METHODS FOR DIMENSION REDUCTION

[1]Shail kumari, [2]Vishiv panchal, [3]Gurpreet Kaur
[1,2]Student, [3]Assistant Professor
[1,2]Department of Computer Science Engineering
Mahavir Swami Institute of Technology, Sonipat, India

Abstract - Dimensionality Reduction (DR) is the pre-processing step to remove redundant features, noisy and irrelevant data, in order to improve learning feature accuracy and reduce the training time. Dimensionality reductions techniques have been proposed and implemented by using feature selection and extraction method. Principal Component Analysis (PCA) one of the Dimensions reduction techniques which give reduced computation time for the learning process. In this paper presents most widely used feature extraction techniques such as EMD, PCA, and feature selection techniques such as correlation, LDA, and forward selection have been analyzed based on high performance and accuracy. These techniques are highly applied in Deep Neural networks for medical image diagnosis and used to improve classification accuracy. Further, we discussed how dimension reduction is made in deep learning. Important ways; they can be separated into techniques that apply to supervised or unsupervised learning and into techniques that either entail feature selection or feature extraction. In this paper, an overview of dimension reduction techniques based on this organization is presented and representative techniques in each category are described.

1. INTRODUCTION

Real-world data, such as speech signals, digital photographs, or fMRI scans, usually has a high dimensionality. In order to handle real-world data adequately, its dimensionality needs to be reduced. Dimensionality reduction is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data. The intrinsic dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data . Dimensionality reduction is important in many domains, since it mitigates the curse of dimensionality and other undesired properties of high-dimensional spaces. As a result, dimensionality reduction facilitates, among others, classification, visualization, and compression of high-dimensional data. Traditionally, dimensionality reduction was performed using linear techniques such as Principal Components Analysis (PCA) and factor analysis. However, these linear techniques cannot adequately handle complex nonlinear data. Therefore, in the last decade, a large number of nonlinear techniques for dimensionality reduction have been proposed. In contrast to the traditional linear techniques, nonlinear techniques have the ability to deal with complex nonlinear data. In particular, for real-world data, the nonlinear dimensionality reduction techniques may offer an advantage, because real-world data is likely to be highly nonlinear. Previous studies have shown that nonlinear techniques outperform their linear counterparts on complex artificial tasks. For instance, the Swiss roll dataset comprises a set of points that lie on a spiral-like two-dimensional manifold within a three-dimensional space. A vast number of nonlinear techniques are perfectly able to find this embedding, whereas linear techniques fail to do so. In contrast to these successes on artificial datasets, successful applications of nonlinear dimensionality reduction techniques on natural datasets are scarce. Beyond this observation, it is not clear to what extent the performances of the various dimensionality reduction techniques differ on artificial and natural tasks (a comparison is performed in , but this comparison is very limited in scope with respect to the number of techniques and tasks that are addressed). Motivated by the lack of a systematic comparison of dimensionality reduction techniques, this paper presents a comparative study of the most important linear dimensionality reduction technique (PCA), and twelve front ranked nonlinear dimensionality reduction techniques. The aims of the paper are to investigate to what extent novel nonlinear dimensionality reduction techniques outperform the traditional PCA on real-world datasets and to identify the inherent weaknesses of the twelve nonlinear dimenisonality reduction techniques. The investigation is performed by both a theoretical and an empirical evaluation of the dimensionality reduction techniques. The identification is performed by a careful analysis of the empirical results on specifically designed artificial datasets and on the real-world datasets. Next to PCA, the paper investigates the following twelve nonlinear techniques: (1) multidimensional scaling, (2) Isomap, (3) Maximum Variance Unfolding, (4) Kernel PCA, (5) diffusion maps, (6) multilayer autoencoders, (7) Locally Linear Embedding, (8) Laplacian Eigenmaps, (9) Hessian LLE, (10) Local Tangent Space Analysis, (11) Locally Linear Coordination, and (12) manifold charting. Although our comparative review includes the most important nonlinear techniques for dimensionality reduction, it is not exhaustive. In the appendix, we list other important (nonlinear) dimensionality reduction techniques that are not included in our comparative review. There, we briefly explain why these techniques are not included. The outline of the remainder of this paper is as follows. In Section 2, we give a formal definition of dimensionality reduction. Section 3 briefly discusses the most important linear technique for dimensionality reduction (PCA). Subsequently, Section 4 describes and discusses the selected twelve nonlinear techniques for dimensionality reduction. Section 5 lists all techniques by theoretical characteristics. Then, in Section 6, we present an empirical comparison of twelve techniques for dimensionality reduction on five artificial datasets and five natural datasets. Section 7 discusses the results of the experiments; moreover, it identifies weaknesses and points of improvement of the selected nonlinear techniques. Section 8 provides our conclusions. Our main conclusion is that the focus of the research community should shift towards nonlocal techniques for dimensionality reduction with objective functions that can be optimized well in practice (such as PCA, Kernel PCA, and auto encoders).

2. METHODS OF DIMENSIONALITY REDUCTION

There are seven different methods that have been applied in the data analytical space. They are illustrated in the Figure1.In most of the times the huge amount of data in data analytical process does not work well. So dimensionality reduction is applied to the large data. The methods of dimensionality reduction define about the reduction of data elements illustratedinFigure.1.

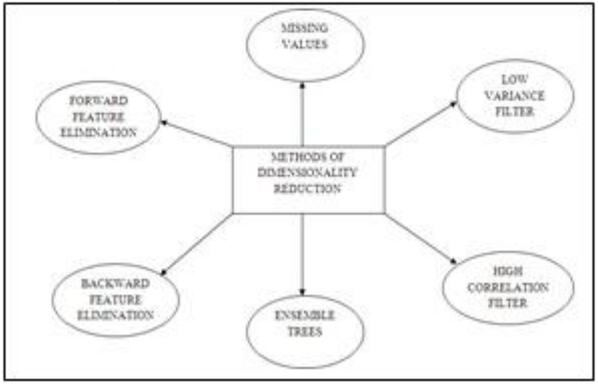

MISSING VALUES
FORWARD FEATURE ELIMINATION
LOW VARIANCE FILTER
METHODS OF DIMENSIONALITY REDUCTION
BACKWARD FEATURE ELIMINATION
ENSEMBLE TREES
HIGH CORRELATION FILTER

Fig. 1. Methods of Dimensionality Reduction

A. Missing values

The data column with enormous amount of missing values is calculated and then they are removed. The numbers of missing values are calculated using either statistical node or group node. The statistical nodes are used for the analysis of data and the group node is used in the techniques of DR. The missing values larger than the threshold are collected and then they are removed. If the threshold value found is larger than the original values then the reduction will be aggressive. The missing values are calculated using the formula which is shown in Figure.2. In KNIME tool the missing value node calculates the ratio of the missing values is calculated by number of missing values divided by the total number of rows.

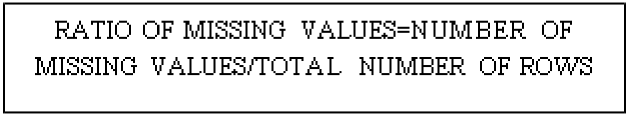

RATIO OF MISSING VALUES=NUMBER OF MISSING VALUES/TOTAL NUMBER OF ROWS

Fig. 2. Formulae for Missing Values

B. Low Variance Filter

The variance of the data is calculated to find the number of information about the data column. In the limit case where the column cells assume a constant value, the variance would be 0 and the column would be of no help in the discrimination of different groups of data. The data column lower than the threshold values of the data are filtered and then they are removed. The filter is applied to the data to find the low variance. First normalization is done to all the columns to find the lower variance.

C. High Correlation Filter

The feature which are given as input are often correlated i.e. the features are dependent to one another and they too have same information. A data column with values highly correlated to those of another data column is not going to add very much new information to the existing pool of input features. To reduce the correlated column, the correlation is measured using the linear correlation node between couple of columns. If any highly correlated column is present between the pair the particular data is removed. Correlation matrix is taken as input to the correlation filter. Filtering highly correlated data columns requires uniform data ranges again which can be obtained with a Normalize node .

D. Ensemble Trees

Ensemble trees are also known as random forest. For effective classification feature, selection is done. One approach to dimensionality reduction is to generate a large and carefully constructed set of trees against a target attribute and then use each attribute’s usage statistics to find the most informative subset of features. The score is calculated using level of the candidates and it defines about the relative attributes which are most predictive. A shallow tree has been generated for ensembling and here each tree is trained on the fraction of all attributes. If the particular attribute is selected as the best split one then the informative features are retained.

E. Backward Feature Elimination

The Backward Feature Elimination loop performs dimensionality reduction against a particular machine learning algorithm. It is a simple iterative method and in each method the selected classification algorithm is performed for number of input features. Then one input feature is removed and the model is trained for (n-1) input features for number of times. The input feature with large number of error rate, after the removal of feature is recognized and they are eliminated. Then they are repeated for number of iterations like n-2, n-3, etc to find the larger error rate. The Backward Feature Elimination Filter finally visualizes the number of features that are kept at each iteration and the corresponding error rate. This method can only be applied to the small number of dataset.

F. Forward Feature Elimination

Similarly to the Backward Feature Elimination approach, a Forward Feature Construction loop builds a number of pre-selected classifiers using an incremental number of input features. The forward feature loop opens with one feature and other feature is added to it for every iterations. Both forward and backward are very expensive and computationally high. Running the optimization loop the best cutoffs in terms of lowest number of columns and best accuracy were determined for each one of the six dimensionality reduction methods and for the best performing model.

3. TECHNIQUES IN DIMENSIONALITY REDUCTION

There are two different major techniques used in dimensionality reduction. They are feature selection and feature extraction.

Feature Selection

In machine learning and statistics, feature selection (FS), also known as variable selection or attribute selection or variable subset selection and it is the process of selecting a subset of relevant features (variables, predictors) for use in model construction. This technique has three different reasons: They are Interpretation of model to make them efficient and simple for the users, Less training time and Reduction of variance for strengthened generalization. Feature selection is applied to the domains with larger features and chooses the feature according to the objective function. A feature selection algorithm can be seen as the combination of a search technique for proposing new feature subsets along with an evaluation measure which scores the different feature subsets .Ever subset of the feature is tested separately to minimize the error or noise rate. Feature selection is classified into three different classes: they are embedded, filter and wrapper method. The classes of feature selection are shown in Figure.3

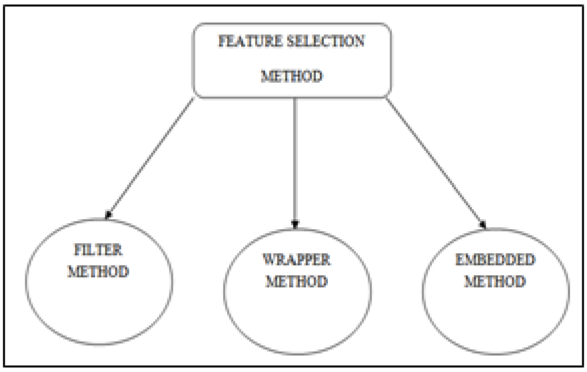


Fig.3. Feature Selection Methods

A.1 Filter Method

In filter method, the variables are selected based on the model used. They are based only on general features like the correlation with the variable to predict . Filter method contains the minimum interesting variables. Other variables will be used for either classification or regression models. These methods work efficiently under time consumption. It is only used as a preprocessing method and they choose the redundant variable because they do not matter the relationship between the variable. The flow of this class is proposed in Figure4. In this all the features are taken and from that the best subsets are chosen and the algorithms for further tasks are applied to it to find the performance.

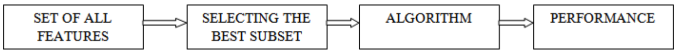


Fig. 4. Filter Method

A.2 Wrapper Method

Unlike the filter approach wrapper method calculates the subset of variables to encounter the interaction between the variable. There are two disadvantages of wrapper method, they are:

- The number of observations are inadequate when the over fittings are increased.
- Larger the variable, the computation time is increased.

Here the subset feature is generated using the search method and the performance is calculated. Wrapper method uses search algorithm for selecting the subset. Alternative search-based techniques are based on targeted projection pursuit which finds low-dimensional projections of the data that score highly: the features that have the largest projections in the lower-dimensional space are then selected. The flow of wrapper method is illustrated in Figure5.

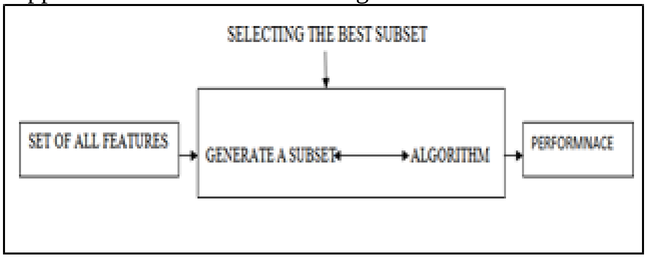


Fig.5. Wrapper Method

Search Approaches

- Exhaustive search
- Best first search
- Simulated annealing
- Genetic algorithm
- Greedy forward selection
- Greedy backward selection
- Scatter search
- Variable neighborhood search

A.3 Embedded Method

Recently, embedded methods have been proposed to reduce the classification of learning. To form the embedded class, the advantages of filter and wrapper classes are used. Many embedded feature selection methods have been introduced during the last few years unifying theoretical framework has not been developed to date. The learning algorithm takes their own variable selection algorithm. The flow is shown in Figure6. This method is varied from other feature selection method.

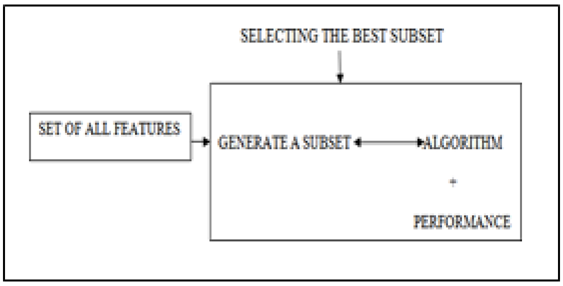


Fig.6. Embedded Method

The parameterized function has been shown in Figure7 which defines the function F to find the vector a, let X be the variable R defines the relation between the variable X and the vector α. Embedded method takes the set of all features as the input and selects the best subset by generating the subset to find the performance of the method. In this method the features are added or removed iteratively to find the subset.

$$\boxed{F: R(\alpha, X) \rightarrow f(\alpha, X)}$$

Fig.7. Formulae for parameterized function

Feature Reduction

In feature reduction, the data of high dimensional space are transformed into lower dimensional space. Transforming the data is done in linear method or nonlinear method. The main linear technique for dimensionality reduction performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized. In order to reduce the features the correlation coefficient is calculated by finding the Eigen vectors and Eigen values. In this model the relevant features for the class is selected by eliminating the redundant features. The feature reduction is done with four aspects which are defined in Figure.10.

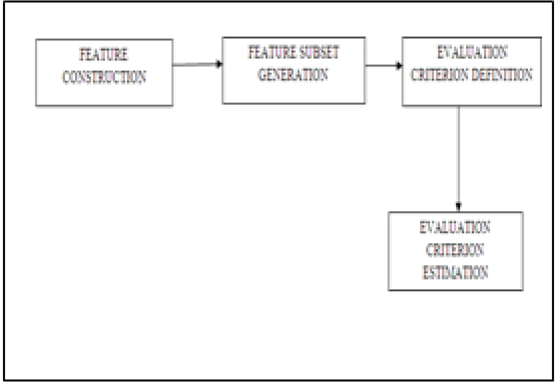

FEATURE CONSTRUCTION
FEATURE SUBSET GENERATION
EVALUATION CRITERION DEFINITION
EVALUATION CRITERION ESTIMATION

Fig. 8. Aspects of Feature Extraction

First, the features are taken/constructed as input and then the subset of the features are extracted from the original features in which the redundant features are eliminated and the similar features are extracted using the algorithms such as PCA, SVD, LDA, etc. Then the filters use criteria not involving any learning machine for example a relevance index based on correlation coefficients or test statistics. Then the filters are used to estimate the feature value.

4. ALGORITHM

Many different algorithms are used for performing in dimensionality reduction process.

PRINCIPAL COPMPONENT ANALYSIS (PCA)

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components [10]. The count of principal component is lower or equal to the original features. The Eigen vector, Eigen value and covariance matrix is found in order to perform PCA.

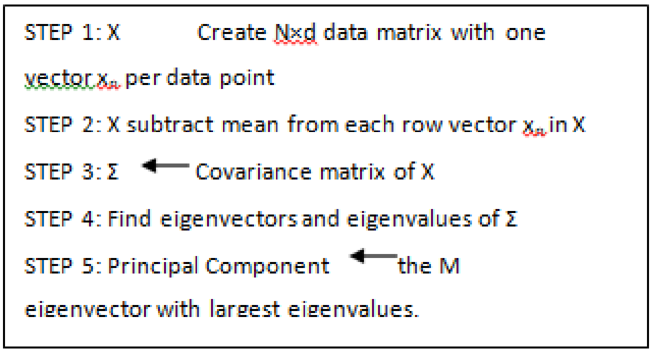
$$\boxed{Z=\Sigma_{\square}^{\square}{}_{=}\ \square_{i}}$$

In which Xi is the ith variables where i=1, 2, 3, 4, p is the linear coefficient of Z which is denoted in matrix and places ad a Ta =1.Here, a1 vector maximizes the variance and here Z is called the first principal component. The matrix C=Cov(X) is the covariance matrix of X. The Eigen value λi is calculated by determining the equation.

$$\boxed{\det(C-\lambda I)}$$

The Eigen vector is defined to be the column of the matrix X.C=A D AT. Where

$$D=\begin{matrix} \lambda 1 & 0 & 0 \\ 0 & \lambda 2 & 0 \\ 0 & 0 & \lambda 3 \end{matrix}$$


STEP 1: X Create Nxd data matrix with one vector x_n per data point

STEP 2: X subtract mean from each row vector x_n in X

STEP 3: Σ ← Covariance matrix of X

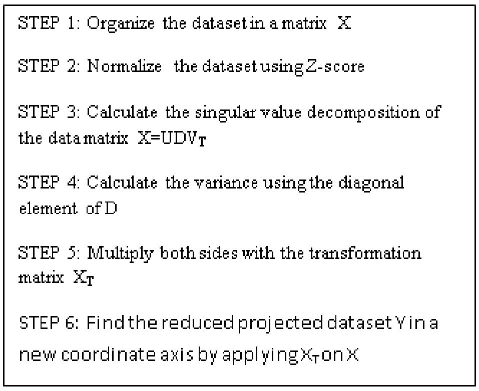
STEP 4: Find eigenvectors and eigenvalues of Σ

STEP 5: Principal Component ← the M eigenvector with largest eigenvalues.

Fig.9. PCA Algorithm

PCA algorithm is applied to large dataset for efficient result. The algorithm is shown in the figure.9. In this algorithm first takes the high dimensional data matrix as input which then calculates the mean for the data matrix and then it is subtracted with each row vector of the data matrix. Then the covariance is found to find the Eigen vector and Eigen value in which the highest. Eigen value column is taken as output and that features are extracted. The principal components are orthogonal because they are the eigenvectors of the covariance matrix which is symmetric and it is sensitive to the relative scaling of the original variables. PCA is used as a tool to compose predictive model. The outcome of this algorithm is defined as factor score.

Singular Value Decomposition (SVD)
In linear model SVD is a type of factorization for the complex model.SVD is most only used in applications like signal processing and statistics. The SVD of an a × b complex or real matrix M is decomposition of the form M = UΣV∗ where U is an a × a real or complex unitary matrix, Σ is an a×b rectangular diagonal matrix with non-negative real numbers on the diagonal and V∗ (the conjugate transpose of V or simply the transpose of V if V is real) is an b × b real or complex unitary matrix and the diagonal entries Σi,i of Σ are known as the singular values of M[11].The a column and b column are called the left and right singular vectors of m. The X is found by multiplying U*D*VT. The data set is organized into matrix and the data are normalized. The SVD is calculated by X=UDVT,.U is calculated by converting row into column matrix, D is the diagonal matrix of U. The transformation matrix XT is multiplied on both the sides of U. Then the new coordinated axis is found by applying XT on X and then the Singular value is decomposed.



```
STEP 1: Organize the dataset in a matrix X
STEP 2: Normalize the dataset using Z-score
STEP 3: Calculate the singular value decomposition of the data matrix $X=UDV_T$
STEP 4: Calculate the variance using the diagonal element of D
STEP 5: Multiply both sides with the transformation matrix $X_T$
STEP 6: Find the reduced projected dataset Y in a new coordinate axis by applying $X_T$ on X
```

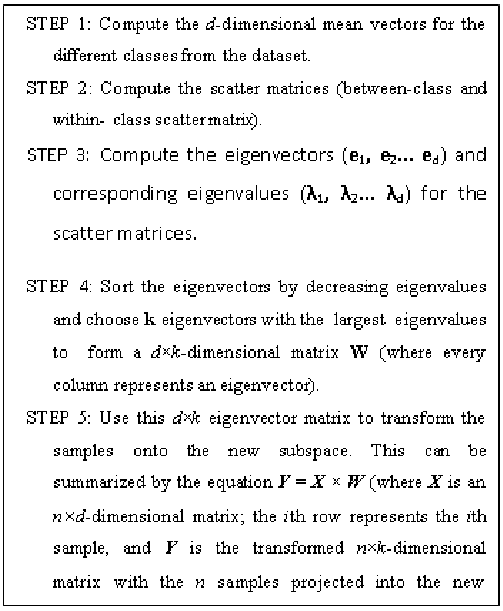
Fig.10. SVD Algorithm

Linear Discriminant Analysis (LDA)

Linear discriminant Analysis (LDA) is a generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events [21]. In this the class can be separated by discriminant analysis. Ronald A. Fisher formulated the Linear Discriminant in 1936 .The discriminates can be found by coefficient. The discriminant function is used to calculate discriminant. In this, Di is linear combination, Xk is predicted variable

$$\boxed{D_i=b_0 \quad \square= \quad b_k X_k}$$

First the matrix is taken as input for different classes then the scatter matrices are computed. The Eigen vectors and Eigen values are calculated for the scatter matrices. The Eigen vectors are sorted by decreasing the Eigen values then the samples are transformed into newer subspace by the formulae Y=X x W where X is n×d dimensional matrix is d x k dimensional matrix.

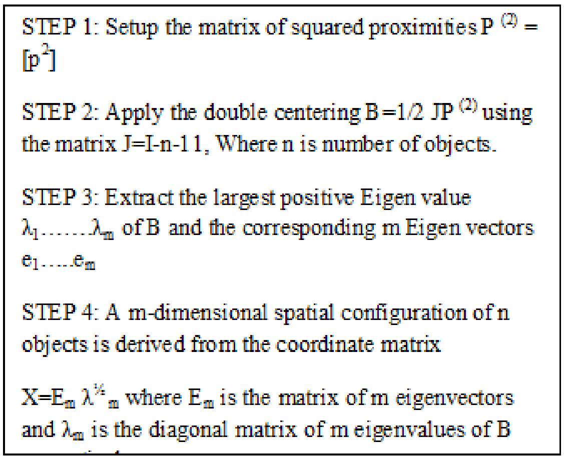


```
STEP 1: Compute the $d$-dimensional mean vectors for the
    different classes from the dataset.
STEP 2: Compute the scatter matrices (between-class and
    within- class scatter matrix).
STEP 3: Compute the eigenvectors ($\mathbf{e_1}, \mathbf{e_2} ... \mathbf{e_d}$) and
    corresponding eigenvalues ($\lambda_1, \lambda_2 ... \lambda_d$) for the
    scatter matrices.
STEP 4: Sort the eigenvectors by decreasing eigenvalues
    and choose $\mathbf{k}$ eigenvectors with the largest eigenvalues
    to form a $d$×$k$-dimensional matrix $\mathbf{W}$ (where every
    column represents an eigenvector).
STEP 5: Use this $d$×$k$ eigenvector matrix to transform the
    samples onto the new subspace. This can be
    summarized by the equation $Y = X \times W$ (where $X$ is an
    $n$×$d$-dimensional matrix; the $i$th row represents the $i$th
    sample, and $Y$ is the transformed $n$×$k$-dimensional
    matrix with the $n$ samples projected into the new
```

Fig.11. LDA Algorithm

Multi-Dimensional Scaling (MDS)

Multidimensional scaling (MDS) is a classical approach for DR used to transform high dimensional space into lower space. MDS transforms the dimensional space by calculating the distance. MDS addresses the problem of constructing a configuration of t points in Euclidean space by using information about the distances between the patterns. To solve the problems of MDS two types of methods have been used. They are metric method and non-metric method. The metric method will produce the original metrics, whereas non-metric method will rank the values of the object. Here, the matrix is taken as input in the proximity and the double centering B is applied to the matrix dataset with the formulae B=1/2 JP (2) and extract the largest Eigen vector value and the particular largest value is derived from the coordinate matrix.



```
STEP 1: Setup the matrix of squared proximities $P^{(2)} = [p^2]$
STEP 2: Apply the double centering $B=1/2\ JP^{(2)}$ using
    the matrix J=I-n-1 1, Where n is number of objects.
STEP 3: Extract the largest positive Eigen value
    $\lambda_1 ....... \lambda_m$ of B and the corresponding m Eigen vectors
    $e_1 ..... e_m$
STEP 4: A m-dimensional spatial configuration of n
    objects is derived from the coordinate matrix
    $X=E_m \lambda^{½}_m$ where $E_m$ is the matrix of m eigenvectors
    and $\lambda_m$ is the diagonal matrix of m eigenvalues of B
```

Fig.12. MDS Algorithm

DIMENSIONALITY REDUCTION IN VARIOUS FIELD

Dimensionality reduction is used in numerous field is as follows.

A. Data Mining

Dimensionality reduction is applied to high dimensional data in order to reduce the features. Features are the attributes of the data which is reduced by certain condition. In practice researchers and practitioners interchangeably use dimension, feature, variable, and attribute. The reduced features are then used to perform data mining techniques. Removal of irrelevant data increases the accuracy of the data. Data dimensionality reduction (DDR) has become an important aspect of data mining since human experts and corporate managers are able to make better use of lower dimensional data compared to high dimensional ones. When preprocessing of the data is done for dimension reduction, it must focus on describing the dataset using minimum number of attributes but it must give the performance comparable to that of original dataset containing all the attributes .The techniques have to be considered which performing retention like dependency and significance.

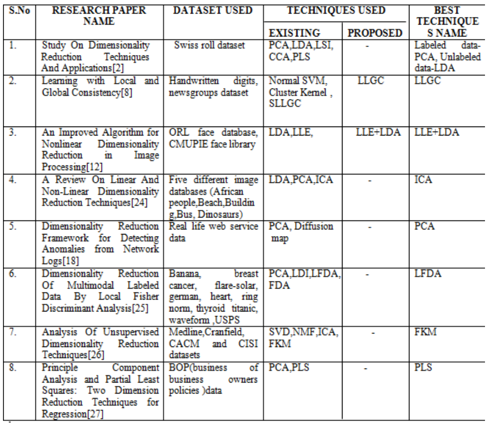
B. Image Processing

Pattern recognition technology in image is an essential branch in information processing, financial perspective and management perspective. It has broad applications and exploring prospects, it has already become a key research projecting pattern recognition and artificial intelligence. Dimensionality reduction like Feature extraction is done to the high dimensional image to distort the image and the recognition rate of the image is reduced. Over-fitting in the training model of the image is recovered by feature extraction. Linear dimension reduction method performs dimensional reduction to high dimensional data through performance target looking for the linear transformation matrix, but since light, expression, posture and factors that make the high dimensional image that have obvious characteristics of nonlinear manifold. Due to curse of dimensionality, high dimensional data leads to lower result. Class label context coming from the spatial configuration of images provides an additional source of classification information and therefore taking this contextual information into account can be beneficial. There are different problems arise in image analysis like

- High dimensionality reduction problem in which it has dimension ranging from hundreds to thousands of factors and they are repeated in different point of time and space.
- Soft dimensionality reduction problem in which the data are not much high and have direct interpretation.
- Visualization problem is one of the techniques to represent the data. There are several techniques. When the time variable is added to the image data there arise two different types of problems like statistical dimensionality reduction and time dependent dimensionality reduction.

C. Network Logs

All networks and systems can be vulnerable to different types of intrusions. When evaluating textual information the features and space/area grows larger and will be inadequate. Diffusion map is a manifold learning method that maps high-dimensional data to a low dimensional diffusion space .The manifold learning method performs the Eigen decomposition of the network data in the matrix form. The textual information is converted into feature space for the reduction. The network data are numerous because of mapping the textual form into the matrix for reduction. To find the coordinate of the system only few variables are needed to describe the data. The amount of log lines that needs to be inspected is reduced. Dimensionality reduction means reducing the complexity of data while preserving some quantity of interest .This is useful for system administrators trying to identify intrusions. The number of interesting log lines is low compared to the total number of lines in the log file. Component analysis methods take the network data as input. The initial weight which is attached to the network is very high and fine tuning i.e. multi-layer encode network is done to reduce the dimensions



S.No	RESEARCH PAPER NAME	DATASET USED	TECHNIQUES USED		BEST TECHNIQUES NAME
			EXISTING	PROPOSED	
1.	Study On Dimensionality Reduction Techniques And Applications[2]	Swiss roll dataset	PCA,LDA,LSI, CCA,PLS	-	Labeled data-PCA, Unlabeled data-LDA
2.	Learning with Local and Global Consistency[8]	Handwritten digits, newsgroups dataset	Normal SVM, Cluster Kernel , SLLGC	LLGC	LLGC
3.	An Improved Algorithm for Nonlinear Dimensionality Reduction in Image Processing[12]	ORL face database, CMUPIE face library	LDA,LLE,	LLE+LDA	LLE+LDA
4.	A Review On Linear And Non-Linear Dimensionality Reduction Techniques[24]	Five different image databases (African people,Beach,Building,Bus, Dinosaurs)	LDA,PCA,ICA	-	ICA
5.	Dimensionality Reduction Framework for Detecting Anomalies from Network Logs[18]	Real life web service data	PCA, Diffusion map	-	PCA
6.	Dimensionality Reduction Of Multimodal Labeled Data By Local Fisher Discriminant Analysis[25]	Banana, breast cancer, flare-solar, german, heart, ring norm, thyroid titanic, waveform ,USPS	PCA,LDI,LFDA, FDA	-	LFDA
7.	Analysis Of Unsupervised Dimensionality Reduction Techniques[26]	Medline,Cranfield, CACM and CISI datasets	SVD,NMF,ICA, FKM	-	FKM
8.	Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression[27]	BOP(business of business owners policies)data	PCA,PLS	-	PLS

Table 1. Comparative analysis on dimensionality reduction

5. CONCLUSION

The main objective of this paper is to provide the overview of the dimensionality reduction. Before performing the reduction the estimation should be made for dimensions. Dimensionality reduction has been used in numerous fields for further techniques. It is used for providing the better result for the data analysis. The guidance should be given to the users handling dimensionality reduction in their data analytics. This paper discussed about the dimensionality reduction methods, techniques used, algorithms, fields where the reduction has been used. Further work is done by performing various dimensionality reduction algorithms and the techniques like clustering, classification, etc can be performed for comparative analysis.

REFERENCE

1. C.O.S. Sorzan, J.Vargas and A. Pascual Montano “A survey of dimensionality reduction techniques”

2. Seven Techniques for Dimensionality Reduction-Open for innovative KNIME.

3. http://www.kdnuggets.com/2015/05/7- methods-data-dimensionality- reduction.html.

4. https://en.wikipedia.org/wiki/Feature_sele ction.

5. Thomas Navin Lal1, Olivier Chapelle, Jason Weston, and André Elisseeff - Learning with Local and Global Consistency-Max Planck Institute for Biological Cybernetics, Tubingen, Germany.

6. https://en.wikipedia.org/wiki/Dimensionality_reduction#Feature_extraction

7. https://en.wikipedia.org/wiki/Principal_component_analysis

8. https://en.wikipedia.org/wiki/Singular_value_decomposition.