NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 4, ISSUE 5, May-2017

A SURVEY ON FEATURE SELECTION TECHNIQUES

KALPANA JAIN
Assistant professor of CSE, College of technology and engineering, Udaipur, India,kalpana_jain2@rediffmail.com

ABSTRACT
Feature selection is a term standard in data mining to reduce inputs to a manageable size for analysis and processing which also focuses on identifying irrelevant information without affecting the accuracy of the classifier. FS selects a subset of relevant features and removes irrelevant and redundant features from the raw data to build a robust learning model. FS is very important, not only because of the curse of dimensionality, but also because of data complexities and the quantities of the data faced by multiple disciplines, such as machine learning, data mining, statistics, pattern recognition and bioinformatics. In recent years, we have seen extensive research in feature selection which has been expanding in depth and in breadth from simple to more advanced techniques, from supervised to unsupervised and semi-supervised feature selection. This paper presents a state-of-art survey of feature selection techniques.

KEYWORDS: Text mining, Text classification, Filter, Wrapper and Feature selection.

INTRODUCTION

Feature selection is considered one of the most crucial pre–processing steps of machine learning(ML) [1]. It contributes by considerably reducing the dimension as well as eliminating inappropriate or redundant data thereby improving the learning accuracy in computational intelligence. The feature selection is fairly significant because with the same training data it may perform better with different feature subsets [2]. The success of machine learning is affected by many factors. Among those factors demonstration and worth of instance data is first and foremost [3]. Sometimes the real life data contain information not so useful for desired purpose. The training stage becomes unfavourable with the existence of noisy, irrelevant and redundant data.

Feature Selection consists of two common aspects [4].

- Search Method a selection algorithm to generate feature subsets that are most advantageous or relevant in model creation.
- Evaluator is basically an evaluation algorithm which decides the goodness of a feature subset and returns the assessment about the correctness of search method.

Usually feature selection methods are classified into two general groups:

According to a paper by John et al in '94 [5] methods which are independent of the inductive algorithm [6] [7] [8] [9] and [10] can be labelled as "filter" models fig. 1 and models which uses inductive algorithms like [11] [12] is classified as "wrapped around" methods fig. 2.

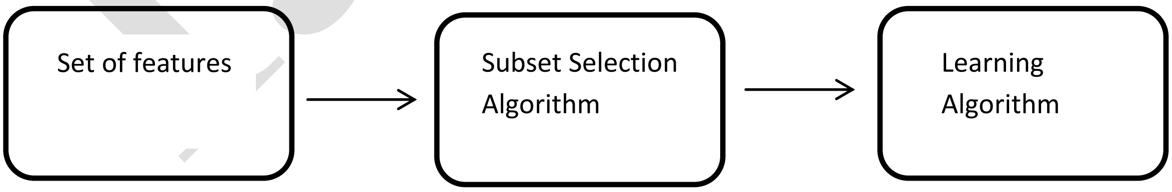

Fig. 1 Filter Method

NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 4, ISSUE 5, May-2017

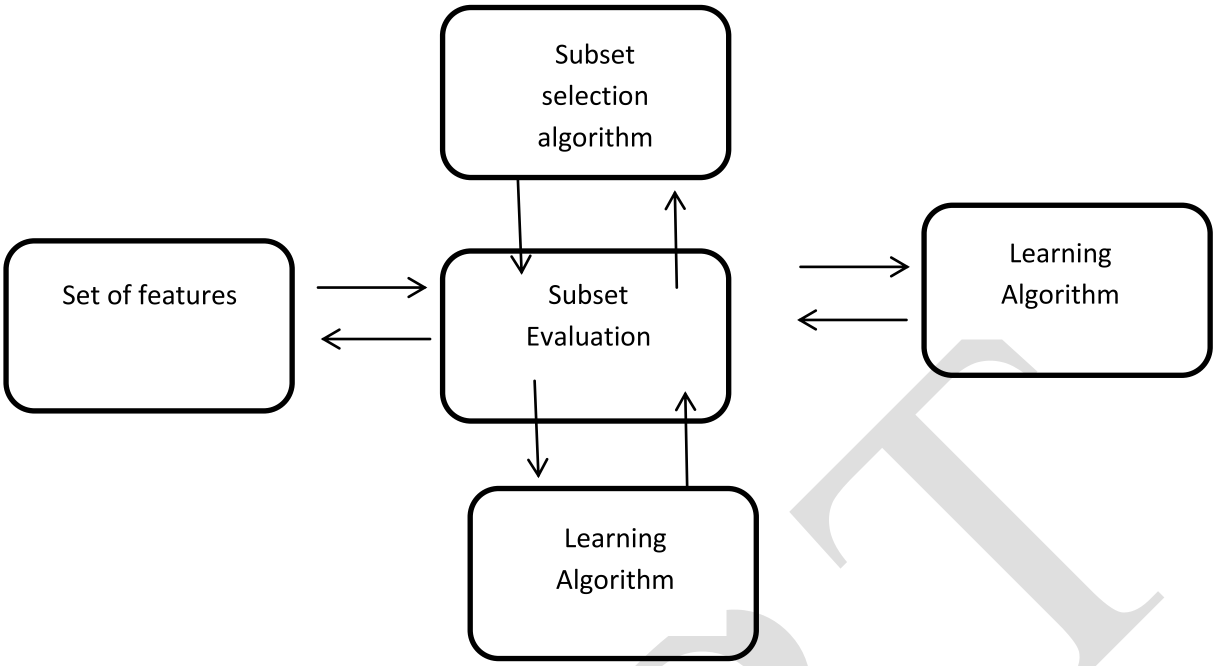

Fig. 2 Wrapper Method

SURVEY ON TYPES OF FEATURE SELECTION
FILTER METHODS

Filter methods picks features according to a performance measure not considering the data employed and can be used only after the features which are best are found. In literature various filter methods are described, a listing of regular methods is presented in Table A, along with the corresponding references to provide details but not all the description can be used for all of data mining tasks. Therefore, the filters methods are also categorised based on the task: regression, classification or clustering.

Table A: Common filter methods

NAME	CLASS	TASK	STUDY
Information gain	univariate, information	classification	13
Gain ratio	univariate, information	classification	14
Chi-square	univariate, statistical	classification	14
Correlation based feature selection (CFS)	multivariate, statistical	classification, regression	14
Fisher score	univariate, statistical	classification	15
Inconsistency criterion	multivariate, consistency	classification	16

Univariate filters evaluate and rate a solo feature, while multivariate filters estimate whole feature subset which depends on the investigate strategies which are as follows:

- Forward selection which starts with an blank set of features and then adds one or more features to the set.
- Backward elimination which starts with the complete set of features and then removes one or more features from the set.
- Bidirectional selection starts can start with both sides (empty set whole set), considering large and small feature subsets.

- Heuristic feature subset selection generates a starting subset using a heuristic algorithm e.g. a genetic_algorithm and then explores it additional.

Various exploration strategies are shown below in Table B.

Table B: Search Strategies

ALGORITHM GROUP	ALGORITHM NAME
Exponential	Branch and bound , Exhaustive search
Sequential	Linear forward selection , Best first
Randomized	Simulated annealing , Random generation

WRAPPER METHOD

Wrapper methods are so called because they wrap a classifier up in a feature selection algorithm [17]. Wrapper methods evaluate subsets according to the performance of classifiers like Naïve Bayes (NB) or Support Vector Machine (SVM) [18],[19], on the other hand for clustering, a wrapper evaluates subsets on the basis of performance of a clustering algorithm like *K*-means [20]. The generation of subset is in the similar way as with filters which is dependent on the investigate strategy, and evaluation is repeated for each subset. Wrapper methods are normally slower than filters methods to find good subsets.
Practically, we can combine any search technique and modelling algorithm to be used as a wrapper, but it is best for greedy investigate strategies and fast modelling algorithms such as Naïve Bayes [21], linear SVM [22], and Extreme Learning Machines [23].

CONCLUSION

Feature selection is an important part of most of the data processing applications including data mining, machine learning and computational intelligence. It helps in removing the irrelevant features and redundant information which affects the accuracy of the model. This paper presents a survey about types of feature selection techniques and processes as discussed by various authors.

REFERENCES

1) Yu, L., Liu, H. *Feature selection for high-dimensional data: a fast correlation based filter solution. Proc. 20th Int"l Conf. Machine Learning, 2003; 856-863.*

2) Blum, A., Langley, P. *Selection of relevant features and examples in machine learning. Artificial Intelligence, 1997; 97:245-271.*

3) Mitchell, T. *Machine Learning. McGraw Hill, 1997.*

4) Karagiannopoulos, M., Anyfantis, D., Kotsiantis, S. B., Pintelas, P. E. *Feature selection for regression problems. The 8th Hellenic European Research on Computer Mathematics & its Applications, HERCMA 2007, 20-22.*

5) [John, et al., 94] John, G.; Kohavi, R.; and Pfleger, K., 1994. *Irrelevant features and the subset selection problem. In Proc. ML-94, pp 121-129. Morgan Kaufmann.*

6) [Kira, et al., 92] Kira, K.; and Rendell, L.A., 1992. *The feature selection problem: Traditional methods and a new algorithm. In Proc. AAAI-92, pp 129-134. MIT Press.*

7) [Almuallim, et al., 91] Almuallim, H.; and Dietterich, T.G. 1991. *Learning with many irrelevant features. In Proc. AAAI-91, pp 547-552. MIT Press.*

8) [Moore et al., 94] Moore A.W; and Lee, M.S., 1994. *Efficient algorithms for minimizing cross validation error. In Proc. ML-94. Morgan Kaufmann.*

9) [Skalak, 94] Skalak, D.B., 1994. *Prototype and feature selection by sampling and random mutation hill climbing algorithms. In Proc. ML-94. Morgan Kaufmann.*

10) Koller, D.; and Shami, M., 1996. *Toward optimal feature selection. In Proc. ML-96. Morgan Kaufmann.*

11) Caruana, R.; and Freitag, D. 1994. Greedy attribute selection. In Proc. ML-94. Morgan Kaufmann.

12) Langley, P.; and Sage, S., 1994. *Induction of selective bayesian classifiers. In Proc. UAI- 94, pp 399-406. Seattle, WA: Morgan Kaufmann.*

13) N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, *"MIFS-ND: A mutual information-based feature selection method", Expert Systems with Applications, vol. 41, issue 14, pp. 6371–6385, 2014.*

14) I. H. Witten and E. Frank, *Data mining: Practical machine learning tools and techniques, San Francisco CA, USA: Morgan Kaufmann, 2011.*

15) R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification, Wiley-interscience, 2012.*

16) H. Liu and R. Setiono, "*A Probabilistic Approach to Feature Selection-A Filter Solution," in: Proc. 13th International Conference on Machine Learning (ICML-1996), Bary, Italy, Morgan Kaufmann, pp. 319–327, 1996.*

17) R. Kohavi and G. John. *Wrappers for feature subset selection. Artificial Intelligence, 1997.*

18) P. S. Bradley and O. L. Mangasarian, "*Feature selection via concave minimization and support vector machines," in: Proc. 15th International Conference on Machine Learning (ICML- 1998), Madison, Wisconsin, USA, Morgan Kaufmann, pp. 82–90, 1998.*

19) S. Maldonado, R. Weber, and F. Famili, "*Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines," Information Sciences, vol. 286, pp. 228–246, 2014.*

20) Y. S. Kim, W. N. Street, and F. Menczer, "*Evolutionary model selection in unsupervised learning," Intelligent Data Analysis, vol. 6, no. 6, pp. 531–556, 2002.*

21) J. C. Cortizo and I. Giraldez, "*Multi Criteria Wrapper Improvements to Naive Bayes Learning," LNCS, vol. 4224, pp. 419–427, 2006.*

22) C. Liu, D. Jiang, and W. Yang, "*Global geometric similarity scheme for feature selection in fault diagnosis," Expert Systems with Applications, vol. 41, issue 8, pp. 3585–3595, 2014.*

23) F. Benoît, M. van Heeswijk, Y. Miche, M. Verleysen, and A. Lendasse, "*Feature selection for nonlinear models with extreme learning machines," Neurocomputing, vol. 102, pp. 111–124, 2013*