A Review on Feature Selection MethodsforHigh Dimensional Data

Senthil Kumar P., Daphne Lopez
School of Information Technology & Engineering,
VIT University, Vellore, India- 632014
senbe@rediffmail.com
daphnelopez@vit.ac.in

***Abstract*— Feature selection has become an important task for effective application of data mining techniquesin real-world high dimensional datasets. It is a process that selects a subset of original features by removing irrelevant and redundant features on the basis of the evaluation criteria without loss of information content. A feature selection method helps to reduce computational complexity of learning algorithm, improve prediction performance, better data understanding and reduce data storage space. Feature selectionhas gained more popularity in data mining and machine learning applications. The general procedure of feature selection process and overview of filter, wrapper and embedded method present in literature form the subject matter of this paper.**

Keyword: Feature Selection, Filter method, Wrapper method and embedded method

I. INTRODUCTION

Environment datasets in real-world are characterized by the large quantity of noise, redundant or irrelevant misleading features that may affect model accuracy.With these factors removed, learning from data mining and machine learning techniques can benefit greatly.A high-dimensional dataset increases the risks that data mining algorithms find wrong patterns that are void in general. Most techniques involve some degree of reduction. This is necessary in order to manage with large amounts of data. Machine learning methods are very difficult to handle with the large number of high-dimensional data found. Data pre-processing is a necessary step in the use of effective machine learning methods. Feature selection is an important techniqueused in data preprocessing.The main aim of feature selection is to determine the minimum number of feature subsets from a problem domain while retaining a high accuracy in representing the original features [1].

When the number of features selected is rather small, chances of information content may be low. On the hand, the presence of noise as also irrelevant data will be highly probable when many features are selected. Hence, feature selection should be on the right selection of subsets, avoiding too large or too small number of features. There are many benefits of an ideal feature selection, such as data visualization,data understanding, reduction of the memory storage and training time, reduce the dimensionality which may improve prediction and classification performance.

Feature selection cannot stand comparison with other dimensionality reduction methods such as Principle Component Analysis (PCA). Feature selection and dimensionality reduction methods reduce the number of features in the dataset. However, dimensional reduction creates a new combination of attributes, while feature selection methods reducethe features present in the dataset without changing them. [2]. Now-a-days many systems in a variety of fields deal with large datasets with high dimensionality. Some of the main areas of feature selection application are classification, prediction, image recognition, rule induction etc.

II. FEATURE SELECTION PROCESS

Feature selection consists of four stages which include subset generation, evaluation, stopping criterion and validation. The subset generation procedure implements a heuristic search method that generates subsets of features for evaluation. Features are added (forward selection), removed (backward elimination) random selection method and weighting method may be considered for the choice of the search starting point at each subset stage. The evaluation function computes the worth of the candidate subset of features formed by the generation procedure and compares this with the candidate found worthy earlier, replacing it if found to be better. A stopping criterion is tested for every iteration to determine whether the feature selection process should be continued or stopped based on the subset generation process. In general stopping criteria like add or remove of features to the subset generation process have not produced a much difference or the maximum number of iteration or a minimum number of features. Eventually validation of the feature selection method is done by different assessment procedure and the results obtained are compared with those got earlier.The general procedure for feature selection process as shown in fig.1

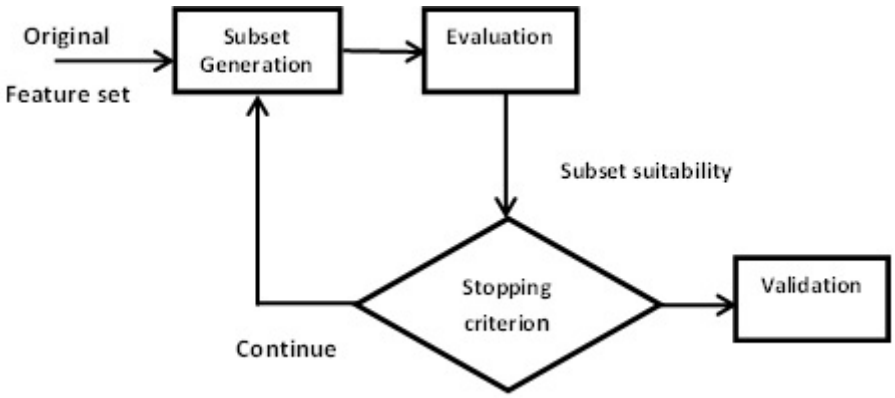

Fig.1. Feature selection process

III. TYPES OF FEATURE SELECTION METHODS

Feature selection methods are broadly classified into three categories: (a).the filter methods, (b) the wrapper methodand (c) embedded methods

A. Filter method

The Filter method use heuristics based on the characteristics of the data handled rather than an algorithm forevaluation of the merit of feature subsets, with the application of a statistical measure for assignment of a rank to each feature. Ranking measure is used to score the features and a threshold is used to eliminate features below the threshold. The Filter method tends to select subsets with all the features and, therefore, a proper threshold is required for the selectionof a subset. It is an independent measure for evaluating features subsets without using any learning algorithms. There are many independent criteria used for evaluation including interclass distance measures, information or uncertainty measures, dependency measures, consistency measure and probability of error measures [3].The method is much faster than wrapper methods, and practical for use on high dimensional datasets. Fig 2. shows the filter method

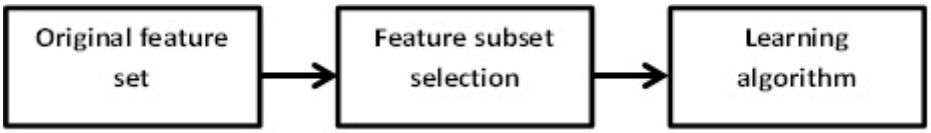

Fig. 2. Filter method

Correlation coefficient:This is a statistical test that measures the linear relationship between two variables. The Pearson Correlation coefficient is defined as

$$R(i) = \frac{\mathrm{cov}(x_i, T)}{\sqrt{\mathrm{var}(x_i) * \mathrm{var}(T)}}$$

Where x_i is the i the variable, T is the target, cov() is the covariance and var() the variance. The value of thecoefficient indicates that the strength of the relationship.

Information gain

Information gain measures the amount of bit of information obtained for category prediction through knowledge of the presence or absence of a term in a document.Such a statistical measure can be used for comparison and the consequentselection of features.

Relief algorithm is based on the filter approach where each feature given in a feature relevance criterion is used for ranking the features.A subset of features is selected on the basis of the threshold value. Relief algorithm can deal with nominal and numerical features. It cannot do removal of redundant features as two predictive, but highly correlated features are both likely to be given high relevance weightings. Its extension is called ReliefF , which solves the problem of noisy and incomplete data. Marko Robnik-˘Sikonja et al.[4] have done theoretically and empirically investigation of the algorithm of the Relief family, dealing with how they work their parameters, how irrelevant and redundant features affect their output, and what kind of dependencies they detect. Focus algorithm generates all feature subsets for determination of the minimum number of features that can provide a consistent labelling of the training data. A specific subset is eliminated when an inconsistency is found. This process continues until a consistent subset is found,stopping when a sufficient consistent solution is found. This algorithm is selects the ideal number of features when the data are noise-less.[5] Some of known filter based feature selection algorithms are Las Vegas filter (LVF) algorithm, Selection Construction Ranking using Attribute Pattern (SCRAP) algorithm, Entropy-Based Reduction (EBR), Fractal Dimension Reduction

(FDR). Huan Liau and Rudy Setiono [6] describes chi2 algorithm asmore effective and useful tool for feature selection and discretization of numeric attributes. The advantages of the filter based feature selection method lies in lesser computational complexity than the wrapper method. It can be scaled to very large dimensional datasets and it can also avoid data overfitting. Some filter methods does not differentiate the features in terms of the correlation to other features. Lack of feature dependency may affect the performance [18].

B. Wrapper method

Wrapper based feature selection method searches for the best subset of features using predetermined accuracy from an induction algorithm for evaluating generated subsets of features. The method may produce improved performance, but is expensive to run. It can reduce by a large number of features. Fig 3.shows the wrapper method

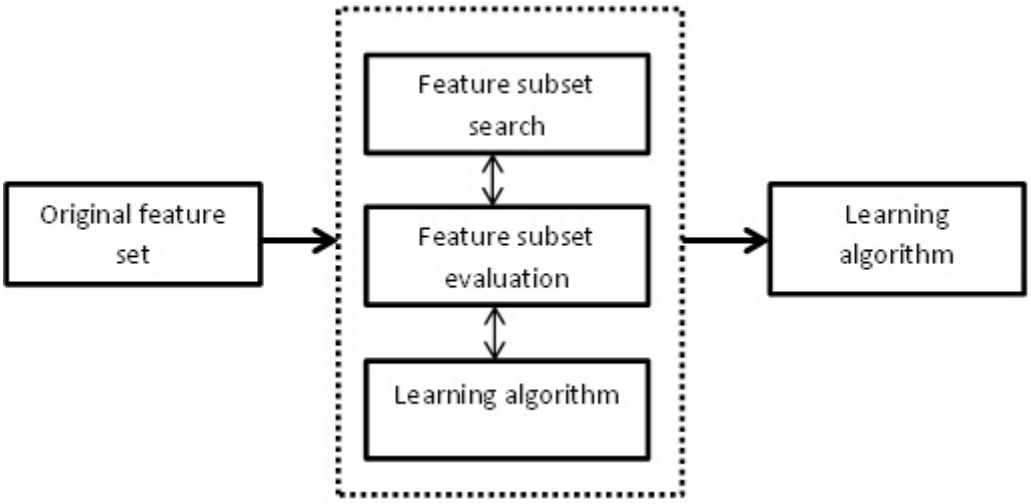

Fig. 3. Wrapper method

This is due to the use of learning algorithms in the evaluation of subsets, some of which can become problems when the datasets are very large. In the wrapper method, optimal subsets generation is the process of the heuristic search algorithm. Many search algorithms can be used for finding a subset of features. Sequential search is iterative in the nature of algorithm starting withan empty feature or complete features and add or remove one feature at a time until the target is obtained.Heuristic search algorithms evaluate different subsets to optimize the best solution. Wrapper methods make utilize he classifier for providing a score the subsets of features based on their predictive power. Li-Yeh Chuang et al [7] propose chaotic binary particle swarm optimization forimplementation the feature selection, in which the K-nearest neighbour (K-NN) method acts as a classifier for evaluating classification accuracy. The wrapper based feature selection method shows good results with respect to the number of subsets of the feature and classification accuracy is also higher compared to other methods. Mohd Shamrie Sainin et al. [8] describe a genetic algorithm based wrapper feature selection using the nearest neighbour distance matrix classifier. R. Setiono et al. [9]have proposed a neural network-based wrapper feature selection that employs the backward elimination method in the search for optimal subsets of features. [17]

C. Embedded method

Embedded method combines the advantages of filter and wrapper approaches. In this method, the feature set is evaluated through the use of both independent criteria and a learning algorithm. The independent criterion is used tochoose the optimal subset for a given cardinality and the learning algorithm selects the finest subset among the optimal subsets across different cardinalities. These methods are suitable for feature selection mainly on high dimensional datasets. Cong Jin et al. [10] propose the attribute selection method, using the input output correlation (IOC) for calculating attributes importance based on ranking. The hybrid of Back Propagation Neural Network (BPNN) and Particle Swarm Optimization (PSO) algorithms is also proposed. PSO algorithm is used for optimization of the weights and thresholds of BPNN, which may avoid the traditional BPNN's inherent flaws and improve generalization performance. The results of the experiment show the proposed attribute selection method based on hybrid algorithms as effectivein data pre-processing.

IV. RELATED WORKS ON FEATURE SELECTION METHODS

Mauricio schiezaroet al. [11] has investigated a feature selection method based on artificial bee colony algorithm for classification of different data sets. The results show that a reduced number of features can attain more classification accuracy compared to that using the complete set of features.M.E . ElAlami [12] present a novel feature selection algorithm from trained by neural network using a genetic algorithm which is meant for

finding the optimal relevant features which maximize the output function of trained artificial neural network.Lei Yu et al. [13] propose a Fast Correlation Based Filter (FCBF) method which can identify relevance and redundancy among relevant features using the concept of predominant correlation. The result shows FCBF method efficiently handling with a high degree of dimensionality reduction.Liuzhi Yin et al. [14]illustrate the feature selection issues and the inefficiency of traditional feature selection method (Correlation coefficient, mutual information and fisher criterion) in the imbalance of data. The authors have proposed two different approaches (i). decomposition of large classes into small psedo-subclasses with relatively uniform size and then computing the goodness of features with the new decomposed data. (ii). the Hellinger distance-based feature selection method. The results show that the proposed two approaches can outperform traditional feature selection methods. Chung-Jui Tu et al [15] propose particle swarm optimization for performance of feature selection and support vector machine with one-versus-rest method serving as a fitness value of PSO for the five classification problem. The results of the proposed method showoptimization of the feature selection process and increase in the classification accuracy compared to other existing feature selection methods. Afef Ben Brahim et al. [16] propose new hybrid feature selection method based on instance learning. Its main task is to change the problem of the small sample size to a tool that allows selection a few subsets of features to be analysed in a filter stage. A cooperative subset search is then proposed with a classifier algorithm as the evaluation system of wrappers. The proposed method outperforms other methods in terms of accuracy and stability of the feature selection.

V. CONCLUSION

This paper gives an overview of feature selection methods found in literature.The basic concept of feature selection and the processinvolved are discussed. The feature selection approaches are categorized into filter, wrapper and embedded methods. The review shows that feature selection as an important step in data mining and machine learning applications. Feature selection is to identify a related subset of features in the datasets and remove any other feature as irrelevant and redundant information content. Data mining algorithms can be activated faster with improved accuracy by using feature selection methods.

REFERENCES

[1] Richard Jensen, Qiang Shen, Computational intelligence and feature selection, IEEE press series on computational intelligence, A John Wiley & Sons, Inc., Publication.2008.
[2] Girish Chandrashekar and Ferat Sahin, A survey on feature selection methods, computer and Electrical Engineering 40(2014)16-28.
[3] Vipin kumar and sonajharia minz, feature selection: A literature review, Smart computing Review, vol 4 No 3 2014.
[4] Marko Robnik-˜Sikonja and Igor Kononenko , "Theoretical and Empirical Analysis of ReliefF and RReliefF", Machine Learning Journal (2003) 53:23-69.
[5] Isabelle Guyon and Andr´e Elisseeff , "An Introduction to Variable and Feature Selection" Journal of Machine Learning Research 3 (2003) 1157-1182.
[6] Huan Liau and Rudy Setiono, Chi: Feature Selection and Discretization of Numeric Attributes, IEEE,1995.
[7] Li-Yeh Chuang, Cheng-Hong Yang, Jung-Chike Li, "Chaotic maps based on binary particle swarm optimization for feature selection", Applied Soft Computing 11, (2011), 239– 248.
[8] Mohd Shamrie Sainin and Rayner Alfred, "A Genetic Based Wrapper Feature Selection Approach Using Nearest Neighbour Distance Matrix", 3rd Conference on Data Mining and Optimization (DMO), Selangor, Malaysia, 2011.
[9] R. Setiono and H. Liu. Neural network feature selector. IEEE Trans. Neural Net. 8(3): 645–662. 1997.
[10] Cong Jin, Shu-Wei Jin, Li-Na Qin, "Attribute selection method based on a hybrid BPNN and PSO algorithms" Applied Soft Computing 12 (2012), 2147–2155.
[11] Mauricio schiezaro and HelioPedrini, "Data Feature selection based on Artificial Bee Colony algorithm", EURASIP Jounal on Image and Video Processing, 47,2013.
[12] M.E.ElAlami, " Filter model for feature subset selection based on genetic algorithm", Knowledge-Based Systems22(2009),356-362.
[13] Lei Yu and Huan Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", Proceedings of the Twentieth International Conference on Machine Learning, Washington DC, 2003.
[14] Liuzhi Yin, Yong Ge, Keli Xiao, Xuehua Wang, Xiaojun Quan," Feature selection for high-dimensional imbalanced data", Neurocomputing 105 (2013) 3–11.
[15] Chung-Jui Tu, Li-Yeh Chuang, Jun-Yang Chang, and Cheng-Hong Yang, " Feature Selection using PSO-SVM", IAENG International Journal of Computer Science, 33:1,2007.
[16] Afef Ben Brahim, Mohamed Limam, "A hybrid feature selection method based on instance learning and cooperative subset search", Pattern Recognition Letters 69 (2016) 28–34.
[17] Pramod Kumar Singh , Dimension Reduction Methods in High Dimensional Data Mining, Available: www.sau.ac.in/~mlta/studymaterial/22-12-2013/PKSingh.pdf
[18] Artur J. Ferreira, Mário A.T. Figueiredo, Efficient feature selection filters for high-dimensional data, Pattern Recognition Letters 33 (2012) 1794–1804

AUTHOR PROFILE

Senthil Kumar P is working as Assistant Professor (SG) at VIT University, India. His research interests include Data Mining and time series analysis.

Daphne Lopez, Professor, VIT University, India has vast experience in teaching, research and industry. To her credit there are a good number of publications in international conferences and journals.