International Journal of Electrical, Electronics and Computer Engineering 6(1): 37-44(2017)

ISSN No. (Online): 2277-2626

A literature review on feature Selection in Big Data

Konika Verma and Pravin Malviya***

*M. Tech. Scholar, Department of Computer Science Engineering, SBITM, Betul, (Madhya Pradesh), India.

*Assistant Professor, Department of Computer Science Engineering, SBITM, Betul, (Madhya Pradesh), India.

Corresponding author: Konika Verma, v.konika@gmail.com)

(Received 07 January 2017 Accepted 08 February, 2017)

(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Feature selection and feature extraction are the two popular methods in Big Data. In today's world the data which we have is very big, in terms of velocity, variety, volume etc. So many feature selection and feature extraction methods have been proposed to obtain the relevant feature or feature subsets in the literature to achieve their objectives of classification and clustering. This paper introduces the concepts of feature relevance, general procedures, evaluation criteria, and the characteristics of feature selection and feature extraction. A detailed overview, arrangement, and comparison of existing feature selection are also done. Also discussed the guidelines for user to select a feature selection algorithm without knowing the information of each algorithm. We conclude this paper with real world applications, challenges, and future research directions of feature selection.

Keywords: Big Data, feature selection, classification, clustering, real world application.

I. INTRODUCTION

We all know that today's era is of Big Data, which we can also call it as high dimensional data, these type of data is available on that place where there is high velocity, huge volume, different varieties of data and many more parameters are there which differentiate them from normal data or non-high dimensional data. The challenge for us is processing of this data because without processing we cannot analyze this data, so processing of Big data is essential. Feature selection and feature extraction are the most important techniques for data pre-processing. Feature selection is mainly for relevant feature selection or attributes selection by removing irrelevant, redundant or noisy data. Feature extraction is for transforming high dimensional space to low dimensional space. Feature selection provides us the selective feature using these below approaches [2]:

(a) The specified size of the subset of features that optimizes an evaluation measure

(b) The smaller size of the subset that satisfies a certain restriction on evaluation measures.

(c) In general, the subset with the best commitment among size and evaluation measure

In the process of feature selection, irrelevant and redundant features or noise in the data may be hinder in many situations, because they are not relevant and important with respect to the class concept such as microarray data analysis [3]. When the number of samples is much less than the features, then machine learning gets particularly difficult, because the search space will be sparsely populated. Therefore, the model will not able to differentiate accurately between noise and relevant data [4]. There are two major approaches to feature selection. The first is Individual Evaluation, and the second is Subset Evaluation. Ranking of the features is known as Individual Evaluation [5]. In Individual Evaluation, the weight of an individual feature is assigned according to its degree of relevance. In Subset Evaluation, candidate feature subsets are constructed using search strategy.

The general procedure for feature selection has four key steps as shown in Figure 1.

(a) Subset Generation

(b) Evaluation of Subset

(c) Stopping Criteria

(d) Result Validation

Subset generation is a heuristic search in which each state specifies a candidate subset for evaluation in the search space. Two basic issues determine the nature of the subset generation process. First, successor generation decides the search starting point, which influences the search direction. To decide the search starting points at each state, forward, backward, compound, weighting, and random methods may be considered [7].

Second, search organization is responsible for the feature selection process with a specific strategy, such as sequential search, exponential search [9, 10] or random search [11]. A newly generated subset must be evaluated by a certain evaluation criteria. Therefore, many evaluation criteria have been proposed in the literature to determine the goodness of the candidate subset of the features. Base on their dependency on mining algorithms, evaluation criteria can be categorized into groups: independent and dependent criteria [8]. Independent criteria exploit the essential characteristics of the training data without involving any mining algorithms to evaluate the goodness of a feature set or feature. And dependent criteria involve predetermined mining algorithms for feature selection to select features based on the performance of the mining algorithm applied to the selected subset of features. Finally, to stop the selection process, stop criteria must be determined. Feature selection process stops at validation procedure. It is not the part of feature selection process, but feature selection method must be validate by carrying out different tests and comparisons with previously established results or comparison with the results of competing methods using artificial datasets, real world datasets, or both.

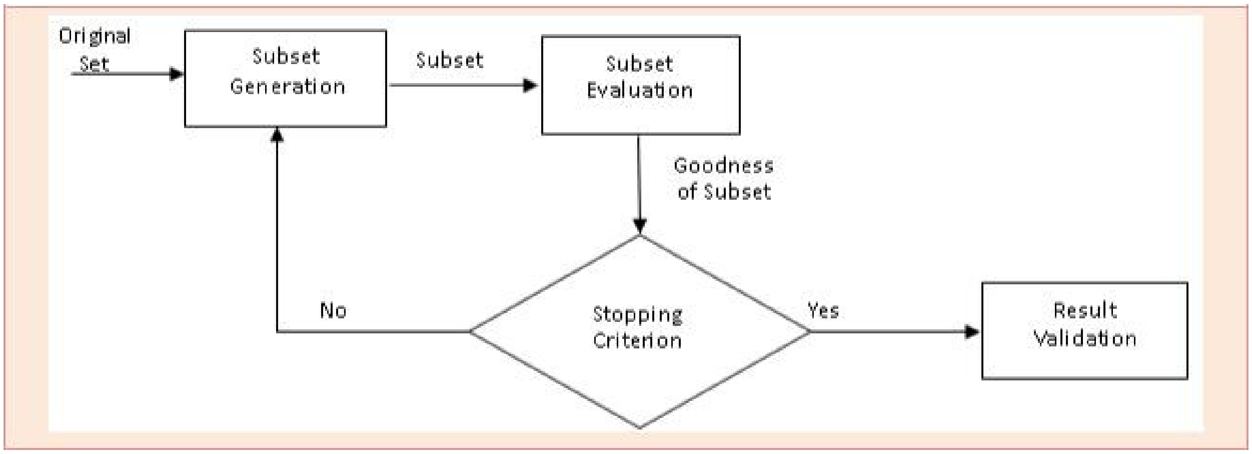

Fig. 1. Four key steps for the feature selection process [3].

The relationship between the inductive learning method and feature selection algorithm infers a model. There are three general approaches for feature selection. First, the *Filter Approach* exploits the general characteristics of training data with independent of the mining algorithm [6]. Second, the *Wrapper Approach* explores the relationship between relevance and optimal feature subset selection. It searches for an optimal feature subset adapted to the specific mining algorithm [12]. And third, the *Embedded Approach* is done with a specific learning algorithm that performs feature selection in the process of training.

A. Development in feature selection

Many, feature selection methods have been proposed in the literature, and their comparative study is a very difficult task. Without knowing the relevant features in advance of the real data set, it is very difficult to find out the effectiveness of the feature selection methods, because data sets may include many challenges such as the huge number of irrelevant and redundant features, noisy data, and high dimensionality in term of features or samples. Therefore, the performance of the feature selection method relies on the performance of the learning method. There are many performance measures mentioned in the literature such as accuracy, computer resources, ratio of feature selection, etc. Most researchers agree that there is no so-called “best method” [6]. Therefore, the new feature selection methods are constantly increasing to tackle the specific problem (as mentioned above) with different strategies.

(i) To ensure a better behaviour of feature selection using an ensemble method.

(ii) Combining with other techniques such as tree ensemble and feature extraction.

(iii) Reinterpreting existing algorithms.

(iv) Creating a new method to deal with still-unresolved problems.

(v) To combine several feature selection methods.

Many comparative studies of existing feature selection methods have been done in the literature, for example, an experimental study of eight filter methods (using mutual information) is used in 33 datasets , and for the text classification problem, 12 feature selection methods are compared . The capability of the survival Relief algorithm (sRelief) and tuned sRelief approach are evaluated in [46]. Seven filters, two embedded methods, and two wrappers are applied in 11 synthetic datasets (tested by four classifiers), which are used for comparative study of feature selection performances in the presence of irrelevant features, noise in the data, redundancy, and the small ratio between the number of attributes and samples [6]. Related to the high-dimensional dataset (in both samples and attributes), the performance of feature selection methods are studied for the multiple-class problem .

In a theoretical perspective, guidelines to select feature selection algorithms are presented, where algorithms are categorized based on three perspectives, namely search organization, evaluation criteria, and data mining tasks. In [2], characterizations of feature selection algorithms are presented with their definitions of feature relevance. In the application perspective, many real-world applications like intrusion detection [33, 31], text categorization [32], DNA microarray analysis [33], music information retrieval [35], image retrieval [36], information retrieval [37], customer relationship management [38], Genomic analysis [33] and remote sensing [39] are considered.

B. Defining Feature Relevance

The optimal feature subset is a subset of all relevant features. Therefore, the relevance of the features must be properly defined according to their relevance. In the literature, features are classified by their relevancy with three qualifiers: irrelevant, weakly relevant, and strongly relevant. A graphical representation is shown in Figure 2 [34]. Many definitions have been proposed to answer a question “relevant to what?” [18]. Therefore, in this section, the definition of the relevance of the feature is presented as suggested in the literature, and the degree of relevance is suggested as well.

Fig. 2. A view of feature relevance [34].

C. Feature Selection

Feature selection is the process of selecting relevant features, or a candidate subset of features. The evaluation criteria are used for getting an optimal feature subset. In high-dimensional data (number of samples <<number of features), finding the optimal feature subset is a difficult task [13]. There are many related problems that are shown as NP-hard [12, 14]. The data with number of features, there exists 2^N candidate subset of features.

D. General Approach for Feature Selection

There are three general approaches for feature selection.

Filter Approach. The filter approach incorporates an independent measure for evaluating features subsets without involving a learning algorithm. This approach is efficient and fast to compute (computationally efficient). However, filter methods can miss features that are not useful by themselves but can be very useful when combined with others. The graphical representation of the filter model is shown in Fig. 3.

Wrapper Approach. The filter and wrapper approach can only be distinguished by the evaluation criteria. The wrapper approach uses a learning algorithm for subset evaluation.

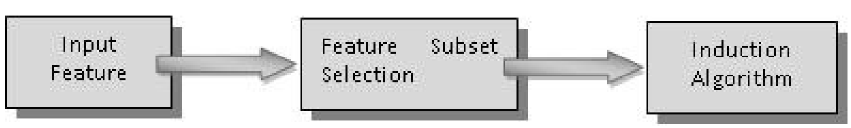

Fig. 3. A filter method.

A graphical representation of the wrapper model is shown in Figure 4.A different wrapper algorithm can be generated by varying the subset generation and subset evaluation measure (using dependent criterion). The wrapper approach selects an optimal subset that is best suited to learning algorithm. Therefore, the performance of the wrapper approach is usually better.

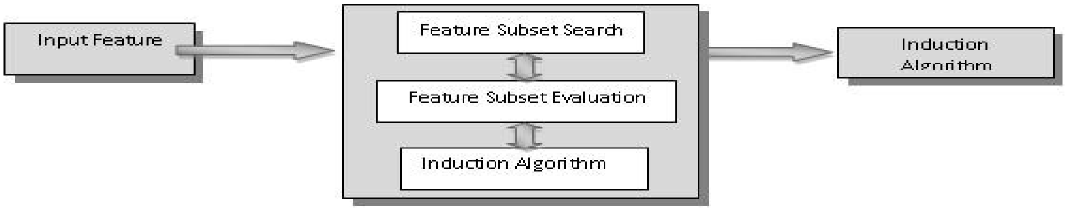

Fig. 4. A filter method.

Embedded Approach. This approach interacts with learning algorithm at a lower computational cost than the wrapper approach. It also captures feature dependencies. It considers not only relations between one input features and the output feature, but also searches locally for features that allow better local discrimination. It uses the independent criteria to decide the optimal subsets for a known cardinality. And then, the learning algorithm is used to select the final optimal subset among the optimal subsets across different cardinality.

E. Categorization and Characteristics of Feature Selection Algorithms

In the literature, a large number of feature selection algorithms are available. Each algorithm can be different in order to inner mechanism and commonalities. Huan Liu and Lei Yu [8] proposed a three-dimensional categorization framework, More algorithms are introduced to strengthen the categorization. Search strategy and evaluation are two dominating factors in the feature selection algorithm. Therefore, mentioned factors are used as two dimensions in the framework. In search strategy, sub-categorization is done - namely complete, sequential, or random corresponding to the data mining task (classification and clustering). Algorithms are categorized as filter, wrapper, and embedded under evaluation criteria. Further categorization of a filter is done in distance, information, dependency and consistency. Wrapper and embedded algorithms are also categorized into predictive accuracy and filter+wrapper, respectively.

A space of characteristics of feature selection algorithms according to their criteria, namely search organization, generation of successors, and evaluation measure is presented in Figure 5 [2].

F. Application of Feature Selection in Real World

During data collection, many problems are often encountered such as a high dependency of features, too many features, or redundant and irrelevant features. To deal with the mentioned problem, feature selection provides a tool to select a feature subset or feature to learn algorithms effectively.

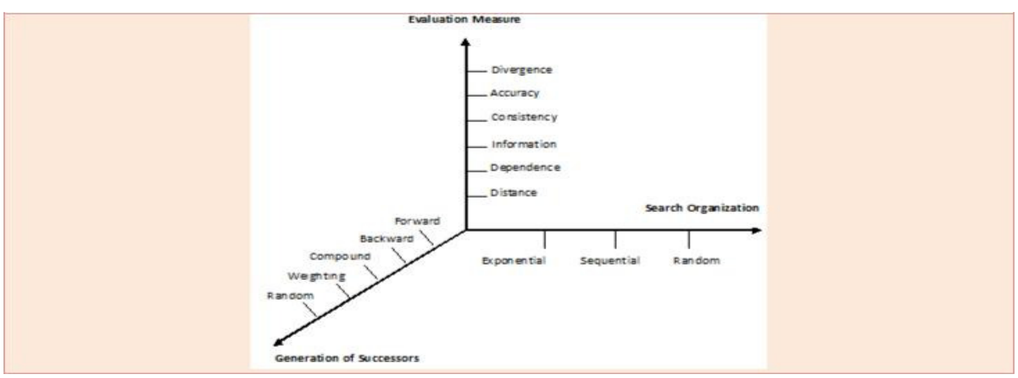

Fig. 5. A space of characteristics of feature selection algorithms [2].

Therefore, in the literature, the applications of feature selection are used frequently in many research areas.

G. Text Categorization

The massive volume of online text data on the Internet such as emails, social sites, and libraries is increasing. Therefore, automatic text categorization and clustering are important tasks.

A major problem with text classification or clustering is the high dimensionality of the document features. A moderate size text document may have hundreds of thousands of features. Therefore, feature selection (dimension reduction) is highly enviable for the efficient use of mining algorithms.

In the literature, many applications of feature selection techniques are effectively used in the area of text mining. Feature selections using the information Gain Ratio (GR) is used for lyrics and poems for text data classification. Many feature selection techniques are used for feature reduction, then evaluated and compared to the classification problem [1, 2, 3, 34].

H. Remote Sensing

Feature selection is one of the important tasks in the remote sensing image classification. In paper [52], the challenges and various issues in feature selection and hyper spectral remote sensing image analysis is explained. In [53], pre-processing techniques have been proposed for hyper spectral images in which feature extraction and feature selection have been emphasized as important components in hyper spectral image classification. Feature selection guided by evolutionary algorithms has been proposed, and use a self-adaptive differential evolution for feature subset generation. Generated feature subsets are evaluated by the wrapper method with the help of fuzzy k-nearest neighbour classifier [54]. Shijin Li, Hao Wu, Dingsheng, and Wan Jiali Zhu have developed a hybrid approach for feature selection using support vector machine and genetic algorithm [55]. They have used the wrapper method to select the optimal number of features in order to obtain better accuracy. In [56], a novel technique has been proposed to select a subset of bands from a hyper spectral image to improve the performance of the classification. It utilizes spatial and spectral information simultaneously to improve the discrimination capability of the classifier [9].

I. Intrusion Detection

In this modern age, information sharing, distribution, or communication is widely done by network-based computer systems. Therefore, the security of the system is an important issue protecting communication networks from intrusion by enemies and criminals. One of the ways to protect communication networks (computer systems) is intrusion detection. Feature selection plays an important role to classifying system activity as legitimate or an intrusion. In [48], data mining techniques and feature selection techniques are used for intrusion detection. In this paper they did a comparative study about techniques, their advantages, and disadvantages. In [49], there is a systematic data mining framework that constructs an intrusion detection model for analyzing audit data. In this work, a large data set is used for an analysis of the frequency patterns. These patterns are guided to select system features for automatic learning using additional statistical and temporal features.

J. Genomic Analysis

A large quantity of genomic and proteomic data is produced by microarray and mass spectrometry technology for understanding of function of an organism, and the behaviour, dynamics, and characteristics of diseases. Tens of thousands of genes are measured in a typical microarray assay and mass spectrometry proteomic profile. Special data analysis is demanded because of the high dimensionality of the microarray data. One of the common ways to handle high dimensionality is identification of the most relevant features in the data. Therefore, in the literature, feature selection has been done successfully on full microarray data. In [50] the Filter, Wrapper, and Embedded methods have been used for feature selection and dimensionality reduction. The techniques covered by them are the most effective for proteomics data and genomic analysis. In [51], comparative studies of 8 feature selection for classification task and their combinations have been done based on gene expression data. It is also shown that classification accuracy can be significantly boosted by a small number of genes by using a feature selection method.

K. Image Retrieval

Recently, the amount of image collections from military and civilian equipment has increased. To access the images or make use of the information, images should be organized in a way that allows effective browsing, retrieving, and searching. As stated in [34], content-based image retrieval is scalable for the large size of images, but it is also cursed by high dimensionality. Therefore, feature selection is an important task for effective browsing, searching, and retrieval. In [54], content-based image retrieval is proposed that annotates images by their own colours, textures, and shape.

L. Challenges and Future Direction

Forward vs Backward Selection. In the literature, it is argued that backward elimination is less efficient than forward selection. To defend backward selection, it is said that forward selection finds weaker subset of features, because weaker features are not assessed while subset selection. Moreover, the computational complexity forward feature selection method is less than backward feature selection. Pros of the forward greedy feature selection method are that it is computationally efficient and does not over fit. Cons, errors made in the early stage by forward greedy feature selection method are do not correct later stages. Backward greedy feature selection has corrections of errors by looking at all the models, but it starts with non-over-fit or sparse model. Both methods have their own pros and cons for feature selection. Therefore, a combination of forward greedy and backward greedy feature selection has been presented that does not over-fit, is computationally efficient, is error corrected by backward greedy step later, and that is made in the early stage in order to trade off. For future research, error correction, over-fitting, and computational efficiency can be considered as features of effective algorithms.

M. Feature Selection with Large Dimensional Data

Recently, the amount of data collections have increased in the form of text documents, images, videos, and medical data that cause the high dimensionality of the data. Dimensionality in the range of hundreds is called high-dimensional data [8]. Recently, feature selection has been applied to tens or hundreds of thousands of features. Moreover, feature selection is cursed by high dimensionality Many feature selection algorithms have higher time complexity about dimensionality, therefore the scalability of feature selection is a difficult problem. A filter approach has less computational complexity than a wrapper approach, because it uses independent subset evaluation criteria for subset evaluation. A filter approach is more scalable than the wrapper, so is preferred to a wrapper approach for feature selection. In literature, the embedded approach [32] has been proposed to utilize the qualities of the filter and wrapper approach high dimension environment. The embedded method has similar time complexity as the filter approach. To handle the high dimensional data, an efficient correlation-based filter algorithm has been proposed . The inference of the above discussion is that future research must be concentrated on low time complexity with high scalability feature selection algorithms. There is a great research opportunity to develop algorithms using sequential and random search strategies for clustering and classification tasks respectively.

N. Subspace Searching and Instance Selection

In clustering, many clusters may exist in different subspaces for small dimensionality with overlapped or non-overlapped dimensions. Subspace searching is not only the feature selection problem. It is finding many subspaces in which feature selection finds one subspace. In literature, many algorithms (subspace clustering) have been developed. Therefore, there is a requirement for efficient subspace search algorithms for clustering. In instance selection, sampling methods have been developed to search for a set of instances that can perform in a focused way.

O. Feature Selection with Sparse Data Matrix

A relatively high percentage of variables that do not have actual data are called sparse data. There are two types of Sparsity namely Controlled Sparsity and Random Sparsity. Controlled Sparsity is a range of values of one or more than one dimension that have no data. Random Sparsity, in contrast, is empty values scattered throughout the data variable. In a business context, many individual transactions are recorded in the application such as market basket analysis, direct-mail marketing, insurance, and health care [8]. These types of data collections have a sparse matrix with a large number of attributes. Some other sparse data are commonly available through computer and internet web technology such as HTML, XML, emails, news, and customer reviews. Video stream data is also increasing rapidly with high dimensionality via surveillance cameras, sensors, and web streaming. Feature selection from labelled or unlabelled sparse data is a difficult task, because many feature selection techniques are not suitable for high dimensional sparse data. It is not advised to modify feature selection algorithms for sparse data [8]. Therefore, it is a requirement of future research to develop efficient feature selection algorithms for sparse data.

P. Scalability and Stability of Feature Selection

The scalability of feature selection algorithms is an important issue for online classifiers, because of the rapid growth of the dataset sizes. A large dataset cannot be loaded in the memory for the single data scan. Full dimensionality of the data must be scanned for feature selection. It is very tough to get a feature relevance score without considering sufficient density around each sample. Therefore, the scalability of feature selection algorithms is a big challenge.

To solve this problem, some methods have tried to overcome by memorizing only important samples or summaries. More attention is required on the scalability of feature selection algorithms. The results of classification cannot be trusted if a different set of features are drawn for the same problem in each iteration. That means feature selection algorithms should be very stable (less sensitive). Well-known feature selection algorithms have less stability. Therefore, it is required for developed algorithms with stability and high classification accuracy.

II. CONCLUSION

We comprise many definitions of feature relevance, feature selection, and optimal feature subsets. The general procedure of feature selection is described with subset generation, evaluation of subsets, and stopping criteria. Three general approaches of feature selection methods, namely filter, wrapper and embedded methods, are described in detail and their pseudo code is also presented. The categorization and characteristics of feature selection are reviewed, and the interesting facts regarding the advantages and disadvantages of feature selection methods to handle the different characteristics of the real world applications are enumerated. The three dimensional categorization of feature selection algorithms give an insight of future challenges and research directions.

REFERENCES

[1]. A. Kalousis, J. Prados, M. Hilario, "Stability of Feature Selection Algorithms: a study on high dimensional spaces," *Knowledge and information System*, vol. **12**, no. 1, pp. 95-116, 2007.

[2]. Luis Carlos Molina, Lluís Belanche, À ngela Nebot, "Feature Selection Algorithms: A Survey and Experimental Evaluation," in *Proc.of ICDM, pp. 306-313, 2002.*

[3]. M. Dash, H. Liu, "Feature Selection for Classification," *Intelligent Data Analysis,* Elsevier, pp. 131-156, 1997.

[4]. F. Provost, "Distributed data mining: scaling up and beyond," *Advances in distributed data mining*, Morgan Kaufmann, San Francisco, 2000.

[5]. I. Guyon, A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. **3**, pp. 1157-1182, 2003.

[6]. Verónica Bolón-Canedo, Noelia Sánchez-Maroño, Amparo Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and Information Systems,* vol. **34**, no. 3, pp 483-519, Mar. 2013.

[7]. J. Doak, "An evaluation of feature selection methods and their application to computer security," *Technical report*, Davis CA: University of California, Department of Computer Science, 1992.

[8]. Huan Liu, Lei Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *IEEE Transaction on Knowledge and Data Engineering*, 2005.

[9]. P. Narendra, K. Fukunaga, "A Branch and Bound Algorithm for Feature Subset Selection," *IEEE Transactions on Computer*, vol. **26**, no. 9, pp. 917-922, 1977.

[10]. J. Pearl, "Heuristics," *Addison-Wesley*, 1983.

[11]. H. Liu, H. Motoda, "Feature Selection for Knowledge Discovery and Data Mining," *Kluwer Academic Publishers*, London, GB, 1998.

[12]. Ron Kohavi, George H. John, "Wrapper for Feature Subset Selection," *Artificial Intelligence*, Elsevier, pp. 273-324, 1997.

[13]. R. Kohavi, G.H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. **97**, no. 1-2, pp. 273-324, 1997.

[14]. A.L. Blum, R.L. Rivest, "Training a 3-node neural networks is NP-complete," *Neural Networks*, vol. **5**, pp. 117-127, 1992.

[15]. M. Kudo, J. Sklansky, "A Comparative Evaluation of medium and large–scale Feature Selectors for Pattern Classifiers," in *Proc. of the 1st International Workshop on Statistical Techniques in Pattern Recognition*, pp. 91-96, 1997.

[16]. M. A. Hall, "Correlation–based Feature Selection for Machine Learning," *PhD thesis*, University of Waikato, 1999.

[17]. R. A. Caruana, D. Freitag, "Greedy Attribute Selection," in *Proc. of the 11th International Conference on Machine Learning*, New Brunswick, NJ, Morgan Kaufmann, pp. 28-36, 1994.

[18]. A. L. Blum, P. Langley, "Selection of Relevant Features and Examples in Machine Learning," *Artificial Intelligence on Relevance*, vol. **97**, pp. 245-271, 1997.

[19]. R. A. Caruana, D. Freitag, "Greedy Attribute Selection," in *Proc. of the 11th Int. Conf. on Machine Learning*, New Brunswick, NJ, Morgan Kaufmann, pp. 28-36, 1994.

[20]. K. Wang, D. Bell, F. Murtagh, "Relevance Approach to Feature Subset Selection," *Kluwer Academic Publishers*, pp. 85-97,1998.

[21]. H. Liu, H. Motoda, "Feature Selection for Knowledge Discovery and Data Mining," *Boston: Kluwer Academic Publishers,* 1998.

[22]. P. M. Narendra, K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Transaction on Computer,* vol. 2**6**, no. 9, pp. 917-922, 1977.

[23]. G. Brassard, P. Bratley, "Fundamentals of Algorithms," *Prentice Hall*, New Jersey, 1996.

[24]. D. Koller, M. Sahami, "Toward Optimal Feature Selection," in *Proc. of the 13th International Conference on Machine Learning,* Bari, IT, Morgan Kaufmann, pp. 284-292, 1996.

[25]. H. Almuallim, T.G. Dietterich, "Learning boolean concepts in the presence of many irrelevant features," *Artificial Intelligence*, **69**(1-2):279–305, 1994.

[26]. M. Ben-Bassat, "Pattern recognition and reduction of dimensionality," *Handbook of statistics-II*, North Holland, pp. 773-791, 1982.

[27]. M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proc. of the 17th International Conference on Machine Learning*, pp. 359-366, 2000.

[28]. P. A. Devijver, J. Kittler, "Pattern Recognition – A Statistical Approach," *Prentice Hall*, London, GB, 1982.

[29]. http://en.wikipedia.org/wiki/Divergence_%28statistics%29

[30]. 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.

[31]. J. C. Gomez, E. Boiy, M. F. Moens, "Highly discriminative statistical features for email classification," *Knowledge and Information Systems*, 2011.

[32]. 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.

[33]. 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.

[34]. C. E. Shannon, W. Weaver, "The Mathematical Theory of Communication," *University of Illinois Press*, ISBN 0-252-72548-4, 1949.

[35]. 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.

[36]. O. Egozi, E. Gabrilovich, S. Markovitch, "Concept-based feature generation and selection for information retrieval," in *Proc. of the twenty-third AAAI conference on artificial intelligence*, pp 1132-1137, 2008.

[37]. K. S. Ng, H. Liu, "Customer retention via data mining," *AI Review*, vol. **14**, no. 6, pp. 569-590, 2000.

[38]. C. De Stefano, F. Fontanella, C. Marrocco, "A GA-Based Feature Selection Algorithm for Remote Sensing Images," *Applications of Evolutionary Computing* , Lecture Notes in Computer Science, *vol.* 4974, pp 285-294, 2008.

[39]. V. Kumar, S. Minz, "Poem Classification using Machine Learning Approach," *International Conference on Soft Computing for problem Solving (SocPro 2012)*, Dec. 2012.

[40]. V. Kumar, S. Minz, "Mood Classification of Lyrics using SentiWordNet", in *Proc. of ICCCI-2013*, Jan. 2013.

[41]. V. Kumar, S. Minz, "Multi-view Ensemble Learning for Poem Data Classification using SentiWordNet," in *Proc. of 2nd International Conference on Advanced Computing, Networking, and Informatics (ICACNI-2014)*, 2014.

[42]. T. Lappas, K. Pelechrinis, "Data Mining Techniques for (Network) Intrusion Detection Systems," *Department of Computer Science and Engineering UC Riverside, Riverside CA,* 92521, 2007.

[43]. 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.

[44]. M. Hauskrecht, R. Pelikan, M. Valko, J. Lyons-Weiler, "Feature Selection and Dimensionality Reduction in Genomics and Proteomics," *Fundamentals of Data Mining in Genomics and Proteomics*, pp. 149-172, 2007.

[45]. 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. Article (CrossRef Link)

[46]. H. Liu, H. Motoda, L. Yu, "Feature selection with selective sampling," in *Proc. of the Nineteenth International Conference on Machine Learning*, pp. 395-402, 2002.

[47] V. Kumar, S. Minz, "Multi-view Ensemble Learning for Poem Data Classification using SentiWordNet," in *Proc. of 2nd International Conference on Advanced Computing, Networking, and Informatics (ICACNI-2014)*, 2014.

[48] T. Lappas, K. Pelechrinis, "Data Mining Techniques for (Network) Intrusion Detection Systems," *Department of Computer Science and Engineering UC Riverside, Riverside CA,* 92521, 2007.

[49] M. Hauskrecht, R. Pelikan, M. Valko, J. Lyons-Weiler, "Feature Selection and Dimensionality Reduction in Genomics and Proteomics," *Fundamentals of Data Mining in Genomics and Proteomics*, pp. 149-172, 2007.

[50] H. Abusamra, "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,* Procedia Computer Science, *vol.* **23,** *pp.* 5-14,2013.

[51] X. Jia, Bor-Chaen, Melba M. Craford, "Feature mining for hyper spectral image classification," in *Proc. of the IEEE*, vol. **101**, no. 3, Mar. 2013.

[52] L. Kauo, "Nonparametric weighted feature extraction for classification" *IEEE Transactions on Geoscience and Remote Sensing*, vol. **42**, no. 5, May 2004.

[53] A. Ghosh, A. Datta, S. Ghosh, "Self-adaptive differential evolution for feature selection in hyper spectral image data," *Applied Soft Computing*, 2013.

[54] S. Li, H. Wu, D. Wan, J. Zhu, "An effective feature selection method for hyper spectral image classification based on genetic algorithm and support vector machine" *Knowledge Based Systems*, vol. **24**, pp. 40-48, 2011.

[55] Y.-Q. Zhao, L. Zhang, S. G. Kong, "Band-Subset-Based Clustering and Fusion for Hyperspectral Imagery Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. **49**, no. 2, Feb. 2011.