A Review Paper on Feature Selection Methodologies and Their Applications

Shweta Srivastava1, Nikita Joshi2, Madhvi Gaur3
Assistant Professor(CSE Department) ABES Engineering College, Ghaziabad, India.

Abstract
Feature selection is the process of eliminating features from the data set that are irrelevant with respect to the task to be performed. Feature selection is important for many reasons such as simplification, performance, computational efficiency and feature interpretability. It can be applied to both supervised and unsupervised learning methodologies. Such techniques are able in improving the efficiency of various machine learning algorithms and that of training as well. Feature selection speed up the run time of learning, improves data quality and data understanding.

Keywords
Feature Selection, Supervised, Search strategies, Unsupervised.

1. INTRODUCTION

Feature selection is used to satisfy the common goal of maximizing the accuracy of the classifier, minimizing the related measurement costs; improve accuracy by reducing irrelevant and possibly redundant features; reduce the complexity and the associated computational cost; and improve the probability that a solution will be comprehensible and realistic [13]. Feature selection is one of the stage for preprocessing the data to reduce the dimensionality [10]. It selects a subset of the existing features without any transformation. It can be said as a special case of feature extraction.

2. FEATURE SELECTION METHODOLOGIES

When we use feature selection, smaller number of features are extracted which means fewer model parameters. It improves the generalization capabilities and reduces complexity and execution time. Feature Selection methodologies can be categorized as: supervised and unsupervised feature selection methods.

Supervised Feature Selection Method: A feature selection method requires a search strategy to select the subset of attributes and an objective function to evaluate the selected subset of features.

Objective function is divided in three categories:

1. Filters: It consists of algorithms that are built in the adaptive systems for data analysis (predictors) [9]. They use an evaluation function that relies on properties of the data. Distance based and margin based criterion can be used for filters.

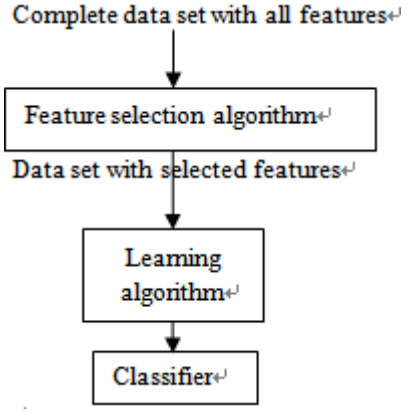

Fig.1. Filter method

2. Wrapper: The algorithms of this method are wrapped around the adaptive systems providing them subsets of features and receiving their feedback (usually accuracy). These wrapper approaches are aimed at improving results of the specific predictors they work with [9]. It utilizes the classifier as black box to find the subset of features based on their predictive power [14].

3. Embedded: They are based on performance evaluation metric calculated directly from the data, without direct reference to the results of any data analysis systems.

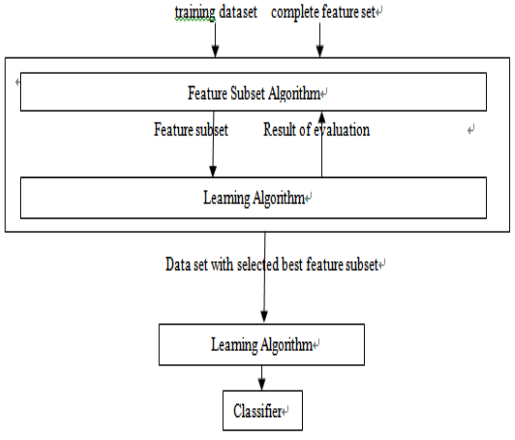

Fig.2. Wrapper method

Manuscript received May 5, 2014
Manuscript revised May 20, 2014

3. *Embedded:* They are based on performance evaluation metric calculated directly from the data, without direct reference to the results of any data analysis systems.

Filter Methods	Wrapper methods	Embedded methods
Filter methods appear to be a probably less optimal.[4]	Wrapper methods are superior alternative in supervised learning problems. [4]	The performance of embedded method degrades if more irrelevant features are inserted in target set.
Filter methods execute faster than wrapper methods.	Wrapper methods execute slower than filter methods.	Embedded method is faster than the wrapper methods.
The result of filter method exhibits more generality than that of wrapper approach.	There is lack of generality in wrapper methods as it is tied to some classifier.	It also lacks generality as it is also dependent on some classification algorithm.
Filter method has more tendencies to select large subset of data.	Wrapper method is more accurate than filter methods as it achieves better recognition rates rather than that of filters.	They are least prone to over fitting.
Computational cost is less for large data set in filter method.	Computational cost is more for large data set in filter method.	Computational cost is less in comparison to wrapper methods.
Independent of classification algorithm.	Dependent on classification algorithm.	Dependent on classification algorithm.

Table 1: Comparison of filter, wrapper and embedded methods

Search Strategy: A search strategy is required to select the candidate subsets and objective function evaluates each of the candidate subset. Search strategy involves exhaustive, heuristic and randomized searching algorithms [15][16]. The time complexity is exponential in terms of dimensionality for exhaustive search and quadratic for heuristic search. The complexity can be linear to the number of iterations in a random search [15]. There are several hybrid algorithms also.

Exhaustive search: It evaluates a number of subsets that grows exponentially with the dimensionality of the search space. It finds the 2N combinations of all m features. Each of the subset is evaluated by objective function and a measure of goodness is returned to the search algorithm. Branch and bound based algorithms are used for exhaustive search. There are several versions of branch and bound algorithm i.e. Basic BB (slowest), Enhanced BB (fastest), Fast BB, BB with partial prediction [17].

Unsupervised feature selection: The procedure of arranging the objects into natural classes whose members are similar to each other, identified by a given metrics. Unsupervised feature selection is particularly difficult due to the absence of class labels for feature relevance estimation. Unsupervised feature selection is a less constrained search problem without class labels, depending on clustering quality measures and can achieve many equally valid feature subsets. Feature redundancy and relevance is measured with respect to clusters instead of classes. The objective of clustering is to maximize intra cluster similarity and minimizing inter cluster similarity. For example: Agglomerative and partitioned clustering. Filter and wrapper approach are helpful in unsupervised feature selection also.

3. APPLICATION OF FEATURE SELECTION IN VARIOUS FIELDS

Text classification:

Heuristic search: The number of subsets evaluated by BFF is less (even much less) than that needed by the branch and bound algorithm [18]. They require large computational time. Heuristic search is problematic when the data has highly correlated features. Heuristics help to reduce the number of alternatives from an exponential number to a polynomial number [19]. For example: Sequential forward selection, sequential backward elimination and bidirectional search.

Forward Selection: Forward selection considers the subset to be empty initially and keeps on adding one feature at a time until the best feature subset is obtained.

Backward Elimination: Backward selection takes complete set of features as input and keeps on removing one attribute at a time until the most appropriate subset of features is obtained.

Randomized search strategy: It performs randomized exploration of the search space where next direction is a sample from a given probability [20]. For example: genetic algorithm.

	Accuracy	Complexity	Advantages	Disadvantages
Exhaustive Search	It always finds the best possible solution.	Exponential	It is highly accurate.	Complexity is high.
Heuristic Search	This technique is good if no backtracking is needed.	Quadratic	It is simple and fast.	Backtracking is not possible.
Randomized Search	It is good with proper control parameters.	Generally low	It is designed to escape local minima.	It is difficult to choose good parameters.

Table 2: Comparison of Exhaustive, Heuristic and Randomized Search

There are various challenges related to automated text classification such as:

1. An appropriate data structure is to be selected to represent the documents.
2. An appropriate objective function is to be chosen to optimize to avoid overfitting and obtain good generalization. Along with it algorithmic issues arising as a result of the high formal dimensionality of the data are to be dealt with [1].

Genre classification:

Metadata such as filename, author, size, date, track length and genres are the common features used to classify and retrieve genre documents. On the basis of these data, the classification is infeasible, so the feature selection step is required. In case of genre classification feature selection is a process where a segment of an audio is characterized into a compact numerical representation [2]. Feature selection is done to reduce the dimensionality of the data as a preprocessing step prior to classification due to high dimensionality of the feature sets.

Microarray data analysis:
1. Almost all bioinformatics problems have the number of features significantly larger than the number of samples (high feature to sample ratio datasets) [3]. For example: Breast cancer classification on the basis of microarray data. Though the information about all the genes is not required in case of such classification.
2. Content analysis and signal analysis in genomics also require feature selection.

Software defect prediction:
There are various software quality assurance attributes such as reliability, functionality, fault proneness, reusability, comprehensibility etc [6]. It is a critical issue to select most appropriate software metrics that likely to indicate fault proneness.

Sentiment analysis:
Sentiment analysis is capturing favorability using natural language processing. It is not just a topic based categorization. It deals with the computational treatment of opinion, sentiment, and subjectivity in text. It is useful in recommendation systems and question answering [8]. To decide about the positivity or negativity of the opinion on the basis of various features such as term presence, feature frequency, feature presence, term position, POS tags, syntax, topic and negation etc. All of the features are not required in each and every case. So feature selection need to be performed.

Stock market analysis:
There are hundreds of stock index futures. Along with it financial data including the stock market data is too complex to be searched easily [11]. In particular, the existence of large amount of continuous data may cause a challenging task to explicit concepts extraction from the raw data due to the huge amount of data space determined by continuous features [12]. So it is necessary to reduce the dimensionality of data and irrelevant factors before searching.

Image Retrieval
Feature selection is applied to content based image retrieval to allow efficient browsing, searching and retrieving [22]. Content based image retrieval is to index the images on the basis of their own visual contents (i.e. color, shape, texture etc.) instead of text based keyword indexing. The biggest problem for content based image retrieval is large amount of images in database [21].

IV. CONCLUSION

This paper provides a comprehensive overview of various characteristic of feature selection. Feature selection as a preprocessing step in very large databases collected from various applications. More work is required to overcome limitations imposed as it is costly to visit high dimensionality data multiple times or accessing instances at random times. In unsupervised feature selection, several clusters may exist in different subspaces of small-small dimensionality, with their sets of dimensions overlapped or non-overlapped. As most existing feature selection algorithms have quadratic or higher time complexity about N, so it is quite difficult to handle high dimensionality. Therefore, more efficient search strategies and evaluation criteria are needed for feature selection with large dimensionality.

References:

[1] Anirban Dasgupta,Petros Drineas, Boulos Harb, Vanja Josifovski, Michael W. Mahoney; “Feature Selection Methods for Text Classification”,KDD- ACM.,2007.
[2] Shyamala Doraisamy, Shahram Golzari, Noris Mohd. Norowi, Md. Nasir B Sulaiman, Nur Izura Udzir; ” A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music” Proceedings of ISMIR, 2008.
[3] Gianluca Bontempi, Benjamin Haibe-Kains; “Feature selection methods for mining bioinformatics data”,
[4] Luis Talavera; “An evaluation of filter and wrapper methods for feature selection in categorical clustering”, Proceedings of the 6th international conference on Advances in Intelligent Data Analysis, 2005.
[5] Yvan Saeys, Inaki Inza, Pedro Larranaga; ”A review of feature selection techniques in bioinformatics”, Oxford Journals, 2007.
[6] N.Gayatri, S.Nickolas, A.V.Reddy; “Feature Selection Using Decision Tree Induction in Class level Metrics Dataset for Software Defect Predictions”; Proceedings of the World Congress on Engineering and Computer Science ; Vol I; 2010.
[7] Tim O’Keefe, Irena Koprinska;” Feature Selection andWeighting Methods in Sentiment Analysis”; Proceedings of the 14th Australasian Document Computing Symposium, Sydney, Australia; 2009.
[8] Bo Pang, Lillian Lee; ”Opinion Mining and Sentiment Analysis”; Foundations and Trends in Information Retrieval, Vol. 2, Nos. 1–2 pp- 1–135, 2008.
[9] Wlodzislaw Duch; “Filter methods”, Springer- Feature Extraction Studies in Fuzziness and Soft Computing Volume 207, 2006, pp 89-117.
[10] Chih- Fong Tsai; “Data pre-processing by genetic algorithms for bankruptcy prediction”, IEEE International Conference on Industrial Engineering and Engineering Management, Pages- 1780 - 1783 , 2011.
[11] Kyoung-jae Kim, Ingoo Han; “Genetic algorithms approach to feature discretization in artificial neural networks for the

prediction of stock price index"; Expert Systems with Applications, Elsevier Science Ltd; 2000.
[12] Liu, H., & Setiono, R.; "Dimensionality reduction via discretization"; Knowledge-Based Systems, 9 (1), 67–72; 1996.
[13] Steppe. J., K.W. Bauer, "Feature Saliency Measures", Computers & Mathematics with Applications, Vol 33, No. 8, pp. 109-126; 1997.
[14] L.Ladha et al., "Feature Selection Methods and Algorithms"; International Journal on Computer Science and Engineering (IJCSE), Vol. 3. No. 5, pp. 1787-1797; 2011.
[15] Roberto Ruiz, Jos´e C. Riquelme, and Jes´us S. Aguilar-Ruiz; "Heuristic Search over a Ranking for Feature Selection"; IWANN, LNCS 3512, pp. 742–749; 2005.
[16] Yao-Hong Chan; "Empirical comparison of forward and backward search strategies in L-GEM based feature selection with RBFNN"; International Conference on Machine Learning and Cybernetics (ICMLC), Vol. 3, pp-1524 - 1527 ; 2010.
[17] P. Somol, P. Pudil; "Feature Selection Toolbox"; Pattern recognition, Published by Elsevier Science Ltd; 2002.
[18] Pingfan Yan; Tong Chang ; "Best First Strategy for Feature Selection"; 9th International Journal on Pattern recognition, Vol.2, pp-706-708; 1988.
[19] Manoranjan Dash, Huan Liu; "Consistency based feature selection"; Published by Elsevier
[20] Computer Science (Artificial Intelligence); 2003.
[21] http://www.di.unipi.it/~bacciu/teaching/IIA2012/lect3-exploratory-hand.pdf
[22] T. Hastie, R. Tibshirani, and J. Friedman; "The Elements of Statistical Learning"; Springer, 2001.
[23] Huan Liu and Lei Yu;" Toward Integrating Feature Selection Algorithms for Classification and Clustering"; IEEE journal of Knowledge and data engineering; volume 17, issue 4; pp. 491-502; 2005

Shweta Srivastava, birth place is Faizabad and date of birth is 2-July-1986. She has done schooling from Canossa Convent Girls Inter College, Faizabad, Uttar Pradesh, India, B. Tech. (IT) from JSS Academy of Technical Education, NOIDA, Uttar Pradesh Technical University, Uttar Pradesh, India (2008) and M. Tech. (CSE) from Jaypee Institute of Information Technology, NOIDA, JIITU, Uttar Pradesh, India (2011). She has three years of teaching experience. Currently she is working as an ASSISTANT PROFESSOR in CSE department of ABES Engineering College, Ghaziabad, Uttar Pradesh, India. She has worked as a LECTURER in JSSATE, NOIDA, Uttar Pradesh, India for one year. She has done two years TEACHING ASSISTANTSHIP in Jaypee Institute of Information Technology, JIITU, NOIDA, Uttar Pradesh, India during M. Tech.