IMPACT: International Journal of Information Engineering and Technology (IMPACT: IJIET)
ISSN(E): Applied; ISSN(P): Applied
Vol. 1, Issue 1, Jun 2015, 25-36

FEATURE SELECTION: AN OVERVIEW

PRITI GUPTA[1] & OMDUTT SHARMA[2]

[1]Professor, Department of Statistics, M. D. University, Rohtak, Haryana, India

[2]Research Scholar, Department of Statistics, M. D. University, Rohtak, Haryana, India

ABSTRACT

Variables and feature selection turn out to be the center of attention to a large extent of study in different areas of application for high dimensional data. It is the process in which unsuitable and superfluous features are rejected from high dimensional data which is unimportant regarding the job to be carried out. There are various reasons such as generalization, presentation, computational effectiveness and feature understandability for which feature selection is crucial. The purpose of feature selection is in three collapses; increasing the forecasting functioning of the predictors; make available more rapidly and rate efficient predictors; make available a healthier accepting of the fundamental method that generate the figures. In data analysis procedure feature selection is useful, since it show which features are essential for calculation, and how these features are interrelated.

KEYWORDS: Feature Selection, Forward Selection, Backward Selection, Clustering, Filters, Wrappers, Embedded, Hybrid, Search Strategies, Exhaustive, Heuristic, Random, Stopping Criteria, Validation

INTRODUCTION

In the past few years, the quantity of high dimensional data that be present and is explicitly accessible on the internet has significantly raises. Hence, with the massive input features machine learning methods have troubles in performance which have airs a motivating challenge for researchers. Thus preprocessing of that data is crucial, in order to employ machine learning methods efficiently. It is also known as variable selection, attribute selection, Feature reduction, variable reduction, dimensionality reduction or feature subset selection in machine learning and statistics. Dimensionality reduction has been a productive area of study and growth since last four to five decades and exposed extremely valuable in eliminating avoidable and repetitive features, rising effectiveness in learning tasks, increasing learning performance. In many applications such as image identification, text classification, consumer relationship administration and others data has become progressively bigger in both number of occurrence and number of features in last few years. High dimensional data can hold high scale of immaterial and superfluous information which may seriously disgrace the act of learning algorithms. Hence, when facing high dimensional data feature selection turn into very compulsory for machine learning tasks.

To eliminate noisy (irrelevant) and redundant attributes an individual fashionable technique is used known as dimensionality reduction. Feature extraction and feature selection are two approaches of dimensionality reductions. Features are projected into a new space with lesser dimensionality in former approach. In contrast, to choose a minute subset of features that minimize repeated features and maximize significant features to the object later one's approach is used. Both techniques are capable of mounting learning performance, diminishing computational complexity, constructing improved generalizable models and diminishing mandatory storage space. In feature selection technique a subset of features is choose from the original feature set without any alteration and maintains the physical meaning of the original features. This property has its significance in various realistic applications for example finding significant genes to a specific disease and building a sentiment dictionary for sentiment analysis. Generally the research in this problem spotlight on making and choosing subsets of features that is handy to construct a high-quality predictor.

State of Art

In past, various feature selection schemes were projected by many researchers. Their relative study is a demanding charge. It is very tricky job to expose the usefulness of the feature selection method, lacking the awareness of relevant features of the real data sets beforehand, since the data sets may perhaps consist of various troubles like the massive amount of immaterial and superfluous features, corrupt data and numerous quantities of features. So the implementation of the feature selection methods is dependent on the execution of learning methods. In the past there are various functioning measure mentioned like computer assets, proportion of feature selection, precision etc. The majority of researchers concur that there is no hypothetically “best method” [1]. For that reason the novel feature selection techniques are regularly rising to transaction with the particular crisis with dissimilar strategies:

- To bear out an improved methods of feature selection using an assembly method [2]
- Linking with supplementary approaches such as feature removal [3]
- Explaining the recent algorithms [4]
- Constructing a new technique to attempt the still undetermined problems [5]
- To merge various feature selection techniques [6].

In past, there had been done various comparative studies of offered feature selection techniques, for e.g. an investigational study of 7 filters, 2 embedded methods and 2 wrappers are applied in 11 artificial datasets for comparative study of feature selection performances in the existence of immaterial features, noisy data, repetitions and the small proportion among the numbers of attributes and samples [1]. The performance of feature selection methods analyzed for the several category problems associated to the large volume dataset (in both samples and attributes) [5, 7]. In the application viewpoint, many realistic applications as disturbance detection [8], text classification [9], DNA microarray examination [10], music information recovery [11], image renewal [12], client association management [13], Genomic studies [10] are measured.

Feature Selection

Feature Selection is a course of action usually used in machine learning, in which relevant features or a subset of the useful or unrepeated features selected from offered data. The high accuracy is contributes by the best subsets which have low dimensions. There are two approaches:

- **Forward Selection** in this approach at starting there is no attributes and adds them one by one, at each step adds that attributes which decreases the error the most, the process is continue until any additional addition does not appreciably diminish the error.

- **Backward Selection** in this approach at starting there is all the features and eliminate them one by one, at each step eliminate that features which diminish the error in the majority (or increases it merely to some extent), the process is continue until any additional elimination increases the error drastically

Subset Selection in Feature Selection

Take a set X of |X| features, let us indicate Y_n be the set of the entire probable subset of dimension n, here n characterize the desired quantity of features. Suppose C(Z) be a criterion function the evaluates features subset Z $\in Y_n$. Without any loss of generality, let us regard an advanced value of C to signify an improved feature subset. Then the problem of feature selection can be formulized as: Search the subset $\check{Z}_n$ for which C ($\check{Z}_n$) $= \max_{Z \in Y_n} C(Z)$. Pre assumed that a proper criterion function has been select to estimate the efficiency of feature subsets, feature selection is used to search a crisis that identify a finest feature subset depend on the selected metrics. The choice of 'n' perhaps a difficult task based on problem individuality, except the 'n' value can be optimized as component of the search method. Specifically the monotonicity property is needed in most favorable feature selection methods. Let us consider two subsets A_1 and A_2 of the feature set X and a criterion C which estimate each subset A_i. The monotonicity conditions require the following:

$$A_1 \subset A_2 \;=>\; C(A_1) \leq C(A_2)$$

Thus a minor significance of feature selection criterion is yield by estimating the feature selection criterion on a subset of features of a specified set.

Generally, four fundamental steps ladder for feature selection are consider, specifically creation of subsets, estimation of subsets, norm of stopping, and validation of outcomes (shown below). Creation of subset is a explore method using the definite search approach [14]. The nature of subset generation processes resolve by two crucial issues. First decides the search preliminary point, which influences the search direction. Second is accountable for the feature selection process with a precise tactic. A created subset feature is estimated with earlier finest feature subset in assessment norm. There are numerous estimation norm have been projected in the literature to establish the reliability of the applicant subset of the features. These norms can be classified into groups: dependent and independent norm, according to their need on mining algorithms [15]. In feature selection to choose a subset of features, dependent norm engage prearranged mining algorithms to pick features depend on the act of the mining algorithm function. To assess the reliability of a feature set or features, the essential characteristics of the training data exclusive of concerning any mining algorithms is developed by independent norms. If new one is better feature than earlier best feature then it replaced the existing one. Until some certain stopping criterion is satisfied this process is continue. The developed finest feature subset needs to be validated, after the stopping criterion. By using simulated data or real data set validation can be done [16]. Feature selection process should be certify by implementation of special investigation and comparison with formerly recognized outcomes or comparison with the consequences of opposing methods using synthetic datasets, real datasets or both.

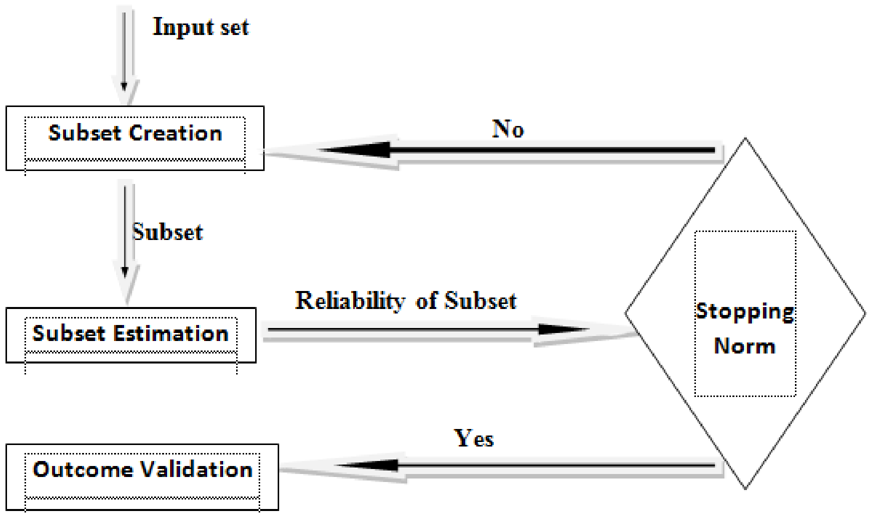

Figure 1: Four Key Steps for the Feature Selection Process

Feature Selection Classification Regarding Selection Criteria

Feature selection techniques classified according to selection criterion as:

- **Filter Methods:** It composed of algorithm namely construct in the favorable organization for data investigation [17]. Its use and estimate the task that depends on characteristics of data. For filters methods distance dependent and scope dependent criterion can be used. The pictorial demonstration of the filter model is exposed as:

Figure 2: The Feature Filter Model

Filter technique is proficient and speedy in interpretation. Those features which are not helpful alone but be able to exceptionally handy when associated with others feature can be overlook by filter methods. Various filter algorithms possibly constructed by changing the subset producer and estimation measure.

- **Wrapper Methods***:* The algorithms of wrapper approach are covered around the favorable systems granted them subsets of features and getting their response. For subset evaluation this methods employ a learning algorithm. Thus the evaluation criterion is a way to distinguish the filter and wrapper approach. A pictorial illustration of the wrapper model is revealed as:

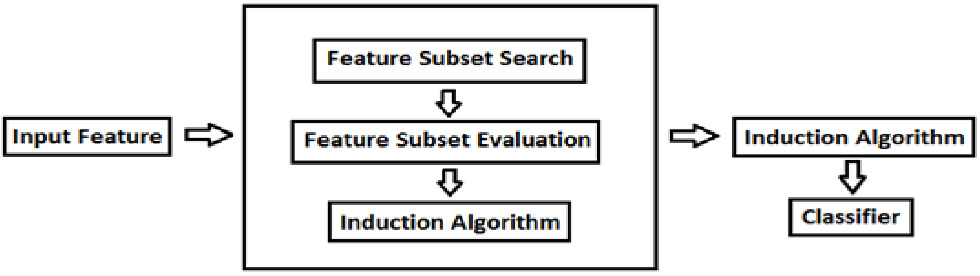

Figure 3: The Wrapper Model

These wrapper approaches are designed at improving result of the particular predictors they work with [17]. To search the subset of attribute based on their prophetic power it employ the classifier as black box [18]. This technique picked a finest subset that is most excellent to learning algorithms. So, their routine is generally better.

- **Embedded Methods***:* To project model estimation process by assimilate the feature selection method [19]. This technique interrelated with learning algorithms at an inferior computational rate than the wrapper approach. For a known cardinality, to settle on the optimal subsets this approach employs the independent criteria. Consequently to the result of any data from the data analysis systems without directly reference they are rooted in performance evaluation metric calculated openly. A graphical representation of the embedded model is shown as:

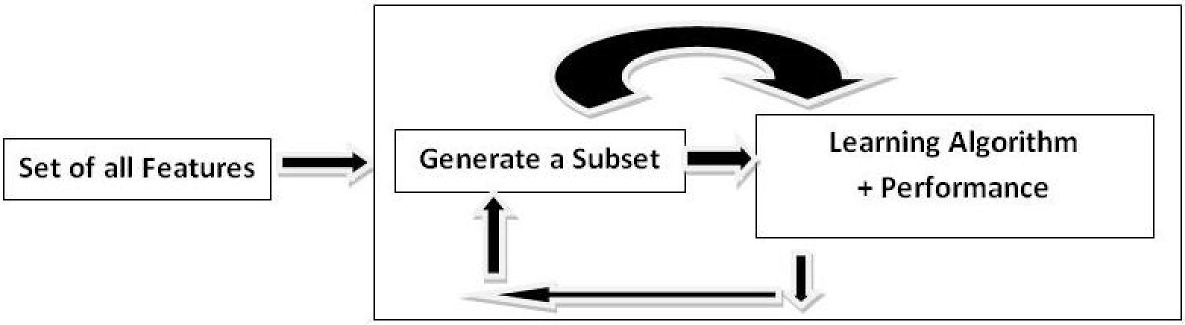

Figure 4: Embedded Model for Feature Selection

- **Hybrid Approach***:* To merge the advantage of more than one of the enumerated methods [20]. To manage the abundant data hybrid approach have newly been projected. These approaches chiefly concentrate on merging filter and wrapper algorithms to accomplish finest achievable routine with a choosy learning algorithm with the time complexity analogous to that of the filter algorithms.

Table 1: Comparison of Filter, Wrapper and Embedded Methods

Filter Methods	Wrapper Methods	Embedded Methods
Almost certainly less optimal	In supervised learning crisis these methods are better substitute	Performance degrades if additional extraneous features are include in objective set
High tendencies to choose bulky subset of data	Achieves improved identification rate than filters	Least prone to over fitting
Outcome exhibits more simplification than that of wrapper approach	Be short of simplification as it is fixed to some classifier.	Also lack simplification owing to reliance on some classification algorithm
Autonomous of classification algorithm	Reliant on classification algorithm	Reliant on classification algorithm
Implement faster than wrapper methods.	Implement slower than filter and embedded methods.	Faster than wrapper methods.
For bulky data set computational cost is less.	More than filter methods.	Less in comparison to wrapper methods.

Feature Selection Classification Regarding Problem Knowledge

Respecting a priori knowledge of the primary probability structure may be there are two fundamental classes of situations as:

- **A Little Prior Information is Exists**: It is at least acknowledged that probability density functions are unimodal.

In proper filter or wrapper surroundings recommended either the fresh forecast-supporting B&B algorithms for finest search [21] or sub-optimal search method for this kind of condition.

- **In Absence of Prior Information**: In this case, even it can't presume that probability density functions are unimodal. For these circumstances either a wrapper-based result using suboptimal search process can be establish accurate or offered the dimension of target data is adequate, it is feasible to pertain an implanted mixture-based process that are dependent on resembling mysterious category-restrictive probability density functions by predetermined combination of a unique form.

Feature Selection Classification Regarding Optimality

Feature subset selection procedure can be divided into fundamental families as:

- **Optimal Methods***:* Exhaustive search methods are example of this type, which is practicable for merely low dimensional issues and speed up methods, regularly assemble upon the Branch and Bound principle [21]. All optimal methods can be anticipated noticeably sluggish for large volume data situation.
- **Suboptimal Methods***:* The optimality of the elected subset for computational efficiency is effectively operated by them. They consist of example Best Individual Features, Sequential Forward and Backward Selection, Genetic algorithms, Random methods and mainly the Floating and Oscillating algorithms.

Even though, it is computationally excessive that the exhaustive search assured the optimality of a resolution in various rational problems. In definite circumstances several techniques are supplementary appropriate, various are more proper in additional circumstances depending on our understanding of problems. That's why in rising novel techniques nonstop labors are required to wrap the mainstream of position which know how to come across in practice.

Searching Approach

To estimate all of the candidate subset by choosing the applicant subsets and objective function a searching approach is essential. Exhaustive, heuristic and randomized searching algorithms are engaged in search approaches [22]. The time complexity of exhaustive search and heuristic search is exponential and quadratic respectively, in conditions of dimensionality. There are various hybrid algorithms too.

Exhaustive Search: It also known as exponential search. It is an optimized search in which finest result is assured. Yet most favorable searches should not be exhaustive. It evaluates numerous subsets that raising exponentially through the dimensionality of the searching area. Without of irritability, to diminish the search space of optimal solution different heuristic functions can be familiarized. Branch and Bound based algorithms are designed for exhaustive search [23]. The order of search space is $O(N^2)$.

Heuristic Search: The quantity of subsets estimated of branch and bound algorithm is more than the desired by BFF [24]. In this a large computational time is required. When the data has extremely interconnected heuristic search is challenging. Sequential forward selection, sequential backward elimination and bidirectional search are some example of heuristic search.

Random Search Strategy: Random search initiate with random selected subset. It performed randomized searching of the searching area wherever next track is a model from a given probability. The order of the search space is

$O(N^2)$. Genetic algorithm is an example of random search algorithm.

Table 2: Comparison of Random, Exhaustive and Heuristic Search

	Complexity	Accuracy	Disadvantages	Advantages
Random Search	Usually low	It is good with appropriate control parameters.	Difficult to pick good parameters	Intended to get away local minima
Exhaustive Search	Exponential	It constantly finds the best promising solution.	Complexity is high	Extremely accurate
Heuristic Search	Quadratic	If no backtracking is required this is good	Backtracking is not possible	Simple and fast

Methodologies of Feature Selection

Unimportant features are removed when we employ feature selection which returns fewer model parameters. These methods improve the simplification ability and ease complexity and diminished implementation time. Supervised and Unsupervised feature selection methods are two categories of feature selection methodologies.

Supervised Feature Selection: To estimate the elected subsets of features by picking the subset of features and an objective functions a feature selection method involves a search strategy. Here two approaches are reviewed for selecting features in circumstances where features must be notable from attributes because both emerge concurrently in the same system:

Nested Subset Methods: Features are remove as a component of the learning process by various learning machines. These contain neural network whose internal nodes are feature extractors. Consequently, node reducing method such as OBD are feature selection algorithms. In this problem Gramschmidt Orthogonalization is obtainable as a substitute to OBD [25].

Direct Objective Optimization: Kernel methods hold an embedded feature area exposed by the core extension:

$k(x, x^1) = f(x)*f(x^1)$;

Here f(x) is a feature vector of probably unbounded dimension. Generalization may improve by picking these implanted features, but executing time remains same. In this problem, for choosing embedded core features a method was projected in the condition of the polynomial kernel, using the framework of minimization of the l_0 norm [26].

- **Unsupervised Feature Selection Methods:** The methods of organizing the items into ordinary categories whose components are parallel to each others, recognized by a specified measure. As a result of the nonexistence of category tag for feature significance assessment unsupervised feature selection is principally tricky. Based on clustering superiority measures unsupervised feature selection is less inhibited search problem exclusive of category tag and can accomplish several uniformly suitable feature subsets. Occasionally, no object ‘y’ is afforded, but concerning to a defined standard one still would like to pick a set of nearly all important features. The various applications like saliency, smoothness, density, entropy and reliability are still handy in various features ranking criterion. Compared to others, a significant feature has a wide range or elevated variance. If the division of examples is identical feature has high entropy. For a time series, if on average its local curvature is sensible a feature is smooth. A feature is extremely interrelated with various extra features then it is in high density area.

Application Areas of Feature Selection in Real World

Throughout data collection, a lot of troubles are frequently encountered for instance elevated dependency of attributes, a lot of attributes, disused and unrelated features. Feature selections supply an instrument to choose a feature subset or feature to find out an algorithm efficiently to sort out the stated difficulty. Consequently, the application of feature selection is used regularly in lots of research fields in the literature.

Text Classification: The gigantic dimensions of online content on the internet such as emails, social sites, and libraries are mounting. Consequently, automatic content classification and grouping are essential responsibilities. Because of the numerous amounts of the document features a key crisis arises with content categorization or grouping. For the proficient employ of mining algorithms selection is greatly desirable. There are a variety of challenges associated to automated text classification such as:

- A suitable information composition is to be preferred to characterize the document.
- A suitable objective function is elect to optimize to evade over-fitting and attain good generalization. In conjunction with its algorithmic topics arising due to the high prescribed volume of the records is to be pact with [27].

Genre Classification: Numerous data for example name of file, writer, dimension of data, date and genres are the universal features familiar with categorize and regain genre document. On the strength of these figures, the feature selection process is mandatory because the categorization is infeasible. Features selection is a procedure where a section of an audio is illustrate into a dense numerical illustration in circumstances of genre categorization [28].

Remote Sensing: In the remote sensing image classification, feature selection is an essential job. The hyper spectral remote sensing image classification and different problems and challenges in feature selection are explained in [29].

Sentiment Analysis: Sentiment analysis is controlling compatibility via usual language dealing out. It is not just an issue dependent classification. It treats with the computational behavior of attitude, attitude and bias in content. It is helpful in approval systems and queries [30]. Feature selection should to be carrying out because all of the features are not compulsory in each and every circumstance.

Genomic Analysis: For accepting the performance of an individual, the lively performance and distinctiveness of diseases and an outsized of genomic and proteomic records formed by microarray and mass spectrometry technology. Almost all bioinformatics issues contains dimension of features extensively more than the number of samples, for example categorization of Breast cancer on account of microarray data. Even though, in case of such classification information about all the genes is not essential. As a consequence of the elevated volume of the microarray data an exceptional records investigation is insisted. To grip lofty data, one of the frequent traditions is recognition of the nearly all significant features in statistic. Consequently on complete microarray data feature selection has been prepared effectively. Relative analysis of eight feature selection for taxonomy jobs and their permutation depends on gene expression data have been completed [31].

Intrusion Detection: In present days, sharing of information, communication or circulation is extensively finished by network-supported computer systems. For that reason, the protection of the communication networks system from interruption by enemies and criminals is a key problem. To protect information network intrusion detection is one of the traditions. An essential responsibility is plays by feature selection to classifying system activity as justifiable or an invasion. An intrusion recognition model for evaluating review data that created by an organized data mining framework is analyzed [32]. For this job to examine the frequency patterns a huge data set is required. These models are directed to choose organization features for usual learning via added statistical and temporal features.

Image Retrieval: Freshly, the quantity of image collections from military and civilian tackle has enlarged. In text based image retrieval to accept professional browsing, searching and regaining data feature selection is useful [33]. Content-based image retrieval is measurable for the huge size of images, although it also cursed by high dimensionality.

Exemplary Future Research Directions

- In recent times, in the form of text, images, videos and medical data, the quantity of information gathering have enlarged; that grounds the elevated volume of the data. Feature selection is bothered by elevated volume of data. Various feature selection algorithms comprise higher time complexity regarding dimensionality 'N', hence the measurability of feature selection is a complicated hitch. As a result the future research must be rigorous on low time complexity with high measurability feature selection algorithms. There is a splendid research prospect to expand algorithms via in order and casual search tactics for grouping and categorization responsibilities correspondingly.
- Correction of errors, over-fitting and efficiency of calculation can be considered as characteristic of effective algorithms in different forward, backward and their combination technique is a significant issue for future research.
- Because of the rapid expansion of the data size, for online classifiers, the scalability of feature selection algorithm is a key issue. In outsized dimensional data around each sample, it is extremely hard to obtain a feature significance score exclusive of allowing for satisfactory density. Consequently, the scalability of feature selection algorithms is a giant challenge. So extra awareness is required to done the scalability of feature selection algorithms.
- Searching of subspace is not merely considered as the feature selection crisis. It is a technique of feature selection in which one subspace of relevant features is find out from many subspaces or high dimensional features. Hence, an effective subspace search algorithms is required for clustering.
- If distinct set of features are drawn; in each iterations, the consequences of categorization can't be trusted for the same problems. It implies that feature selection should be very stable, but the stability of existing feature selection algorithms has less. Thus, it is mandatory to construct algorithms with consistency and high categorization precision.

CONCLUSIONS

An inclusive outline of different trait of feature selection is presented by in this paper. The novel progress in variables and features selection has tackled the crisis from the realistic opinion of increasing the performance of predictors. We contain many definitions of feature selection. The labeling and individuality of feature selection are analyzed, and the fascinating information concerning the merits and demerits of feature selection procedures to grip the unusual characteristics of the genuine applications are specified. As feature selection is a preprocessing step in huge records composed from different applications. Consequently additional effort is essential to conquer restricted force as it is expensive to call elevated volume data many times. For feature selection with elevated volume data more professional strategies of search and criteria of evaluation are desirable. An approach of future challenges and research guidelines are provided by the three dimensional categorization of feature selection algorithms.

Further down the road, association must be made among the problems of attributes and feature selection and those of investigational blueprint and dynamic learning, in an attempt to shift away from obtaining data in the direction of investigational data, and to tackle crisis of causality conclusion.

REFERENCES

1. V. Bolón-Canedo, N. Sȧnchez-Maroῄo, A. Alonos-Betanzos, “A review of feature selection methods on synthetic data”, Knowledge and Information Systems, vol. 34, no. 3, pp 483-519, Mar.2013.
2. V. Bolón-Canedo, N. Sȧnchez-Maroῄo, A. Alonos-Betanzos, “An ensemble of filters and classifiers for microarray data classification”, Journals Pattern Recognition, vol. 45, pp 531-539, 2012.
3. I. Vainer, S. Kraus, G. A. Kaminka, H. Slovin, “Obtaining scalable and accurate classification in large scale spatio-temporal domains”, Knowledge Information System, 2010.
4. Y. Sun, J. Li, “Iterative RELIEF for feature weighting”, in Proc. of the 21st international conference on machine learning, pp913-920, 2006
5. B. Chidlovskii, L. Lecerf, “Scalable feature selection for multi-class problems”, Machine Learning and knowledge Discovery Databases, vol. 5211, pp 227-240, 2008.
6. Y. Zhang, C. Ding, T. Li, “Gene selection algorithm by combining relief and MRMR”, BMC Genomics 9(Suppl 2): S27, 2008.
7. J. Hua, W. Tembe, E. Dougherty, “Performance of feature- selection methods in the classification of high-dimension data”, Journal of Pattern Recognition, vol. 42, no. 3, pp. 409-424, 2009.
8. W. Lee, S. J. Stolfo, K. W. Mok, “Adaptive intrusion detection: a data mining approach”, Artificial Intelligent Reviews, vol. 14, no. 6, pp 533-567, 2000.
9. G. Forman, “An extensive empirical study of feature selection metrics for text classification”, Journal of Machine Learning Research, vol. 3, pp. 1289-1305, 2003.
10. L. Yu, H. Liu, “Redundancy based feature selection for microarray data”, in Proc. of the 10th ACM SIGKDD conference on knowledge discovery and data mining, pp 737-742, 2004.

11. P. Saari, T. Eerola, O. Lartillot, “Generalizability and simplicity as criteria in feature selection: application to mood classification in music”, IEEE Transaction on Audio Speech Language Processing, vol. 19, no. 6, pp 1802-1812, 2011.
12. J. G. Dy, C. E. Brodley, A. C. Kak, L. S. Broderick, A. M. Aisen, “Unsupervised feature selection applied to content-based retrieval of lung images”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 25, no. 3, pp. 373-378, 2003.
13. K. S. Ng, H. Liu, “Customer retention via data mining”, AI Review, vol. 14, no. 6, pp 569-590, 2000.
14. H. Liu, H. Motoda, “Feature Selection for Knowledge Discovery and Data Mining”, Boston: Kluwer Academic Publishers, 1998.
15. H. Liu, L. Yu, “Toward Integrating Feature Selection Algorithms for Classification and Clustering”, IEEE Transaction on Knowledge and Data Engineering, 2005.
16. M. Dash, H. Liu, “Feature Selection for Classification”, Intelligent Data Analysis, Elsevier, pp. 131-156, 1997.
17. W. Duch, “Filter Methods”, Springer – Feature Extraction Studies in Fuzziness and Soft Computing, vol. 207, pp 89-117, 2006.
18. 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.
19. I. Guyon, A. Elisseeff, “An introduction to variable and feature selection”, Journal of Machine Learning Research, vol. 3, pp 1157-1182, 2003.
20. M. Sebban, R. Nock, “A hybrid filter/ wrapper approach of feature selection using information theory”, Pattern Recognition, vol. 35, pp 835-846, 2002.
21. P. Somol, P. Pudil, J. Kittler, “Fast branch and bound algorithm for optimal feature selection”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 26(7), pp 900-912, 2004.
22. R. Ruiz, C. Riquelme, S. Aguilar-Ruiz, “Heuristic Search over a ranking for feature selection”, IWANN, LNCS3512, pp 742-749, 2005.
23. P. Somol, P. Pudil, “Feature Selection Toolbox”, Pattern recognition, Published by Elsevier Science Ltd., 2002.
24. P. Yan, T. Chang, “Best first strategy for feature selection”, 9th International Journal on Pattern recognition, vol. 2, pp 706-708, 1988.
25. H. Stoppiglia, G. Dreyfus, R. Dubois, Y. Oussas, “Ranking a random feature for variable and feature selection”, JMLR, vol. 3, pp 1399-1414, 2003.
26. J. Weston, A. Elisseeff, B. Schoelkopf, M. Tipping, “Use of the zero norm with linear models and kernel methods”, JMLR, vol. 3, 1439-1461, 2003.
27. A. Dasgupta, P. Drineas, B. Harb, V. Josifovski, M. W. Mahoney, “Feature selection methods for text-classification”, KDD-ACM, 2007.

28. S. Doraisamy, S. Golzari, N. M. Norowi, Md. N. B Sulaiman, N. I. Udzir, “A study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music”, Proceeding of ISMIR, 2008.

29. X. Jia, B. Chaen, M. M. Craford, “Feature mining for hyper spectral image classification”, In Proc. of the IEEE, vol. 101, no. 3, 2013.

30. B. Pang, L. Lee, “Opinion Mining and Sentiment Analysis”, Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-135, 2008.

31. H. Abusamara, “ A comparative study of feature selection and classification methods for gene expression data of glioma”, In Proc. of 4th International Conference on Computational Systems- Biology and Bioinformatics, Proceida Computer Science, vol. 23, pp. 5-14, 2013.

32. W. Lee, S. J. Stolfo, K. W. Mok, “Adaptive intrusion detection: a data mining approach”, AI Review, vol. 14, no. 6, pp 533-567, 2000.

33. T. Hastie, R. Tibsirani, J. Friedman, “The Elements of Statistical Learning”, Springer, 2001.