A Review on Dimensionality Reduction Techniques

Priyanka Jindal
Department of CSE, GJU S&T,
Hisar, Haryana-125001

Dharmender Kumar
Department of CSE, GJU S&T,
Hisar, Haryana-125001

ABSTRACT

Progress in digital data acquisition and storage technology has resulted in exponential growth in high dimensional data. Removing redundant and irrelevant features from this high-dimensional data helps in improving mining performance and comprehensibility and increasing learning accuracy. Feature selection and feature extraction techniques as a preprocessing step are used for reducing data dimensionality. This paper analyses some existing popular feature selection and feature extraction techniques and addresses benefits and challenges of these algorithms which would be beneficial for beginners..

Keywords

Feature Selection, Feature Extraction, Principal Component Analysis (PCA), Filter methods, Wrapper Methods.

1. INTRODUCTION

Nowadays, the growth of high-throughput technologies has resulted in exponential growth in data with respect to both dimensionality and sample size. Efficient and effective management of this data becomes challenging. It become a major problem for users that how to find useful or needed information from this vast amount of data quickly and easily. Information retrieval technology can help people to partially solve this problem.

High dimensional data is the data which contains hundreds or thousands of features. Many of these features in such dataset contains useful information for understanding the data but it also contains a lot of irrelevant or redundant features which reduce the performance, quality of data and computational efficiency [1]. DOROTHEA [2] is such a dataset used for drug discovery consist of about 1 lakh features and 2000 instances. To avoid these problems, a pre-processing step called ‘dimensionality reduction’ is used to remove these irrelevant and redundant features.

Dimensionality reduction is preprocessing in high dimensional data analysis, visualization and modeling. It has become the focus of much research in areas of applications including text processing of internet documents, gene expression array analysis, combinatorial chemistry and especially medical field [3]. Removing irrelevant features from data not only allows search algorithm to run faster and accurate but also results in high accuracy and increased computational efficiency.

A motive of this paper is to present a review on various dimensionality reduction approaches. The rest of the paper is organized as follows: Section II presents a brief review on dimensionality reduction approaches existing in literature: feature selection and feature extraction and various algorithms available for each method. Finally conclusion part is given to motivate the researchers.

2. REVIEW ON DIMENSIONALITY REDUCTION TECHNIQUES

High dimensional data is problematic due to high computational cost and memory usage [4]. There are two types of dimensionality reduction techniques in literature in data mining: 1) Feature Selection and 2) Feature Extraction / Transformation.

Feature selection [5] is the process of identifying and removing irrelevant and redundant information from data. It helps us to find good feature subsets related to target concepts. It describes several tools and techniques for reducing dimensionality of data.

Feature extraction [6] creates new feature sets that are more significant from original features by applying some transformations. The new feature sets have comparatively low dimensional subspace preserving most of the relevant information.

Feature extraction and selection methods can be used in isolated or in combination with the aim to improve performance such as estimated accuracy, visualization and comprehensibility of learned knowledge. Among the two methods of dimensionality reduction, feature extraction is more general.

2.1 Feature Selection

Feature selection [3], [7] selects a subset of original features, without any loss of useful information**.** It is a heuristic search process of selecting descriptors that are most effective in characterizing a specified domain. It addresses the particular task of finding a subset of given features that are useful to solve the domain problem.

The aim of feature selection is to select a subset of features as small as possible. Before applying data mining tasks, this steps needs to be applied as a pre-processing step. Feature selection offers various advantages [8] as:

- Reducing storage requirement.
- Avoiding over fitting.
- Facilitating data visualization.
- Speeding up execution of mining algorithms.
- Reducing training times.

Feature selection process consists of four steps: feature subset generation, subset evaluation, stopping criterion and result validation. It uses various search methods like complete, sequential and random search to generate small feature subset. The goodness of the generated subset is evaluated using an evaluation criterion. If the newly generated subset is better than the previous subset, it replaces the previous subset with the best feature subset. The above two processes are repeated until the stopping criterion is reached. The final best feature subset is then validated by using different tests [9]. The best feature subset is the one with least number of dimensions that most contribute to learning accuracy. The whole process of feature selection is summarized in figure 2 given below.

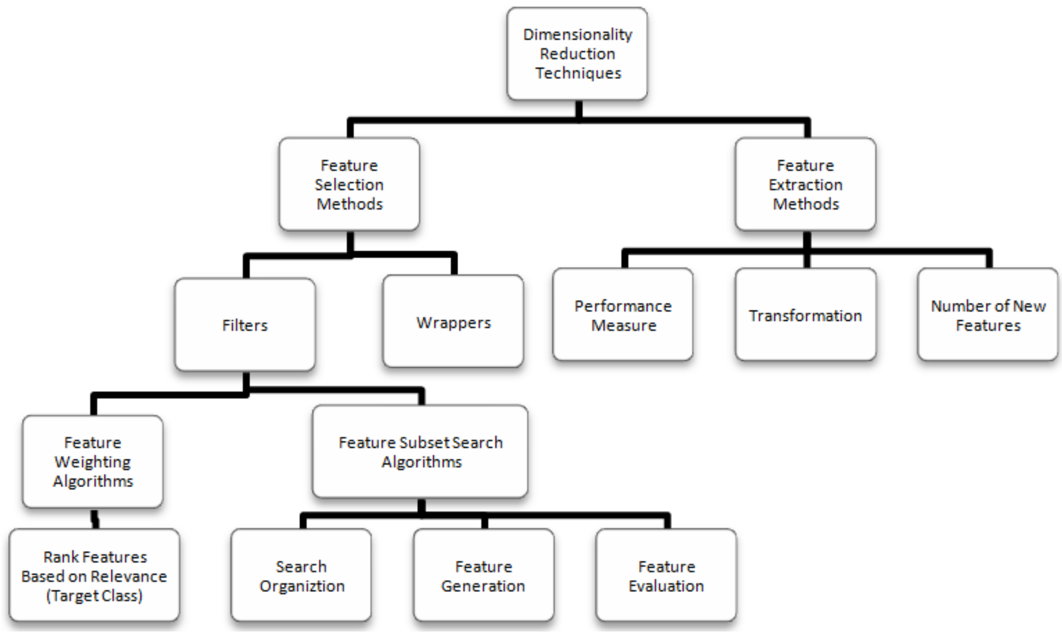

Fig 1: Hierarchal Structure of Dimensionality Reduction Techniques

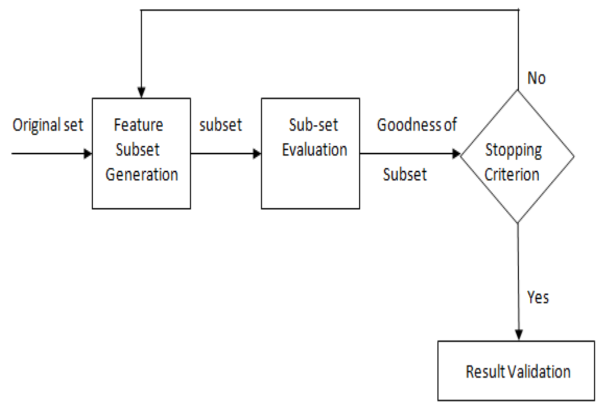

Fig 2: Feature Selection Process

Feature selection methods are classified into three classes:

- Filter methods,
- Wrapper methods and
- Embedded/hybrid methods.

Relying on characteristics of data, **Filter methods** [10] evaluate features without utilizing any classification algorithm. Filter models uses statistical properties of variables to filter out poorly informative variables. These models can be either 'Univariate' or 'multivariate'. In the Univariate [11] scheme each feature is ranked independently of feature space while the multivariate scheme evaluates features in a batch way. Filter models are easily scalable to very high dimensional datasets, computationally simple and fast, the drawbacks of filter models is that they totally ignores the effect of selected feature subset on performance of induction algorithm [12].

Wrapper methods [13] utilize a predetermined learning algorithm to evaluate the quality of selected features and offer a simple and powerful way to address the problem of feature selection. The accuracy measured by this algorithm is very high as this method considers the interaction between feature subset searches and model selection. But this method is computationally expensive and lacks generality. This method is more demanding than filter methods.

Hybrid models [14] are combination of both filter models and wrapper models. They include the features of 1) filter models: they are less computationally intensive. 2) Wrapper models: they include the interaction with model construction. A block diagram for hybrid model is presented in fig 3.

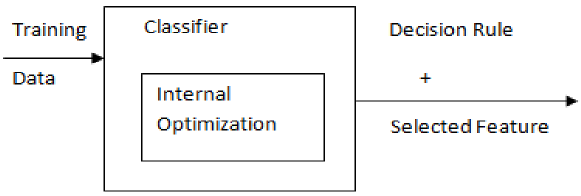

Fig 3: Hybrid Model for Feature Selection

The features of these three methods are summarized in table 1 given below

Table 1: Comparison of Feature Selection Methods

Method	Features	Algorithm
Filter approach	- Suitable for very large features. - Computationally simple and fast. - posses generality	- Focus - Relief -Correlation Coefficient - Bw-ratio - mRmR - Interact
Wrapper approach	- High accuracy. - Lacks generality. - Considers interaction between feature subset searches and model	-Forward selection -Backward elimination -Evolutionary

	selection.	local selection
Hybrid approach	- Less computation than wrapper methods. - Considers feature dependencies.	-Two phase local selection -Sequential forward selection

Some of the algorithms for feature selection techniques are mRmR, Relief, Correlation Coefficient, CMIM, BW-ratio, evolutionary local selection algorithm, two-phase feature selection algorithm, interact etc. The following table provides a comparative analysis of various feature selection algorithms.

Table 2: Various Feature Selection Algorithms

Algorithm	Type	Approach used	Features	Drawbacks
Interact [15]	Filter	Uses symmetric uncertainty and backward elimination approach	-Improves accuracy	-mining performance decreases as dimensionality increases.
Relief[16]	Filter	Relevance evaluation	-scalable to dataset with high dimensionality	-cannot eliminate redundant features.
Fast clustering based feature selection [17]	Filter	Graph theoretic clustering method used for clustering	-Highly reduced dimensionality.	-works well for microarray data only.
Correlation- based Feature Selection[18]	Filter	Uses symmetric uncertainty	-handles both irrelevant and redundant features. -prevents re-introduction of redundant features.	-works on smaller datasets. -cannot handle numeric class problems.
Condition Dynamic Mutual [19] Information	Filter	Mutual information	-better performance	-Sensitive to noise
Affinity propagation sequential feature selection [20]	Wrapper	Affinity propagation clustering algorithm applied to get clusters. SFS applied for each cluster o get best subset.	-faster than sequential feature selection.	-accuracy is not so high.
Evolutionary local selection[21]	Wrapper	K-Means algorithm used for clustering.	-Covers a large space of possible feature combination.	-cluster quality decreases as the number of features increases.
Two phase feature selection[22]	Hybrid	(filter) Artificial Neural Network Weight Analysis used to remove irrelevant features. (Wrapper) Genetic Algorithm used to remove redundant features.	-handles both irrelevant and redundant feature. -improves accuracy.	
Hybrid feature selection[23]	Hybrid	(Filter) Mutual information (Wrapper) Wrapper based feature selection which uses Shepley value.	-improves accuracy.	-high computation cost for high dimension data.

Feature selection method also has some drawbacks:

- Search for subset of relevant features introduces an additional layer of complexity in modeling task.
- Consume more time for learning.

2.2 Feature Extraction/Transformation

Feature extraction [24] is a method to obtain new features from available features by applying some transformations to reduce complexity and to give a simple representation of data representing each variable in feature space as a linear combination of input variables. Feature extraction is more general method than feature selection methods. Various feature extraction methods are Principal component analysis (PCA), Non-Linear principal component analysis, independent component analysis etc.

The most popular and widely used feature extraction approach is **Principal Component Analysis** (PCA) [25] introduced by Karl. Principal component analysis (PCA) consists into an orthogonal transformation to convert samples belonging to correlated variables into samples of linearly uncorrelated variables. It can project the data from the original space into a lower dimensional space in an unsupervised manner. The main reason for the use of PCA concerns the fact that PCA is a simple nonparametric method used to extract the most relevant information from a set of redundant or noisy data.

Principle components (PC) are new variables with two properties: 1) each PC is a linear combination of the original variables; 2) the PC's are uncorrelated to each other and also the redundant information is removed [12]. Main application areas of PCA include data compression, image analysis, visualization, pattern recognition, regression, and time series prediction.

PCA suffers from some limitations:

- It assumes that the relationships between variables are linear.
- Interpretation of PCA is only meaningful if all of the variables are assumed to be scaled at the numeric level. It lacks a probabilistic model structure which is important in many contexts such as mixture modeling and Bayesian decision.

3. CONCLUSION

In this paper, various widely used dimensionality reduction techniques are investigated with the purpose of how effectively these techniques can be used to achieve high performance of learning algorithms that ultimately improves the predictive accuracy of various classifiers. Among all the existing dimensionality reduction algorithms, some algorithms involve in removal of either irrelevant or redundant features while others involve in removal of both types of features. These algorithms develop small feature subset consisting of either same types of features as in the original feature set or derive some new features from original features depending on the need. Our study concludes that due to availability of wide variety of variables, data distributions and objectives, still there is no feature selection algorithm that is universally better than others after a lot of research. So, there is substantial need to count on criteria for feature selection which should involve in selection of best feature subset in certain situations.

4. REFERENCES

[1] Kashif Javed, Haroon A. Babri and Mehreen Saeed, "Feature Selection based on Class-Dependent Densities for High Dimensional Binary Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 24, No. 3, pp. 465-477, 2012.

[2] I. Guyon, S. Gunn, A. Ben-Hur and G. Dror, "Feature Selection Challenge", *In Advances in Neural Information Processing Systems Conference (NIPS)*, MIT Press, pp. 545-552, 2003.

[3] Kajal Naidu, Aparna Dhemge and Kapil Wankhade, "Feature Selection Algorithm for Improving the Performace of Classification: A Survey", *In proceedings of fourth international conference on Communication Systems and Network Technology*, pp. 468-471, 2014.

[4] G. K. Janecek, G. F. Gansterer, M. A. Demel, G. F. Ecker "On the Relationship between Feature Selection and Classification Accuracy", *In Proceeding of workshop and conference on New Challenges for Feature Selection*, pp. 40-105, 2008.

[5] A. H. Shahana, V. Preeja, " Survey on Feature Subset Selection for high dimensional data", *In proceedings of International Conference on Circuit, Power and Computing Technologies*, pp. 1-4, 2016.

[6] H. Motoda and H. Liu, "Feature Selection, Extraction and Construction", *In proceedings of Conference on Knowledge Discovery and Data Mining*, Taipei, Taiwan, pp. 67–72, 2002.

[7] Monica Rogati, Yiming Yang, "High-performing Feature Selection for Text Classification," *In proceedings of 11th International Conference on Information and Knowledge Management*, McLean, Virginia, USA, pp.659 – 661, 2002.

[8] Zilin Zeng, Hongjun Zhang, Rui Zhang and Youliang Zhang, "Hybrid Feature Selection Method based on Rough Conditional Mutual Information and Naïve Bayesian Classifier", *ISRN Applied Mathematics*, Vol. 2014, Article Id 382738, pp. 1-11, 2014.

[9] K. Sutha, Dr. J. Jebamalar Tamilselvi, " A Review of Feature Selection Algorithms for Data Mining Techniques*", International Journal of Computer Science and Engineering*, Vol. 7, No. 6, pp. 63-67, 2015.

[10] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.

[11] Jiliang Tang, Salem Alelyani and Huan Liu, "Feature Selection for Classification: A Review", *In Data classification: Algorithms and Applications*, pp. 1-37, 2014.

[12] M. Ramaswami and R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining", *Journal of Computing*, Vol. 1, No. 1, pp. 7-11, 2009.

[13] R. Kohavi and G.H. John, "Wrappers for Feature Subset Selection", *Artificial Intelligence*, Vol. 97, No. 1-2, pp. 273–324, 1997.

[14] Samina Khalid, Tehmina Khalil and Sharmila Nasreen, "A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning", *In Proceedings of*

Conference on Science and Information, pp. 372-378, 2014.

[15] Zheng Zhao and Huan Liu "Searching for Interacting Features" *In proceedings of 20th International joint Conference on Artificial Intelligence*, pp. 1156-1161, 2007.

[16] K.Kira and L.A Rendell, "The Feature Selection Problem: Traditional Methods and A New Algorithm," *In proceedings of 10th National Conference on Artificial Intelligence*, pp.129-134, 1992.

[17] Qinbao Song, Jingjie Ni and Guangtao Wang, "A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No.1, pp. 1-14, 2013

[18] Mark A. Hall and Lloyd A. Smith, "Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper", *In proceedings of the Twelfth International Conference on Florida Artificial Intelligence Research Society*, Vol. 235, pp. 235-239, 1999.

[19] Wang Liping, "Feature Selection Algorithm Based On Conditional Dynamic Mutual Information", *International Journal on Smart Sensing and Intelligent Systems*", Vol. 8, No. 1, pp. 316-337, 2015.

[20] Kexin Zhu and Jian Yang, "A Cluster-Based Sequential Feature Selection Algorithm", *In proceedings of 9th International Conference on Natural Computation*, pp. 848-852, 2013.

[21] Y.Kim, W.Street, and F.Menczer, "Feature Selection for Unsupervised Learning Via Evolutionary Search," *In proceedings of 6th International Conference on Knowledge Discovery and Data Mining*, pp 365-369, 2000.

[22] B.M Vidhyavathi, "A New Approach to Feature Selection for Data Mining", *International Journal of Computational Intelligence Research*, Vol.7, No. 3, pp. 263 – 269, 2011.

[23] Jihong Liu, "A Hybrid Feature Selection Algorithm for Data sets of thousands of Variables" *In proceedings on 2nd International Conference on Advanced Computer Control*, pp. 288-291, 2010.

[24] Prashant K. Aher, Swapnil D. Daphal and Alice N. Cheeran, "Analysis of Feature Extraction Techniques for Improved Emotion Recognition in Presence of Additive Noise", *In proceedings of International Conference on Computation System and Information Technology for Sustainable Solutions*, pp. 350-354, 2016.

[25] Ali Sophian, Gui Yun Tian, David Taylor and John Rudlin, "A feature Extraction Technique Based on Principal Component Analysis for Pulsed Eddy Current NDT ", *NDT & E International*, Vol.36 , No. 1 , pp. 37-41, 2003.