A SURVEY OF DIMENSIONALITY REDUCTION AND CLASSIFICATION METHODS

Nebu Varghese[1], Vinay Verghese[2], Prof. Gayathri. P[3] and Dr. N. Jaisankar[4]

[1]B.Tech Student, School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu, India
nebugetsmail_1@hotmail.com

[2]B.Tech Student, School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu, India
vin070@gmail.com

[3]Assistant Professor, School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu, India
pgayathri@vit.ac.in

[4]Professor, School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu, India
njaisankar@vit.ac.in

ABSTRACT

Dimensionality Reduction is usually achieved on the feature space by adopting any one of the prescribed methods that fall under the selected technique. Feature selection and Feature extraction being the two aforesaid techniques of reducing dimensionality, the former discards certain features that may be useful at a later stage whereas the latter re-constructs its features into a simpler dimension thereby preserving all its initial characteristics. The sole 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 depending upon the given set of parameters and varying conditions as described, under each algorithm's usage statistics. This paper also presents guidelines where in, selection of the best possible algorithm for a specific instance can be determined with ease when a condition arises where in two or more algorithms may be suitable for executing the aforementioned task.

KEYWORDS

Feature selection, Feature extraction & Dimensionality Reduction

1. INTRODUCTION

The objective of data mining is to process, classify and select useful data from a given dataset. The information gathered for data mining procedures may consist of erroneous data where in certain quantities may display missing values or even may convey inappropriate information that misleads the user because of the alterations that have been made due to irrelevant inputs. Data pre-processing is an approach where these irregularities are detected, analyzed and rectified by applying certain behavioral predictions and modifications to alter the inapplicability present in the data set. Dimensionality reduction being one of the predominant techniques of data pre-processing is necessitated in order to incorporate a systematized structure into the information set, before it is being sent into the mining processor. It is also an efficacious approach towards downsizing of data which can abet in effective storage and retrieval of data. Some of its applications include its efficient usage in the areas of text mining, pattern matching, image processing, micro-array data analysis and image retrieval. Dimensionality reduction can

be implemented using various algorithms like SVD, SVM, ICA, PCA etc. A descriptive analysis of eight algorithms used for this process has been provided and some are being compared with one another to judge its strengths or weaknesses with respect to its equivalent. Data pre-processing also acts as one of the pivotal procedures in various machine learning projects.

2. ALGORITHMS

2.1. Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) is a gene selection procedure performed to decrease dimensionality of data. SVD is a matrix factorization method, and comes under linear vector algebra. In data analysis applied on gene expressions, the primary objectives of applying SVD are identification and extraction of the structural constitution within the data and also relating to significant associations involving gene expressions [1].

The key method of SVD is to calculate the eigenvalues and eigenvectors of the covariance matrix from the complete sample-gene matrix. The eigenvalues or singular values will help deduce variation in the corresponding eigenvectors; If the singular value is greater, the respective eigenvector will contain a higher variability. Generally some of the initially appearing eigenvectors that illustrate higher unpredictability are chosen as prime Principal Components (PCs) in order to reduce the data into a smaller quantity of dimensions. Since the remaining feature elements are removed, information is lost while the original data is recovered. This loss of data is utilized to obtain the feature genes [2].

A tool known as clustering is generally used to analyse the data. Groups of genes that show comparable expression profiles are discovered by Clustering. There are two categories that come under the well-known clustering methods and are known as the Model-free and Model-based methods [3]. In the model-free clustering methods, there is no probabilistic model prescribed for the data, and clusters are prepared either by iteratively dividing genes to form a tree or optimizing a certain target function. The model-based clustering methods build clusters assuming that the data follows a mixture dependent distribution. Among the model based methods, the mixture-Gaussian distribution method is comparably simpler in terms of computation. The SVD method is applied hereon.

In gene expression data, SVD is known to overcome two difficulties. The primary one is about parameter estimation, which becomes challenging when the size of a cluster is small or dimension of the data, is high. The secondary reason depends upon the normality assumption on gene expression levels, which is rarely satisfied by real data. First SVD is applied on the data, and then a technique known as probit transformation is done. Certain results have shown that the transformation using SVD can normally be applied on both types of data sets, with or without scattered genes and can be found useful. And after application of probit transformation, there is an improvement in the working of the model-free clustering methods seen in the data sets which contain scattered genes.

2.2. Principal Components Analysis (PCA)

The PCA is a statistical data analysis method that transforms the initial set of variables into an assorted set of linear combinations, known as the principal components (PC), with specific properties with respect to variances. This condenses the dimensionality of the system while maintaining information on the variable connections [4]. The analysis is done on a data set by calculating and analysing the data covariance matrix, its eigenvalues along with its respective eigenvectors systematized in descending order. Dimensionality Reduction is a process used in Data Mining where the numbers of random variables under consideration are reduced.

Dimensionality Reduction is broadly categorized as Feature Selection where a subgroup of all the features is selected and Feature Extraction where the existing features are combined and a

new subset of the combinations is created. Principal Components Analysis (PCA) is one of the common techniques used under Feature Extraction. PCA uses a signal based representation criterion where the purpose of feature extraction is to represent the samples accurately in a lower dimensional space whereas the alternate technique, Linear Discriminant Analysis (LDA) deploys a classification based approach. PCA performs dimensionality reduction whilst maintaining maximum feasible arbitrariness in the high-dimensional space. It can be seen as a data visualization method since high dimensional datasets can be condensed to a lower dimension (2D OR 3D) and then plotted using graphs or visualized using charts.

In a case of dimensionality reduction of two highly dimensional picture databases, Simple Principal Components Analysis (SPCA) underwent testing and assessment with other techniques. SPCA depicted a quick convergence rate and stability to the reorganization of the samples compared with other methods [5].

Simple PCA (SPCA) method produced estimated solutions without the requirement of calculating a variance-covariance matrix followed by its diagonalization. It is independent of training parameters like neural networks. Predominantly, for highly dimensional datasets, SPCA is much quicker than other existent procedures.

The principal methods used for PCA are the Matrix method and the Data method. In the Matrix method, all of the data contained in the datasets are employed to calculate the variance-covariance structure and express it in the form of a matrix. The matrix is further disintegrated and a diagonalization technique is applied. Data methods on the other hand, work directly with the data. In SPCA, the data oriented approach is taken so there is no issue of computing the matrix and also, no learning parameters are required.

PCA is also used to lower the dimensionality of the data set before clustering happens. Using PCA before cluster analysis could abet in an improved drawing out of the cluster organization in the data set. Given that Principal Components (PCs) are ordered yet not varying together and, the first few PCs that consist of maximum variations in the data are typically used in cluster analysis. PCA is used to examine how well the genes that were extracted depict the normal variance between the data, thereby getting rid of the variance due to any other sources to the highest potential degree. Finally, PCA can furthermore be used as a platform to contrast between the performances of various different methods used to establish normal variance [6].

The major drawback observed in PCA is that it gives no consideration to class separability because it does not account for the class label of the feature vector [7]. PCA just performs a coordinate rotation that aligns the coordinate axes transformed earlier, along the directions of maximum variance. There is no assurance that the directions of maximum variance will comprise of features worthy enough for discrimination.

2.3. Support Vector Machines (SVM)

SVM is a recently developed technique used for classification suggested by Vapnik, which was consecutively applied to several domains. SVM is applied to microarray cancer data which comprises of several gene expressions. SVM is applied after many steps after analysis to finally classify cancer tissues as part of an integrated algorithm. The main aim of SVM is to obtain an optimal separating hyper-plane with the maximum margin (w) and a real value for organization of data. Practically, the data may not be linearly separable, and few deviations on classification are permitted on the learning data sets [8].

SVM method is one of the most imminent tools used for analysis of gene expression data from DNA microarrays. This is mainly possible since the SVM method is specifically appropriate for certain specific data sets where the number of samples is much smaller in comparison to the number of features (genes). Moreover, the SVM Technique can also be used for gene selection.

The advantages of using the SVM model are its Optimality and it results in a distinctive solution. SVM used along with a form of cross-validation, which gives unique assessment of generalization capability, formulates a procedure of features (genes) selection known as recursive feature replacement (RFR) [1].

SVM and SVD are basically used to analyse gene expression data even though they are associated with a separate category of computational techniques: the SVM technique is an illustration of a supervised method while the SVD technique belongs to a class of unsupervised methods. The scope of application of these methods for various functions namely: clustering, classification, feature selection and modelling of dynamics of gene expression is considered and examined. By using the weights of a classifier to produce a feature ranking, the SVM technique is presently one of the best-known classification methods with computational advantages over the other existing ones [9].

A gene selection algorithm was proposed by Guyon et al employing an SVM method that deploys a Recursive Feature Elimination (RFE). The SVM-RFE method, starts with the gene set consisting of the complete gene set, and eliminates iteratively the gene that is least significant to the classifier from the gene set. Gene ranking score is selected as a criterion to measure the significance of the gene for classification and it is defined by the sum square of the weight vector of the SVMs, and is calculated separately [10]. The gene expression profiles between early and late stage primary CRCs (colorectal cancer) were compared, as well as the gene expression profiles between metastatic CRCs and late stage primary CRCs. Highest classification accuracy (100%) was achieved with smaller gene subsets, by using the SVM-T-RFE method. Results proved to illustrate an improvement than that of previous studies.

2.4. Independent Component Analysis (ICA)

Independent component analysis (ICA) is a computational technique used for splitting an assorted signal into its reduced subcomponents. A simple practice of ICA is the "cocktail party problem", wherein the fundamental speech signals are divided from a sample data comprising of individuals conversing together within a room. Usually this predicament is interpreted by considering the absence of time delays or echoes. An imperative note that is to be taken into consideration is that if N sources are present, at least N estimations (e.g. microphones) are required to mine the primal signals.

This methodology can be approached with an alternative technique known as the Regularized Whitening method to decrease the dimension to a limited set of independent sources or dormant variables, which later can be used in downstream analysis. Unlike PCA, which only analyses covariance, the ICA algorithm is capable of employing higher order statistics which may contain substantial complementary data. This is predominantly valued when the dispatch of data differs significantly from its prevalent parameters, which for at least two reasons is typically the situation when dealt with microarray expression data. Most ICA algorithms require "whitened" data, by means of an identity covariance matrix. This is more than a statistical prerequisite, where in the algorithm tries to disintegrate the data beyond its first two moments which is the ultimate goal of PCA and its associated methods [11].

ICA can also be used as an augmented version of the PCA based method. With the regularized whitening technique, dimension was reduced to a smaller set of independent sources or latent variables, which then can be used in subsequent discriminant analysis. 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 [11].

A different method is to acquire the benefits of Mother Nature that are arbitrarily brought about through the usage of random initial projection vectors in the ICA. As an outcome to these procedural parameters, three algorithms were devised for the ICA-DR (Independent Component

Analysis – Dimensionality Reduction) procedure. The first being ICA-DR1, is developed by using the VD (virtual dimensionality) parameter along with a principle for component prioritization and selection. The second one, called ICA-DR2 executes the ICA as an unsystematic algorithm with randomness characterized by random initial projection vectors. As a result, the ICA-DR2 spontaneously determines a desired set of ICs for DR without consideration of criteria. The last algorithm is called ICA-DR3 which incorporates the use of a characteristically designed initialization algorithm in association with the VD to produce an applicable group of initial projection vectors to substitute the random projection vectors used by the ICA to generate each of the ICs [12].

A diverse scenario has been experienced, in the instance of dimensionality reduction of hyper spectral data where the physical sources do not occur. Here, the ICA is only used to discover the projection where all the anticipated components are "the most independent" with respect to negentropy. Independent Component Analysis is strongly associated to Projection Pursuit (PP), a statistical data analysis tool designed to decrease the dimensionality of multivariate data sets by recognizing "interesting projections" in accordance to a projection directory. PP has lately been recommended for analysing hyper spectral images both for supervised classification and for unsupervised data analysis. The key dissimilarity is that ICA concurrently checks all the modules and predicts the directions where all the projected components are "the most independent" in the sense of a degree to independence. On a conflicting note, regular PP consecutively searches for a projection. When a direction is established, data are projected against the subspace orthogonal to the latter and this is carried on. The selection of the two methods should be overseen by the application [13].

2.5. Canonical Correlation Analysis (CCA)

In statistics, Canonical Correlation Analysis, as presented by Harold Hoteling, is a method of creating sense out of cross-covariance matrices. If we take into consideration, two groups of variables and their correlations amongst the variables, then canonical correlation analysis will facilitate us to discover linear combinations of the variables and those variables which have the highest correlation with each other. CCA can also be used to create a model equation which relates two sets of variables, for instance a set of performance measures and a set of descriptive variables, or a set of outputs and a set of inputs.

CCA could accomplish class prediction with a certain primal set of samples belonging to standard classes (such as AML and ALL) to categorize new, anonymous samples. Their initial leukaemia dataset comprised of 38 bone marrow samples (27 ALL, 11 AML) acquired from acute leukaemia patients at the time of diagnosis. The primary concern was to establish if there were genes whose expression form was intensely associated with the class distinction to be predicted. To establish whether the perceived correlations were stronger than those projected, a technique called neighbourhood analysis was developed. The second problem was on how to make use of an assorted set of common samples to construct a "class predictor" proficient of allocating a fresh sample to one out of two classes. A system which applies a permanent subset of "informative genes" (selected based on their connection with the class distinction) and creates an estimation based upon the expression level of these genes in a fresh sample was developed [14].

Canonical Correlation Analysis (CCA) and its regularized version, (RCCA), are techniques for blending two modalities. CCA has been used to discover linear relationships between the pixel values of images and the text committed between these images. RCCA has been used to study expressions of genes measured in liver cells and associate them with concentrations of hepatic fatty acids in mice. CCA is a simple procedure but it suffers from over fitting when the modalities have large quantities of dimensions. RCCA is an amendment to CCA that prevents over fitting but this procedure is computationally very expensive. CCA lacks regularization and thus the conforming covariance matrices have defective inverses. For this purpose, the

embedding modules are not orthogonal but are extremely correlated to each other and yield the same information. RCCA overcomes this regularization problem but still does not create an equivalent level of refinement between patient classes [15].

CCA is an approach of determining the linear affiliation involving two multidimensional variables. It selects two bases, one for each variable, that are ideal with regard to correlations and, simultaneously, it establishes the corresponding correlations. Essentially, it finds the two bases in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized. This is the most significant dissimilarity between CCA and ordinary correlation analysis which greatly rely on the principle in which the variables are defined [16].

DNA microarrays assess expression of numerous thousands of genes. In several organisms, the consequences of drugs on gene expression could be probed by using Microarray classification. As contrasted to sequencing, the microarrays for gene expression analysis are economical in terms of computation. An optimal subgroup of features from the feature set can be discovered by deploying certain feature selection procedures. Using these procedures an optimal subset of features (S) needs to be determined from a general set of N features, resulting in the highest classification accuracy. This condensed optimal feature subset can be named as the predictor set, |S| and generally as much lower dimension in comparison to the overall set (S << N). CCA/CFA (Cross Modal Factor) features can be used to mine the optimal predictor set that takes into matter, feature enslavements among different genes. These procedures look for the similarities in data sources or statistical associations between various genes [17].

2.6. Locally Linear Embedding (LLE)

Non-linear dimensionality reduction methods are largely categorized into two groups such as the ones that offer a mapping from the high dimensional space to a lower dimensional embedding or vice versa, and the others that just provide a visualization in terms of graphs or charts in lower dimensionalities. In the perspective of machine learning, mapping approaches may be regarded like an initial feature extraction step, which is followed by pattern recognition algorithms.

At present, several existing data dimension reduction techniques are either supervised, which is the case where data requires to be labelled. Here the recommended technique was to practice a reviewed LLE method, which is solely unsupervised and quick as the feature extraction scheme for the analysis of microarray data. Three widely accessible microarray datasets are used to assess the aforementioned technique. The efficaciousness of LLE is assessed by the classification precision unit of an SVM classifier. LLE is initially performed on the expression data to condense the dimensionality from a several thousands to a relatively smaller number. Then SVM classifier is applied, and then the Leave-one-out classifier criterion is used to assess the usefulness of the feature reduction's operation [18].

LLE levels well with the intrinsic abundant dimensionality constant, d, and does not involve a precise gridding of the embedding space. As more and more dimensions are supplemented to the embedding space, the prevailing ones remain unvarying; therefore LLE need not be repeated to compute advanced dimensional embedding. LLE is not bounded to practice manifolds of tremendously low dimensionality or co-dimensionality unlike methods like principal curves and surfaces or additive component models. The intrinsic value of d can alone be projected by analysing a reciprocal cost function, in which restoring weights derived from the embedding vectors Y_{Wi} are applied to the data points X_{Wi} [19].

The local linear embedding algorithm (LLE) is a non-linear dimension-reducing method that is extensively used for its computational plainness and impulsive approach. Primarily, LLE linearly restructures each and every input point from its nearest neighbours and then preserves these neighbourhood relationships in a lower dimensional space [20].

LLE feature extraction procedure is purely unsupervised, and it does not require the presence of class labels. This feature extraction design is a basic linear algebra problem, and it does not contain any training or reiteration process, such that the procedure can be exceptionally quick. The feature reduction on leukaemia and colon datasets evidently outpaces this on lymphoma, and this can mean that LLE feature extraction is more suitable for binary-class feature reduction techniques when compared to multiple (greater than 2) class predicaments [18].

2.7. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a procedure used to overcome dimensionality reduction. It is used mainly in the Small Sample Size (SSS) problem. This issue can in a large set of data. In controlled experiments, the amount of existing cases is relatively rise in medical data sets where there are great amounts of dimensions or features less and the occurrence of features or variables is generally significantly larger than the size of the samples [21].At this point, the small sample size problem (SSS problem) arises.

The data involved in the process can be analysed by two methods namely Feature Selection and Feature Transformation where in, Feature transformation processes generate a new group of features as a result of combining or transforming the existing ones. The various procedures used here are PCA, LDA, and PLS, etc.

A drawback of the original LDA is that at least one non-singular scatter matrix is required for computation. When there are insufficient numbers of samples in the data set, this condition fails, i.e. the data dimension generally exceeds the number of available data values. To overcome the non-singularity restriction, (LDA/GSVD) has been introduced which is basically LDA built on the generalized singular value decomposition. Two-stage approaches have been proposed for dimensionality reduction to decrease computational complexity with no non-singularity constraint that comes with FDA. Also, nowadays GLDA algorithms are being applied to dimensionality reduction techniques [22].

In the Direct Linear Discriminant Analysis method, the dimensionality reduction is carried out in two stages. Firstly, a transformation matrix is calculated to transform the training samples to a specific range space. Then, the dimensionality of these transformed samples is further reduced using certain specific regulating matrices.

This enhanced DLDA method surpassed DLDA and certain other related techniques such as PCA + LDA technique and the OLDA technique [23]. The experimentation was tested and implemented on eight DNA microarray gene expression datasets and its enhanced DLDA technique showed an average classification accuracy of 91.1% which proved to be quite superior compared to the average accuracy of the DLDA technique (79.6%). Additionally, PCA transformations are not needed by LDA to reduce dimensionality like other methods such as PCA + LDA and it has a higher accuracy percentage (above 90%). Some of the drawbacks of the LDA method are that the null space of SB (range space) is ignored in the first stage. Since the inverse of the eigenvalues of SB are made use of in its computation, the eigenvectors related to larger eigenvalues of SB are de-stressed that are further constructive in discriminant analysis. In LDA based multiclass classification techniques where the dimension reducing transformation matrix is directly applied for multiclass classification.

A form of LDA known as Null space based linear discriminant analysis (NS-LDA) is used to extract data in cancer classification. NSLDA initially extracts the first order derivative information of the mass spectrometry profiles. Based on the null-space strategy, NSLDA next reduces the data dimension and obtains the discriminant features at the same time.

Fisher's linear discriminant analysis (LDA) in combination with a genetic algorithm is used to study the spatial system of gene subsets [24]. To test the effectiveness of the proposed LDA-based GA, widespread experiments were carried out on seven public datasets and the final

results were compared with seven other prominent algorithms. High prediction accuracy (96% - 100%) was achieved with a minor quantity of informative (test) genes (usually < 20).

2.8. Partial Least Squares Regression (PLS REGRESSION)

Partial least squares regression (PLS regression) is a mathematical approach that allows certain relation to principal components regression, instead defining hyper-planes of high variance amongst the response and independent variables, PLS discovers a linear regression model by predicting the estimated variables and the perceivable variables to an alternate space. Since both the X and Y data are projected to fresh spaces, the PLS family of techniques are commonly known as bilinear factor models.

An alternative meaning for PLS is projection to latent assemblies, but the term partial least squares continues to be used in several domains. Despite the inventive purposes were in the social sciences, PLS regression continues to be extensively used in chemo-metrics and associated fields to this day. It is furthermore used in fields such as bioinformatics, sensometrics, neuroscience and anthropology. In a distinct manner, PLS path modelling is regularly applied in the fields of econometrics, marketing, social sciences and strategic management.

PLS regression has been lately known as a means to simplify and pool characteristics from principal component analysis (PCA) and multiple regressions. It is predominantly functional when we want to forecast a group of dependent variables from a (very) large set of independent variables (i.e., predictors). The aim of PLS regression is to estimate Y variables from X variables and to explicate their collective assortment. If the collection of predictors is enormous in comparison to the number of observations, X becomes singular hence the regression method is not achievable. One tactic is to disregard some predictors or also, a method called principal component regression is performed to execute a principal component analysis (PCA) of the X matrix and then use the principal components of X as regressors on Y. The problem here is to choose an optimal feature subset. One probable scheme is to retain only the first few components. But these are selected to explain X reasonably than Y, and therefore, nothing will assure the principal components, which "explain" X, that are relevant for Y [25].

Typical statistical policies in different classification methods do not execute properly when the dimension, p, is larger compared to the gene sample size, N. Dimension reduction of the higher (p) dimensional space by PCA is realized by fabricating principal components (PCs), which are the linear combinations of the novel p predictor/ explanatory variables. In PCA, dimension reduction is accomplished by creating linear combinations that make best use of the variance-based objective function, namely variable (X_w).Similarly PLS works well with an independent function based purely on covariance. [26].

In PCA, the individual weights are collected independent of the response variable; thus, the assortment of subcomponents with large variance may not essentially be analytical of the response. Thus, for estimation of survival times using gene expression data, a method recommended was PLS dimension reduction, which takes advantage of a covariance criterion. Dimension reduction using PCA totally disregards the response data and its corresponding censoring statistics. PLS integrates the response variable during the dimensionality reduction procedure [27].

In particular, PLS surpasses PCA in the case of microarray gene expression data. Although, PCA can be fairly competent if one selects the predictors (genes) which can be predicted from the response classes before pre-deploying PCA. PLS only consists of indicating the amount of gene components whereas PCA necessitates choosing the K gene components [28].

3. CONCLUSIONS

The sole purpose of this survey was to provide a vivid analysis on different popular and predominant algorithmic approaches that are used to execute dimensionality reduction. The Singular Value Decomposition (SVD) method emphasizes on a gene selection technique whereas the Principal Component Analysis (PCA) method focuses on converting the values into a set of linear combinations. The Support Vector Machines (SVM) (supervised) is compared to the Singular Value Decomposition method (unsupervised) and thus their efficiencies are discussed within their applicable expanses.

PCA and Linear Discriminant Analysis (LDA) also fall under a peculiar category of feature transformation where in the former uses a statistical signal criterion whereas the latter uses a classification model. The Partial Least Squares (PLS) method can also be categorized under the same roof of transformation and is compared to PCA where in the former uses a linear regression model whereas the latter stresses on the use of maximum variance calculated. The Locally Linear Embedding (LLE) technique is a manifold learning methodology and thus falls under the non-linear practices of dimensionality reduction.

The Canonical Correlation Analysis (CCA) technique abets in discovering linear combinations from two sets of variables and thereby approximates correlations amongst these variables. The Independent Component Analysis (ICA) method works on the principle of additive subcomponents and separation of singular units from a large multivariate source. Thus dimensionality reduction algorithms can be executed onto a specific dataset with a particular problem, depending upon their usage statistics and parameters under which their applicable conditions are satisfied.

REFERENCES

[1]   Krzysztof Simek, Krzysztof Fujarewicz, Andrzej Swierniak, Marek Kimmela, Barbara Jarzab, Malgorzata Wiench, Joanna Rzeszowska, "Using SVD and SVM methods for selection, classification, clustering and modeling of DNA microarray data," Engineering Applications of Artificial Intelligence (17) 2004, Science Direct 0952-1976.

[2]   Jagdish Chandra Patra, Lei Wang, Ee Luang Ang, and Narendra S. Chaudhari, "Neural Network-Based Analysis of DNA Microarray Data," Proceedings of International Joint Conference on Neural Networks, Montreal,Canada 2005, IEEE 0-7803-9048-2.

[3]   Faming Liang, "Use of SVD-based probit transformation in clustering gene expression profiles," Computational Statistics & Data Analysis (51) 2007, Science Direct 0167-9473.

[4]   S. Bicciato, A. Luchini, C. Di Bello, "Disjoint PCA Models For Marker Identification And Classification Of Cancer Types Using Gene Expression Data,"2002, IEEE 0-7803-7557-2.

[5]   Matthew Partridge, Rafael Calvo, "Fast Dimensionality Reduction and Simple PCA,"2006.

[6]   Vinay Nadimpally , Mohammed J. Zaki, "A Novel Approach to Determine Normal Variation in Gene Expression Data," SIGKDD Explorations, Vol. 5 Issue 2,2002.

[7]   Ricardo Gutierrez-Osuna, "Introduction to Pattern Recognition".

[8]   Zne-Jung Lee, "An integrated algorithm for gene selection and classification applied to microarray data of ovarian cancer," Artificial Intelligence in Medicine (42) 2008, Elsevier 0933-3657.

[9]   Isabelle Guyon, Jason Weston, Stephen Barnhill, M.D., "Gene Selection for Cancer Classification using Support Vector Machines," Barnhill Bioinformatics 2001.

[10]   Xiaobo Li, Sihua Peng, Jian Chen, Bingjian Lu, Honghe Zhang, Maode Lai, "SVM-T-RFE: A novel gene selection algorithm for Identifying metastasis-related genes in colorectal cancer using

gene expression profiles," Biochemical and Biophysical Research Communications 2012,ScienceDirect S0006-291X(12)00134-9.

[11]   Han Liu, Rafal Kustra, "Dimension Reduction of Microarray Data with Penalized Independent Component Analysis," 2002.

[12]   Jing Wang, Student Member, IEEE, and Chein-I Chang, Senior Member, "Independent Component Analysis Based Dimensionality Reduction With Applications in Hyper spectral Image Analysis," IEEE Transactions On Geoscience And Remote Sensing Vol. 44, 2006, IEEE 0196-2892.

[13]   M. Lennon, G. Mercier, M.C. Mouchot, L. Hubert-Moy, "Independent Component Analysis as a tool for the dimensionality reduction and the representation of hyper spectral images."

[14]   T. R. Golub, D. K. Slonim, Tamayo, Huard,M. Gaasenbeek, J. P. Mesirov, Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, D. Bloomfield, E. S. Lander, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," Sciencemag Vol 286,1999.

[15]   Abhishek Golugula, George Lee, Stephen R. Master, Michael D. Feldman, John E. Tomaszewski, and Anant Madabhushi, "Supervised Regularized Canonical Correlation Analysis: Integrating Histologic and Proteomic Data for Predicting Biochemical Failures," EMBS 2011, IEEE 978-1-4244-4122-8.

[16]   Magnus Borga, "Canonical Correlation Tutorial," June 2001.

[17]   Girija Chetty, Madhu Chetty, "Multiclass Microarray Gene Expression Classification based on Fusion of Correlation Features".

[18]   Shi Chao, Chen Lihui, "Feature Dimension Reduction for Micro-array Data Analysis Using Locally Linear Embedding,"2004.

[19]   Sam T, Roweis, Lawrence K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Sciencemag Vol 290, 2000.

[20]   Y. Goldberg, Y. Ritov, "LLE with low-dimensional neighbourhood representation," Pattern Recognition, 2009.

[21]   Lei Zhu, Bin Han, Lihua Li, Shenhua Xu, Hanzhou Mou, Zhiguo Zheng , "Null space LDA based feature extraction of mass spectrometry data for cancer classification," 2009, IEEE 978-1-4244-4134-1.

[22]   Hyunsoo Kim, Barry L. Drake, Haesun Park, "Multiclass classifiers based on dimension reduction with generalized LDA,".Pattern Recognition (40) 2007, ScienceDirect 0031-3203.

[23]   Kuldip K. Paliwal , Alok Sharma, "Improved direct LDA and its application to DNA microarray gene expression data,". Pattern Recognition Letters (41) 2010, ScienceDirect 0167-8655.

[24]   Edmundo Bonilla Huerta, Beatrice Duval, Jin-KaoHao, "A hybrid LDA and genetic algorithm for gene selection and classification of microarray data," Neurocomputing (73) 2010, ScienceDirect 0925-2312.

[25]   Herve Abdi, "Partial Least Squares Regression," 2003.

[26]   D.V.Nguyen, David Rocke, "Tumour Classification by PLS using micro-array gene expression data," Bioinformatics Vol 18, 2002.

[27]   D.V.Nguyen, "Partial least squares dimension reduction for microarray gene expression data with a censored response," Mathematical Biosciences (193) 2005, ScienceDirect 0025-5564.

[28]   D.V.Nguyen, David Rocke, "On partial least squares dimension reduction for microarray-based classification: a simulation study," Computational Statistics & Data Analysis (46) 2004, ScienceDirect 0167-9473.