Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.9, No.1, 2018 www.iiste.org

A Review on Dimension Reduction Techniques in Data Mining

Omprakash Saini*
PG Scholar, CSE, VITS, Bhopal, India

Prof. Sumit Sharma
HOD CSE, VITS, Bhopal, India

Abstract
Real world data is high-dimensional like images, speech signals containing multiple dimensions to represent data. Higher dimensional data are more complex for detecting and exploiting the relationships among terms. Dimensionality reduction is a technique used for reducing complexity for analyzing high dimensional data. There are many methodologies that are being used to find the Critical Dimensions for a dataset that significantly reduces the number of dimensions. They reduce the dimensions from the original input data. Dimensionality reduction methods can be of two types as feature extractions and feature selection techniques. Feature Extraction is a distinct form of Dimensionality Reduction to extract some important feature from input dataset. Two different approaches available for dimensionality reduction are supervised approach and unsupervised approach. One exclusive purpose of this survey is to provide an adequate comprehension of the different dimensionality reduction techniques that exist currently and also to introduce the applicability of any one of the prescribed methods that depends upon the given set of parameters and varying conditions. This paper surveys the schemes that are majorly used for Dimensionality Reduction mainly high dimension datasets. A comparative analysis of surveyed methodologies is also done, based on which, best methodology for a certain type of dataset can be chosen.
Keywords: Data Mining, Dimensionality Reduction, Clustering, feature selection; curse of dimensionality; critical dimension

I. INTRODUCTION

Data mining is task of extracting data from larger dataset. Main objective of data mining is to collect, process, classified and grab (select) useful data from available dataset. Real world data like speech signals, images are high dimensional that is hard to analyze. Various techniques of high dimensional data clustering find many applications in real time applications. Former clustering algorithms are biased while it is applied with high-dimensional datasets. Main objective of dimensionality reduction is to search for small set of important feature that describes larger dataset. Many methods use global dimensionality reduction techniques that over fits the problems of higher dimensions. The solution for higher dimensional data is to reduce dimensions of data before applying clustering techniques. Dimensionality reduction is to solve two different kinds of problems. The one is used in extracting a feature vector from an original object, and the other is to reduce the dimensionality of a high-dimensional feature vector already extracted [1]. It will reduce input dimensions as well as data from dataset. Feature extraction is one the special form of dimensionality reduction, used for extracting important features subset from available dataset. When analyzing larger dataset with higher dimensions, it is necessary to transform into smaller and more manageable dataset [2].

Different approaches have been used for dimensionality reduction. They are supervised approaches and unsupervised approaches [3]. When some discriminate analysis uses class information called as supervised approach while some analysis do not use class information referred to as unsupervised approaches. Some supervised approach methods are like LDA (Linear Discriminate Analysis), NN (Neural Network) used for dimensionality reduction. Clustering is an unsupervised approach that do not use label information. Unsupervised methods like PCA (Principal Component Analysis), ICA (Independent Component Analysis), SVD (Single Value Decomposition), KPCA (Kernel Principal Component Analysis), Fourier analysis (FA) etc. are used for reducing dimensions of data set. In this review paper, all these techniques for dimensionality reduction of data set has been discussed and compared. This study will be useful to select method for particular application.

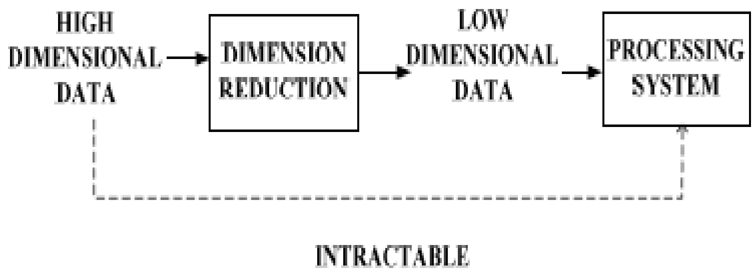

Fig. 1 Process of Dimensionality Reduction

In order to achieve accuracy in classification of such data, we require identifying and removing irrelevant and insignificant dimensions. The process of reducing dimensions is referred as Dimensionality Reduction. It is a crucial pre-processing step in Data Mining to improve computational efficiency and accuracy. Dimensionality reduction provides benefits such as improved dataset classification accuracy, increased computational efficiency and better visualization of dimensions.

The survey is focused on studying the methodologies that are used to reduce dimensions in a dataset without compromising on classification accuracy. It has been observed that the methodologies that have been adopted for reducing features are dataset specific, i.e., in each surveyed paper, the reduction scheme is applied to a specific set of datasets but not generalized.

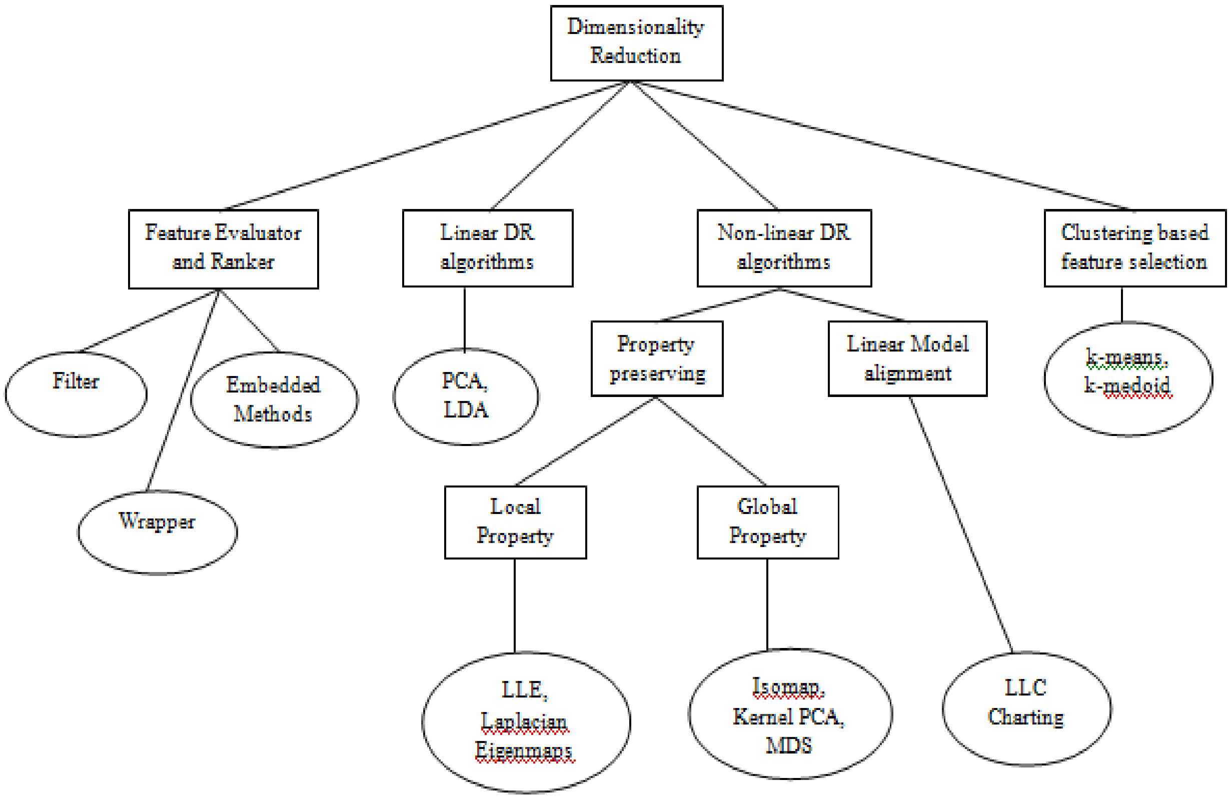

Fig.2. Categorization of dimensionality reduction techniques

The algorithms that are used for feature reduction are categorized as Feature Ranker, Feature Evaluator, Dimensionality Reduction and Clustering Algorithms. Each of these types of algorithms has their own advantages and disadvantages. Depending upon the combinations of algorithms used, mainly 4 types of methodologies has been identified. Fig. 2 shows the broad categorization of various types of methodologies that are used for dimensionality reduction.

II. LITERATURE SURVEY

The field of Data Mining has a different behavior towards Big Data. It can deal with data-sets having size gigabytes or even TBs. The main concern over here is that the algorithms which are used in data mining operations work on small data sets and do not give better results on large data sets. To work efficiently with large data sets, the algorithms must have high scalability. Clustering high dimensional data has always been a challenge for clustering techniques. Clustering is unsupervised classification of patterns (observations, data items, or feature vectors) into teams (clusters). The drawbacks of clustering have been addressed in several contexts by researchers in several disciplines and so reflect its broad charm and quality in concert of the steps in exploratory data analysis. Clustering is useful in several exploratory pattern analysis, grouping, decision making and machine learning situations including data mining, document retrieval, image segmentation and pattern classification. Adil Fahad, et al [18] performed a survey on clustering algorithms for Big Data. They have categorized 24 Clustering Algorithms as Partition-based, Hierarchical-based, Density based, Grid-based and Model-based. Depending on the size of datasets, handling capacity of noisy data and types of datasets, Clusters are formed and the complexities of algorithms are calculated. They concluded that no clustering algorithm performs well for all the evaluation criteria. The entire clustering algorithm suffers from Stability problem. MacQueen [19] defined a technique for partitioning N-dimensional population into k-sets, which they named as K-means. They successfully concluded that k-means is computationally feasible and economical and has been a successful implementation for differentiating the data within a class. S. Nazim presented a comparative review of dimensionality reduction techniques in regard with information visualization. The survey analyzed some DR methods supporting the concept of dimensionality reduction for getting the visualization of information with minimum loss of original information. As we deal with Big Data. The issue of stability of clusters comes into picture. The theories [20] state that k-means does not break down even for arbitrarily large samples of data. The focus is on the behavior of stability of clusters formed by k-means algorithm-Means is closely related to principal component analysis [21]. The outcomes subject with regard to effectiveness of the solution obtained from k-means. Unsupervised dimensionality reduction and unsupervised learning are associated closely [22]. The result provides new perception towards the observed quality of output obtained by PCA-based data reduction.

Bara'a Ali Attea et al. [23] discovered that performance of clustering algorithms degrades with more and more overlaps among clusters in a data set. These facts have motivated to develop a fuzzy multi-objective particle swarm optimization framework (FMOPSO) in an innovative fashion for data clustering, which is able to deliver more effective results than state-of-the-art clustering algorithms. To ascertain the superiority of the proposed algorithm, number of statistical tests has been carried out on a variety of numerical and categorical real life data sets.

Suresh Chandra Satapathy et al. [24] introduced an idea of an algorithm that can combine dimensionality reduction techniques of weighted PCs with AUTO-PSO for clustering. The intention behind it was to reduce complexity of data sets and speed up the Auto PSO clustering process. A significant improvement in total runtime has been achieved. Moreover, the clustering accuracy of the dimensionality reduction technique i.e. AUTO-PSO clustering algorithm is comparable to the one that uses full dimension space.

Li-Yeh Chuang et al. [25] invented an improved particle swarm optimization based on Gauss chaotic map for clustering. Gauss chaotic map provides the significant chaos distribution to balance the exploration and exploitation capability for search process. This easy and fast function generates a random seed processes, and further improve the performance of PSO due to their unpredictability. In the experimental analysis, the eight different clustering algorithms were compared on six test data sets. The results indicated that the performance of the proposed one is appreciably better than the performance of other existing algorithms.

Xiaohui Cui et al. [26] presented a Particle Swarm Optimization (PSO) document clustering algorithm. Unlike, localized searching of the K-Means algorithm, PSO clustering algorithm used to perform a globalized search in the entire solution space. In the experiments conducted, they have applied the K-Means PSO, and hybrid PSO clustering algorithm on four different text document data sets. From the comparative results, the hybrid PSO algorithm can generate more compact clustering results than the K-Means algorithm

II. MAJOR CHALLENGES AND ISSUES IN DIMENSIONALITY REDUCTION

As the dimensionality of dataset increases, the volume of the space increases so fast that the available data become sparse. Generally, this data is not distributed uniformly over the search space i.e., usually a larger percentage of the training data resides in the corners of the feature space which is more difficult to classify than that near the centre. In order to obtain a statistically sound and reliable result, the amount of training data needed to support the result often grows exponentially with the dimensionality. Hence, high dimensionality leads to a problem known as "Curse of Dimensionality" that specifically makes it difficult to perform classification on a dataset having a large number of dimensions [4].

Dimensionality reduction is used for downsizing input data that is more relevant for further analysis. Reduced dataset preserves much of variance from larger dataset and without any loss of important features. It will also become easy to detect and use from real world data. Comparison of FA, PCA and wavelet analysis are applied to stable combustion and It was found that frequent characteristic were of similar qualitatively and quantitatively [5]. But when unstable and transient combustion is applied, FA and PCA are not feasible, only

Wavelet analysis is capable of revealing dynamic frequency component. FA and PCA are having limitation as they are linear methods and work efficiently with steady and structured phenomena. PCA and 2DPCA requires more computational time and memory usage because it requires whole training data to extract vectors. Also it is linear method, so it is not able to represent the non-linear data effectively and efficiently [6]. So in this paper a new online non-linear method is adopted that can easily handle non-linear data by applying kernel method to PCA. Kernel method is having disadvantages that it takes entire data set with heavy computational load and it needs to refer to previously acquired data to update Eigen vectors. So, I2DKPCA is adopted to solve the problem. PCA is used for dimensionality reduction, based on synchronized covariance which is not always effective for some cases [7][8]. This paper proposes asynchronous methods for dimensionality reduction. A new semi-supervised approach is developed that combines supervised (LDA) and unsupervised learning (K-means) approach with dimensionality reduction to improve K-means clustering performance so it can perform well in new space. Some of issues available are [9]:

- PCA is a most widely used linear dimensionality reduction technique. As it is a linear method it can work with linear data only and not work with real data efficiently because of complexity and high-dimensionality. PCA works with structured and steady dataset [5].
- ICA is an unsupervised dimensionality reduction technique. ICA has high computational complexity that relates to data independence measure [9].
- LDA is having issue that lack of the sample data per class does degrade the classification performance as significantly due to the generalization of decision for arbitrary data with noise regularization. Robustness improvement is pursued as the other critical issue in LDA for better classification performance in noisy environment [10].

III. EXISTING APPROACHES

Many research areas like science, engineering, astronomy, biology, economics, sensing networks etc. are having greater amount of information available from various observation and experiments. Mining of data means extracting data from this information. This real world data is with larger dimensions and complex to analyze. So, for better analyzing dimensions need to be reduce. Various available methods for dimensionality reduction are described here:

A. Linear Discriminant Analysis

LDA is a widely used technique for dimensionality reduction. During some medical dataset experiments, growth of amount in existing cases is more where dimensions are greater or less features and occurrence of features are significantly larger than the size of sample [10]. Number of independent features, relative to which the data is described, LDA creates linear combination of these which yields the larger mean differences between described classes. Mathematically two measures have been defined for all class samples [10]:

- Within-class scatter matrix
- Between-class scatter matrix

Main goal of LDA is to maximize the between-class measure while minimize within-class measure [11]. To do this is to maximize the ratio det |sb| / det |sw|. Advantage of this is that it has been proven that if Sw is a non-singular matrix then this ratio is maximized when the column vector of projection matrix w, are the Eigen vector of Sw-1Sb

B. Neural Network

Neural Network is a method that uses supervised approach for analysis. NN is a supervised dimensionality reduction method used for study of human brain neurons. A neuron in the brain receives input from other neurons through its dendrites [12].

- The perceptron receives several input values (x0 – xn).
- Each input connection has weight (w0 – wn) in the range of 0 – 1.
- The threshold unit then sums the input, and if the sum exceeds that threshold value, a signal is send to output, otherwise no signal is sent.

Neural network learning is also referred to as connectionist learning because of input/output connection between them. The simplest neuron network contains three layers: the input layer, one hidden layer and one output layer.

For each net input unit computation, each input connected to the unit is multiplied by its corresponding weight w, and this is summed.

Given a unit *j* in a hidden or output layer, the net input, *Ij*, to unit *j* is

$$Ij = \sum_{I} w_{ij} O_j + \Theta_j \quad (3)$$

Where wij = weight of connection from unit I in the previous layer to unit j,
Oj = i^{th} output from previous layer
Ѳj = bias of unit Training time require for Neural network

Ɵj is larger so, it only useful for some application where it is feasible [13]. They have been condemned for their poor interpretability. Advantages of neuron network include higher tolerance of noisy data as well as their ability to classify some patterns on which they have not been trained. Neuron network adopts parallelization technique for speeding up the computational process.

C. Principal Component Analysis

PCA is most widely used linear method for dimensionality reduction. The PCA is a statistical data analysis method that transforms the initial set of input variables into various set of linear combinations, referred as the principal components (PC). This PC contains specific properties with respect to variances. This reduces the dimensionality of the system while retaining information on the variable connections [15]. Steps that PCA performs are [14][15]:

- The given high-dimensional input data are normalized as each attribute falls within same range. This is to ensure that all attributes with larger domains will not dominate attributes with smaller domain.
- PCA calculate k orthogonal vector which provides a basis for normalized input. The input data are linear combination of PC.
- This principal component is sorted in decreasing order of their strength or significance.
- Because of this sorting of principal component, the size of data can be reduced by excluding weaker component meaning that PCs with lower variance. It should be possible to reconstruct a better approximation of original input data by using these strongest principal components. PCA is used in various domains like face recognition, Image Compression, Microarray Gene Expression, coin classification, seismic series analysis. Main problem with PCA is that it is not efficiently represent larger and nonlinear distribution of input data.

D. Kernel Principal Component Analysis

In order to solve the problem of non-linearity, various different approaches along with kernel functions have also been studied as extensions to the PCA. To convert the nonlinear distribution of input data to linear distribution, kernel PCA maps the samples into high-dimensional kernel space before conducting PCA [6]. Basic principle of KPCA is to transform original input vectors to a high dimensional feature space F with a nonlinear function and then to calculate the linear PCA in feature space [13]. Given a set of input vectors x1, x2…..,xm ϵ Rn and then the covariance matrix in F is given by,

$$C_F = 1/m \sum i=1m\ \phi(xi)\phi(xi)T \qquad (4)$$

Some important drawbacks of Kernel PCA are that the kernel matrix size is proportional to the square of the number of instances in the input dataset. Also, Kernel PCA mainly focuses on retaining large pair wise distances. Kernel PCA computes principal eigen vector for kernel matrix rather than covariance matrix. The kernel matrix K having the data points xi. The entries in the kernel matrix are defined by kij = k(xi,xj),where k is kernel function. There is direct reformulation of PCA in kernel space as kernel matrix is most similar to data point production in high-dimensional space which is created using kernel function. Kernel PCA has been applied successfully to different domains like face recognition, speech recognition, novelty detection etc.

E. Independent Component Analysis

ICA is a computational technique mainly used for splitting an assorted signal into its reduced subcomponents. Considering the absence of time delays, usually this kind of difficulty is interpreted. If there are N sources present, at least N estimations are required to mine the original signals like microphones [13]. The ICA algorithm is capable of employing higher order statistics which may contain some essential complementary data unlike PCA, which only analyses covariance. Most ICA algorithms require an identity covariance matrix; this is more than a statistical criterion.

Definition: ICA of random vector x consist of finding linear transformation s = Wx so that the component si are a independent as possible, in the sense of maximizing some function F (s1,..,sm) that measures independence.

ICA can also be used as an augmented version of the PCA based method. Dimension was reduced to a smaller set of independent sources or latent variables, which then can be used further discriminant analysis by using regularized whitening technique. The components of the collaborating matrix can themselves be examined to gather more thoughts from a larger perspective, about the genetic keystones of the procedure that generated the data required for further processing [16]. ICA is used for discovering the projection where all the expected components are “the most independent”. ICA checks all the modules in parallel and forecasts the directions where all the projected components are most independent to each other in the sense of an independent degree.

F. Single Value Decomposition

Singular Value Decomposition is gene selection procedure for reducing dimensionality. SVD is method that uses matrix factorization and that comes under linear vector algebra. Main purpose of applying Singular value decomposition is identification and structural constitution extraction inside the data and also involves gene expression that relates to some important associations. Main goal of singular value decomposition is to compute Eigen values and Eigen vectors of covariance matrix from sample gene matrix. To infer variation in corresponding Eigen vector, these Eigen values or singular values are used. Initially appearing eigenvectors are chosen as principal component that illustrate higher unpredictability. By removing some feature element, information may lose while original input set is preserved. In order to obtain feature genes some loss of information is utilized [17].

One of the advantages of using single value decomposition is that an algorithm can be easily computed. Also used in kinematics and dynamics of robot manipulators. But it is not having ability of handling larger dataset. SVD of a set A of N k-dimensional vector is

$$A=U \sum V_T \tag{5}$$

Where, A = N×k data matrix composed of the N k dimensional vectors,
U = N×k orthonormal matrix,
∑ = k×k diagonal matrix of Eigen values,
V = k×k orthonormal basis matrix.

SVD can solve problem of gene expression data in parameter estimation. When the size of cluster is small or higher dimension data it becomes challenging task. Singular value decomposition is applied to dataset then probit transformation performed on that result. From these results we can conclude that SVD transformation can be applied on both with or without scattered genes and becomes beneficial. All these techniques discussed here are compared in Table I with some parameters.

TABLE 1: Comparison of different Dimensionality Reduction Methods

PARAMETERS	ICA	KPCA	LDA	NN	PCA	SVD
Data pre-processing	Not required	For large data set required	Not required	Required	Not required	Required
Dataset type	Multivariate Signals	Eigen values	-	Numeric	Eigen values	Multivariate data, gene expression data
Fault Tolerance	Sensitive to Fault	More sensitive to fault compare to PCA because of non-linear behavior	Less sensitive	Sensitive to fault	Less Sensitive due to Linear Nature	Less sensitive
Important parameters	Statistics Transformation S=Wx	Orthogonal Linear Transformation , Feature Space	-	Weighted connection	Orthogonal Linear Transformation	Singular values
Large data set handling ability	Good	Moderate	Good	Medium	Good	Not good
Multi-dimensional data set ability	Good	Very good	Good	Not good	Good	Not good
Overfitting	Problem for high dimension data with insufficient sample size	Problem for large data set	Happens when trailing set is small or set is with insignificant PCA dimensions	Not feasible for Longer training time and having poor interpretability	Problem for large data set	Only for singular values
Required Parameters	Source Matrix, weight matrix (w)	Loading vector (w), data matrix with column wise zero empirical mean, covariance matrix of data set	Sb = Between class matrices, Sw =with class scatter matrices	Input/ Output units each having weights associated with	Loading vector (w), data matrix with column wise zero empirical mean, covariance matrix of data set	Dataset values
Training	Not required	Most required	Not required	Required	Required	Not required
Training time	Slightly moderate compare to other model	Very High	Less than PCA	High	High	Moderate

IV. CONCLUSION

The sole purpose of this paper is to provide information on different dimensionality reduction techniques. Ultimate goal of performing dimensionality reduction is to improve the accuracy and efficiency. This is to improve the throughput, for minimizing error rate, and for decreasing complexity of computational work. This study work will be useful to select the method for particular application based on characteristics of dataset and method i.e. Data type, Training time, Large dataset handling capability, Multi-dimensional dataset, Important parameters, Fault tolerance etc. This work gives clear idea of comparison of all available dimensional reduction techniques. It is concluded that in order to select a scheme to reduce dimensionality, we should consider the type of dataset and specific requirement of machine learning algorithm. Table I can be referred for this purpose. A combination of schemes may also be used to overcome the disadvantages of one scheme over another.

References

[1.] Holmes, Michael, Alexander Gray, and Charles Isbell. "Fast SVD for large-scale matrices." In Workshop on Efficient Machine Learning at NIPS, vol. 58, pp. 249-252. 2007.

[2.] L. Cao, K. Chua, W. Chong, H. Lee, and Q. Gu, "A comparison of pca, fKPCAg and fICAg for dimensionality reduction in support vector machine," Neurocomputing, vol. 55, no. 12, pp. 321 – 336, 2003

[3.] Liu, Chien-Liang, Wen-Hoar Hsaio, Chia-Hoang Lee, and Fu-Sheng Gou. "Semi-supervised linear discriminant clustering." IEEE Transactions on cybernetics 44, no. 7 (2014): 989-1000.

[4.] Aggarwal, Charu C., Alexander Hinneburg, and Daniel A. Keim. "On the surprising behavior of distance metrics in high dimensional spaces." In ICDT, vol. 1, pp. 420-434. 2001.

[5.] Choi, Yonghwa, Seiichi Ozawa, and Minho Lee. "Incremental two-dimensional kernel principal component analysis." Neurocomputing 134 (2014): 280-288.

[6.] De Lathauwer, L., B. De Moor, J. Vandewalle, and Blind Source Separation by Higher-Order. "Singular value decomposition." In Proc. EUSIPCO-94, Edinburgh, Scotland, UK, vol. 1, pp. 175-178. 1994.

[7.] Delac, Kresimir, Mislav Grgic, and Sonja Grgic. "Independent comparative study of PCA, ICA, and LDA on the FERET data set." International Journal of Imaging Systems and Technology 15, no. 5 (2005): 252-260.

[8.] Fodor, Imola K. A survey of dimension reduction techniques. No. UCRL-ID-148494. Lawrence Livermore National Lab., CA (US), 2002.

[9.] Bashiri, M., and A. Farshbaf Geranmayeh. "Tuning the parameters of an artificial neural network using central composite design and genetic algorithm." Scientia Iranica 18, no. 6 (2011): 1600-1608.

[10.] Zhu, Lei, Bin Han, Lihua Li, Shenhua Xu, Hanzhou Mou, and Zhiguo Zheng. "Null space LDA based feature extraction of mass spectrometry data for cancer classification." In Biomedical Engineering and Informatics, 2009. BMEI'09. 2nd International Conference on, pp. 1-4. IEEE, 2009.

[11.] Van Der Maaten, Laurens, Eric Postma, and Jaap Van den Herik. "Dimensionality reduction: a comparative." J Mach Learn Res 10 (2009): 66-71.

[12.] Raducanu, Bogdan, and Fadi Dornaika. "A supervised non-linear dimensionality reduction approach for manifold learning." Pattern Recognition 45, no. 6 (2012): 2432-2444.

[13.] Jeong, Seungdo, Sang-Wook Kim, Kidong Kim, and Byung-Uk Choi. "An effective method for approximating the euclidean distance in high-dimensional space." In DEXA, pp. 863-872. 2006.

[14.] Tsai, Flora S., and Kap Luk Chan. "Dimensionality reduction techniques for data exploration." In Information, Communications & Signal Processing, 2007 6th International Conference on, pp. 1-5. IEEE, 2007.

[15.] Bicciato, S., A. Luchini, and C. Di Bello. "Disjoint PCA models for marker identification and classification of cancer types using gene expression data." In Molecular, Cellular and Tissue Engineering, 2002. Proceedings of the IEEE-EMBS Special Topic Conference on, pp. 98-99. IEEE, 2002.

[16.] Li, Hailin. "Asynchronism-based principal component analysis for time series data mining." Expert systems with applications 41, no. 6 (2014): 2842-2850.

[17.] Varghese, Nebu, Vinay Verghese, P. Gayathri, and N. Jaisankar. "A survey of dimensionality reduction and classification methods." International Journal of Computer Science and Engineering Survey 3, no. 3 (2012): 45.

[18.] Fahad, Adil, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil, Albert Y. Zomaya, Sebti Foufou, and Abdelaziz Bouras. "A survey of clustering algorithms for big data: Taxonomy and empirical analysis." IEEE transactions on emerging topics in computing 2, no. 3 (2014): 267-279.

[19.] MacQueen, James. "Some methods for classification and analysis of multivariate observations." In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, no. 14, pp. 281-297. 1997.

[20.] Shamir, Ohad, and Naftali Tishby. "Model Selection and Stability in k-means Clustering." In COLT, pp. 367-378. 2008.

[21.] Ding, Chris, and Xiaofeng He. "Principal component analysis and effective k-means clustering." In Proceedings of the 2004 SIAM International Conference on Data Mining, pp. 497-501. Society for Industrial and Applied Mathematics, 2004.

[22.] Ding, Chris, and Xiaofeng He. "K-means clustering via principal component analysis." In Proceedings of the twenty-first international conference on Machine learning, p. 29. ACM, 2004.

[23.] Attea, Bara'A. Ali. "A fuzzy multi-objective particle swarm optimization for effective data clustering." Memetic Computing2, no. 4 (2010): 305-312.

[24.] Satapathy, Suresh Chandra, and Anima Naik. "Efficient clustering of dataset based on particle swarm optimization." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) ISSN (2013): 2249-6831.
[25.] Chuang, Li-Yeh, Yu-Da Lin, and Cheng-Hong Yang. "An improved particle swarm optimization for data clustering." In Proceedings of the International MultiConference of Engineers &Computer Scientist 2012 I, IMECS. 2012.
[26.] Cui, Xiaohui, and Thomas E. Potok. "Document clustering analysis based on hybrid PSO+ K-means algorithm." Journal of Computer Sciences (special issue) 27 (2005): 33.