A Literature Review of Feature Selection Techniques and Applications

Review of feature selection in data mining

S.Visalakshi
Research Scholar
Department of Computer Science
Avinashilingam Institute for Home Science and Higher Education for Women
Coimbatore-43, Tamilnadu, India
visaraji@gmail.com

V. Radha
Professor
Department of Computer Science
Avinashilingam Institute for Home Science and Higher Education for Women
Coimbatore-43, Tamilnadu, India
radhasrimail@gmail.com

***Abstract* — Water is the elixir of life. It is a vital component of human survival. Water should be purified for better and healthy style life of all living and non-living things. The quality of water plays an important role for all living beings. Water used for drinking purpose should be colourless, odourless and free from excess salts. Detecting such a variety of contamination from the drinking water becomes a challenging task. Feature selection acts as a significant role in identifying irrelevant features and redundant features from large dataset. Feature selection is a preprocessing course of action universally used for large amount of data. Feature selection concepts instruct us, to pick a subset of features or catalog of attribute or variables which helps to build an efficient model for describing the selected subset. Other than selecting the subset, it also congregate some other purposes, such as dimensionality reduction, compact the amount of data which are required for learning process, progress in predictive accuracy and increasing the constructed models. The main aim of this work is to investigate about the concept of feature selection, various criterions of feature selection methods and some existing methods are discussed from 1997 till 2014 and address the issues and challenges of feature selection.**

Keywords—data mining, feature selection, filter method, wrapper method, forward selection, backward elimination method, supervised learning, unsupervised learning.

I. INTRODUCTION

A general problem in machine learning is to identify a group of relevant features for constructing the classification model. The most important job of feature selection is to eliminate inappropriate and redundant data which facilitate to improve the performance of learning algorithms. Feature Selection (FS) is a spirited and productive field of research area in machine learning, data mining, pattern recognition, etc. Liu et al. (2010) state three main advantages of FS namely, eradicating inappropriate and unnecessary features, making the data mining task easier and increasing accuracy rate and simplifying the conditional model and making it more understandable [15]. According to Friedman (1997), for an accurate classifier, it is required to diminish both bias and variance of model [4]. Munson and Caruana (2009) states that the feature selection process helps us to find the greatest bias-variance trade-off point [16]. The basic concept of feature selection is given by different authors from their own angles. But finally, they instinct the same point. The subsequent list enumerates the feature selection in different ranges are listed below:

1. Idealized: find out the minimal subset [4],
2. Classical: Select a subset from number of features which is M<N[17], and
3. Improving the prediction accuracy: select relevant features or reduce the size of bulk of without affecting the accuracy metric [13].

The main benefit of feature selection during the process of learning, decrease in size of data, improvement in prediction accuracy, pull out important features, understanding the attributes or variables easily and finally execution time will be reduced. The final two aspects satisfy the general definition of data mining. Data mining is a term to describe the process of extracting relevant interesting patterns and relationship from large databases. Normally, feature selection process, will search for the subset and discover the best among 2^N of N features based on evaluation criteria. To select the features the supervised, unsupervised and semi-supervised learning techniques are used. Supervised feature selection method requires large amount of labeled training data. Unsupervised feature selection works very well with unlabeled training data. Finally, the semi-supervised is generally categorized into two major groups. The foremost one is based on cluster assumption whereas the succeeding set is based on multiple assumptions [27].

The paper is structured as follows: Section 2 describes about the general process of selecting features. Section 3 explained about the evaluation criteria of feature selection methods. Section 4 analyze about the various approaches and methods available in feature selection. Section 5 conclude the work and proposes future work.

II. GENERAL PROCESS OF FEATURE SELECTION

The features are selected using the four basic steps:

978-1-4799-3975-6/14/$31.00 ©2014 IEEE

1. Generation process- create the next candidate subset,
2. evaluation task-estimate the subset,
3. Stopping condition, and
4. Validation procedure-confirm whether the subset is suitable for problem statement.

In the first step, principally it produces the subset of features for estimation process. The search process will start with null features, all related features or with random subset of features. According to null features and all related features, iteratively the features are added or removed, but, with random selection of subset features, both are possible, i.e., features might be added or removed or produced randomly.

The generation step uses different approaches to pass through the available subsets. In addition to this process, instead of searching all possible subsets, it can stop, once it reaches certain number of features or it arrives with iterative numbers or once the optimum solutions are reached (Dash and Liu, 1997) [3]. John et al. (1994) [11] list two searching algorithms: Forward Selection and Backward Elimination. Other searching algorithms are, greedy hill-climbing algorithm (Yang and Honavar (1998) [24] and best first search (Hall (2000)) [6]. A new idea for feature selection was proposed by Yu and Liu (2004) [26]. It initiates with individual feature selection and then eliminates the irrelevant features. Later, subset selection is preformed to remove redundant features. To improve the performance of feature selection, scale back the search space for the subset selection. Suppose the generation procedures generate 2^N for N number of features even for medium sized N, this problem is handled effectively with different approaches such as complete, heuristic and random.

In the second step, evaluate the finest subset produced and this value will be judged against with the preceding top. If it invent to be superior than the before one then the new value will be replaced with previous subset. A valuation task will measure the judicious skill of feature selection methods into two extensive categories of methods. They are *filter* and *wrapper* methods. Filter method is free from inductive algorithm, and wrapper method utilize inductive algorithm as the evaluation function. The filter and wrapper methods are described in the following sections.

In third step, stopping criteria will be based on first step. It checks for adding up features or removing the features and if any features is not producing better subset and also ensure whether the evaluation function obtained the exact optimal subset. The loop continues until the stopping criterion is satisfied. Finally, the feature selection processs stop the progress of by giving the selected subset of features to a validation procedure. The validation is not done in feature selection but it is good to confirm whether the selected features are valid or not [11].

III. Evaluation Criterion methods

The variable evaluation is broadly categorized into two methods, namely, *Filter* and *Wrapper* methods. The main objective of evaluation is to compute the selective capability of feature or subset to make out the class label. The evaluation functions are classified into five categories which were analyzed briefly by several authors [11].

A. Filter Method

This method relies on general uniqueness of the data to be evaluated and pick feature subset not including any mining algorithm. Filter method uses the exact assessment criterion which includes distance, information, dependency and consistency [15]. The filtering approach is depicted in the following Fig. 1. In Fig.1, the feature are selected using ranking and statistical techniques, the selected feature subset are evaluated for quality check and this approach independent of using Machine Learning (ML) algorithm. The filter method uses the principal criteria of ranking technique and uses rank ordering method for variable selection. The reason for using ranking method is simplicity, produce good and relevant features. Ranking method will filter out irrelevant features before classification process starts.

The main advantage of using feature ranking is that it is computationally less and avoids the over fitting problems. Some of the ranking methods are Mutual Information, Pearson correlation criteria, Chi-square test, Correlation coefficients, etc. Filter process is faster than wrapper method. The advantage of filter method are, it is easily scalable with high dimensional datasets, computationally simple, fast in process and independent of data mining algorithm. So, feature selection will be performed only once and different classifiers are used for evaluation process. Main disadvantage of this technique are, it will ignore the interaction between classifiers and most proposed techniques will consider each feature separately which leads to degrade the classification performance compared to wrapper [20].

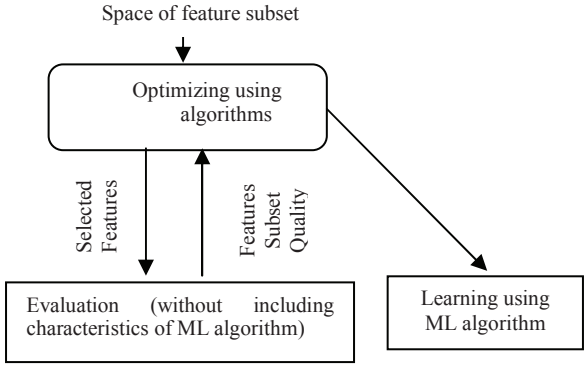

Fig.1: Filter based approach

B. Wrapper Method

The wrapper needs one set of mining algorithm and uses the performance as evaluation criteria. This method searches for feature which is suited for mining algorithm and aims to improve the mining performance. To evaluate the features, the predictive accuracy used for classification and goodness of cluster is evaluated using clustering [15]. The flow of wrapper based approach is shown in Fig.2, According to Fig.2, the evaluation of each feature depends on machine learning algorithm. The search procedure of all promising feature subsets is defined, and various subsets of features are generated and evaluated.

The quality of attribute is measured by using machine learning algorithm and applied to each subset feature. The advantage of wrapper is that, interaction will take place between feature subset search and model selection. The disadvantage is high risk of overfitting compared to filter method and is also computationally demanding in nature [14].

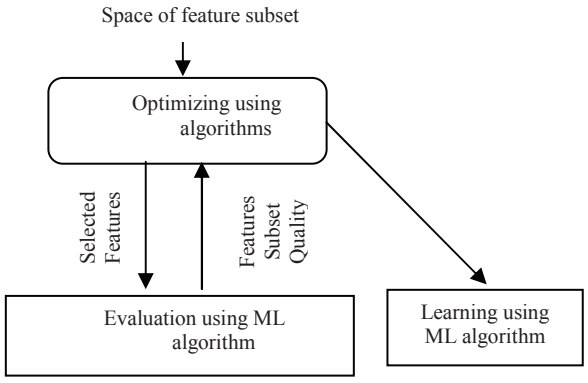

Fig.2. Wrapped based approach

IV. STATE OF THE ART

In this section, the papers are reviewed according to the year of publication. The existing and proposed methods of feature selection are analyzed briefly.

M.Dash and H. Liu [3] reviewed many existing methods from 1970. They identified four different steps for feature selection methods and classify the existing methods as generation procedure and evaluation functions. Different benchmark datasets (CorrAL (32 instances), modified Part 3+3 (binary classes + Boolean features) and Monk3) are used in comparative study. The advantages and disadvantages of each method are explained clearly. They compared the evaluation functions such as distance measure, information measure, dependence measure, consistency and classification error with three different general metrics, namely, generality, time complexity and accuracy. The guiding principles are given for applying feature selection methods to corresponding data types and domain characteristics. The generation procedures are categorized into three (complete, heuristic and random) and evaluation is categorized by five (distance, information, dependence, consistency and classifier error rate measure). Testing by combining the existing method of both general procedure and evaluation functions with new combination method to improve the feature selection concepts effectively are tried. Finally, it also proposes future research work for newcomers to this field.

Xiaohua Hu and Nick Cercone [23] applied and integrate different feature selection methods such as discretization, generalization and rough set for telecommunication customer database. Integrating these three methods helps to reduce the data horizontally and vertically. The basic concept of discretization and generalization are analyzed briefly with corresponding algorithm. The merits and demerits of both methods were analyzed. The first phase of the work will decrease greatly the number of tuples which are considered in further process of feature selection. The second phase, using rough set theory the relevant attributes, is chosen from the dataset. The rough set theory based feature selection algorithm proposed a filter rough set algorithm. Finally, reduced subset will transform into different knowledge, based on rule which in turn based on knowledge discovery algorithm. Combining all the work, the author proposed a discovery system known as DBROUGH-II. Dataset, used in this experiment, was telecommunication customer data. The proposed approaches reduce huge datasets horizontally and vertically. Data mining algorithm is used in vast application. Hence, the future work will apply the proposed approach with other applications.

Laetitia Jourdan, Clarisse Dhaenens and El-Ghazali Talbi [13] experiment that genetic algorithm work well with high dimensional dataset. The main aim of the work is to find out genetic features and environmental factors that lead to multi-factorial disease. Data mining tool is used for analysis and later propose 2-phase approach based on Genetic Algorithm (GA). The proposed approach evaluated using two different databases, i.e., artificial database used to validate the proposed approach. In the first phase, GA and clustering are used to handle the size of database problem (discover real interaction between locus gene). Then, the final result of GA gave as input to second phase. In this phase, k- mean clustering algorithm is used to improve the execution time with selected features. The same approach is applied in real time dataset and the result satisfies the biologist that the algorithm seems robust and it can isolate the interesting features from database.

Dunja Mladenic and Marko Grobelnik [2] presented the machine learning algorithm to extract text data. Real world database, which was collected from dataset, are used to evaluate the proposed learning algorithm. The high number of features is reduced with the help of stop-list, and instead of using whole document, the descriptive of document are implemented with the help of pruning concept. Text data are analyzed by using word sequence, and new approach was proposed to frame a word sequence effectively. The authors briefly describe the search strategy for text document. Feature subset selection approaches are discussed briefly and different scoring measures are also discussed in brief. The experimental results show that the proposed learning approach works better with Odds Ratio compared to other scoring measures like information gain, mutual information and cross entropy.

Riyaz Sikora and Selwyn Piramuthu [18] propose a new design of techniques and tools, based on genetic algorithm for handling massive databases. Self-adaptive feature selection concept, along with wrapper based feature selection and hausdroff distance measure, are used. The basic concept of feature selection process and Genetic algorithm are described elaborately. The main purpose of using Hausdorff distance is to measure computer vision and graphics application and excel in discriminate properties. The GA performs better compared to evolutionary algorithms even database size increase in future. The authors analyze previous work in multiclass classification, and multiclass learning methods.

Tao Li, Chengliang Zhang and Mitsunori Ogihara [22] solve the problem that exists when building multiple classifiers for tissue classification. This research paper compares various feature selection methods with classification methods. The study shows that multi-class classification

problem is very difficult to understand compared to binary sample classification. The artificial and real-time dataset are used for experiments. The results are good for database with small number of classes. When number of classes increase, prediction accuracy goes down. From the above study, no single feature selection scheme is available to handle different kinds of databases. In future, design a feature selection method which will consider the correlative between features and whether ensemble methods can apply to gene expression classification.

Oliver Buchtala, Manuel Klimek, and Bernhard Sick [17] demonstrate that Evolutionary Algorithm (EA) will perform feature and model selection simultaneously for Radial Basis Function networks (RBF) classifiers. To reduce optimization effort, various techniques are integrated to improve the EA. This paper addresses and experiments four different data mining problems. They are intrusion detection in computer networks, biometric signature verification, customer acquisition and optimization of chemical production. Experiments are evaluated with this proposed approach with the above four data mining problems and prove that evolutionary optimization of RBF network architecture applies to wide range of problems which exist in different data mining field. The result shows improvement in classification rates, speed in fitness evaluation, integration of soft constraints and temperature based control of EA. The authors' future work is to inculcate this evolutionary approach to filter based feature selection and heuristic model selection techniques.

Bo Wang, Kening Gao, And Bin Zhang [1] applied Rough Set Feature Selection Algorithm for discrepancy of information related to objects and it becomes a great challenge to perform inductive learning algorithm. To solve this problem, rough set theory introduces a new tool to solve the above problem. The general process of rough set feature selection is analyzed briefly. The toys information system database is used for experiment process. The rough set feature selection algorithm is used during preprocessing phase. After the reduction process, the relevancy was calculated for feature reduction. Results show that rough set feature selection algorithm is capable of reaching high efficient in subset feature estimation, looking for minimal reduction of NP-hard problem. This algorithm is feasible in nature but still needs further research to improve effectiveness and speed for high dimensional database.

Kian Hong Quah and Chai Quek [12] proposed a Monte Carlo Evaluative Selection (MCES) for non-linear system and found solutions for benchmark problems. This approach helps to improve the estimation and relevancy measures. To evaluate the proposed novel feature selection algorithm, four sets (audio-detection problem, sonar classification, automobile MPG prediction and stock marked prediction) of application and problems are taken into concern. Generally there is an assumption that wrapper based feature selection is capable of extracting the relevant features effectively, but in particle, the assumption is wrong. The second problem is, during classification, the interruption of bias will be introduced because of pre-defined threshold value, which in turn affects the classification accuracy. The third problem is, while integrating feature selection with induction algorithm, there is increase in execution cost. To address the above problems, the authors propose a novel feature selection algorithm, named MCES. The experimental result shows that MCES is capable of handling decoupling process effectively. It can handle both classification and regressive task, ability to identify both irrelevant and correlated features and finally there is high improvement in classification accuracy.

Hongwen Zheng and Yanxia Zhang [8] dealt with the problem of high-dimensional database in astronomical field. Increase in data becomes an intensive and demanding problem for extracting relevant information from database. To extract relevant information, a study has been done in feature selection, feature extraction, taxonomy of feature selection, comparison made between filter and wrapper methods. The computational time and accuracy are considered for evaluation of astronomical database. In general, filter method reduces the computational cost and wrapper increase in classification accuracy. For training a system, any learning algorithm is chosen, but selecting a suitable filter method is very important. To conclude, this work states that high dimensional database is resolved by using the feature selection methods.

Sven F. Crone, Nikolaos Kourentzes [21] address the problem which exists in artificial neural network architecture for time series data and propose a universal methodology which combines both filter and wrapper method for feature selection, automatic feature evaluation, construction and transformation. Comparisons have been made with bench mark methods for setting synthetic time series ESTSP'08 competition dataset. The problem faced from the first step of collecting input vector, where it must capture all characteristics of complex time series, including components of deterministic, cycle and seasonality, interacting with linear and non-linear pulses, level of shifts, structural breaks and different noise will raise for different time series data and also varies from season to season. No, suitable methodology for filter based feature selection method have been proposed to handle inconsistency across varying data condition. Likewise, the wrapper based method is also having problems in handling various parameters such as computational power, absence of valid and reliable evaluation, and no appropriate methodology will be opt to handle unknown time series frequencies and multiple overlying seasonality. Hence the authors point out general challenges to determine desirable properties of existing methodology of feature selection to specify input vector of neural network with feature evaluation, feature construction, feature transformation and network architecture selection. To address all these issues, the authors propose a unique methodology which combines both filter and wrapper methods. The experimental result shows that the proposed methodology outperforms in set of benchmark methods and achieves a good ranking for time series dataset.

Shih-Chieh Chena, Shih-Wei Linb, and Shuo-Yan Chou [19] proposed ensemble Scatter Search (SS) approach to select an appropriate parameter value for the applications. The popular classifiers such as Decision Tree (DT), Back Propagation Network (BPN) and Support Vector Machine (SVM) are ensemble to improve the classification accuracy rate. A brief review of three classifiers is presented and benefits of ensemble are also enlisted. The SS approach was

proposed in 1977 for population based and uses to solve small problems. This paper briefly describes the procedure for proposed SS ensemble approach. The proposed SSDT, SSBPN and SSSVM algorithm generate solution for diversification generation method, improvement method, reference set update method and subset generation method. To assess the performance of proposed SS ensemble approach, 18 dataset from UCI machine learning repository were used. From the experimental result, the classification accuracy rate is the best among other classifiers. To conclude the research work, the proposed approach gives better results compared to other 6 classifiers. Future work also proposed that SS based approach can be applied to meta-heuristic and for real-world problems and other genetic algorithm and particle swam optimization can also ensemble to achieve high classification accuracy rate.

Girish Chandrashekar and Ferat Sahin [5] made a comparative study of various techniques and methods present in feature selection. The main objective of feature selection is reducing the computation time, improving prediction performance and creates better understanding of various machine learning applications. The goal of this paper is to provide clear idea of elimination of irrelevant features in turn applies to many real time applications. The feature selection methods such as filter, wrapper and embedded methods are briefly analyzed. The Support Vector Machine (SVM) and Radial Basis Function (RBF) classifiers are used for wide range of applications. Totally there are 7 datasets are breast cancer, diabetes, lonospehere, liver disorders and medical from UCI machine learning repository, fault mode data and discrete fault mode are real time datasets. The validation method such as 2-fold cross validation and k-fold cross validation are used to evaluate the feature selection process. The overall study of this paper concludes that feature selection helps to understanding the data, enhancement of generalization and identification of irrelevant variables.

Hui Li et.al [9] create novel wrapper feature selection based SVM for Financial Distress Identification (FDI) using SVM models such as Linerar SVM (LSVM), Polynomial SVM (PSVM), Gaussian SVM (GSVM) and Sigmond SVM (SSVM) and also the authors explained about SVM based FDI. The statistics based wrapper feature selection of SVM and different levels of process are concise. To evaluate the proposed approach, the China based FDI dataset (668 samples) are applied. The experimental result shows that GSVM/LSVM with Statistics-based Wrapper Feature Selection improves the predictive performance compare to other SVM models.

Lan Bai, Zhen Wang, Yuan-Hai Shao and Nai-Yang Deng [14], proposed feature selection method based on twin support vector machine (TWIN) called (FTSVM) and feature ranking SVM (RTSVM). The linear SVM construct single hyperplane which corresponds to single weight for each features and linear Twin Support Vector Machine (TWSVM) construct two hyperplanes which correspond to two weights for each feature but proposed FTSVM uses feature selection matrix for each feature. Hence it leads to solve the multi-objective mixed-integer programming problem by greedy algorithm. The proposed novel method has two forms, they are linear FTSVM and nonlinear. The proposed feature selection method and feature ranking method are elaborated briefly. The algorithm for both methods is given in this paper. To analyze the performance of both proposed method namely RTSVM and FTSVM with state of the art L_1-TWSVM and L_2-TWSVM. Different UCI datasets (Aus, Ech, Hea, Hep, lon, Son, Spe and Wpb) more than 2500 samples are used to evaluate the performance efficiency of FTSVM and RTSVM. It applied to both linear and nonlinear cases. It works effectively in nonlinear case compare to linear.

The author enhanced the feature selection process with Parzen window (He Deng-chao, Hao Wen-ning, Chen Gang and Jin Da-wei [7]. This proposed algorithm named Parzen window improved feature selector (PWIFS) uses conditional mutual information for the evaluation of feature selection to overcome the problem of feature redundant and uses Parzen window to estimate the density function on the basis of probability. The general introduction about information theory, Parzen window density and how to calculate the conditional mutual information with parzen window are explained. To evaluate the proposed algorithm five dataset from UC-Irvine repository are used. PWIFS can select the features which enhance the performance rate and also this algorithm was applied to continuous variables. The PWIFS exhibits better performance compare to other conventional methods such as Parzen Window Feature selectior (PWFS), Mutual information feature Selector (MIFS) and MIFS-U.

Jimin Lee, Nomin Batnyam and Sejong Oh [10], proposed a new feature selection method based on R-Value. The R-Value is a measure and used to capture the overlapped region among the classes in features. The basic concept of filter, wrapper methods and feature selection algorithm with distance discriminate and Relief are discussed. The step by step algorithm to calculate the $R(F_k)$ are clearly explained. To evaluate the RFS algorithm several types of datasets are used and the dataset belongs to microarray data. To evaluate the feature selection, three classifiers such as k-Nearest Neighbor (KNN), Naïve Bayes (NB) and Support Vector Machine (SVM) are used. The experimental result shows that, the proposed feature selection method produced less time complexity and the classification accuracy is increased.

The author proposed multi annotator approach which helps to build annotation wrapper for search result records to identify frequent websites using apriori algorithm [25]. This work divided into three phases. The first phase resultant records are identified and divide those into different groups. Second phase, annotate each group and aggregate different annotations for the purpose of predicting final annotation label. The third phase, search site are automatically annotated with the help of wrapper annotation. The architecture of proposed design diagram is given and gave a clear idea about each phase. The six basic annotators are used along with new annotator. Three different domains (Book, Game and Music) are used for experiment purpose. The main feature of proposed annotator is capable to retrieve the result from web database and utilize both LIS and IIS web database from three domains. The experimental result shows that proposed technique taking less time for searching process. To evaluate the proposed technique, precision and recall are considered. Clustering based shifting method is applied to overcome the

problem of automatic aligned problem. The new approach works well even with new queries.

V. CONCLUSION

The main target of feature selection is to select the relevant subset and get rid of the irrelevant and redundant information from the database. This paper mainly focused on basic concept of feature selection, process of feature selection, evaluation criteria and various approaches used in feature selection which are applicable for all databases. From the state of art, feature selection methods are flexible and capable of providing a solution to any kind of problem faced when performing feature selection. From literature study, instead of using filter and wrapper methods separately, better embed both methods for selecting relevant features, improves in classification accuracy, speed up the process and reduce error rate. Still there is a scope that genetic algorithm and ensemble (filter and wrapper) are capable of handling multi-dimensional dataset. In future, it is planned to embed wrapper and filter method to extract relevant features for detecting contamination and from where the attacks are generated and giving alarm when contamination is detected in drinking water.

REFERENCES

[1] Bo Wang, Kening Gao, and Bin Zhang, "Algorithm of Feature Selection For Inconsistent Data Preprocessing Based Rough Set," International Journal of Information and Systems Sciences, vol. 1, no. 3-4, pp. 311-319, 2005.

[2] Dunja Mladenic and Marko Grobelnik, "Feature selection on hierarchy of web documents, Decision Support Systems," Elsevier, vol. 35, pp. 45– 87, 2003.

[3] M. Dash and H. Liu, "Feature Selection for Classification," Intelligent Data Analysis, Elsevier, vol. 1, pp. 131-156, 1997.

[4] J.H. Friedman, "On bias, variance, 0/1loss, and the curse-of-dimensionality," Data Mining and Knowledge Discovery, vol. 1, no. 1, pp. 55–77, 1997.

[5] Girish Chandrashekar and Ferat Sahin, "A survey on feature selection methods," Computers and Electrical Engineering, vol. 40, pp. 16–28, 2014.

[6] M.A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," Proceedings of Seventeenth International Conference on Machine Learning (ICML), Stanford, CA, Morgan Kaufmann, San Mateo, CA, pp. 359–366, 2000.

[7] He Deng-chao, Hao Wen-ning, Chen Gang and Jin Da-wei, "An Improved Feature Selection Algorithm Based on Parzen Window and Conditional Mutual Information," Proceedings of the 2nd International Symposium on Computer, Communication, Control and Automation (ISCCCA-13), pp. 0334-0337, 2013.

[8] Hongwen Zheng and Yanxia Zhang, "Feature selection for high-dimensional data in astronomy," Advances in Space Research, Elsevier, vol. 41, pp. 1960–1964, 2008.

[9] Hui Li, Chang-Jiang Li, Xian-Jun Wu and Jie Sun, "Statistics-based wrapper for feature selection: An implementation onfinancial distress identification with support vector machine," Applied Soft Computing, Elsevier, vol.19, pp. 57–67, 2014.

[10] Jimin Lee, Nomin Batnyam and Sejong Oh, "RFS: Efficient feature selection method based on R-value," Computers in Biology and Medicine, Elsevier, vol. 43, pp. 91–99, 2013.

[11] John, G.H, Kohavi, R and Pfleger, K, "Irrelevant features and the subset selection problem," Proceedings of the Eleventh International Conference on Machine Learning, pp. 121–129, 1994.

[12] Kian Hong Quah and Chai Quek, "MCES: A Novel Monte Carlo Evaluative Selection Approach for Objective Feature Selections," IEEE Transactions on Neural Networks, vol. 18, no. 2, pp. 431-448, 2007.

[13] Laetitia Jourdan,Clarisse Dhaenens and El-Ghazali Talbi,, "A Genetic Algorithm for Feature Selection in Data-Mining for Genetics," MIC'2001 - 4th Metaheuristics International Conference, pp. 29-33, 2001.

[14] Lan Bai, Zhen Wang, Yuan-Hai Shao and Nai-Yang Deng, "A novel feature selection method for twin support vector machine," Knowledge-Based Systems, Elsevier, vol. 59, pp. 1–8, 2014.

[15] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever evolving frontier in data mining," In Proc. The Fourth Workshop on Feature Selection in Data Mining, vol. 4, pp. 4–13, 2010.

[16] M. Munson and R. Caruana, "On feature selection, bias-variance, and bagging," Machine Learning and Knowledge Discovery in Databases, pp. 144–159, 2009.

[17] Oliver Buchtala, Manuel Klimek, and Bernhard Sick, Member, IEEE, "Evolutionary Optimization of Radial Basis Function Classifiers for Data Mining Applications," IEEE Transactions on Systems, Man, and Cybernetics—PART B: Cybernetics, vol. 35, no. 5, pp. 928-947, 2005.

[18] Riyaz Sikora and Selwyn Piramuthu, "Efficient Genetic Algorithm Based Data Mining Using Feature Selection with Hausdorff Distance," Information Technology and Management, Springer, vol. 6, pp. 315–331, 2005.

[19] Shih-Chieh Chena, Shih-Wei Linb, and Shuo-Yan Chou, "Enhancing the classification accuracy by scatter-search-based ensemble approach," Applied Soft Computing, Elsevier, vol. 11, pp. 1021-1028, 2011.

[20] Sunita Beniwal and Jitender Arora, "Classification and Feature Selection Techniques in Data Mining," International Journal of Engineering Research & Technology, vol. 1, no. 6, ISSN: 2278-0181, pp. 1-6, 2012.

[21] Sven F. Crone, Nikolaos Kourentzes, "Feature selection for time series prediction – A combined filter and wrapper approach for neural networks," Neurocomputing, Elsevier, vol. 73, pp. 1923–1936, 2010 .

[22] Tao Li, Chengliang Zhang and Mitsunori Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression," Bioinformatics, vol. 20, no. 15, pp. 2429–243, 2004.

[23] Xiaohua Hu and Nick Cercone, "Data Mining via Discretization, Generalization and Rough Set Feature Selection," Knowledge and Information Systems, Springer Verlang, vol. 1, pp. 33-60, 1999.

[24] J. Yang and V. Honoavar, Feature Extraction, Construction and Selection : A data Mining Perspective, chapter 1, Feature Subset Selection Using a Genetic Algorithm, , H. Liu and H. Motoda Eds, Kluwer academic publishers edition, pp. 117–136, 1998.

[25] V.Yogam and K.Umamaheswari, "Automatic Annotation Wrapper Generation and Mining Web Database Search Result," International Journal of Innovative Research in Science, Engineering and Technology, vol. 3, No. 3, ISSN: 2319-8753, March 2014.

[26] L. Yu and H. Liu, "Redundancy Based Feature Selection for Microarray Data," Proc. 10th ACM SIGKDD Conf. Knowledge Discovery and Data Mining, pp. 737-742, 2004.

[27] Zenglin Xu, Member, IEEE, Irwin King, Senior Member, IEEE, Michael Rung-Tsong Lyu, Fellow, IEEE, and Rong Jin, "Discriminative Semi-Supervised Feature Selection Via Manifold Regularization," IEEE TRANSACTIONS ON NEURAL NETWORKS, vol. 21, no.7,pp.1033-1047, 2010.