ijecs open access www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 7 Issue 12 December 2018, Page No. 24439-24447
Index Copernicus Value (2015): 58.10, 76.25 (2016) DOI: 10.18535/ijecs/v7i12.05

Lung Cancer Feature Selection using Minimum Spanning Tree Based Clustering.

Kadam Pallavi R.[1], Dr. Satish Kumar Varma.[2]

Department of Information Technology,
Pillai College Of Engineering New Panvel, India

Abstract-
The recent increase of data poses a major challenge in data extracting. High dimensional data contains high degree of irrelevant and redundant information. Feature selection is the process of eliminating such irrelevant and redundant data set with respect to the task to be performed. Several features selection techniques are used to improve the efficiency and performance of various machine learning algorithms. There are several methods that have been proposed to extract features from such high dimensional data. This paper proposes Clustering based extended Fast Feature Selection method to extract features from high dimensional data. The proposed algorithm Semi-supervised learning which is useful to partition the data in appropriate clusters. Also, it selects the most frequent feature subset from the input.

Keywords- Feature Selection, Search strategies, machine learning, clustering, relevance, semi-supervised learning.

I. Introduction

In recent years, social media services are used very widely that allow people to communicate and express themselves conveniently and easily. The pervasive use of social media generates high dimensional data. This creates new challenges to the task of data mining such as classification and clustering processes. The one of the approaches for handling such large scale and high dimensional data is Feature Selection.

Feature selection methods have been used from long years in the field of statistics and pattern recognition with the wide spread use of machine learning techniques [1]. Feature selection methods are needed when there are too much data to be processed efficiently by machine learning algorithms, or when some of the features are costly to acquire and hence the minimum number of features are preferred [2].

Feature Selection methods are used to satisfy some common goals such as eliminating irrelevant and redundant data and improving result comprehensibility, maximizing the accuracy of classifier, reducing dimensionality, and helping to avoid slow execution time of learning algorithms [3].

According to their working principles, the feature selection methods are divided into following categories [3].

1. The methods which select the best subset of features that has a certain number of features

2. The methods which select the best subset of features according to their own principles, independent of outside measures

A. Feature Selection Methods:

According to interaction with learning algorithms, feature Selection methods are divided into several classes [6].

1) *Filter method:* It is the feature selection method which works independent from the learning algorithm. It consists of algorithms built in the adaptive systems for data analysis [4]. They use an evaluation function which relies on properties of data. Filter methods are fast, scalable and can

be used with any learning algorithm effectively [5] .

Distance based and margin-based criterion are considered useful for filters.

2) *Wrapper method:* This type of feature selection method uses learning algorithm to guide its search process to weigh features. The algorithms of wrapper method are wrapped around the adaptive systems providing them subsets of

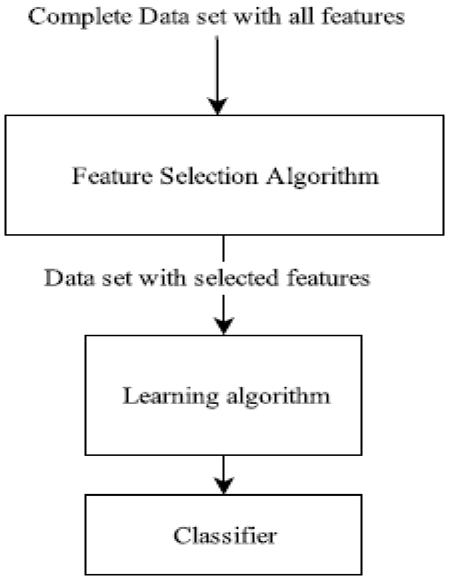

Fig. 1 Filter Method

Embedded methods are faster than wrappers and make the most efficient selection for the learning algorithm that they collaborate with [10].

B. Feature Selection Groups:

features and receiving their feedback [8]. These types of approaches aim to improving results of the specific predictors they work with.

3) *Embedded method:* In this type of method, a feature selection is embedded into a learning algorithm and optimized for it. They are based on performance evaluation metric calculated directly from data and have no direct reference to the results of any data analysis systems

Within the filter model, different feature selection algorithms can be further classified into two groups namely feature weighting algorithms and subset search algorithms. This classification is based on whether they evaluate the goodness of features individually or through feature subsets [11].

1) *Feature weighting algorithms:* These assign weights to each feature and rank them based on their relevance to the target concept. A feature is good and will be selected if its weight of relevance is greater than a threshold value. The algorithm called Relief is based on this criterion [12].

2) *Subset search algorithms:* These searches through candidate feature subsets which are guided by a certain evaluation measure which captures the goodness of each subset [13]. An optimal subset is selected when search ends.

In recent years, the application of cluster analysis is more effective than traditional feature selection algorithms, with respect to the filter feature selection methods [15].

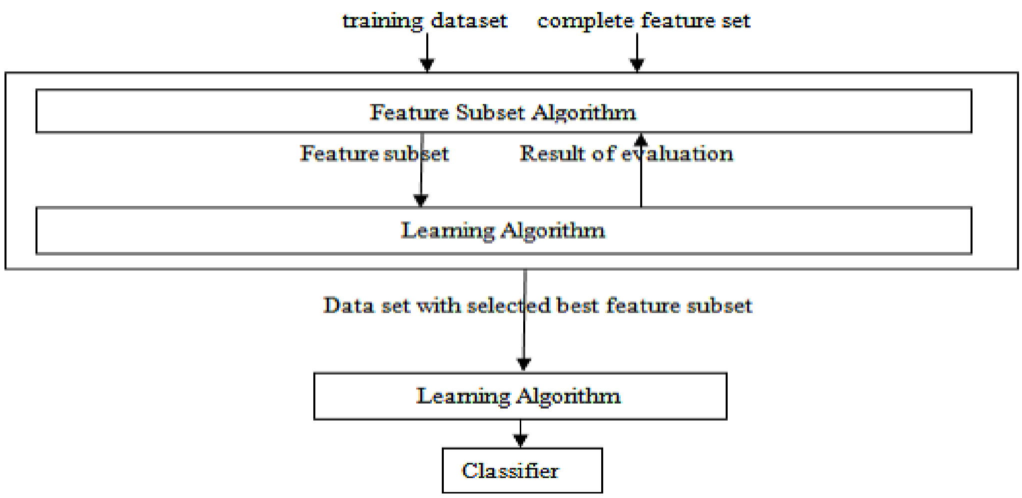

Fig. 2 Wrapper method

The graph-theoretic methods have been used in many applications those use cluster analysis. Graph theoretic clustering works in following steps: Compute a neighborhood graph of instances and after that delete any edge which is much larger or much shorter than its neighbors [9]. The result gained is forest and each tree in that forest is a cluster. In proposed work we apply graph theoretic clustering methods to select features. For that, Minimum Spanning Tree based clustering technique is used. Based on MST we propose the algorithm which is extended FAST clustering-based algorithm. Like FAST it works in two steps. In first step relevant features are divided into cluster. In second step most, representative feature is selected from clusters. The proposed algorithm is based on semi-supervised learning. It finds best suitable subset. In addition to that it also finds most frequent feature subset which can be used for further reference. It is well improved for feature selection from text and image data.

This clustering-based algorithm has high probability of producing a subset of useful and independent features. The proposed algorithm tested upon publically available image microarray and text data sets. Also, it is compared with two well known feature selection algorithms. The proposed algorithm selects most useful and relevant features.

II. Related Work

Feature selection is the process of finding and removing irrelevant and redundant features. As irrelevant features are not useful in contributing to predictive accuracy and redundant features provide the same information which is already present in other selected features. There are several algorithms, some can effectively remove irrelevant features but fail to deal with redundant features [12].

Relief is one of the algorithms which rely on relevance evaluation [12]. The key idea of this algorithm is to estimate the relevance of features by considering how well their values distinguish between the instances of the same and different classes that are near to each other. Relief randomly samples a number (m) of instances from the training set and up-dates the relevance estimation of each feature based on the difference between the selected instance and the two nearest instances of the same and opposite classes. The time complexity of Relief is O(mMN) where M is number of instances and N is number of features in the data set. However, the major drawback of Relief is it does not deal with removing redundant features. As long as features are seen relevant to the class concept, they will all be selected even though many of them are highly correlated to each other.

Another subset search algorithm called CFS exploits heuristic search. But this algorithm does not have strong scalability to deal with high dimensional data.

An efficient well-known algorithm named Fast Correlation-Based Filter Solution (FCBF) can effectively identify both irrelevant and redundant features with less time complexity than subset search algorithms [11].

The algorithm FCBF#, extended FCBF, has different search strategy than FCBF and it can produce more accurate classifiers for size k subset selection problem. FCBF# selects best subset of features from the full set by applying backward elimination. This algorithm is good alternative for feature selection from images and text data [16]. But the time complexity if FCBF and FCBF# is somewhat large.

In recent years, Minimum Spanning Tree (MST) based Clustering algorithms are mostly used for feature selection, because they do not assume that data points are grouped around centers or separated by a regular geometric curve [9]. Recent algorithm named as Fast Clustering based Feature Selection Algorithm (FAST) is based on this MST strategy [19]. FAST algorithm works like following:

1. Features are divided into different clusters by using graph theoretic clustering methods.
2. From each cluster, the most representative feature that is strongly related to target class is selected and it results final feature subset.

As features in the different clusters are relatively different and independent, the clustering based FAST produces useful and independent features with high probability. The significant drawback of FAST is some of the unlabelled data is not divided in particular cluster in which it should be, as FAST uses supervised learning. It affects the representative feature selection process to be held after clustering and subsequently it affects the resultant feature subset.

The proposed algorithm is extended version of FAST. It uses semi supervised learning to label the features. It is the combination of supervised as well as unsupervised learning. Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. As unlabeled data also can be classified by using labeled data it is useful to divide the data into their appropriate clusters.

In addition to finding best feature subset relevant to target concept, it also determines most frequent feature subset by using a well known Apriori algorithm. Like FAST, the proposed algorithm does not limit to some specific types of data.

III. Feature Subset Selection

A. Framework:

The proposed algorithm is composed of two connected components of irrelevant feature removal and redundant feature elimination.

1. Irrelevant Feature Removal: It obtains features relevant to the target concept by eliminating irrelevant ones.
2. Redundant Feature Elimination: It removes redundant features from relevant ones via choosing representatives from different feature clusters, and thus produces the final subset. Redundant feature elimination is some what complex than irrelevant feature removal.

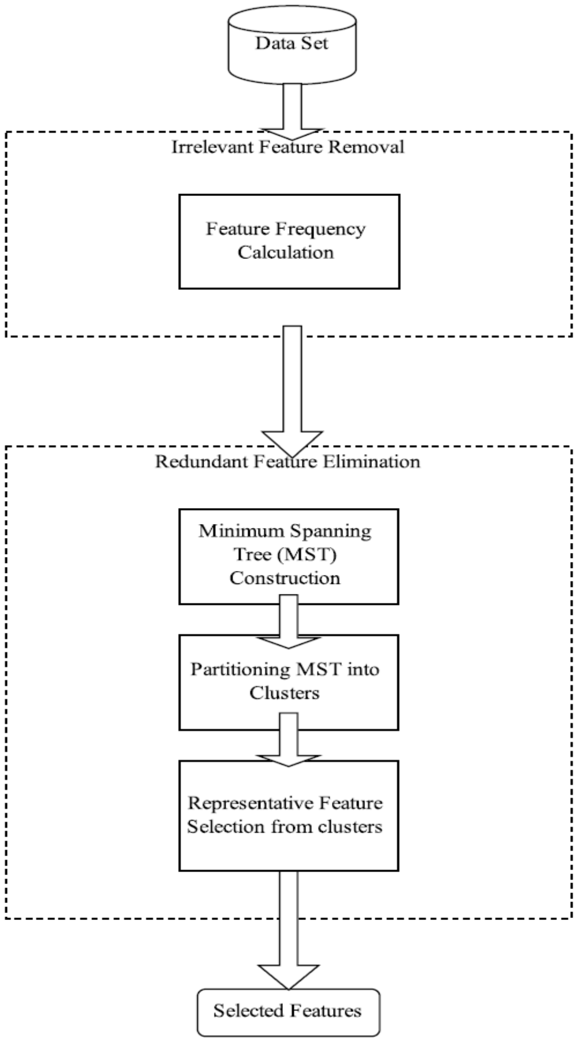

Fig. 3 Framework of the proposed algorithm

B. Definitions:

For this concept some definitions of relevant and redundant features are used. Then the definitions based on variable correlation are provided.

Suppose F is the full set of features, $F_i \in F$ be a feature, $S_i=F-\{F_i\}$ and $S_i' \subseteq S_i$. Let s_i' be a value assignment of all features in S_i', f_i a value-assignment of feature F_i and *c* a value assignment of the target concept *C*. The relevant feature can be defined as follows. These definitions are defined by Yu and Liu in [5].

Definition 1 (Relevant Feature): F_i is relevant to the target concept C if and only if there exists some s_i', f_i and c, such that, for probability p $(S_i'=s_i', F_i=f_i)>0$,
$p\ (C=c \mid S_i'=s_i, F_i=f_i) \neq p\ (C=c \mid S_i'=s_i)$, Otherwise, feature F_i is irrelevant feature.
It indicates that there are two kinds of relevant features due to different S_i'

1. When $\mid S_i'=S_i$, from the definition we can know that F_i is directly relevant to the target concept
2. When $S_i' \nsubseteq S_i$, from the definition we may obtain that $p(C|S_i, F_i)=p(C \mid S_i)$.

Most of the information which is in redundant features is already there in other features. As a result, redundant features have no contribution in getting better interpreting ability to the target concept. It is based on Markov blanket. The definitions are introduced as follows.

Definition 2 (Markov blanket): Suppose, given a feature $F_i \in F$, let $M_i \subset F(F_i \notin M_i)$, M_i is said to be a Markov blanket for F_i if and only if

$$p(F-M_i-\{F_i\}, C|F_i, M_i)=p(F-M_i-\{F_i\}, C|M_i).$$

Definition 3 (Redundant feature): Assume S be a set of features, a feature in S is redundant if and only if it has a Markov Blanket within S.

The feature redundancy is measured in terms of feature correlation and feature relevance is measured in terms of feature-target concept correlation. The correlation between feature values or feature values and target classes is nonlinear. The symmetric uncertainty (SU) is used to decide the relation between two features or feature and target class and is derived from the mutual information by normalizing it to the entropies of feature values or feature values and target classes, and has been used to evaluate the goodness of features for classification [19].

Therefore, symmetric uncertainty can be chosen as the measure of correlation between either two features or a feature and the target concept.

The symmetric uncertainty is defined as

$$SU(X,Y) = \frac{2 \times Gain(X|Y)}{H(X) + H(Y)}. \quad (1)$$

Where

1. H(X) is the entropy of a discrete random variable X. Suppose p(x) is the prior probabilities for all values of X, H(X) is defined by

$$H(X) = -\sum_{x\in X} p(x) \log_2 p(x). \quad (2)$$

2. Gain(X|Y) is the amount by which the entropy of Y decreases. It reflects the additional information about Y provided by X and is called the information gain [] which is given by

$$\begin{aligned} Gain(X|Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X). \end{aligned} \quad (3)$$

   Where H(X|Y) is the conditional entropy which quantifies the remaining entropy (uncertainty) of a random variable X given that the value of another random variable Y is known.

Suppose p(x) is the prior probabilities for all values of X and p(x|y) is the posterior probabilities of X given the values of Y, H(X|Y) is defined by

$$H(X|Y) = -\sum_{y\in Y} p(y) \sum_{x\in X} p(x|y) \log_2 p(x|y). \quad (4)$$

Information gain is a symmetrical measure. Means the amount of information gained about X after observing Y is equal to the amount information gained about Y after observing X. This ensures that the order of two variables will not affect the value of the measure.

Symmetric uncertainty treats a pair of variables symmetrically, it compensates for information gain's bias toward variables with more values and normalizes its value to the range [0,1]. A value 1 of SU (X, Y) indicates that knowledge of the value of either one completely predicts the value of the other and the value 0 reveals that X and Y are independent of each other.

Given: SU (X, Y) - symmetric uncertainty of variables X and Y, then

T-Relevance - relevance between a feature and the target concept or class C,

F-Correlation - correlation between a pair of features,

F-Redundancy – feature redundancy, and

R-Feature – representative feature of a cluster

can be defined as follows. These definitions are defined by Song, Ni and Wang in [19].

Definition 4 (T-Relevance): The relevance between the feature $F_i \in F$ and the target concept C is referred to as the T-Relevance of F_i and C, and denoted by $SU(F_i,C)$. If $SU(F_i,C)$ is greater than a predetermined threshold θ, we say that F_i is a strong T-Relevance feature.

Definition 5 (F-Correlation): The correlation between any pair of features F_i and $F_j (F_i,F_j \in F \wedge i \neq j)$ is called the F-Correlation of F_i and F_j, and denoted by $SU(F_i,F_j)$

Definition 6 (F-Redundancy): Let $S = \{F_1,F_2,...,F_i,...,F_{k<|F|}\}$ be a cluster of features. If $\exists F_j \in S$, $SU(F_j,C) \geq SU(F_i,C) \wedge SU(F_i,F_j) > SU(F_i,C)$

is always corrected for each $F_i \in S(i \neq j)$, then F_i are redundant features with respect to the given F_j (i.e. each F_i is a F-Redundancy).

Definition 7 (R-Feature): A feature $F_i \in S=\{F_1,F_2,...,F_i,...,F_k\}(k < |F|)$ is a representative feature of the cluster S (i.e. F_i is a R-Feature) if and only if, $F_i=\text{argmax}_{F_j\in S}SU(F_j,C)$.

This means the feature, which has the strongest T-Relevance, can act as an R-Feature for all the features in the cluster.

According to the above definitions, feature subset selection can be a process that identifies and retains the strong T-Relevance features and selects R-Features from selective feature clusters. The heuristics are that

1. Irrelevant features have no/weak correlation with target concept;

2. Redundant features are assembled in a cluster and a representative feature can be taken out of the selective cluster.

C. *Algorithm:*

The proposed algorithm works in following three steps to remove redundant features.

1. Constructing the minimum spanning tree (MST) from a weighted complete graph
2. Partitioning MST into a forest and in the forest each tree representing a cluster
3. Selection of representative features from the clusters

For a data set D with m features $\{F_1, F2, Fi...,F_m\}$ and class C, we apply the semi supervised learning to label the data. Then we compute the frequency of each feature F-Frequency in the first step. Here we apply Apriori algorithm to find out most frequent feature subset. To find the best subset suitable to target concept, we compute T-Relevance $SU(F_i,C)$ value for each feature F_i $(1 \leq i \leq m)$. The features whose $SU(F_i,C)$ values are greater than a predefined threshold θ comprise the target-relevant feature subset $F'=\{F_1',F_2',...,F_k'\}$ where $(k\leq m)$.

In the second step, we first calculate the F-Correlation $SU(F_i',F_j')$ value for each pair of features F_i' and F_j' $(F_i',F_j' \in F' \wedge i \neq j)$. Then, viewing features F_i' and F_j' as vertices and $SU(F_i',F_j')(i \neq j)$ as the weight of the edge between vertices F_i' and F_j', a weighted complete graph G=(V,E) is constructed where $V = \{F_i' | F_i' \in F' \wedge i \in [1,k] \}$ and $E = \{(F_i',F_j') \mid (F_i',F_j' \in F' \wedge i,j \in [1,k] \wedge i \neq j)\}$. As symmetric uncertainty is symmetric further the F-Correlation $SU(F_i',F_j')$ is symmetric as well, thus G is an undirected graph.

The complete graph G shows the correlations among all the target-relevant features. Unfortunately, the constructed graph G is very dense as it has k vertices and k(k-1)/2 edges. For high dimensional data, edges with different weights are strongly interwoven. Moreover, the decomposition of this dense complete graph is NP-hard. Thus for graph G, we build a MST, which connects all vertices such that the sum of the weights of the edges is the minimum. For that we use the well-known Prim's algorithm. The weight of edge $((F_i',F_j')$ considered as F-Correlation $SU(F_i',F_j')$.

The third step after building the MST is, we first remove the edges $E=\{(F'i,F'j) \mid E = \{(F_i',F_j') |(F_i',F_j' \in F' \wedge i,j \in[1,k] \wedge i \neq j)\}$, whose weights are smaller than both of the T-Relevance $SU(F_i',C)$ and $SU(F_j',C)$, from the MST. Each deletion results in two disconnected trees T1 and T2. Assuming the set of vertices in any one of the final trees to be V(T), we have the property that for each pair of vertices $(F_i',F_j' \in V(T))$, $SU(F_i',F_j')\geq SU(F_i',C) \vee SU(F_i',F_j') \geq SU(F_j',C)$ always holds. From Definition 6 is can be seen that this property guarantees the features in V(T) are redundant.

After removing all the unnecessary edges, a forest Forest is obtained. Each tree $T_j\in$Forest represents a cluster that is denoted as $V(T_j)$ which is the vertex of T_j also. We chose selective clusters to select the features. The features in each cluster are redundant, thus for each selective cluster we choose representative feature F_R whose T-Relevance is greatest. All F_R from each selective cluster comprise the final feature subset.

The algorithm CBFAST is shown below.

Algorithm 1:

Inputs: D $(F_1,F2,...,F_m, C)$ - the given data set
θ - the T-Relevance threshold
Δ – the F-frequency threshold

Output: S - selected feature subset.

//==== Part 1: Irrelevant Feature Removal ====

1. Apply **algorithm 2** to calculate F-frequency and add F into S
2. for i = 1 to m do
3. T-Relevance = SU(F_i,C)
4. if T-Relevance > then
5. S = S ∪ {F_i};

//==== Part 2: Minimum Spanning Tree Construction ====

6. G = NULL; //G is a complete graph
7. for each pair of features {F_i',F_j'} ⊂ S do
8. F-Correlation = SU(F_i',F_j')
9. Add F_i' and/or F_j' to G with F-Correlation as the weight of the corresponding egde.
10. minSpanTree = Prim (G); //Using Prim's Algorithm to generate the minimum spanning tree

//==== Part 3: Tree Partition into clusters and Representative Feature Selection from clusters====

11. Forest = minSpanTree
12. for each edge E_{ij} ∈ Forest do
13. if SU(F_i',F_j') < SU(F_i',C) ∧ SU(F_i',F_j') < SU(F_j',C) then
14. Forest = Forest - E_{ij}
15. S = ϕ
16. for each tree T_i ∈ Forest do
17. $F_R^j = \mathrm{argmax}_{F_k' \in T_i} SU(F_k',C)$
18. S = S ∪ {F_R^j};
19. return S

Algorithm 2: FP algorithm for irrelevant feature removal

C_k: Candidate itemset of size k
L_k: frequent itemset of size k

Join set: C_k is generated by joining $L_{k-1}$with itself
Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset

Inputs: D (F_1, F2,...,F_m, C) - the given data set
Output: feature subset
L_1= {frequent items};
for(k= 1; L_k!=∅; k++) **do begin**
C_{k+1}= candidates generated from L_k;
for each transaction t in database do
increment the count of all candidates in C_{k+1}that are
contained in t
Lk+1= candidates in Ck+1with min_support
end
return ∪kLk;

D. Data Source and Experimental Setup:

To evaluate the performance, the proposed algorithm is to be operated on some publically available data sets that cover the range of application domains such as text, image, and microarray data. The features involved vary from 30 to 40000.

To evaluate the effectiveness of the algorithm is compared with another feature selection algorithm like FCBF, CFS. Also, its search efficiency is compared with FAST also.

E. Results:

The module one differentiates between relevant and irrelevant features related to target concept. Here target concept is Risk1Yr. Figure 4.1 shows the relevant features selected and irrelevant features rejected by method.

Feature Classification After T-Relevance Calculation

Selected Features (Relevant Features)

Feature Name	T-Relevance
WEIGHIN	1.5083
TOTALCW2	1.572
AGE	1.4706
TOTALCW6	1.6047
TOTALCW4	1.5886
STAGE	1.5374
TOTALCIN	1.5575

Unselected Features (Irrelevant Features)

Feature Name	T-Relevance
TRT	1.4014
ID	1.2979

Fig. 4 Relevant Features

Figure 4.2 and 4.3 shows the MST generated by Kruskals and Prims algorithm.

SR.NO	To	From	Value
1	0	2	0.2392
2	0	1	0.2007
3	0	4	0.2589
4	1	6	0.2656
5	0	3	0.2505
6	0	5	0.2619

Fig. 4.2 MST by Kruskal

SR.NO	To	From	Value
1	2	6	0.2667
2	5	6	0.2688
3	1	6	0.2656
4	3	6	0.2669
5	0	6	0.2657
6	4	6	0.268

Fig. 4.3 MST by Prim

After removing redundant features, the final feature subset selected by both Prim's and Kruskal's algorithm is shown in figure 4.4 and 4.5.

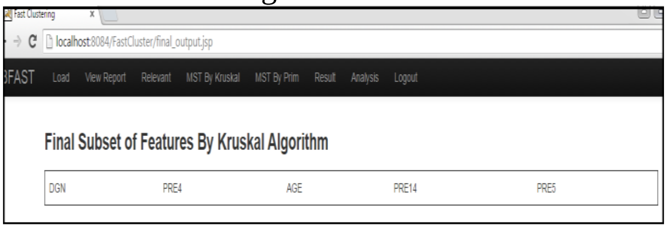

Fig 4.4 Final Feature subset by Kruskal

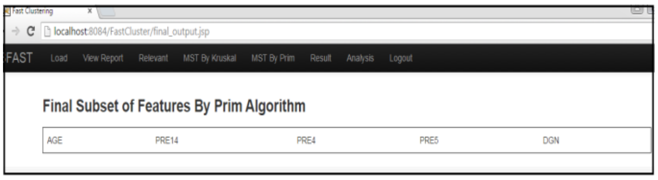

Fig 4.5 Final Feature subset by Prim.

In figure 4.6, the graph is plotted which shows the running time of Kruskal's and Prim's algorithm. From this graph we can say that the time complexity of generating MST by Kruskal's algorithm is less than that of using Prim's. Thus, it reduces total time complexity of this algorithm.

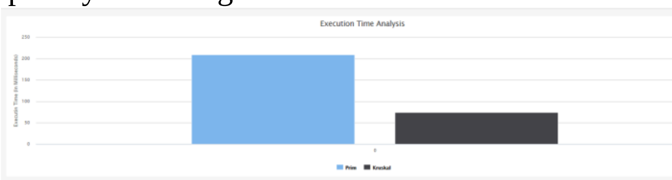

Fig4.6 Execution Analysis of Kruskal and Prim

Generally other compared algorithms achieve significant reduction of dimensionality by selecting only a small portion of the original features. The Proposed algorithm on average obtains the best proportion of selected features.

The proposed algorithm is best suitable for microarray data and performance is highly improved for image data.

IV. Conclusion

In this paper, a novel clustering-based feature subset selection algorithm for high dimensional data is proposed. The algorithm involves steps like, eliminating irrelevant features, constructing MST, partitioning MST and selecting representative features from each cluster.

As while removing irrelevant features, most useful and most frequent features are retained, it results the subset of useful features. The resultant feature subsets contain the features highly correlated with the class, yet uncorrelated with each other. The proposed algorithm on average obtains the best proportion of selected features than the other algorithms. Others reduce dimensionality significantly by selecting only a small portion of the original features. The proposed algorithm does not limit for any particular data and is best suitable for microarray data and also highly improved for image data.

Acknowledgement

We would like to thank to the editors and the anonymous reviewers for their insightful and helpful comments and suggestions.

References

[1] John, G., Kohavi, R., Pfleger, K. "Irrelevant features and the subset selection problem". In Proceedings of the Eleventh International Machine Learning Conference (1994) 121–129. New Brunswick, NJ: Morgan Kaufmann.

[2] Saeys, Y., Inza, I., Larrañaga, P. "A review of feature selection techniques in bioinformatics". Bioinformatics, 23(19) (2007), 2507– 2517.

[3] Liu H., Motoda H. "Computational methods of feature selection". Taylor, (2008).

[4] Guyon, I., Elisseef, A. "An introduction to variable and feature selection". J. Machine Learning Res. 3 (2003) 1157–1182.

[5] Molina, L.C., Belanche, L., Nebot, A. "Feature Selection Algorithms: A Survey and Experimental Evaluation". ICDM, (2002).

[6] Setiono, R., and Liu, H. “A probabilistic approach to feature selection-a fitler solution”. In Proceedings of International Conference on Machine Learning, 319-327 (1996).

[7] Blum, A.L., Langley, P. “Selection of relevant features and examples in machine learning”. Artificial Intelligence (1997) 245-271.

[8] Senliol, B. “Feature Selection Methods on High Dimensional Space”. Undergraduate Thesis for Computer Engineering Program in ITU, 2008.

[9] Shilpa N. Dande, Prof. Parinita Chate, “A Fast Clustering-Based Feature Subset Selection Algorithm for High Dimensional Data”, International Journal of Innovative Research and Studies, Volume 3 Issue 3, March 2014.

[10] Peng, H.C., Long, F., Ding, C. “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and minredundancy”. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8, pp.1226–1238, 2005.

[11] L. Yu and H. Liu. “Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution”. In Proceedings of The Twentieth International Conference on Machine Leaning (ICML-03), 856-863, Washington, D.C., August 21-24, 2003.

[12] Yao-Hong Chan; “Empirical comparison of forward and backward search strategies in L-GEM based feature selection with RBFNN”. International Conference on Machine Learning and Cybernetics (ICMLC), Vol. 3, pp- 1524 - 1527 ; 2010.

[13] P. Somol, P. Pudil; “Feature Selection Toolbox”. Pattern recognition, Published by Elsevier Science Ltd; 2002.

[14] A.GowriDurga, A.GowriPriya, “Feature Subset Selection Algorithm for High Dimensional Data using Fast Clustering Method”, IJCAT International Journal of Computing and Technology, Volume 1, Issue 2, March 2014.

[15] Forman G., “An extensive empirical study of feature selection metrics for text classification”, Journal of Machine Learning Research, 3, pp 1289-1305, 2003.

[16] Dettling, M., Bühlmann, P. “Supervised Clustering of Genes”. Genome Biology (2002), 3: research 0069.1–0069.15.

[17] Blake, C., & Merz, C. “UCI repository of machine learning databases”. (http://www.ics.uci.edu/»mlearn/MLRepository.html.) (1998).

[18] Computer Science (Artificial Intelligence); 2003.

[19] Qinbao Song, Jingjie Ni and Guangtao Wang, “A Fast Clustering-Based Feature Subset Selection Algorithm for High Dimensional Data”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL: 25 NO: 1 YEAR 2013.