm DWFS-CKN on microarray dataset. The algorithm is compared with the existing three algorithms Gini Index, MRMR and Relief-F. The DWFS-CKN gives the better accuracy and execute with less time.

In [25] the author optimizes the wrapper based incremental feature subset selection algorithm and present the embedded version of incremental feature subset selection algorithm which consumes the less memory and less space than the existing incremental FSS when it is applied on the high dimensional dataset.

The author[26] in their study of datamining eleven feature selection methods and three fuzzy modelling methods were analyzed and performance were compared on different medical data set. The mutual correlation based feature subset selection produce higher accuracy on Wisconsin breast cancer dataset. The four statistical criteria WTS1, WTS2, FCS1 and FCS2 with ANFIS produce the better accuracy on the Pima Indian diabetes datasets.

The author [27] use the fuzzy based nonlinear transformation method on small high dimensional medical dataset's for features extraction which gives the better classification performance than the PCA and KPCA methods. In [28] the author introduces the feature selection method based on fuzzy entropy measure which is tested on the similarity classifier. The method was tested on the different medical datasets like dermatology, Pima-Indian diabetes, breast cancer Parkinson's and dermatology. The maximum mean accuracy of 98.28% was achieved with dermatology dataset on multi class.

The author [29] introduces the feature selection method based on fuzzy entropy measure which is tested on the similarity classifier. The method was tested on the dermatology dataset. The improved mean classification accuracy of 98.83% was achieved on multi class.

In [30] the author introduces a classification model with feature selection DAMC which is embedded based, they also introduce the application model based on DAMC-CD which

is designed for binary dataset produce significantly high classification accuracy.

7.2. Result Analysis of feature selection algorithms

The existing study shows that most of the feature selection algorithms are applied on synthetic dataset and few of them use the real data. From the analysis number of feature selection algorithm uses ten to hundreds of features. Only few algorithms use the high dimensional data with thousands of features. When the algorithm speaks about the computational time it forgets to deal with the cost constraint. Also few algorithms are designed to support high dimensional medical datasets with multiclass. Most of the feature selection algorithms are not clearly mention the accuracy in percentage. Table1 list out the mean accuracy of the feature selection algorithm irrespective of the feature selection method and the maximum number of features used in the algorithm which is also represented graphically in figure1. The analysis result concludes that when the feature are increase the accuracy of an algorithm get affected and is decreased. It is a challenging task to keep the accuracy in the highest level even if the features are increased.

Table 1. Average accuracy of feature selection algorithm and the maximum number of attributes

Category	No of Attributes	Accuracy In %
1	10-99	98
2	100-999	97
3	1000-9999	95

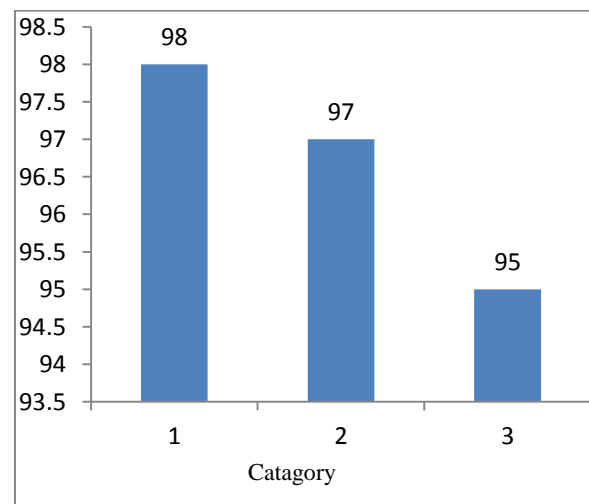


Fig 1. mean accuracy of the feature selection algorithm with maximum number of features

8. DRAWBACKS OF FEATURE SELECTION ALGORITHMS

We have different kinds of feature selection algorithms. Some of them based on filter method and some based on wrapper method and some are based on embedded method. Not all the feature selection support multiclass dataset. Some method support only binary dataset. When the feature selection is applied on high dimensional medical dataset the algorithm which selects the appropriate and

optimal features is forget to increase the accuracy constraint of the classifier. The feature selection algorithm which supports both binary dataset and the multiclass dataset sometimes produces high accuracy on the binary dataset but gives low accuracy when it is used in the multiclass data set. The feature selection algorithm must support multiclass dataset and produce high accuracy when applied on classification.

9. CONCLUSION

This study of feature selection algorithms of large survey shows that the feature selection algorithm consistently improves the accuracy of the classifier. Each feature selection methodology has its own advantages and disadvantages. The dataset with larger attributes use the wrapper methods with lesser improvement in accuracy. Inclusion of larger attribute causes the reduction of accuracy. Each algorithm has different behavior which shows relaying single algorithm for different dataset is infeasible. The feature selection algorithms are one which decides the accuracy of the classification of different datasets. The feature selection algorithm must select the relevant features and also remove the irrelevant and inconsistent features which cause the degradation of accuracy of the classification algorithms. The classification and feature selection algorithms must support both binary as well as multiclass datasets. When the feature selection algorithm is applied on high dimensional dataset to select the appropriate features the algorithm not only produce the best accuracy but also it has to reduce the time and space. The best feature selection algorithm satisfies the constraint like improvement in accuracy ,consumption of less space and time and effective support of high dimensional with multi class dataset. The future improvement from this study is to design a feature selection algorithm for high dimensional multiclass dataset with considerably improvement in accuracy with less space and time requirement.

10. REFERENCES

- [1] Ellen pitt, Richi nayak, "The use of various data mining and feature selection methods in the analysis of a population survey dataset", *Australlian computer society inc* 2007.
- [2] L.Latha, T.deepa, "Feature selection methods and algorithms", *International journal on computer science and engineering*, Vol. 3 No. 5 May 2011.
- [3] C.Daisy, B.Subhulaksmi, S.Baskar, N.Ramraj, "Efficient Dimensionality Reduction Approaches for Feature selection", *IEEE computer society*, 2007.
- [4] Grzegorz Ilczuk and Alicja Wakulicz-Deja, "Selection of Important Attributes for Medical Diagnosis Systems", *Springer-Verlag Berlin Heidelberg* 2007.
- [5] Yue Huang, Paul McCullagh, Norman Black, Roy Harper, "Feature selection and classification model construction on type 2 diabetic patients' data" *Artificial Intelligence in Medicine*, Elsevier, 2007.
- [6] Zhenyu Chen, Jianping Li, Liwei Wei, "A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue", *Artificial Intelligence in Medicine*, Elsevier, 2007.
- [7] Mark Hall, "A decision tree-based attribute weighting filter for naive Bayes", *Science Direct, Elsevier*, 2007.
- [8] Piyushkuma, A. Mundra, Jagath C. Rajapakse, "Gene and sample selection for cancer classification with support vectors based statistic", *Neurocomputing*, Elsevier, 2010.
- [9] Tzu-Tsung Wong, Liang-Hao Chang, "Individual attribute prior setting methods for naive Bayesian classifiers", *Pattern Recognition*, Elsevier, 2010.
- [10] Tzu-Tsung Wong, "A hybrid discretization method for naive Bayesian classifiers", *Pattern Recognition*, Elsevier, 2011.
- [11] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", *International Journal of Computer Applications*, Volume 47– No.10, 2012.
- [12] A.Sudha, P.Gayathri, N.Jaisankar, "Effective Analysis and Predictive Model of Stroke Disease using Classification Methods", *International Journal of Computer Applications*, Volume 43– No.14, 2012.
- [13] K.Rajesh, Shela Anand, "Analysis of SEER Dataset for Breast Cancer Diagnosis using C4.5 Classification Algorithm", *International Journal of Advanced Research in Computer and Communication Engineering*, 2012.
- [14] Hui-Ling Chen, Chang-Cheng Huang, Xin-Gang Yu, Xin Xu, Xin Sun, Gang Wang, Su-Jing Wang, "An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach", *Expert Systems with Applications*, Elsevier, 2012.
- [15] Liangxiao Jiang, Zhihua Cai, Harry Zhang, Dianhong Wang, "Not so greedy: Randomly Selected Naive Bayes", *Expert Systems with Applications*, Elsevier, 2012.
- [16] Senthilkumar Balasubramanian and Umamaheswari, "Novel preprocessing in the computer aided deduction of breast cancer", *Journal of Computer Science* 2012, 8 (12), 1957-1960.
- [17] Santanu Ghorai, Anirban Mukherjee, Sanghamitra Sengupta, and Pranab K. Dutta, "Cancer classification from gene expression data by NPPC ensemble", *IEEE/ACM transaction on computational biology and bio informatics*, Vol.8, No.3, 2011.
- [18] H. Liu, H. Motoda, Lie Yu, "A selective sampling approach to active feature selection", *Science Direct, Elsevier*, 2004.
- [19] J.Liang, S. Yang, A. Winstanley, "Invariant optimal feature selection a distance discriminant and feature ranking based solution", *Science Direct, Elsevier*, 2008.
- [20] C.Daisy, S. Basskar, N. Ramraj, J. Saravanand Koori, p. Jeevanandam, "A novel information theoretic interact algorithm for feature selection using three machine learning algorithms", *Expert system with applications*, Elsevier, 2010.
- [21] Sabastian Malonado, Richard Weber, Jayanta Basak, "Simultaneous feature selection and classification using kernel-penalized support vector machine", *Information Science*, Elsevier, 2011.

- [22] Sushmita Paul, Praipta Maji, "Rough set based gene selection algorithm for microarray sample classification", International conference on methods and models in computer science", IEEE, 2010.
- [23] Hui-Ling Chen, Bo Yang, Jie Liu, Da-You Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis", Expert systems with Applications, Elsevier, 2011.
- [24] Jayachidra and Punithavalli, "Distinguishability based weighted feature selection using column wise K-neighborhood for the classification of gene microarray dataset", American journal of applied science, 2014.
- [25] Pablo Bermejo, José A. Gámez, José M. Puerta, "Speeding up wrapper feature subset selection with naïve Bayes classifier", Knowledge-based systems, Elsevier, 2014.
- [26] Sean N. Ghazavi, Thunshun W. Liao, "Medical data mining by fuzzy modeling with selected features, Artificial Intelligence in Medicine, Elsevier, 2008.
- [27] Der-Chiang Li, Chiao-Wen Liu, Susan C. Hu, "A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets", Artificial Intelligence in Medicine", Elsevier, 2011.
- [28] Pasi Luukka, "Feature selection using fuzzy entropy measures with similarity classifier", Expert system with applications, Elsevier, 2011.
- [29] Cesar Iyakaremye, Pasi Luukka, David Koloseni, "Feature selection using Yu's similarity measure and fuzzy entropy measures", IEEE, 2012.
- [30] Minseok Seo, Sejong Oh, "A novel divide and merge classification for high dimensional datasets", Computational biology and chemistry", Elsevier, 2013.