

Usage of Principal Component Analysis (PCA) in AI Applications

Somia B. Mohammed Ahmed Khalid Saife Eldin F. Osman Rasha Gaffer .M. Helali
College of Higher Education Department of Computer Science Computer Science Department Computer Science Department
The National Ribat University Najran University Emirates College for Science and Technology University of Bisha
Khartoum , Sudan Najran , KSA Khartoum , Sudan Bisha, KSA

Abstract— Principal Component Analysis (PCA) is a powerful statistical technique for variable reduction, It used when variables are highly correlated. PCA becomes an essential tool for multivariate data analysis and unsupervised dimension reduction. PCA incorporated with AI techniques to improve performance of many applications like image processing, pattern recognition, classification and anomaly detection. The goal of this survey is to provide a comprehensive review of the literature related to Principal Component Analysis (PCA).

Keywords— *Principal Component Analysis, feature reduction, classification, features extraction.*

INTRODUCTION

Investigating the relationship between variables is a favorite research activity of social science. They often want to explore the structure of a large body of data. To understand this data have to be condensed in one way or another, and the row data have to be combined to form summary which are more easily comprehended [2]. Among the most popular methods to achieve such consideration and summarization is Principal Component Analysis [2].

Principal Component Analysis (PCA) is an important method in machine learning due to its twofold nature. PCA reduces the dimensionality of the dataset, which takes the dimensions that encode the most important information and removes the dimensions that encode the least important information. By reducing the number of dimensions, the data utilizes less space, thus allowing classification on larger datasets in less time. Further, by taking only the salient dimensions, PCA projects the dataset onto dimensions that hold the most meaning, thus drawing out patterns in the dataset [3]. PCA is a useful statistical technique that has found application in fields such as face recognition and image compression and is a common technique for finding patterns in data of high dimension. But a major problem in mining scientific data sets is that the data is often high dimensional. When the number of dimensions reaches hundreds or even thousands, the computational time for the pattern recognition algorithms can become prohibitive. In many cases there are a large number of

features representing the object. One problem is that the computational time for the pattern recognition [4].

I- MATHEMATICAL MODEL FOR PRINCIPLE COMPONENT ANALYSIS (PCA)

Principal component analysis as mentioned above 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 [5].

The number of principal components is less than or equal to the number of original variables [5]. 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, also known as Karhunen-Loeve (KL) transformation is basically a statistical technique used in image recognition and classification [6]. The main emphasis of PCA in image processing is to transform the 2D image into 1D feature vector in subspace. The authors that have used PCA for image/ face recognition such as: Jose and Gottumukkal *et al.* [7] proposed the local and global feature extraction method by exploiting Modular Principal Component Analysis (MPCA) that incorporates Modular PCA and 2D PCA method.[6]

Principal component analysis is a statistical tool used to analyze data sets. The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of large number of interrelated variables, while retaining as much as possible of the variation present in the data set [22]. The mathematics behind principle component analysis is statistics and is hinged behind standard deviation, eigenvalues and eigenvectors. The entire subject of

statistics is based around the idea that you have this big set of data, and you want to analyze that set in terms of the relationships between the individual points in that data set [23].

Principal Component Analysis (PCA, also called Karhunen-Loeve transform) is used for dimensionality reduction techniques of data analysis and compression. It is based on transforming a relatively large number of variables into a smaller number of uncorrelated variables by finding a few orthogonal linear combinations of the original variables with the largest variance. The first principal component of the transformation is the linear combination of the original variables with the largest variance; the second principal component is the linear combination of the original variables with the second largest variance and orthogonal to the first principal component and so on. In many data sets, the first several principal components contribute most of the variance in the original data set, so that the rest can be disregarded with minimal loss of the variance for dimension reduction of the data [18, 19]. The transformation works as follows.

Given a set of observations x_1, x_2, \dots, x_n , where each observation is represented by a vector of length m , the data set is thus represented by a window $X_{n \times m}$

$$X_{n \times m} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} = [x_1, x_2, \dots, x_n] \quad (1)$$

The average observation is defined as

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

The deviation from the average is defined as

$$\Phi_i = x_i - \mu \quad (3)$$

The sample covariance matrix of the data set is defined as

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T = \frac{1}{n} \sum_{i=1}^n \Phi_i \Phi_i^T = \frac{1}{n} A A^T \quad (4)$$

$$\text{Where } A = [\Phi_1, \Phi_2, \dots, \Phi_n] \quad (5)$$

To apply PCA to reduce high dimensional data, eigenvalues and corresponding eigenvectors of the sample covariance matrix C are computed. We choose the k eigenvectors having the largest eigenvalues. Often there will be just a few large eigenvalues, and this implies that k is the inherent dimensionality of the subspace governing the "signal" while the remaining $(m - k)$ dimensions generally contain noise [19].

We form a $m \times k$ matrix U whose columns consist of the k eigenvectors. The representation of the data by principal components consists of projecting the data onto the k dimensional subspace according to the following rules [28]

$$Y_i = U^T(x_i - \mu) = U^T \Phi_i \quad (6)$$

Hotelling (1933)[20] initially developed PCA to explain the variance-covariance structure of a set of variables by linearly combining the original variables. The PCA technique can account for most of the variation of the original p variables via k uncorrelated principal components, where $k \leq p$. Restated, let $x = x_1, x_2, \dots, x_p$ be a set of original variables with a variance-covariance matrix Σ . Through the PCA, a set of uncorrelated linear combinations can be obtained in the following matrix:

$$Y = A^T x \quad (7)$$

Where $Y = (Y_1, Y_2, \dots, Y_p)^T$, Y_1 is called the first principal component, Y_2 is called the second principal component and so on; $A = (a_{ij})_{p \times p}$ and A is an orthogonal matrix with $A^T A = I$. Therefore, X can also be expressed as follows:

A Mathematical Model Based on Principal Component Analysis for Optimization of Correlated Multiresponse Surfaces

$$X = A Y = \sum_{j=1}^p A_j Y_j \quad (8)$$

Where $A_j = [a_{1j}, a_{2j}, \dots, a_{pj}]^T$ is the j th eigenvector of Σ . Consequently, the secondary variables have following characteristics [21]:

$$\sum_{j=1}^p a_{ij}^2 = 1 \quad \forall i = 1, \dots, p \quad (9)$$

$$\sum_{k=1}^p a_{ik} a_{jk} = 0 \quad \forall i, j | i \neq j \quad (10)$$

Each secondary variable can be obtain from a linear combination of original variables. The first secondary variable covers maximum deviation existing in original variables.[26]

$Var(Z_1 = p_1' Y) = p_1' \Sigma p_1$ would be maximized subject to the constraint that $p_1' p_1 = 1$. It was shown that the characteristic vector associated with the largest root of the following equation is the optimal solution for p_1 and the largest root λ_1 is the variance of Z_1 .

$$|\Sigma - \lambda I| = 0 \quad (11)$$

The k th secondary variable covers maximum deviation, which is not covered by $k-1$ th one.

If the solution of Equation (5) is expressed as $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$ such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, the k th component would be the characteristic vector associated to λ_k . these secondary variables are independent.

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set [24]. This is achieved by transforming to a new set of variables, the principal components (PC's), which are uncorrelated and which are ordered so that the first few retain most of the variation present in all of the original variables[27]

II- PRINCIPAL COMPONENT ANALYSIS FOR CLASSIFICATION

PCA is widely used in the field of image processing feature reduction, feature extraction, anomaly detection, classification and pattern recognition. The following section presents some studies combines PCA with classification methods to improve the performance of these methods.

Masoud Mazloom and Shohreh Kasaei in [8] presents a hybrid approach to increase the face recognition accuracy using a combination of Wavelet, PCA, and Neural Networks. They apply a combination of wavelet transform and PCA [8]. Another attempt was done in [9] by Ganesh Linge and Meenakshi Pawa. They presented two methodologies for the face recognition; the first one is feature extraction and second is the feed forward back propagation neural network. The feature extraction is with Principal Component Analysis and classification with the help of neural network. Another study addressed same issue done by S. R. Barahate and Yadavrao at 2010. They develop a computational model to identify the unknown person's face by comparing characteristics of face to those of known individuals. Principal Component Analysis, based on information theory concepts, seek a computational model that best describe a face [10]. In [11] author investigates the use of PCA as part of model to recognize and classify sensory data. Principal component analysis used successfully for feature extraction and classification.

J. Novakovic, S. Rankov, made different attempt. They presented a comparison between several classification algorithms with feature extraction on real dataset. Principal Component Analysis (PCA) has been used for feature extraction with different values of the ratio R, evaluated and compared using four different types of classifiers on two real benchmark data sets. They reached to the result that feature extraction is especially effective for classification algorithms that do not have any inherent feature selections or feature extraction build in, such as the nearest neighbor methods or some types of neural network[12].

Recently, in 2016 Kiranjeet Kaur presented a study of Using PCA and support vector machine Classification for Heart Disease Prediction system. The main goal of their work is to develop an efficient heart disease prediction system using feature extraction and SVM classifier that can be used to predict the occurrence of disease [13].

III- PCA FOR FEATURE REDUCTION

PCA is one of the most fundamental tools of dimensionality reduction for extracting effective features from high dimensional vectors of input data [14,15]. 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 [16]. Annie George in [17] presented a study used Principal Component Analysis (PCA) for dimensionality reduction in Anomaly detection model.

The major drawback observed in PCA is that it gives no consideration to class reparability because it does not account for the class label of the feature vector [16]. 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 [16].

VI- PCA FOR ANOMALY DETECTION

Ling Huang, et. Al. proposed a method for discovering anomalies that combines distributed tracking and Principal Component Analysis (PCA). Their method was shown to work well empirically in highly aggregated networks, that is, those with a limited number of large nodes and at coarse time scales [29].

Daniela Brauckho, et. Al. [30] try to apply the popular PCA method in real world anomaly detection. They found that direct application of the PCA method results in poor performance in terms of ROC curves; they investigated the problem and found that the main source of the problem is the bias coming from correlation in prediction error terms. After a detailed theoretical analysis, it appears that the correct framework is not the classical PCA but rather the Karhunen-Loeve expansion. They have presented the KL expansion and have provided a Galerkin method for developing a predictive model. This method has thereafter been applied to data traces from the Switch network and we have shown that an important improvement is attained when temporal correlation is considered.

P. Rameswara Anand and Tulasi Krishna Kumar.K [31] propose an online over-sampling principal component analysis (osPCA) algorithm to address interactive visualization of anomalies. Their algorithm is using mixture models and the EM algorithm for anomaly detection, however their ideas can be generalized to anomaly detection in other probabilistic settings. They are implement their ideas in the SGI MineSet product as a mining plug-in using the MineSet visualizers.

V- PCA ADVANTAGES AND LIMITATIONS

The advantages of PCA are it is suitable for visualization of complex data and capturing the variation in data as possible. PCA suffers from the drawbacks of not coping well with high dimensional data and scaling up to large data set due to its prohibitive computational complexity ($O(N.d^2)$). Another shortcoming is that classical PCA may not perform well in terms of recognition for applications where local region based features have discriminant information (e.g. facial expressions, pose, illuminations, and change detection applications)

VI- CONCLUSIONS

This paper provides a comprehensive review of the literature related to Principal Component Analysis (PCA). Where, the study shows PCA as an essential tool for multivariate data analysis and unsupervised dimension reduction. Also, PCA incorporated with AI techniques to improve performance of many applications like image processing, pattern recognition, classification and anomaly detection.

REFERENCES:

- [1] Shobha R. Patil and Sanjay Pandey, "Principal component analysis: a survey", Global Journal of Mechanical Engineering and Computational Science- ISSN 2249-3468 GJMECS Vol.1 (3),.
- [2] KROONENBERG, PIETER M., "THREE-MODE PRINCIPAL COMPONENT ANALYSIS: THEORY AND APPLICATIONS. ",VOL. 2. DSWO PRESS, 1983.
- [3] Peter Wei , A Study of Principal Component Analysis on Classifiers Using Histogram of Gradients Features, final report available at: <http://www.contrib.andrew.cmu.edu/~pwei/papers/FinalReport.pdf>.
- [4] Witten, I. H., and E. Frank. 1999., "Data Mining: Practical Machine Learning tools and techniques with Java implementations", Morgan Kaufman.
- [5] Jackson, J.E. (1991). *A User's Guide to Principal Components* (Wiley).
- [6] Pereira J., Cavalcanti G., and Ren T., "Modular Image Principal Component Analysis for Face Recognition," in Proceedings of International Joint Conference on Neural Networks, USA, pp. 2481-2486, 2009.
- [7] Gottumukkal R. and Asari K., "An Improved Face Recognition Technique Based on Modular PCA Approach," Pattern Recognition Letters, vol. 25, no. 4, pp. 429-436, 2004.
- [8] Face Recognition using Wavelet, PCA, and Neural Networks, Proceeding of the First International Conference on Modeling, Simulation and Applied Optimization, Sharjah, U.A.E. February 1-3, 2005
- [9] Linge, Ganesh, and Meenakshi Pawar. "Neural Network Based Face Recognition Using PCA." International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4011-4014
- [10] Barahate, S. R., and J. Saturwar. "Face recognition using PCA based algorithm and neural network." *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*. ACM, 2010.
- [11] Burka, Zak, "Perceptual audio classification using principal component analysis" (2010). Thesis. Rochester Institute of Technology. Accessed from
- [12] Novakovic, Jasmina, and Sinisa Rankov. "Classification Performance Using Principal Component Analysis and Different Value of the Ratio R." *International Journal of Computers Communications & Control* 6.2 (2011): 317-327.
- [13] Kiranjeet Kaur, Lalit Mann Singh. Heart Disease Prediction System Using PCA and SVM Classification, International Journal of Advance Research, Ideas and Innovations in Technology, www.ijariit.com.
- [14] Smith, Lindsay I. "A tutorial on principal components analysis." *Cornell University, USA* 51 (2002): 52.
- [15] CHEN Bo, Ma Wu, "Research of Intrusion Detection based on Principal Components Analysis", Information Engineering Institute, Dalian University, China, Second International Conference on Information and Computing Science, 2009.
- [16] Varghese, Nebu, et al. "A survey of dimensionality reduction and classification methods.", *International Journal of Computer Science and Engineering Survey (IJCSSES)* 3.3 (2012): 45.
- [17] George, Annie. "Anomaly detection based on machine learning: dimensionality reduction using PCA and classification using SVM." *International Journal of Computer Applications* 47.21 (2012).
- [18] I.T. Jolliffe, "Principal Component Analysis", 2nd Ed., Springer Verlag, NY, 2002.
- [19] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification", China Machine Press, Beijing, 2nd edition, 2004.
- [20] Hotelling, H., 1933, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, 24(7), 498-520.
- [21] Timm, N. H., 2002, *Applied Multivariate Analysis*, Springer-Verlag, New York.
- [22] I.T. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer series in statistics 2002, page 1-3.
- [23] Lindsay I Smith, A tutorial on Principal Components Analysis, February 26, 2002, page 2-8
- [24] Jolliffe, I. T., 2002, *Principal Component Analysis*, 2nd edition, Springer
- [25] jolliffe , I . T . (1986) . *Principal Component Analysis* . New York : Springer-Verlag
- [26] Bashiri, Mahdi, and Taha Hossein Hejazi. "A mathematical model based on principal component analysis for optimization of correlated multiresponse surfaces." 19.3 (2012): 223-239.
- [27] Jacobsson, Martin. "Forecasting commodity futures using Principal Component Analysis and Copula." (2015).
- [28] Altaher, A., Ramadass, S., Abdelrahman, N., & Khalid, A. (2010). *Network Anomaly Detection and Visualization using Combined PCA and Adaptive Filtering*. LJS Publisher and IJCSIS Press.
- [29] HUANG, L., NGUYEN, X. L., GAROFALAKIS, M., JORDAN, M., JOSEPH, A.D., AND TAFT, N. In-network PCA and anomaly detection. In NIPS (2006).
- [30] Daniela Brauckho_, Kav_e Salamatian, Martin May. Applying PCA for Tra_c Anomaly Detection: Problems and Solutions. Proceeding of IEEE INFOCOM 2009,, Apr 2009, Rio de Janeiro, Brazil. pp.2866-2870, 2009.
- [31] P. Rameswara Anand,, Tulasi Krishna Kumar.K, "PCA Based Anomaly Detection", International Journal of Research in Advent Technology, Vol.2, No.2, February 2014 E-ISSN: 2321-9637