A New Approach of Feature Selection Technique for Improving Performance of the Supervisor Technique

Shankara Gowda S R[2]
Assistant Professor,
Dept. of CS&E,
Don Bosco Institute of Technology,
Bangalore, India.

Dr. Nanda Kumar A N[1]
Professor & Head,
Dept. of IS&E,
New Horizon College of Engineering,
Banglore, India.

Rashmi B R[3]
M.Tech Scolar,
Dept. of CS&E,
Don Bosco Institute of Technology,
Bangalore, India.

Abstract - **Feature selection encompasses pinpointing a subsection of the most beneficial features that yields well suited results as the inventive entire set of features. A feature selection algorithm may be appraised from both the good organization and usefulness points of view. Although the good organization concerns the time necessary to discover a subsection of features, the usefulness is related to the excellence of the subsection of features. In this paper a fast clustering based feature Selection algorithm is proposed which removes redundant features by clustering and completely removes irrelevant features by statistical measures to select the most significant features from the dataset. The Proposed feature selection technique is compared with the existing techniques and the results obtained will demonstrate and prove the better performance.**

Keywords - Dataset, Data mining, feature extraction, feature selection, feature clustering.

I. INTRODUCTION

In the past decades, the data dimensionality involved in machine learning and data mining tasks has increased explosively. Data with extremely high dimensionality has presented severe challenges to existing learning methods, known as the curse of dimensionality [25]. With the continuation of a large set of features, a learning model tends to over fit and their learning execution savages. To address the problem of the curse of dimensionality, dimensionality reduction techniques have been studied, which form an significant branch in the machine learning research area. Feature selection is one of the most used techniques to decrease dimensionality among practitioners. It aims to choose a small subset of the relevant features from the unique ones according to certain relevance evaluation criterion [23], which usually leads to enhanced learning performance, e.g. higher learning accuracy, lower computational cost, and better model interpretability. Feature selection has been successfully applied in many real applications, such as pattern recognition, text categorization, Image processing, bioinformatics, and so forth.

According to whether the label information is utilized, different feature selection algorithms can be classified into supervised, unsupervised, or semi-supervised algorithms. With respect to different selection strategies, feature selection algorithms can also be categorized as being of the filter, wrapper, hybrid, and embedded models. Feature selection algorithms of the filter model are independent of any classifier. They estimate the relevance of a feature by studying its characteristics using certain statistical criteria.

Relief, Fisher score, CFS, and FCBF are among the most representative algorithms of the filter model. On the other hand, algorithms belonging to the wrapper model utilize a classified as selection criteria. In other words, they select a set of features that has the most discriminative power using a given classifier, such as: SVM, KNN, and so on. An example of the wrapper model is the FSSEM, 1SVM [1]. Other examples of the wrapper model could be any combination of a preferred search strategy and given classifier. Since the wrapper model depends on a given classifier, cross-validation is usually required in the estimation process. They are in general more computationally exclusive and prejudiced to the chosen classifier. Therefore, in real applications, the filter model is more popular, especially for problems with large datasets. However, the wrapper model has been empirically proven to be better, in terms of classification accuracy, to those of a filter model. Because of these reason in every model, the hybrid model was proposed to overcome any issues between the channel and wrapper models. Mainly it incorporates the statistical criteria, as filter model does, to choose a few hopeful elements subsets with a given cardinality. Second, it picks the subset with the most elevated characterization exactness [4]. Thus, the hybrid model usually achieves both similar accuracy to the wrapper and similar efficiency to the filter model. Hybrid model feature selection algorithms includes: BBHFS [3], HGA [3].

Finally, the embedded model performs feature selection in the given learning time. In other words, it performs model fitting and feature selection simultaneously. Examples of embedded model include C4.5 [4], BlogReg [2], and SBMLR [1]. Based on different types of outputs, most feature selection algorithms can be categorized as one of the three classes: subset selection [5], which returns a subset of selected features identified by the index of the feature; feature weighting [5], which proceeds weight matching to each feature; and the hybrid of subset selection and feature weighting, which returns a ranked subset of features. Commonly, a feature selection method comprises of four fundamental steps [4], to be specific, subset generation, subset evaluation, stopping criterion, and result validation. In the initial step, a candidate feature subset will be chosen in view of a given inquiry approach, which is sent, in the second means, to be assessed by evaluation criterion. The subset that best fits the assessment rule will be chosen over all the candidates that have been assessed after the ceasing foundations are met. In the last step, the picked subset will be approved utilizing domain knowledge or validation set.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICACT - 2016 Conference Proceedings

Various feature selection algorithms are proposed in recent days, which majorly concentrate only on removing the irrelevant features present in the data set where they do not deal with removing the redundant features from the training dataset. The redundant features in the large data set reduce the accuracy of the classifiers. This paper proposes a rank based feature selection algorithm named as unsupervised learning with ranking based feature selection. This algorithm selects the most significant features from the dataset and removes the redundant and irrelevant features. The unsupervised learner takes in the preparation dataset utilizing expectation maximization function and gatherings the components into clusters and these clustered features are positioned in view of the $\chi 2$ measure inside the cluster. The noteworthy features from every cluster are picked taking into account the threshold function.

The rest of the paper is organized as follows. The Second Section gives the insight about the already existing techniques and the related study is carried out. The Third section will elaborate the proposed cluster and rank based feature selection technique. The Fourth section demonstrates the results obtained to prove the efficiency of the proposed work with graphs and finally the conclusion.

II. BACKGROUND WORK

Feature selection is intend at a selecting a subset of features by eliminating irrelevant or no predictive information. It is a procedure of selecting a subset of unique features according to exact criteria. Irrelevant features do not provide to the accuracy and redundant features mostly provide the information which is already there in other features. There are numerous feature selection algorithm present, some of them are helpful at removing irrelevant features but not effective to handle redundant features. So far some of the other can remove irrelevant feature while taking care of redundant features [1]. FAST algorithm comes in to second group. One of the feature selection algorithms is Relief [3], which weighs every feature according to its capability to discriminate instances under different targets based on distance-based criteria function. Though, Relief is ineffective at removing redundant features as two predictive but highly correlated features are expected both to be highly weighted [4]. Relief-F [5] extends Relief, enabling this technique to work with noisy and partial data sets and to deal with multiclass problems, but still cannot identify redundant features. Redundant features also influence the accuracy and speed of learning algorithm; hence it is necessary to eliminate it. CFS [6], FCBF [7], and CMIM [9] are examples that take into consideration the redundant features. CFS [6] is accomplished by the hypothesis that a good feature subset is one that include features highly connected with the target, yet uncorrelated with each other. FCBF ([7], [8]) is a fast filter method which can distinguish relevant features as well as redundancy among relevant features without pair wise correlation analysis. CMIM [9] iteratively picks features which increases their shared information with the class to predict, conditionally to the response of any feature already picked.

Several feature selection for clustering methods have been proposed in the literature. Some algorithms hold text data, while others hold streaming data. Still others are capable of handling different kind of data. In this section, we will discuss different methods with respect to data types they can handle.

A. Spectral Feature Selection (SPEC)

The Spectral Feature selection (SPEC) algorithm is a unified framework that enables the joint study of supervised and unsupervised learning, we will utilize SPEC in this work as an example of filter-based unsupervised feature selection methods. SPEC [8] estimates the feature relevance by estimating feature consistency with the spectrum of a matrix derived from a resemblance matrix S. SPEC utilizes the Radial-Bases Function RBF as a resemblance function between two samples xi and xj.

Graph G will be constructed from S and adjacency matrix W will be constructed from G. Then, degree matrix $\sum D$ will be calculated from W. $\sum D$ is diagonal matrix where $\sum Dii = Pnj = 1Wij$. Given $\sum D$ and W Laplacian matrix L and the normalized Laplacian matrix L are computed The main idea behind SPEC is that the features consistent with the graph structure are assigned related values to instances that are close to each other in the graph. Therefore, these features should be relevant since they perform similarly in each similar group of samples (i.e. clusters). Motivated by graph hypothesis that states that graph structure information can be captured from its spectrum, SPEC studies how to choose features indicated to the structure of the graph G induced from the samples similarity matrix S.

The weight of each feature fi in SPEC is evaluated using three functions. These functions were derived from the normalized cut function with the spectrum of the graph, and extended to their more general forms. In this chapter, we will not explain these functions in detail; therefore, we refer the reader to [8] for more details. We assume here that each function takes feature vector fi and returns the weight based on the normalized Laplacian L.

B. Feature Weighting K-Means

K-means clustering is one of the most well-liked clustering techniques. It has been widely used in data mining and machine learning problems. Large amount of *k*-means variations was proposed to handle feature selection [9, 6, 3, and 7]. Most of these variations begin by clustering the data into k clusters. Then, it assigns weight to each feature. The feature that decreases within-cluster distance and increases between-cluster distance is chosen, hence, gets higher weight. For example, an entropy weighting *k*-means (EWKM) was projected for subspace clustering. It concurrently minimizes the within-cluster scattering and maximizes the negative weight entropy in the clustering process. EWKM calculates the weight of each feature in each cluster by counting the weight entropy in the objective function of *k*-means. The subsets of features matching to each cluster are, then, preferred based on that weight. Thus, EWKM allows subspace clustering where the set of preferred features may differ from one cluster to another. In addition, [7] proposed feature weighting *k*-means clustering using generalized Fisher ratio that minimizes the ratio of the average of within-cluster distortion over the normal between-cluster distortion. In this algorithm, numerous candidate clustering are generated and the one with the minimal Fisher ratio is firm to

be the final cluster. Similarly, [8] proposed∑ another variation of feature weighting *k*-means (W-*k*-means) that measures the weight of each feature based on its variance of the within-cluster distance.

Input:
D: dataset
N: number of samples
ψ : the objective function Eq. (0.7)
Initialize:
C: apply k-means on D to obtain initial
clusters Z: choose k centroids
w: initialize the weight of each feature so that
   t: 0

Output:
C: the clustering
Z: the centroids
W: the features weights
1. **while** not stop **do**
2. Fix Z and w and solve ψ with respect to C.
3. Stop when no changes occur on C
4. Fix C and w and solve ψ with respect to Z.
5. Stop when no changes occur on Z.
6. Fix C and w and solve ψ with respect to w.
7. Stop when no changes occur on w.
8. t=t+1
9. **end while**

The above Algorithm illustrates the process of W-*k*-means. It, iteratively, minimizes by fixing two parameters at each step and solve with respect to the third one. If there is no change in after the minimization, the algorithm is said to be converged.

C. Feature Selection Based on Information Gain

In this feature selection technique, the information gain measure is applied on the training dataset to identify the significant features based on information gain value of the individual features [10] in terms of entropy. The entropy value of each feature of the training dataset "TD" is calculated and ranked based on the information gain value [17].

D. Feature Selection Based on Gain ratio

In this feature selection technique, the information gain ratio *GR (f)* is calculated for each feature of the training dataset "*TD*" to identify the significant feature based on the information present in the features of the "*TD*" [17].

E. Feature Selection Based on symetric uncertainity

This technique uses the correlation measure to select the significant feature from the training dataset "*TD*". In addition to that, the symmetric uncertainty "*SU*" is calculated using the entropy measure to identify the similarity between the two features *fi* and *fj* [21, 22].

F. Feature Selection Based on Unsupervised learning algorithm

Unsupervised learning is formally known as clustering algorithm. This algorithm groups similar objects with respect to the given criteria like density, distance, etc. The objects present in a group are highly similar than the outliers. This technique is applied for selecting the significant features from the training dataset. Each unsupervised algorithm has its own advantages and disadvantages that determine the application of each algorithm. This paper utilizes expectation maximization (EM) clustering technique to select the significant features by identifying the independent features in order to remove the redundant features from a training dataset "*TD*" [23−25].

G. Decision Tree based classifier

This tree based classifier constructs the predictive model using decision tree. Basically, the statistical tool information gain is used to learn the dataset and construct the decision tree. Information gain is computed for each attribute present in the training dataset "*TD*". The feature with higher information value is considered as the root node and the dataset is divided into further levels. The information gain value is computed for all the nodes and this process is iterated until a single class value is obtained in all the nodes [20, 21].

H. Supervised Learning Algorithm

The supervised learning algorithm builds the predictive model "*PM*" by learning the training dataset "*TD*" to predict the unlabeled tuple "*T*". This model can be built by various supervised learning techniques such as tree based, probabilistic and rule based. This paper uses the supervised learners namely naive bayes (NB), instance based IB1 and C4.5/J48 supervised learners to evaluate and compare the performance of the proposed feature selection algorithm in terms of predictive accuracy and time taken to build the prediction model with the existing algorithms[24] .

III. CLUSTER RANKED FEATRUE SELECTION TECHNIQUE AND STATISTICAL MATRIX FORM(SMF)

The proposed framework is built on spectral graph theory. In the framework, the relevance of a feature is determined by its reliability with the structure of the graph induced from S. The three feature ranking functions (ϕb1 (·), ϕb2 (·) and ϕb3 (·)) lay the foundation of the framework and enable us to derive families of supervised and unsupervised feature selection in a unified manner. We realize the unified framework in Algorithm 1. It selects features in three steps:

(1) building resemblance set S and constructing its graph representation (Line 1-3); (2) evaluating features using the spectrum of the graph (Line 4-6); and (3) ranking features in descending order in terms of feature relevance3 (Line 7-8).

This algorithm follows a sequence of steps to select the significant features from the training dataset. Initially, the training dataset is transposed. Then features are clustered and the features in each cluster are ranked based on the χ^2 value. Then the threshold value is computed to choose highly significant features from each clustered features. All the chosen features from different clusters are combined together as candidate significant features selected from the training dataset

Input: X, γ (.), k, $\hat{\varphi} \in \{ \hat{\varphi}_1 , \hat{\varphi}_2 , \hat{\varphi}_3 \}$
Output: SFSPEC - the ranked feature list

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICACT - 2016 Conference Proceedings

1. construct S, the similarity set from X (and Y);
2. construct graph G from S;
3. build W, D and L from G;
4. **for** each feature vector fi do
5. $\hat{f_i} \leftarrow \frac{D^{\frac{1}{2}}\hat{f_i}}{\|D^{\frac{1}{2}}\hat{f_i}\|}$; $SFSPEC^{(i)} \leftarrow \hat{\varphi}(Fi)$;
6. end
7. ranking SFSPEC in ascending order for $\hat{\varphi}_1$ and $\hat{\varphi}_2$ or descending order for $\hat{\varphi}_3$;
8. return SFSPEC

The time complexity of the proposed technique mainly depends on the cost of building the similarity matrix and the calculation of $°(¢)$. If we use the *RBF* function to build the resemblance matrix and $°(¢)$ is in the form of *Lr*, the time complexity of *SPEC* can be obtained as follow.

First, we need $O(mn2)$ operations to build S*; W; D; L* and *L*. And we need $O(rn3)$ operations to calculate $°(L)$. Next, we need $O(n2)$ operations to calculate *SFSPEC* (*i*) for each feature: transforming $f i$ to b $f i$ require $O(n)$ operations; calculating using b′1, b′2 and b′2 need $O(n2)$ operations4. Therefore, we need $O(mn2)$ operations to calculate scores for *m* features. Last, we needs $O(m\log m)$ operations to rank the features. Hence, the overall time complexity of *proposed technique* is $O_¡ (rn + m)n2 ¢$, or $O_¡(mn2¢$ if $°(¢)$ is not used.

SMF provides a statistical view about data based on data orientation into the dataset, works on ranking model and we get weight of data based on query and classified data. Utilize a standard deviation technique to determine an overall ranking of the feature matrix Modules used are Processing Ability(PA)-In a network there are usually free loaders who download files without sharing any of their resources, which impacts the search performance of coadjutant communities. To prevent free loaders, utilize the PA to differentiate between leeching and enthusiastic user.

Effective Sharing-This feature is based on the observation that the file-sharing among user is extremely unbalanced. Instead of all user participating in file-sharing, only a small number of volunteer user provide most of the resource sharing services in a network.

Index Power (IP)-Here if a user records a large number of file sharing messages in its index, many of the messages may never be used by the mechanism. The Index Power (IP) feature determines the amount of content in a queried user's index and assesses its quality.

IV. PERFORMANCE EVALUATION

We empirically evaluate the performance of *proposed technique*. In the experiments, we compared the algorithms specified as in algorithm with Laplacian Score (unsupervised) and ReliefF (supervised). Laplacian Score and ReliefF are both state-of-the-art feature selection algorithms, comparing with them enables us to examine the efficiency of the algorithms derived from proposed

technique. Four benchmark data sets are used for experiments: Hockbase6, Relathe6, Pie10p7 and Pix10p8.

Hock base & Relathe are text data sets generated from the 20-new-group data: Baseball *vs.* Hockey (Hockbase) and (2) Religion *vs.* Atheism (Relathe). Pie10p & Pix10p are face image data sets containing 10 persons in each. And we sub sample the images down to a size of 100£100 = 10000.

Table 1: Summary of Four Benchmark Dataset

Data set	Instance	Feature	Classes
HOCKBASE	1993	8298	2
RELATHE	1427	8298	2
PIE 10p	210	10000	10
PIX 10p	100	10000	10

	$\hat{\varphi}_1, r$	$\hat{\varphi}_1, r^4$	$\hat{\varphi}_2, r$	$\hat{\varphi}_2, r^4$	$\hat{\varphi}_3, r$	$\hat{\varphi}_3, r^4$
HOCKBASE(Laplacian Score=0.58)						
RBF	0.61	0.61	0.58	0.58	0.60	0.60
DIF	0.74	0.74	0.75	0.74	0.70	0.70
RELATHE(Laplacian Score=0.59)						
RBF	0.63	0.63	0.59	0.59	0.55	0.55
DIF	0.67	0.67	0.67	0.67	0.61	0.61
PIE10P(Laplacian Score=0.74)						
RBF	0.75	0.75	0.74	0.78	0.87	0.86
DIF	0.81	0.81	0.92	0.91	0.91	0.91
PIX10P(Laplacian Score=0.88)						
RBF	0.78	0.78	0.88	0.94	0.93	0.91
DIF	0.79	0.79	0.84	0.85	0.93	0.92

The performance of the proposed FSULR algorithm is analyzed in terms of predictive accuracy, time to build the model and the number of features reduced. The Figure 1 represents the overall performance of feature selection algorithms in producing the predictive accuracy for supervised learners and it is observed that the FSLUR performs better than all the other feature selection algorithms compared. The second and third positions are retained by the FSCon and FSInfo respectively. Figure 2 shows that the proposed FSULR achieves better accuracy than the other algorithms compared for NB and IB1 classifiers. The FSCon achieves better results for J48 classifier as compared to all the other classifiers.
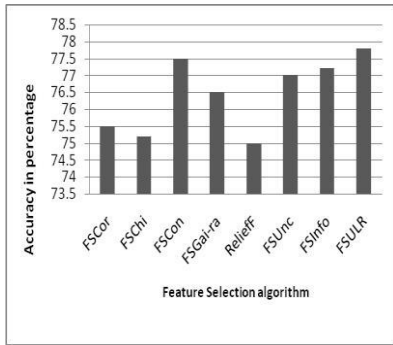
Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICACT - 2016 Conference Proceedings

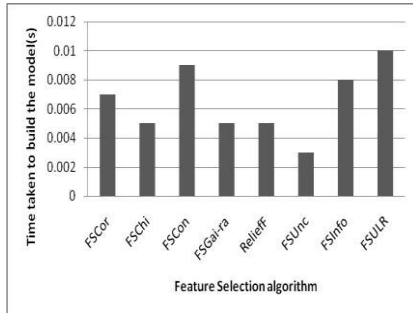Figure 1: Over all prediction accuracy Comparison.

Figure 2: Overall Time Consumption Comparison

CONCLUSION

Feature selection algorithms can be classified as supervised or unsupervised (Liu & Yu, 2005). Recently, an increasing number of researchers paid concentration to developing unsupervised feature selection. The authors proposed an unsupervised feature selection algorithm based on iteratively determining the soft cluster indicator matrix and the feature weight vector. They then comprehensive the algorithm to handling data with class labels. Since the input of the algorithm restricts to covariance matrix, it does not handle general resemblance matrix and cannot be extended as a universal framework for designing new feature selection algorithms and covering existing algorithms.

In this paper, we propose a general framework of spectral feature selection for both supervised and unsupervised learning, which facilitates the combined study of supervised and unsupervised feature selection. We show that some dominant existing feature selection algorithms can be resulting as special cases from the framework; and families of new effective algorithms can also be derived. Extensive experiments display the generality and usability of the proposed framework. It is natural to extend the framework with existing kernel and metric learning methods (Lanckriet et al., 2004) to a variety of applications. Use of Statistical Matrix Form as statistical measure ,can easily fetch the features by their weight and in ascending order.

REFERENCES

[1]   J. Sinno, Q. Y. Pan. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–135, 2010.

[2]   M. R. Rashedur, R. M. Fazle. Using and comparing different decision tree classification techniques for mining ICDDR, B Hospital Surveillance data. *Expert Systems with Applications*, vol. 38, no. 9, pp. 11421–11436, 2011.

M. Wasikowski, X. W. Chen. Combating the small sample class imbalance problem using feature selection. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1388–1400, 2010.

[4]   Q. B. Song, J. J. Ni, G. T. Wang. A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering*, vol. l5, no. 1, pp. 1–14, 2013.

[5]   J. F.Artur, A. T.M´ario. Efficient feature selection filters for high-dimensional data. *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1794– 1804, 2012.

[6]   J. Wu, L. Chen, Y. P. Feng, Z. B. Zheng, M. C. Zhou, Z. H. Wu. Predicting quality of service for selection by neighborhood-based collaborative filtering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 2, pp. 428–439, 2012.

[7]   C. P. Hou, F. P. Nie, X. Li, D. Yi, Y.Wu. Joint embedding learning and sparse regression: A framework for unsupervised feature selection. *IEEE Transactions on Cybernetics*, vol. 44, no. 6, pp. 793–804, 2014.

[8]   P. Bermejo, L. dela Ossa, J. A. G´amez, J. M. Puerta. Fast wrapper feature subset selection in high-dimensional datasets by means of filtre-ranking Original Research Article. *Knowledge-based Systems*, vol. 25, no. 1, pp. 35–44, 2012.

[9]   S. Atulji, G. Shameek, V. K. Jayaramanb. Hybrid biogeography based simultaneous feature selection and MHC class peptide binding prediction using support vector machines and random forests. *Journal of Immunological Methods*, vol. 387, no. 1-2, pp. 284–292, 2013.

[10]   H. L. Wei, S. Billings. Feature subset selection and ranking for data dimensionality reduction. *IEEE Transactions on* Pattern Analysis and Machine Intelligence, vol. 29, no. 1, pp. 162–166, 2007.

[11]   X. V. Nguyen, B. James. Comments on supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognition*, vol. 46, no. 4, pp. 1220–1225, 2013.

[12]   A. G. Iffat, S. S. Leslie. Feature subset selection in large dimensionality domains. *Pattern Recognition*, vol. 43, no. 1, pp. 5–13, 2010.

[13]   M. Hall. Correlation-based Feature Selection for Machine Learning, Ph.D dissertation, The University of Waikato, New Zealond, 1999.

[14]   Y. Lei, L. Huan. Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, vol. 5, no. 1, pp. 1205–1224, 2004.

[15]   M. Dash, H. Liu, H. Motoda. Consistency based feature selection. In *Proceedings of the 4th Pacific Asia Conference* on Knowledge Discovery and Data Mining, Kyoto, Japan, pp. 98–109, 2000.

[16]   M. Robnik-Sikonja, I. Kononenko. Theoretical and empirical analysis of Relief and RReliefF. *Machine Learning*, vol. 53, no. 1-2, pp. 23–69, 2003

[17]   H. Uˇguz. A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. *Knowledge-based Systems*, vol. 24, no. 7, pp. 1024–1032, 2011

[18]   M. Dash, H. Liu, H. Motoda. Consistency based feature selection. In *Proceedings of the 4th Pacific Asia Conference on Knowledge Discovery and Data Mining,* Kyoto, Japan, pp. 98–109, 2000

[19]   S. Yijun, T. Sinisa, G. Steve. Local-learning-based feature selection for high-dimensional data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1610–1626, 2010

[20]   W. Peng, S. Cesar, S. Edward. Prediction based on integration of decisional DNA and a feature selection algorithm RELIEF-F. *Cybernetics and Systems*, vol. 44, no. 3, pp. 173–183, 2013

[21]   H. W. Liu, J. G. Sun, L. Liu, H. J. Zhang. Feature selection with dynamic mutual information. *Pattern Recognition*, vol. 42, no. 7, pp. 1330–1339, 2009.

[22]   M. C. Lee. Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, vol. 36, no. 8, pp. 10896–10904, 2009.

[23]   P. Mitra, C. A. Murthy, S. K. Pal. Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 301–312, 2002

[24]   J. Handl, J. Knowles. Feature subset selection in unsupervised learning via multi objective optimization. *International Journal of* Computational Intelligence Research, vol. 2, no. 3, pp. 217– 238, 2006

[25]   H. Liu, L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactionson Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491– 502, 2005.