Effectiveness of Searching Techniques in Feature Subset Selection: A Review

P. Thangaraju[1]
Assistant Professor
Department of Computer Applications
Bishop Heber College (Autonomous)
Tiruchirappalli, India

N. Mala[2]
Research Scholar
Department of Computer Science
Bishop Heber College (Autonomous)
Tiruchirappalli, India

***Abstract*- Feature selection in data mining minimizes the complexity involved in selecting meaningful attributes from the given data set. Attribute selection can be majorly categorized into Filter Approach or Wrapper Approach. Filter Approach filters the meaningful attributes from the data set, whereas, wrapper approach creates a wrapper like coverage between the meaningful and meaningless attributes. These two categories can further be classified into heuristics and complete search, meta-heuristic and artificial neural network methods. The objective of this paper is to survey common key steps involved feature selection and to describe more about the insights of feature subset selection proposed by various researchers. Experimentation is done for selecting the best algorithms for attribute subset selection with the rough dataset and the results are discussed.**

Keywords- data mining, feature selection, data set, filter approach, wrapper approach

I. INTRODUCTION

Data mining routines have turned into one of the overwhelming methodologies in information investigation. Data mining is defined as the process of non-trivial extraction of previously unknown and potentially useful information from data stored in databases. Data mining is used to find patterns or item sets hidden within data, and associations among the patterns. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets.

As a theme under the field of regulated learning, classifiers are created and prepared to name new cases as indicated by a set of features got from the data. Nevertheless, utilizing an excess of features in the characterization calculation can be risky, especially if there are irrelevant features. This can prompt over fitting, in which commotion or unessential features may apply undue impact on the arrangement choices as a result of the unassuming size of the preparation information. Also, there may be redundancies in the removed features.

This issue of recognizing the features most important to the characterization errand is known as feature selection: it gives a central venture in the investigation of such kind of information. By selecting just a subset of properties, the expectation exactness can conceivably enhance and more understanding in the way of the forecast issue can be picked up by distinguishing just the qualities that are applicable to the expectation of the ailment finding. Also, the ID of a little set of qualities that is to be sure fit for giving complete biased data, brings about economical indicative examines for just a couple of qualities which may be produced and be generally sent in clinical settings.

As it is discussed above, the filter model relies on general characteristics of the data to evaluate and select feature subsets without involving any mining algorithm. The wrapper model requires one predetermined mining algorithm and uses its performance as the evaluation criterion. It searches for features better suited to the mining algorithm aiming to improve mining performance, but it also tends to be more computationally expensive than the filter model. The hybrid model attempts to take advantage of the two models by exploiting their different evaluation criteria indifferent search stages. Apart from that, all the approaches will be evaluated through the evaluation criterion to prove its efficiency. Evaluation criterion of a feature subset is generally assessed by its applicability and optimality.

The working process of feature subset selection is done by the following steps

1. Original dataset is inputted to the subset generation process.
2. Generated subset is evaluated by the subset evaluation process.
3. Stopping Criterion: If the goodness of the retrieved subset is not satisfied repeat step 1to 3 else jump step.
4. The result then inputted to the result validation process

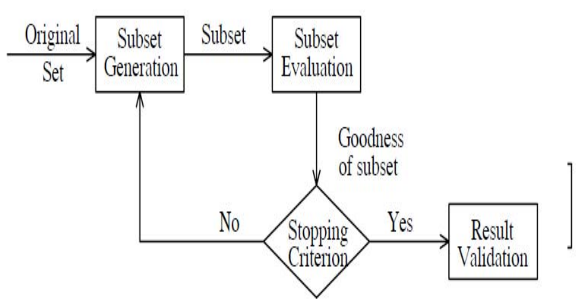

Figure 1. Feature Selection Key Processes

[1] The exhaustive search method (also called enumerative search method) works by considering all possible band combinations by way of calculating their reparability indices. Although this search method guarantees the optimality of solution, it poses the problem of being computationally prohibitive. For a dataset with d features (i.e. bands), 2d – 1 combinations are possible. This method is practicable if the number of bands is less than 10. The use of 10 or more bands would be costly in terms of computational speed. However are of the opinion that advancements in computer technology should eventually render exhaustive search an operational reality. This, including the fact that the datasets considered in this research had less than ten bands, influenced the author's decision to consider this method.

[2] Best first search is an Artificial Intelligence search strategy that allows backtracking along the search path. Best first moves through the search space by making local changes to the current feature subset. However, if the path being explored begins to look less promising, the best first search can back-track to a more promising previous subset and continue the search from there. Given enough time, a best first search will explore the entire search space, so it is common to use a stopping criterion. Normally this involves limiting the number of fully expanded subsets that result in no improvement.

[3]Probabilistic Search: LVF: Las Vegas Filter algorithm adopts the inconsistency rate as the evaluation measure. It generates feature subsets randomly with equal probability, and once a consistent feature subset is obtained that satisfies the threshold inconsistency rate.LVF is fast in reducing the number of features in the early stages and can produce optimal solutions.

[4] Genetic Algorithms (GA) are search algorithms inspired by evolution and natural selection, and they can be used to solve different and diverse types of problems. The algorithm starts with a group of individuals (chromosomes) called a population. Each chromosome is composed of a sequence of genes that would be bits, characters, or numbers. Reproduction is achieved using crossover (2 parents are used to produce 1 or children) and mutation (alteration of a gene or more). Each Chromosome is evaluated using a fitnes function, which dine which chromosomes is highly-fitted in the environment. The processes iterated for multiple times for a number of generations until optimal solution is reached. The reached solution could be a single individual or a group of individuals obtained by repeating the GA process for many runs.

[5] Greedy search is a discrete version of the gradient descent (ascent) algorithm, implements a local search of each repetition. The algorithm starts with an initial network and determines a nearest neighbor graph that improves the network by including, eliminating or inverting an arc in the graph. The process is repeated until there is no neighbor that improves the current solution

[6] Forward selection: This method starts with no variables. Add the variables one by one, at each step adding the feature that has the minimum error. Repeat the above step until any further addition does not signify any decrease in error. Backward selection: This method starts with all variables. It then removes the variables one by one, at each step removing the feature that has the highest error. Repeats the above

[22] The features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Feature extraction can be applied in many data mining applications to improve the predictive accuracy. Transforming the input data Into the set of features is called feature extraction.

II. EXPERIMENT

Searching methods that are discussed above are common for both filter and wrapper approaches. In filter methods, the subset selection procedure is independent of the learning algorithm and is generally a pre-processing step. Obviously, this leads to a faster learning pipeline but it is possible for the criterion used in the pre-processing step to result in a subset that may not work very well downstream in the learning algorithm. In wrapper methods, the subset selection takes place based on the learning algorithm used to train the model itself. Roughly speaking, every subset that is proposed by the subset selection measure is evaluated in the context of the learning algorithm. Obviously, this means that computationally intensive learning algorithms cannot be used.

In this paper, an experiment is conducted using Weka, a famous data mining tool to endorse the effective methods on both the approaches. A data set that contains the information about various countries and their flags is obtained from UCI machine learning repository [23]. The data set on the whole contains 194 instances and 30 attributes, with which 10 attributes are numeric-valued. The remainder is either Boolean or nominal valued. The attribute names and their corresponding description are shown in Table 1

TABLE 1. EXPERIMENTAL DATASET DESCRIPTION

S.No	Attribute Name	Description
1	Name	Name of the country concerned
2	Landmass	1=N.America, 2=S.America, 3=Europe, 4=Africa, 4=Asia, 6=Oceania
3	Zone	Geographic quadrant, based on Greenwich and the Equator; 1=NE, 2=SE, 3=SW, 4=NW
4	Area	in thousands of square km
5	Population	in round millions
6	Language	1=English, 2=Spanish, 3=French, 4=German, 5=Slavic, 6=Other Indo-European, 7=Chinese, 8=Arabic, 9=Japanese/Turkish/Finnish/Magyar, 10=Others
7	Religion	0=Catholic, 1=Other Christian, 2=Muslim, 3=Buddhist, 4=Hindu, 5=Ethnic, 6=Marxist, 7=Others
8	Bars	Number of vertical bars in the flag
9	Stripes	Number of horizontal stripes in the flag
10	Colours	Number of different colours in the flag
11	Red	0 if red absent, 1 if red present in the flag
12	Green	same for green
13	Blue	same for blue
14	Gold	same for gold (also yellow)
15	White	same for white
16	Black	same for black
17	Orange	same for orange (also brown)
18	Main hue	predominant color in the flag (tie-breaks decided by taking the topmost hue, if that fails then the most central hue, and if that fails the leftmost hue)
19	Circles	Number of circles in the flag
20	Crosses	Number of (upright) crosses
21	Satires	Number of diagonal crosses
22	Quarters	Number of quartered sections
23	Sun stars	Number of sun or star symbols
24	Crescent	1 if a crescent moon symbol present, else 0
25	Triangle	1 if any triangles present, 0 otherwise
26	Icon	1 if an inanimate image present (e.g., a boat), otherwise 0
27	Animate	1 if an animate image (e.g., an eagle, a tree, a human hand) present, 0 otherwise
28	Text	1 if any letters or writing on the flag (e.g., a motto or slogan), 0 otherwise
29	Top left	color in the top-left corner (moving right to decide tie-breaks)
30	Bo right	Color in the bottom-left corner (moving left to decide tie-breaks)

III. RESULT AND DISCUSSION

The dataset was inputted several times on to the WEKA [24] tool, to record the performance of various feature subset selection algorithms. The merit of subset of each algorithm was noted and depicted in table 2. The result shows the merit of subset found values are same for the algorithms best first, linear forward search, greedy stepwise except genetic search in filter subset evaluation. On the other hand, the wrapper subset evaluation methods retrieved same value for all algorithms. Figure2. Depicts the pictorial representation of Table 2.

TABLE 2. PERFORMANCE OF FEATURE SUBSET SELECTION ALGORITHMS

Algorithm	**Filter Subset Evaluation**		**Wrapper Subset Evaluation**	
	Total number of Subset Evaluated	**Merit of Subset Found**	**Total number of Subset Evaluated**	**Merit of Subset Found**
Best First Search	396	1.015	141	1
Genetic Search	110	0.0965	211	1
Linear Forward Search	425	1.015	170	1
Subset Size Forward Selection	373	1.015	58	1

It is clearly visible that the chosen algorithms works very close to each other and they produce same results in feature subset selection

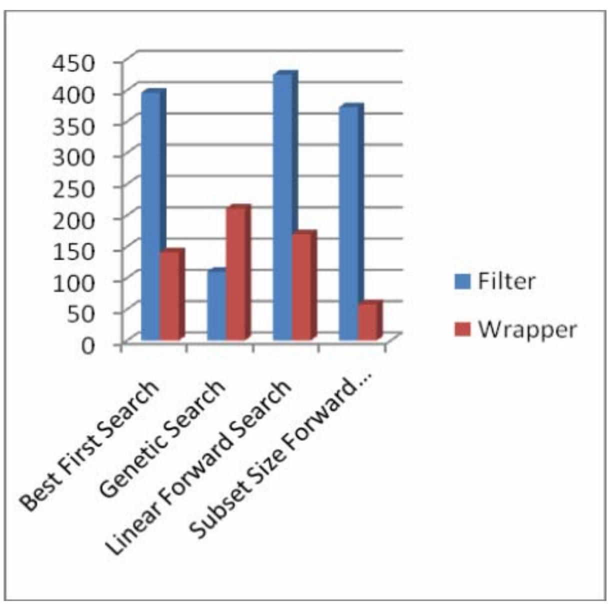

Figure 2.Performance of Feature Subset Selection Algorithms

IV. CONCLUSION

The primary objective of this paper is to review some articles related to feature subset selection. This work portrays how important the attributes selection is in data mining. Not all the collected attributes will be significant to predict the hidden information in large data sources. There may be irrelevant features subsists in the dataset leads to the wrong predictions. Hence, it is more essential to identify the subsets of relevant features. In future, this work can be extended by combining two, best search with partial correlation algorithms together to improve the efficiency in subset selection.

ACKNOWLEDGMENT

I would like to express my special thanks to god and my guide Prof P. Thangaraju. He give full support & encourage for doing my research work successfully and thanks to my god and my dear friends who are all supporting with us.

REFERENCES

[1] Gidudu Anthony and Heinz Ruther, “Comparison of Feature Selection Techniques for SVM Classification”, International Society for Photogrammetry and Remote Sensing, Proceedings, XXXVI/7, 2011
[2] Hall, Mark A. "Correlation-based feature selection for machine learning." PhD diss., The University of Waikato, 1999.
[3] Revathi, K., and T. Kalai Selvi. "Survey: Effective Feature Subset Selection Methods and Algorithms for High Dimensional Data." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 12, December 2013
[4] Aziz, Amira Sayed A., Ahmad Taher Azar, Mostafa A. Salama, Aboul Ella Hassanien, and SE-O. Hanafy. "Genetic algorithm with different feature selection techniques for anomaly detectors generation." In Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on, pp. 769-774. IEEE, 2013.
[5] Jantawan, Bangsuk, and Cheng-Fa Tasai. "A Comparison of Filter and Wrapper Approaches with Data Mining Techniques for Categorical Variables Selection."*vol* 2: 4501-4508.
[6] Kumari, Binita, and Tripti Swarnkar. "Filter versus wrapper feature subset selection in large dimensionality micro array: A review." (2011).
[7] Karegowda, Asha Gowda, M. A. Jayaram, and A. S. Manjunath. "Feature subset selection problem using wrapper approach in supervised learning."International journal of Computer applications 1, no. 7 (2010): 13-17.
[8] Kohavi, Ron, and George H. John. "Wrappers for feature subset selection."Artificial intelligence 97, no. 1 (1997): 273-324.
[9] Dash, Manoranjan, and Huan Liu. "Feature selection for classification."Intelligent data analysis 1, no. 3 (1997): 131-156.
[10] Guyon, Isabelle, and André Elisseeff. "An introduction to variable and feature selection." The Journal of Machine Learning Research 3 (2003): 1157-1182.
[11] Cooper, Gregory F., and Edward Herskovits. "A Bayesian method for the induction of probabilistic networks from data." *Machine learning* 9, no. 4 (1992): 309-347
[12] Nocedal, Jorge, and Stephen J. Wright. "Numerical OptimizationSpringer." New York (1999).
[13] Friedman, Nir, Dan Geiger, and Moises Goldszmidt. "Bayesian network classifiers." *Machine learning* 29, no. 2-3 (1997): 131-163.
[14] Song, Qinbao, Jingjie Ni, and Guangtao Wang. "A fast clustering-based feature subset selection algorithm for high-dimensional data." Knowledge and Data Engineering, IEEE Transactions on 25, no. 1 (2013): 1-14.
[15] Dash, Manoranjan, Huan Liu, and Hiroshi Motoda. "Consistency based feature selection." In Knowledge Discovery and Data Mining. Current Issues and New Applications, pp. 98-109. Springer Berlin Heidelberg, 2000.
[16] Dash, Manoranjan, and Huan Liu. "Consistency-based search in feature selection." Artificial intelligence 151, no. 1 (2003): 155-176.

[17] Liu, Huan, Hiroshi Motoda, and Lei Yu. "A selective sampling approach to active feature selection." Artificial Intelligence 159, no. 1 (2004): 49-74.
[18] Battiti, Roberto. "Using mutual information for selecting features in supervised neural net learning." Neural Networks, IEEE Transactions on 5, no. 4 (1994): 537-550.
[19] Krier, Catherine, Damien François, Fabrice Rossi, Michel Verleysen, and France Chesnay Cedex. "Feature clustering and mutual information for the selection of variables in spectral data." In ESANN, pp. 157-162. 2007.
[20] Zhao, Zheng, and Huan Liu. "Searching for Interacting Features." In *IJCAI*, vol. 7, pp. 1156-1161. 2007.
[21] Karthikeyan, T., and P. Thangaraju. "Analysis of classification algorithms applied to hepatitis patients." International Journal of Computer Applications 62, no. 5 (2013).
[22] Karthikeyan, T., and P. Thangaraju. "PCA-NB algorithm to enhance the predictive accuracy." Int. J. Eng. Tech 6, no. 1 (2014): 381-387.
[23] https://archive.ics.uci.edu/ml/datasets.html
[24] www.cs.waikato.ac.nz/ml/weka/