Faculty of Computational Sciences & Informatics - Academic City University College, Accra, Ghana
SMART Scientific Projects & Research Consortium (SMART SPaRC)
Sekinah-Hope Foundation for Female STEM Education
ICT University Foundations USA
IEEE Computer Society, Nigeria Section

37th International Science Technology Education Arts Management & Social Sciences (iSTEAMS) Cross-Border Conference - Accra Ghana 2023

Robust Principal Component Analysis: The prospects in Multi-View Learning for Facial Images

[1]Anichebe, Gregory E., [1]Mwasomola, [1]Tumaini Simon, [1] Ezeora, Nnamdi Johnson, [1] Ezugwu, Assumpta Obianuju, Uzo, Blessing Chimezie, [1&2]Nzeh, Royransom Nzeh, [1]Asogwa, Caroline, [1]Uzo, Izuchukwu & [1]Abhadiomhen, Stanley Ebhohimhen
[1]Department of Computer Science, University of Nigeria, Nsukka, Nigeria
[2] Jiangsu University, China,
E-mail: gregory.anichebe@unn.edu.ng, tumaini1405@gmail.com, nnamdi.ezeora@unn.edu.ng, assumpta.ezugwu@unn.edu.ng, blessing.uzo@unn.edu.ng, royransom.nzeh@unn.edu.ng, caroline.asogwa@unn.edu.ng, izuchukwu.uzo@unn.edu.ng, stanley.abhadiomhen@unn.edu.ng

ABSTRACT

High dimensional data has the tendency to increase the computation complexity involved in training machine learning models. As a result, it is always beneficial transforming high dimensional data into low dimensional space to increase the learning speed, reduce computational space and generally, influence the performance or accuracy of the learning algorithm. Moreover, computer vision problems such as face image recognition with potentially large dimensional data tends to profit from dimensionality reduction. However, camera poses or effects such as lighting during image capturing, face direction, wearing of sunglasses or hat may impair the accuracy of face recognition. Therefore, providing the learning algorithm with complimentary knowledge ability is necessary. For that reason, multiview learning provides the mechanism required to use complementary knowledge in specific views to improve the learning task. This paper explored common problems linked with high dimensional data and demonstrates how Principal Component Analysis (PCA) can be applied together with Multiview learning to improve the performance of learning algorithms.

Keywords: Principal Component Analysis, Multiview Learning, Dimensionality Reduction

Proceedings Citation Format
Anichebe, Gregory E., Mwasomola, Tumaini Simon, Ezeora, Nnamdi Johnson, Ezugwu, Assumpta Obianuju, Uzo, Blessing Chimezie, Nzeh, Royransom Nzeh, Asogwa, Caroline, Uzo, Izuchukwu & Abhadiomhen, Stanley Ebhohimhen (2023): Robust Principal Component Analysis: The prospects in Multi-View Learning for Facial Images. Proceedings of the 37th iSTEAMS Multidisciplinary Cross-Border Conference. 30th October – 1st November, 2023. Academic City University College, Accra, Ghana. Pp 123-130. dx.doi.org/10.22624/AIMS/ACCRACROSSBORDER2023P8

1. INTRODUCTION

In computer vision, high dimensional data is a double-edged sword because on one end, it provides an opportunity to explore a wide variety of relationships that exist amongst variables and on the other end, it increases the computational complexity that is required to train machine learning models [1]. Nevertheless, projecting a given data from high dimensional space into a low dimensional space without losing important information will not only reduce the computation complexity but, it will also help the learning model to make accurate predictions.

Moreover, it is often difficult visualizing high dimensional data. For example, if we consider a dataset that contains grayscale images represented in a vector form with 1024 dimensions, one question that comes to mind is how do we visualize such data easily? To be honest, it is usually very difficult to visualize high dimensional data however, dimensionality reduction methods such as PCA [2] with likely application area which includes face image recognition, handwritten digit recognition, and gene expression etc. may be utilized to reduce the dimensions of the dataset to aid easy visualization.

Furthermore, several research works [1-4] have proved that applying dimensionality reduction techniques of face images can reduce the complexity of training machine learning models and at the same time, increase the recognition accuracy. On the other hand, factors such as, lighting during image capturing, face direction, wearing of sunglasses or hat may impair recognition accuracy of face images [4]. Thus, it is paramount to employ a complementary knowledge from multiple views to enhance learning accuracy. In light of sense, the studies in [5] and [6] suggested that multi-view learning can be used to extract knowledge from different views.

For example, different camera poses may constitute different views that can be merge by using techniques such as fusion and co-learning [7]. Accordingly, this report will present the basic background concept in section 2 and thereafter, in section 3, it presents a brief rational for dimensionality reduction. Next, it provides an overview on multiview learning in section 4. In section 6, it illustrates with example how PCA can be used for dimensionality reduction.

2. BASIC BACKGROUND CONCEPT

More often, complex machine learning algorithms such as face image recognition models are trained using datasets with large feature. At times, some of these features are either correlated or some are less important than others and may cause the model to Overfit. Therefore, dimensionality reduction has become a widespread phenomenon that is utilized in applied machine learning to improve performance. As a result, this section explains dimensionality reduction and also describes some techniques that may be applied in dimensionality reduction.

2.1. What is dimensionality reduction?

Dimension reduction can be described as the transformation of original data (high dimensional) to a low dimensional space without losing important information [8]. The concept of dimensionality reduction may be compared to the traditional file compression (zip) which utilizes the lossless file compression method.

However, in lossless file compression, the procedure is completely revisable for the reason that, no feature is lost after decompression. In contrast, some features of data will be lost in dimensionality reduction as a result of compression. Nonetheless, dimensionality reduction hopes that the lost features may represent noise that renders no influence at all to the performance of the learned model. Therefore, the goal of dimensionality reduction is to minimize information loss while retaining variability of the data.

For example, given a set of data with p features, $\mathbf{X_1}, \mathbf{X_2}, \ldots\ldots \mathbf{X_n}$ that may be correlated. We can compute the linear transformation that maps the correlated features into uncorrelated (the linear combination of the original p features) as follows:

$$\mathbf{G} \in \mathfrak{R}^{\mathbf{p*d}}: \ \mathbf{x} \in \mathfrak{R}^{\mathbf{p}} \ \rightarrow \mathbf{y} = \mathbf{G^T}\,\mathbf{x} \in \mathfrak{R}^{\mathbf{d}} \ \ (\mathbf{d} \ll \mathbf{p}),$$

Where
d, the reduced demission after the transformation is much lesser than p (original features) See Figure 1 for illustration of this procedure.

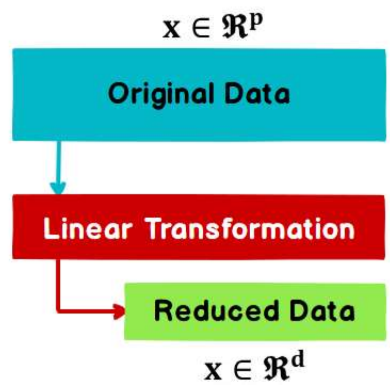

Figure 1: Linear Transformation Of High Dimensional Data Into Low Dimensional Space

2.2 Techniques for Dimensionality Reduction

There are various methods that can be utilized in dimensionality reduction. For example, Linear Discriminant Analysis (LDA) [8] may be used for supervised dimensionality reduction, and Principal Component Analysis (PCA) [2] can be used for compressing unsupervised data while Kernel Principal Component Analysis (KDA) [9] is used for nonlinear dimensionality reduction. This subsection will describe each of these methods briefly.

LDA is a generalized version of the Fishers Linear Discriminant method [10]. It can be used in supervised learning to obtain the principal component of large dataset by reducing the number of features assuming they are not correlated. By incorporating output labels, LDA reduces dimensionality while enhancing class separability, benefiting tasks like face recognition. PCA is mainly useful for compressing unsupervised data using techniques that reduces the dimension of data by finding new set of uncorrelated features (d) smaller than the original set of features (p).

PCA is widely applied in reducing dimensions in textual contents, speech recognition systems and images. PCA procedure is described in section 5. KDA, on the other hand, uses kernel to extend PCA and can be used for nonlinear dimensionality reduction through a procedure that begins by projecting the nonlinear data into a higher dimensional space. Next, it uses PCA to reduce the dimension of the projected data into the low dimensional space. Although, KDA maybe affective in dimensional reduction, it is generally computationally expensive considering the procedure described above. Figure 2 provides a tabular form of the various dimensionality reduction techniques.

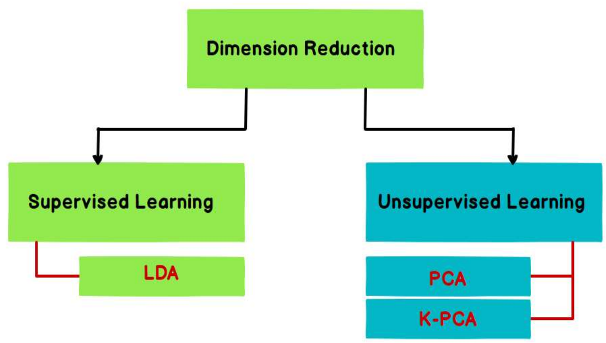

Figure 2: Dimensionality reduction techniques

3. WHY DIMENSIONALITY REDUCTION?

Most complex machine learning problems utilizes huge datasets, especially when training models using sound, textual contents or images. These high dimensional data sources are often associated with large sets of features, which can increase the space and time complexity needed to train machine learning models. Furthermore, having exceptionally large set of features may cause the model to overfit the datasets because they can end up adding more noise that is likely to reduce the accuracy of the model.

One important method often utilized to improve performance in complex machine learning tasks includes the reduction in the dimension after retaining variability of the data. Variability is retained by projecting correlated features into uncorrelated axes that reduces over-fitting. Also, as mentioned before, the complexity involved with visualizing large dimensional data equally makes the concept of dimensionally reduction very important in machine learning.

4. OVERVIEW OF MULTIVIEW LEARNING

Most traditional unsupervised leaning models tend to identify patterns in single view dataset. This contrast the currently obtainable in real-world scenario, where data such as images may be captured from multiple sources. This method of learning has become important considering that factors such as camera poses or effects such as lighting during image capturing, face direction, wearing of sunglasses or hat can impair the accuracy of face recognition in single view mode.

Recently, machine learning models are now utilizing multiple view [7] data sources to achieve similar result as the single view. For example, if we consider a dataset with p features $(X_1, X_2 \ldots\ldots X_v)$ obtained from a combination of views such as, face poses etc., we can define multi-view learning as merging the knowledge obtained from those different views to find correlated knowledge. Thereafter, we can use the complementary knowledge in specific views to enhance learning task. Additionally, there are different strategies such as translation, fusion, alignment, co-learning and representation learning that maybe utilized to deal with merging multi-view data. Also, multi-view clustering tends to group different data samples in different views by fusing the knowledge across those views to seek better accurate result.

5. RECENT WORK ON PCA AND MULTI VIEW LEARNING

PCA has been widely used over the years to improve the performance of learning models. For example, in face images where dimensions can be very large, we have shown in section 4 how PCA, relying on multi view dataset, can be used to reduce the dimension from n dimensional features to K desired principal component. This paper provides a brief overview of some recent research works on PCA and multi view learning below.

Firstly, in paper [8], a Linear Principal Component Discriminant Analysis for supervised and unsupervised learning is presented. They considered some objectives which includes, projecting high dimensional data into a low dimensional space with maximum variance. Next, they established a framework to analyze the data through simulation to evaluate the performances of their proposed methods. Though they obtained satisfactory result which shows the robustness of their system, however, their method didn't cover the kernel approach required for nonlinear dimensionality reduction.

Considering the above problem, the study in [11] presented a Nonlinear Feature Extraction based on Fisher Discriminant Analysis to extract nonlinear feature using the kernel method. In the same line, Alkandari and S.J. Aljaber [4] presented a system that automatically identifies facial image. Their method is based on current expectations arising from the continuous growth in Artificial intelligence that requires an increase in training and recognition speed of learned models.

The above research utilizes the single view dataset which fall short of current obtainable real-world scenario where data such as images may be captured from multiple sources. Along this line, Ding et al. [7] observed that factors such as camera poses or effects such as lighting during image capturing, face direction, wearing sunglasses or hat may impair the accuracy of face recognition. Hence, they proposed multiple view learning method that uses data from multiple sources, such as face poses to achieve better result than the single view models. In a more recent work by Huang et al. [5] and Sun et al. [6], they presented a system based on generalized Eigenvalue Proximal SVM.

6. PCA PROCEDURES

PCA is a very powerful algorithm that can be used to solve dimensionality problems. To solve dimensionality problems, PCA algorithm analyses the dimensions of the datasets to find those points with optimum variance. Next, it creates new sets of uncorrelated point such that, the linear relationship between the uncorrelated and correlated feature is maintained. In this section, the PCA algorithm is described below.

6.1 Mean Normalization

Assuming we have face images from different views as training dataset (X) with m number of training examples and n number of features, we can represent these features as an n dimensional vector with the dimension of each training example as $\mathrm{m} \times \mathrm{n}$ which represent the pixel intensity of the image at each point. Suppose we want to reduce the dimension of the training examples from n dimension to k desired number of dimensions, then we can compute the linear transformation of n to k using:

$$G \in \Re^{n*k}: \; x \in \Re^{n} \;\; \rightarrow y = G^{T} x \in \Re^{k} \;\; (k \ll n) \quad (1)$$

However, to begin the linear transformation, PCA requires that the features are all in the same range. Hence, we perform mean normalization for all the features as follows: Firstly, we compute the mean for each feature dimension and store the value after subtracting it from the dataset. Next, the standard deviation of each feature is computed, each feature is the divided by its standard deviation.

$$x' = \frac{x - \bar{x}}{\sigma} \qquad (2)$$

We can implement the above equation quickly by using built-in functions in MATLAB and octave as follows:

```
mu = mean(X);
sigma = std(X);
X_norm = (X. - mu). /sigma            (3)
```

In summary, mean normalization ensures all features have zero mean (same range of values).

6.2 Compute the Covariance Matrix

Next, we create the covariance matrix using the normalized dataset as follows:

$$\text{Conv}= \frac{1}{m} \sum_{i=1}^{n} (x^{i})(x^{i})^{T} \qquad (4)$$

The vector form of the equation above can be rewriting as:

$$\text{Conv} = (1/m) * (X' * X), \qquad (5)$$

where m is the number of training example, accordingly, if we compute the inner product of $X^T$and X with the dimension of the transpose of X as $n \times m$ and the dimension of X as $m \times n$, then, the resulting covariance matrix will be an $\mathbf{n} \times \mathbf{n}$ square matrix. Therefore, the covariance matrix measures the degree to which corresponding features from two sets of ordered data move in similar path.

6.3 Compute the Eigen values and Eigen vectors
After computing the covariance matrix of X, PCA uses the variance obtained to reduce the dimension of n features to the desire principal component K. It projects the correlated features from the original dataset into uncorrelated orthogonal axes with optimum variance. Then, we can compute the Eigen value that defines the magnitude of the principal component and Eigen vector that is the projection of n onto K principal component.
We can compute the Eigen vectors of the Covariance matrix using MATLAB and Octave built in functions that computes the Singular Value Decomposition (SVD)

$$[U, S, V] = svd(Conv)$$
$$s.t\ Ax = \lambda x. \qquad (6)$$

where λ is the Eigen value, x is the Eigen vector and A is the original dataset (X)

6.4 Project the data onto the Principal Component
Once we have obtained the Eigen vector of n x 1 dimension, we project the original data onto the principal component by transforming n x 1 → n x K dimensional matrix. The data is projected onto the principal components as follows:

$\mathbf{Ureduce} = \mathbf{U}(:, \mathbf{1}{:}\mathbf{K})$: where K represents desired number of dimensions. The reduced data becomes an m x k dimensional matrix provided by the equation given below:

$$Z = X * Ureduce; \qquad (7)$$

6.5 Reconstruct an Approximation of the Data from the Top K Principal Component
As soon as we have successfully projected the data onto a low dimensional space, we can utilize the same procedure employed in lossless traditional file compression method that reverses compression through decompression without losing any information. Nevertheless, PCA aims to minimize the lost resulting from data compression. Hence, it's probable not to recover the exact same copy of the data, but we can approximately get back the original data by projecting the reduced data in low dimension back to the original high dimensional space $(m\,x\,k * k\,x\,n \rightarrow m\,x\,n)$.

We can reconstruct an approximation of the data from the top K principal components as follows:

$$X\_recover = Z * Ureduce^T \qquad (8)$$

As can be observed from the above, dimensionality reduction based on PCA can reduce the space required for computation. Furthermore, it can reduce the chances of over-fitting while also reducing the time it takes to train learning models.

7. CONCLUSION

Principal Component Analysis (PCA) can reduce the computational complexity when training models with high dimensional data, and it also makes virtualization easier. On the other hand, implementing PCA could be computational expensive. However, the benefits outweigh the negative side. Similarly, it was observed that factors such as movement of part of the face, wearing of glasses or hat may prevent facial recognition. Consequently, multi view learning provides ways to complement knowledge in specific views.

REFERENCES

1. Z.Xingfu and R. Xiangmin. "Two Dimensional Principal Component Analysis based Independent Component Analysis for face recognition", in 2011 International Conference on Multimedia Technology, 2011.
2. M.A. Mansoori and M.R. Casu "Efficient FPGA Implementation of PCA Algorithm for Large Data using High Level Synthesis",in 2019 15th Conference on Ph.D Research in Microelectronics and Electronics (PRIME). 2019.
3. W. Puyati and A. Walairacht, "Efficiency Improvement for Unconstrained Face Recognition by Weightening Probability Values of Modular PCA and Wavelet PCA", in 2008 10th International Conference on Advanced Communication Technology, 2008.
4. A. Alkandari and S.J. Aljaber, "Principle Component Analysis algorithm (PCA) for image recognition", in 2015 Second International Conference on Computing Technology and Information Management (ICCTIM). 2015.
5. P. Huang et al., "Multi-View Learning With Robust Generalized Eigenvalue Proximal SVM". IEEE Access, 2019. 7: p. 102437-102454.
6. S.Sun., X. Xie, and C. Dong, "Multiview Learning With Generalized Eigenvalue Proximal Support Vector Machines". IEEE Transactions on Cybernetics, 2019. 49(2): p. 688-697.
7. Z.Ding, H. Zhao, and Y. Fu, "Multi-view Transformation Learning, in Learning Representation for Multi-View Data Analysis: Models and Applications", Z. Ding, H. Zhao, and Y. Fu, Editors. 2019, Springer International Publishing: Cham. p. 99-126.
8. Y. Pei, "Linear Principal Component Discriminant Analysis", in 2015 IEEE International Conference on Systems, Man, and Cybernetics. 2015.
9. C. Liu, et al,"Design and application of Compound Kernel-PCA algorithm in face recognition", in 2016 35th Chinese Control Conference (CCC). 2016.
10. H. Atasoy, S. Yıldırım, and E. Yıldırım, "Emotion recognition from speech using Fisher's discriminant analysis and Bayesian classifier", in 2015 23nd Signal Processing and Communications Applications Conference (SIU). 2015.
11. X. Xingjian, et al, "Nonlinear Feature Extraction Based on Kernel Adaptive Marginal Fisher Analysis for Target HRRP Recognition", in 2018 14th IEEE International Conference on Signal Processing (ICSP). 2018.
12. A.Ng, Machine Learning [Online]. Available from: https://www.coursera.org/learn/machine-learning/home/welcome. Accessed on: Jan 1, 2020