ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

A Review on Filter Based Feature Selection

K.Mani[1] P.Kalpana[2]

Associate Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, India[1]

Assistant Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, India[2]

ABSTRACT: Feature selection plays a vital role in machine learning and in data mining process. A feature selection process selects a small subset of features from the original feature space. It minimizes redundancy, removes noise and maximizes the relevance to the target in classification. It is a preprocessing phase which improves the accuracy, speed, data quality and understanding. It also serves to reduce dimensionality and computational resources. This paper presents an overview of various methods and techniques in understanding the concepts of filter based feature selection.

KEYWORD: Feature Selection, Preprocessing, Redundancy, Relevance, Dimensionality, Filter Based Feature Selection and Classification.

I. INTRODUCTION

Data mining is a multidisciplinary effort used to extract knowledge from data. The proliferation of large data within many domains poses unprecedented challenges to data mining. Researchers and practitioners are realizing that the feature selection is an integral component for the effective use of data mining tools and techniques [25]. A feature refers to an aspect of the data. Feature Selection (FS) is a method of selecting a small subset of features from the original feature space by following certain criteria i.e., it is a process of selecting *M* features from the original set of *N* features, $M \subseteq N$. It is one of the essential and indispensable data preprocessing techniques in various domains viz., artificial intelligence, data mining and machine learning. It is also known as variable subset selection or attribute selection [2]. It helps us to select the more relevant features and eliminate redundant, irrelevant and noisy data.

A feature which has an influence on the output and its role can't be replaced by the rest is known as relevant feature [7]. The feature relevance is classified into three types. They are strong relevance, weak relevance and irrelevant [20]. A strong relevant feature is always necessary for the optimal subset and it cannot be removed. A feature is said to be a weakly relevant if it is necessary for an optimal subset only at certain conditions. An irrelevant feature is one which is not necessary at all because it does not contribute any information to the target and hence it must be removed [10]. A feature which takes the role of another is said to be redundant. Removing irrelevant and redundant features will potentially give a better generalization, understanding and visualization with less training and testing time**.**

A variable which provides no extra information about the classes and thus serves as a noise for the predictor is said to be dependant. Eliminating them reduces the amount of data and lead to improvement in the classification accuracy. FS also reduces the dimensionality of feature space and also reduces the running time of the learning algorithm and helps to improve the quality of the classification algorithm.

The rest of the paper is organized as follows. Section 2 discusses the various types of feature selection, categories of supervised feature selection and comparison of filter and wrapper. The detail of the feature selection process is depicted in Section 3. Section 4 presents the overview of the filter-based feature selection. The most significant measures used in features selection are discussed in section 5. Section 6 describes the most prominent feature selection algorithms found in the literature and finally section 7 ends with conclusion.

II. FEATURE SELECTION: TYPES, CATEGORIES AND COMPARISON

A. Types of Feature Selection

Feature selection algorithms can be grouped as supervised, unsupervised and semi-supervised based on labeling or un-labeling of the training set. Supervised feature selection estimates the relevance of features guided by the label information but it is a time consuming process. The unsupervised feature selection works on unlabeled data and it is

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

difficult to evaluate the relevance of features [19]. The semi-supervised feature selection works with small subset of labeled information along with large number of unlabeled data [6].

B. Categories of Supervised Feature Selection

Supervised feature selection can further be categorized into filter, wrapper and embedded models [8]. Filter model depends on the characteristics of the data viz., consistency, distance, dependency, information and correlation. It does not intermingle with learning algorithm [13], [18]. Wrapper model uses the specific learning algorithm itself to assess the quality of the selected features [18]. Embedded model uses a learning algorithm which performs feature selection in the process of training [2]. The overview of filter model is shown in Fig. 1.

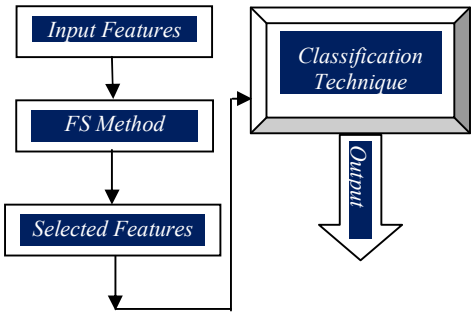

Fig.1. General View of Filter Model for Classification

C. Comparison of Filter and Wrapper Method

The filter is a supervised learning method and it is independent of learning algorithm which uses the characteristics of data to select and evaluate the features. It is more general, faster, requires low computational complexity and gives less accuracy than wrapper method. It does not inherit any bias of a mining algorithm [10], because it is not tied with the learning algorithm. So it does not require re-execution for different learning algorithms. It is a significant choice when the selected features are very large and also computationally efficient. The wrapper is an unsupervised learning model which uses the learning algorithm to select and evaluate features so it requires re-execution for different learning algorithm. This model is less general, slower, requires more computational cost and gives more accuracy than filters because it is tightly coupled with a selected learning algorithm. It is a significant choice when the selected features are very small [4], [15].

III. FEATURE SELECTION PROCESS

The general procedure of feature selection involves four steps [2], which are shown in Fig. 2. The subset generation is a process of heuristic search and it generates subsets of features for evaluation [2]. The most acquainted heuristic search methods are forward selection and backward elimination [8]. The forward selection algorithm starts with an empty set and keeps on adding features whereas the backward elimination starts with all features and keeps on removing features [3], [14], [18]. Based on the evaluation criterion, the generated feature subset is evaluated with the previous best feature subset. If the new subset is found to be superior to the previous best subset, then the existing feature subset is replaced by the new best feature subset. Filter model uses the information or uncertainty, distance or divergence, dependency, probability of error measures, consistency and interclass distance measures as the evaluation criteria whereas the wrapper uses predictive accuracy as the evaluation criteria for classification, and measures of goodness for clustering. The evaluation focuses on selecting the relevant features, eliminating the irrelevant and redundant attributes. This task is repeated until the stopping criterion is fulfilled. The stopping criterion may be any one of the following:

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

i) A predefined number of features are selected
ii) A predefined number of iterations are reached
iii) Addition or deletion of a feature which fails to produce better subset
iv) An optimal subset obtained according to evaluation criteria

Finally the produced optimal feature subset is validated and it can be done using selected subset and the original set with synthetic data or real-world data and the results were compared [2][16].

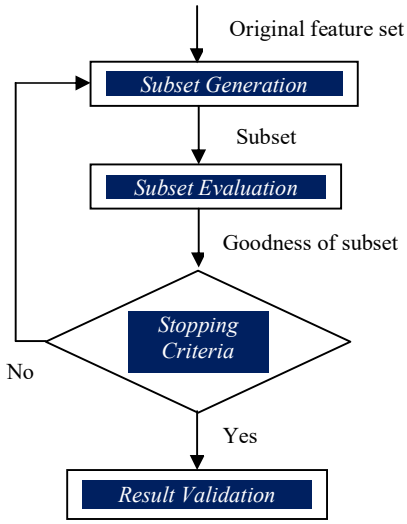

Fig.2. Feature Selection Process

IV. FILTER-BASED FEATURE SELECTION: AN OVERVIEW

A. Types of Filters

The Filter Based Feature Selection (FBFS) relies on the characteristics of the data without employing any classification algorithm [19]. There are two kinds of filter viz., univariate and multivariate. The univariate filter treats and evaluates each feature independently. The multivariate filter evaluates features in the context of others in a batch method. The multivariate filters are theoretically sound but the univariate filters are very simple and fast.

B. Steps in FBFS

The FBFS algorithm consists of two major steps. The first step ranks the features based on certain criteria [19]. The features with highest ranking will be selected in the second step.

C. General View of Filters

The general view of FBFS algorithm is exemplified [10] in Algorithm 1.

Algorithm 1: Filter Based Feature Selection

Input:
A training dataset D with n features $f_1, f_2, \ldots, f_i, \ldots, f_n$, S_0 a subset from which the search should begin, evaluation measure M and δ a stopping condition.

Output:
An optimal subset of features S_{opt}

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Method:

Procedure Filter_Algorithm *(D, S_0, M, δ)*
begin
$S_{opt} \leftarrow S_0$;
$\gamma_{best} \leftarrow$ eval(S_0,D,M) ;
do
begin
$S \leftarrow$ generate(D); //generates a subset for evaluation
$\gamma \leftarrow$ eval(S,D,M); //evaluates the current subset S by M
if γ is better than γ_{best} then
begin
$\gamma_{best} \leftarrow \gamma$;
$S_{opt} \leftarrow S$
end
end
until(δ is reached);
return S_{opt}
end;

V. FBFS METHODS

This section presents the most common measures which are used in the feature selection process.

A. Information Gain (IG):

It is a supervised, univariate, simple, powerful, symmetrical and entropy-based feature selection algorithm. It determines the feature relevance between the attribute and class label [9]. The *IG* for a feature *X* and the class label *Y* is calculated using the formula.

$$IG(X,Y) = H(X) - H(X \mid Y) \tag{1}$$

Where *H(X)* is the entropy of *X* and *H(X|Y)* is the entropy of *Y* after observing *X*.
The entropy of *X* can be calculated as

$$H(X) = -\sum_i P(x_i)\log_2(P(x_i)) \tag{2}$$

The entropy of *X* after observing *Y* can be computed as

$$H(X,Y) = -\sum_j P(y_j)\sum_i P(x_i \mid y_j)\log_2(P(x_i \mid y_j)) \tag{3}$$

This method calculates the *IG* for each and every feature separately and selects *'m'* features as more relevant ones among *'n'* features with high *IG* i.e., it considers a feature *F* with high *IG* as the more relevant one. The main drawback of the algorithm is that it selects the feature with high *IG* which may or may not be more informative. The *IG* cannot handle redundant features [19] because the features are chosen in a univariate way.

B. Gain Ratio (GR):

It is supervised, univariate, non-symmetrical and entropy based measure introduced to eliminate the bias of *IG*. *GR* can be computed as

$$GR = \frac{IG}{H(X)} \tag{4}$$

In order to predict class label *Y*, it needs to be normalized by dividing *IG* by entropy of feature *X* and vice versa. The *GR* value will always fall in the range [0, 1]. *GR=1* specifies that the knowledge of *X* completely predicts *Y* and *GR=0* indicates that *X* and *Y* are uncorrelated. This method favours the features with less value.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

C. Symmetric Uncertainty (SU):

It is a modified information gain measure developed to overcome the drawbacks of *IG* by dividing the sum of the entropies of *X* and *Y* [11] and it is given by [19]

$$SU = 2 \times \frac{IG(X,Y)}{H(X) + H(Y)} \tag{5}$$

SU can take a value in the range [0, 1] due to the correlation factor 2**.** *SU=1* specifies that the knowledge of one feature completely predicts the other. If *SU=0* means *X* and *Y* are uncorrelated. This method also favours the features which have less value.

D. Mutual Information (MI)

It is a univariate and supervised feature weighting method. This method computes mutual information using (6) between each features and the target class label, then the features are ranked accordingly and top *'n'* features are selected.

$$I(A,B) = \sum_i \sum_j \mathrm{P}(a_i, b_j) \log_2 \left(\frac{P(a_i, b_j)}{P(a_i) \times P(b_j)}\right) \tag{6}$$

Cover and Thomas [18] showed the same formula for Mutual Information and Information Gain. But Yang and Pedersen discriminates *IG* and *MI*.

E. Gini Index (GI)

It is a univariate and supervised feature weighting method. It is a measure for quantifying a feature's ability to distinguish between classes. Given *C* classes, *GI* of a feature *f* can be calculated as [15]

$$GI(f) = 1 - \sum_{i=1}^{C} [P(i \mid f)]^2 \tag{7}$$

The maximum value of *GI* is 0.5 for a binary classification. The more relevant features have smaller *GI* values. The main drawback is that it does not eliminate redundant features.

F. Chi-Square

It is one of the common methods of feature selection based on statistics and filter [1] [8]. It is used to test whether the class label or the target is independent of a particular feature or not. Chi-square value for a feature with *'r'* different values and *'c'* number of classes is defined as [1]

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{8}$$

Where O_{ij} is the number of instances with value *'i'* which are in class *'j'*. E_{ij} is the expected number of instances with value *'i'* and class *'j'*. In this method, the initial or null hypothesis H_0 is assumed that the two features are unrelated. If the value of chi-square calculated using (7) is greater than the tabulated value, then the hypothesis H_0 is accepted otherwise it is rejected.

G. Euclidean Distance

This method calculates the correlation between features using Euclidean distance [1] [7] using (9). If there are *'n'* features in a given data set then it calculates the Euclidean distance for each feature p_i with all other remaining *n-1* features. The Euclidean distance between two features such as p_k and q_k is calculated as

$$dist = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2} \tag{9}$$

If the distance between the features is larger, then it is said to be irrelevant. Adding a new feature will not misery the distance between any two samples.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

H. T-test

It measures the relationship between two samples statistically by comparing its mean values and is calculated as

$$t = \frac{\overline{X1} - \overline{X2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{10}$$

Where

$\overline{X_1}$ and $\overline{X_2}$ are the sample means.

S_1^2 and S_2^2 are the standard variances.

n_1 and n_2 are the number of elements in the samples.

I. minimum Redundancy and maximum Relevance (mRmR)

It is a supervised, multivariate and feature-set based feature selection method. It selects the features based on statistical dependency criteria. Due to the difficulty in directly implementing the maximal dependency condition, *mRmR* uses an approximation to maximize the dependency between the joint distribution of the selected features and the class variable. The minimum Redundancy for discrete features and continuous features are defined as [19]
For discrete features: min $W_{1,}$

$$W_1 = \frac{1}{|S|^2} \sum_{i,j \in S} I(i,j) \tag{11}$$

For continuous features: min $W_{c,}$

$$W_c = \frac{1}{|S|^2} \sum_{i,j \in S} |C(i,j)| \tag{12}$$

Where S is a subset of features, $I(i, j)$ and $C(i, j)$ are mutual information and correlation between the features f_i and f_j respectively. The maximum Relevance for discrete features and continuous features are defined as [19],
For discrete features: max $V_{1,}$

$$V_1 = \frac{1}{|S|^2} \sum_{i \in S} I(h,i) \tag{13}$$

For continuous features: max $V_{c,}$

$$V_c = \frac{1}{|S|^2} \sum_{i} F(i,h) \tag{14}$$

Where h is a target class, $I(h,i)$ is the mutual information between feature f_i and the target h, and $F(i,h)$ is the F-statistic.

J. Fisher score

It is a supervised and univariate feature weighting method which picks features that assigns similar value to the samples from the same class and different value to samples from different classes. The evaluation measure used in Fisher Score can be expressed as:

$$S_i = \frac{\sum_{k=1}^{K} n_j (\mu_{ij} - \mu_i)^2}{\sum_{k=1}^{K} n_j \rho_{ij}^2} \tag{15}$$

Where μ_{ij} and ρ_{ij} are the mean and the variance of the i^{th} feature in the j^{th} class respectively. n_j is the number of instances in the j^{th} instance and μ_i is the mean of the i^{th} feature. This method selects the top-m ranked features with large fisher scores. The main drawback of this method is that it cannot handle redundant features because it evaluates features independently [19].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

K. Pearson Correlation Coefficient

It is a supervised and/or unsupervised, univariate method which works only with numeric data. It determines the linear relationship between the variables. It can be computed as [27].

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \tag{16}$$

Where $\bar{x}$ and $\bar{y}$ are the means, σ_x and σ_y are the standard deviation of x_i's and y_i's.

L. Crammer's V

It is a measure of association between two nominal variables which gives a value between 0 and +1. It is otherwise called as Crammer's phi and it is based on Pearson chi-squared statistics [27].

$$V = \sqrt{\frac{\chi^2}{N(K-1)}} \tag{17}$$

Where N denotes the total number of observations and K denotes the number of features or the number of instances whichever is less. The $\chi 2$ is computed as

$$\chi^2 = \frac{\sum_{k=1}^{c} n_k (\mu_k^j - \mu_i)^2}{\sigma_j^2} \tag{18}$$

Where k denotes the class and j denotes the feature.

M. Markov's Blanket Filter (MBF)

To remove the irrelevant features from the feature space S, this filter calculates the Markov's Blanket M of each feature F_i from S. If M is the Markov's Blanket of F_i, *if* F_i is conditionally independent of the subset G of S, M and F_i, then F_i can be removed from the feature space S [1].

N. Random Forest

It was developed by Leo Breiman [15] is a group of un-pruned classification or regression trees and it is made from the random selection of samples of the training data. The features are selected randomly with induction process. The prediction is made by aggregating the predictions of the ensemble. Random forest provides a significant improvement in the performance as compared to single tree classifier c4.5 [15].

O. Kruskal Wallis

Kruskal Wallis [32] is a supervised, univariate and non-parametric feature weighting method which tests whether two or more classes have equal median and provides a value accordingly. If the value is close to zero then the feature is said to be discriminative. This method selects a feature with discriminative information and removes others.

P. Laplacian Score

It is an unsupervised and univariate feature weighting method which selects the features that retain the locality. It evaluates the features independently and picks the top k-features which have the minimal Laplacian score (SC_L) and it cannot handle feature redundancy [25].

Q. SPEC

SPEC is an unsupervised / supervised and univariate feature weighting method. It is an extension for Laplacian Score. In addition to the affinity matrix K, the degree matrix D, this method uses normalized laplacian matrix L [25] [20].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

VI. FBFS ALGORITHMS: LITERATURE REVIEW

This section presents the most significant *FBFS* algorithms used in the literature.

A. Correlation-based Feature Selection (CFS)

The CFS evaluates subsets of features using the following hypothesis [12].

"Good feature subset contain features that are highly correlated with the classification and yet uncorrelated to each other"

It uses a correlation based heuristics to evaluate the merit of a feature subset S consisting of k features [5]:

$$Merit_{S_k} = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{19}$$

Where $\overline{r_{cf}}$ and $\overline{r_{ff}}$ are the mean feature–class and feature–feature correlations respectively. CFS calculates feature–class and feature–feature correlation using symmetrical uncertainty and then it selects a subset of maximum relevant features using the best first search. The main problem with this method is that it cannot handle the problems where the class is numeric.

B. Fast Correlation Based Filter (FCBF)

It is a supervised, multivariate and feature-set based feature selection algorithm. FCBF starts by selecting a set of features S' from S that are highly correlated to the class with $SU \geq \delta$, where δ is the user defined threshold. In FCBF, a feature *'i'* with symmetrical uncertainty to the class *'c'* $SU_{i,c}$ will be called predominant iff $SU_{i,c} \geq \delta$ and $SU_{j,i} \geq SU_{i,c}$ $\forall f_i \in S'$, $j \neq i$. However, if there exists such feature *'j'* where $SU_{j,i} \geq SU_{i,c}$, then the feature *'j'* will be called as the redundant feature for *'i'* and it will be removed [23].

C. Relief

It is a very popular and classic filter method that was invented by Kira and Rendell [29]. It is an instance based feature selection which works by sampling an instance randomly from the data and then finds its nearest neighbor from the same and opposite class. The nearest neighbors are compared to the sampled instance and it is used to update relevance scores of each attribute. This is repeated for a user specified number of instances m. Features with weight, that is greater than the threshold τ, are considered as the relevant to the target variable. [22][26].

D. Relief-F

Kononenko et al., [12] in their algorithm uses Manhattan distance for finding the nearest miss and nearest hit rather than Euclidean distance. It updates the weight vector with the absolute difference between x_i and the nearest hit, and x_i and nearest miss rather than the square of those differences. The algorithm is repeated for 'n' times for each instance rather than 'm' times as in Relief. It searches 'k' nearest hits and misses and the average value contributes to the weight for each feature. The Relief algorithm decomposes the multi nominal into a number of binomial problems where as the Relief searches k-nearest and misses for each different class and their average contributes for updating the weight vector.

E. Las Vegas Filter (LVF)

This algorithm must produce a minimal subset of features which optimizes certain criteria, such as classification accuracy. The LVF scans for a minimal subset of features to describe a given set of supervised training examples $X = < x_1, y_1 >, ... < x_M, y_M >$, where $|x_i| = N$. The subsets are selected uniformly at random with respect to the set of all possible subsets. They are then evaluated based on an inconsistency criterion, which tests the extent to which the reduced-dimension data can separate the class labels. If the newly selected subset is smaller in size and has an equal or lesser inconsistency rate, then the subset is retained [21].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

F. FOCUS

It is a filter algorithm which is based on an exhaustive search strategy. It selects the subsets of features that optimize an evaluation measure [17] and it was introduced by Almuallim and Dietterich [28] and showed that it can detect the necessary and sufficient features in a quasi-polynomial time provided (i) the complexity of the target concept is limited and (ii) there is no noise.

G. One-R

It is a simple algorithm proposed by Holte, which builds one simple rule for each attribute in the training data and selects the rule with the smallest error. It treats all the numerically valued features as contiguous and divides the range of values into several disjoint intervals. It handles missing values by treating "missing" as a legitimate value.

H. Kolmogorov-Smirnov Feature Filter

The K-S Class Correlation-Based Filter (K-S CCBF) algorithm consists of two stages. The first stage ranks the features using SU co-efficient as in FCBF algorithm and the features are ranked in descending order, ending the first stage. In the second stage, redundant features are removed by utilizing the ranking information by comparing the features that are high in the ranking order with all those with lower rankings by checking the condition $S_c(g,h) < \lambda_\alpha$ where $KS_c(g,c)$ is the Kolmogorov-Smirnov statistic and it is computed as follows:

$$KS_C(g,h) = \max_c(KS(g(c),h(c))) \tag{20}$$

and λ_α is the threshold. If this condition is true, redundant feature, lower in ranking is removed from the original feature set *F*. If the condition is not valid then the algorithm leaves both features. After this, only the most significant features are left in the feature subset and all redundant features are removed [24].

I. PRBF (Pearson's Redundancy Based Filter)

This filter consists of two steps. The first step, determines the relevance between the features and class label using the symmetrical uncertainty (SU) as in FCBF. Further filtering is carried out in the second stage for determining redundancy and it is based on the Pearson $\chi 2$ test [30].

J. INTERACT

It uses the same goodness measure as in FCBF i.e., SU but it also includes the consistency contribution (c-contribution). The c-contribution of a feature indication about how significantly is the elimination of that feature will affect consistency. The c-contribution of an irrelevant feature is zero [31]. The algorithm consists of two major parts. In the first phase the features are ranked in descending order based on their SU values. In the second part, features are evaluated starting from the end of the ranked feature list one by one i.e., backward elimination approach. If c-contribution of a feature is less than the threshold then, the feature is removed otherwise it is selected [26].

K. Feature Selection Based on Mutual Correlation

This method, first calculates the mutual correlation between each pair of features. In the second step, the feature which is mostly correlated with others is removed. The most correlated feature is computed using the mean value of the correlation of the features in the feature space *S*. The algorithm is repeated with the remaining feature set until the desired size of feature set is obtained [33].

L. Incremental Usefulness

The definition of Incremental usefulness is stated as follows:

> "Given a sample of data *S*, a learning algorithm L, and a subset of features *X'*, feature x_i is incrementally useful to *L* with respect to *X'* if the accuracy of the hypothesis that *L* produces using the feature set $\{x_i\}\ U\ X'$ is better than the accuracy achieved using just the feature subset *X'* "

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

M. CorrSF and ConsSF

The CorrSF and ConsSF are the two important sub-set search algorithms which exploit sequential forward search and also utilizes the correlation measure or consistency measure to guide the search. The CorrSF is a variation of the CFS algorithm which runs faster and more suitable for high dimensional data than CFS [23].

VII. CONCLUSION

Feature selection has been a most remarkable research issue due to voluminous data with thousands of features and it can be widely used in many domains such as statistics, machine learning, web mining, text mining, image processing and microarray data analysis. The main objective of the feature selection is to build a model to improve the performance of the classifier by minimizing redundancy, removing noisy data and maximizing the relevance. This article has presented a brief summary of the concepts of feature selection, methods and the existing algorithms on FBFS in literature. It also helps us to analyze the methodology behind each algorithm in selecting the more relevant features and removing irrelevant features. Among the algorithms listed above, the individual can choose the desired algorithm by considering the criterion viz., simplicity and stability, number of reduced features, classification accuracy, storage and computational resources for their application in hand.

REFERENCES

1. S.Vanaja and K.Ramesh Kumar, *"Analysis of Feature Selection Algorithms on Classification: A Survey"*, *IJCA*, Vol.96, No. 17, July 2014.
2. Vipin Kumar and Sonajharia, *"Feature Selection: A Literature Review"*, Soft Computing Review, Vol. 4, No. 3, June 2014.
3. Shweta Srivastava, Nikita Joshi and Madhvi Gaur, *"A Review Paper on Feature Selection Methodologies and their Applications"*, IJCSNS, Vol. 14 No.5, May 2014.
4. K.Revathi and T.Kalaiselvi, *"Survey: Effective Feature Subset Selection Methods and Algorithms for High Dimensional Data"*, *IJARCET*, Vol. 2, Issue 12, December 2013.
5. Varuna Tyagi and Anju Mishra, *"A Survey on Different Feature Selection Methods for Microarray Data Analysis"*, *IJCA* (0975-8887), Vol. 67, No. 16, April 2013.
6. Sunita Beniwal and Jitender Arora, *"Classification and Feature selection techniques", IJERT*, Vol.1, Issue 6, August 2012.
7. L.Latha et al., "*Feature Selection Methods and Algorithms",* IJCSE, Vol. 3, No.5, May 2011
8. Jasmina Novakovic, Perica STRBAC and Dusan BILATOVIC, *"Toward Optimal Feature Selection using Ranking Methods and Classification Algorithms"*, Yagoslav Journal of Operations Research, 21(2011), November 1, 119-135, DOI:10.2298/ YJOR1101119N
9. M.Ramaswami and R.Bhaskaran, *"A Study on Feature Selection Techniques in Educational Data Mining"*, Journal of Computing, Issue 1, December 2009, ISSN: 2151-9617.
10. Huan Liu and Lei Yu, "*Toward Integrating Feature Selection Algorithms for Classification and Clustering"*, IEEE transaction on Knowledge and Data Engineering, 2005.
11. Manoranjan Dash and Huan Liu, *"Consistency-based Search in Feature Selection"*, Artificial Intelligence, *Elsevier*, pp.155-176, 2003.
12. Marko Robnik andSikonja Igor Kononenko, *"Theoretical and Empirical Analysis of ReliefF and RReliefF"*, *Machine Learning Journal,* 53:23-69, 2003.
13. B.M.Vidyavathi and Dr.C.N.Ravikumar, *"A Novel Hybrid Filter Feature Selection Method for Data Mining",* Ubiquitous Computing and Communication Journal, Vol. 3, No. 3.
14. Aparna Choudhary and Jai Kumar Saraswat, *"Survey on Hybird Approach for Feature Selection",* IJSR, Online.
15. K.Porkodi. *"Comparison of Filter Based Feature Selection Algorithms: An Overview"*, IJIRTS, Online.
16. Barak Chizi , Lior Rokach and Oded Maimon, *"A survey of feature selection techniques".* In Proceedings of Datenverarbeitungspraxis, 2009, pp. 1888-1895.
17. Michael Doumpos and Athina Salappa, *"Feature Selection Algorithms in Classification Problems: An Experimental Evaluation"*, In proceedings of the 4th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering Databases, 2007.
18. Tarex Amr, "Survey on Feature Selection", Available: http://tarekamr.appspot.com/pdfs/fs022013.pdf
19. Jiliang Tang, Salem Alelyani and Huan Liu. *"Feature Selection for Classification: A Review"*, Available: http:// www. public. asu.edu
20. Pengpeng Lin, *"A Framework for Consistency Based Feature Selection"*, Western Kentucky University, Graduate Studies and Research Master Thesis, 2003.
21. *"Computational Methods of Feature Selection"*, Boca Raton, London, New York, Washington, D.C. CRC PRESS, 2001.
22. Mark A. Hall and Geoffrey Holmes, *"Benchmarking Attribute Selection Techniques for Data Mining"*, University of Waikato, Hamilton, New Zealand.
23. Lie Yu and Huan Liu, "*Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution"*, Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003.
24. Marcin Blachnik, Wlodzislaw Duch, Adam Kachel and Jacek Biesiada, *"Feature Selection for Supervised Classification: A KolmogorovSmirnovClassCorrelation-BasedFilter",* Methods of Artificial Intelligence, Gliwice, pp.33-40, 2009.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

25. Zheng Zhao, Fred Morstatter et. al., *"Advancing Feature Selection Research – ASU Feature Selection Repository",* Available at: http://jiangtanghu.com
26. Noelia S´anchez-Maron˜o, Amparo Alonso-Betanzos, and Mar´ıa Tombilla-Sanrom´an, *"Filter methods for feature selection. A comparative study"*, Available at: https:// www. researchgate. net
27. Saptarsi Goswami and Amlan Chakrabarti, *"Feature Selection: A Practitioner View"*, I.J. Information Technology and Computer Science, 2014, 11, pp. 66-77.
28. K.Sutha and Dr.J.Jebamalar Tamilselvi, *"A Review of Feature Selection Algorithms for Data Mining Techniques"*, International Journal on Computer Science and Engineering (IJCSE) ISSN: 0975-3397, Vol. 7, No.6, June 2015, pp. 63-67.
29. Kenji Kira and Larry A.Rendell, *"The feature selection problem: Traditional Methods and a new Algorithm"*, AAAI-92 Proceedings. Available at: www.aaai.org
30. Jacek Biesiada and Wlodzislaw Duch, *"Feature Selection for High-Dimensional Data: A Pearson Redundancy Based Filter",* Available at: : https:// www. researchgate. net
31. Fuchun Sun, *"Advances in Neural Networks: 5th International Symposium on Neural Networks"*, 2008.
32. Sajid Ali Khan, Ayyaz Hussain, Abdul Basit, and Sheeraz Akram, *"Kruskal-Wallis-Based Computationally Efficient Feature Selection for Face Recognition"*, The Scientific World Journal Volume 2014 (2014).
33. Adrian Pino and Carlos Morell, *"Analytical and Experimental Study of Filter Feature Selection for High Dimensional Datasets",* Eureka-2013, Fourth International Workshop Proceedings, Atlantis press, 2013.

BIOGRAPHY

Mani. K received his MCA and M.Tech from the Bharathidasan University, Trichy, India in Computer Applications and Advanced Information Technology respectively. Since 1989, he has been with the Department of Computer Science at the Nehru Memorial College, affiliated to Bharathidasan University where he is currently working as an Associate Professor. He completed his PhD in Cryptography with primary emphasis on evolution of framework for enhancing the security and optimizing the run time in cryptographic algorithms. He published and presented around 15 research papers at international journals and conferences.

P.Kalpana received B.Sc and M.Sc degrees in Computer Science from Seethalakshmi Ramaswami College, affiliated to Bharathidasan University, Tiruchirappalli, India in 1999 and 2001 respectively. She received M.Phil degree in Computer Science in 2004 from Bharathidasan University. She also received MBA degree in Human Resource Management from Bharathidasan University in 2007. She is presently working as an Assistant Professor in the Department of Computer Science, Nehru Memorial College, Puthanampatti, Tiruchirappalli. She is pursuing PhD degree in Computer Science at Bharathidasan University. Her research interests include Algorithms, Data Pre-processing and Data Mining techniques.