November-December 2019
ISSN: 0193-4120 Page No. 3907 - 3912

Principle Component Evaluation for Analyzing the Design Trends in Fashion Industry

A. Mary Posonia[1], Sai Vaishnavi.R[2], Deepika.P[3]
[1]Associate Professor, Department of Computer Science, Sathyabama Institute of Science and Technology, Chennai -119, India
[2,3] UG Student , Department of Computer Science and Engineering, Sathyabama Institute of Science & Technology, Chennai-119, India
maryposonia.cse@sathyabama.ac.in, soniadelicate@gmail.com

Article Info
Volume 81
Page Number: 3907 - 3912
Publication Issue:
November-December 2019

Article History
Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *19 December 2019*

Abstract:
Principle component analysis has its applications in many fields like finance, data mining, psychology, bio information, etc., as it is used to find patterns by reducing the dimensions of the data. The following paper proposes an idea to apply Principle Component Analysis (PCA) on the attributes of the clothes collected in a synthetic dataset to predict the class of the most used designs, colours, patterns, fabrics, etc. to design new clothes. This will assist the designers to design models that are most liked by the people thus increasing the sales with less effort and investment.

Keywords: *PCA, Eigen vectors, Discrimininator, Imgae processing, Support Vector Machine*

I. Introduction

The data from the real world is obtained through surveys, transactions processing, and many other sources. The data collected in this format cannot be directly used for analysis or prediction hence, has to be cleaned. There are various methods in the cleaning process for the removal of noisy data, missing values, and inconsistent data. The data obtained from this step is generally multidimensional in nature and for the analysis process to be simple and fast we have to reduce the dimensionality of the data this is done using the Principle Component Analysis. PCA is implemented in diverse fields and in their applications to reduce the dimensionality of the data collected. In this paper, we are going to see how the Principal Component Analysis (PCA) works and can be used in the fashion fields. Generally, in the fashion field, Principle Component Analysis is used for automatic classification and detection of fabric defects missed in the visual inspection and overcoming its drawbacks like accuracy, consistency, and speed. It also helps in reducing labour costs, thus increasing product quality and manufacturing efficiency. PCA is also used in practical sizing system to improve sizing for clothing by collecting the anthropometric measurements, this is achieved in such an acceptable manner to the apparel industry. But in this paper, we are proposing an idea to increase sales of designs by using Principle Component Analysis.

II. Related Work

In image processing field principle component analysis is used to detect patterns in training data of complex dimensions by reducing its

November-December 2019
ISSN: 0193-4120 Page No. 3907 - 3912

dimensionality as proposed in paper [1] and calculates the Eigenvectors from the covariance matrix. Where Eigen values represent the whole training set.

In face recognition, Principle Component Analysis (PCA) and 2DPrinciple Component Analysis (2DPCA) are used to reduce the problem of having one training sample per person as proposed in paper [2]. Where PCA is used to reconstruct the training image.

PCA is also used in reducing data storage requirements for a machine learning algorithm as proposed in paper [3]. This is implemented by using differentiation and combining it with an existing similarity detection algorithm. The differentiation is performed using PCA.

Paper [4] proposes Principle Component Analysis (PCA) design based on Hadoop. This Hadoop-based design is used on an embedded heterogeneous platform.

PCA is also used in the real estate field for real property evaluation as proposed in paper [5] by reducing the variability of a dataset. This paper also gives methods for checking the sufficiency of data to PCA, since it is an important parameter for successful PCA implementation.

PCA and Gabor wavelets are also used for the inspection of fabric defects as proposed in paper [6]. This helps in increasing the quality of the fabrics. PCA using singular value detection is used to reduce the dimensions of feature vectors.

The iris regions are classified using PCA and simple vector machines. As proposed in paper [7] the PCA is used to extract the features obtained from the cropped regions. The outputs after applying PCA will be the inputs to the SVM's and ending up with classification.

Finally, this literature review concludes the PCA is used in different fields for different applications with different methods, but PCA is less used in the fashion industry as of now. So we proposed an idea to increase sales in the fashion industry using PCA.

III. Methodology

Data Reduction

The data predominantly contains many dimensions and views and all these values can't be considered while the analysis hence the data has to be reduced in dimension for better understanding and representation as explained by Jolliffe [8]. When the data contain, variables (p) is summarized into smaller sets of (k) derived variables (synthetic, composite). This assists us to represent the data in a more efficient way and also improves the ease of understanding but in this process, there is a loss of significant data (oversimplification). The pictorial representation of data point reduction from p-dimension to k-dimension is mentioned in the figure 1.

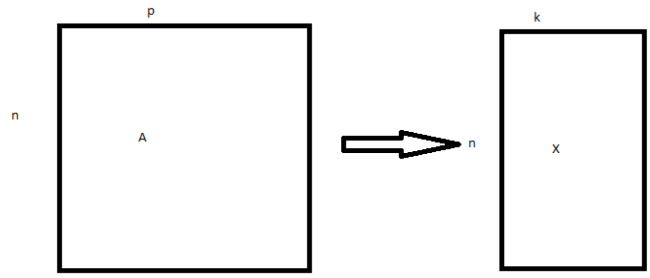

Figure 1: Reduction of data from p dimensions to k dimensions

History of Principle Component Analysis

This method was first used by Hotelling and Pearson in the year 1930s and was implemented in ecology by Goodall in 1953 but the method was named as "factor analysis." Factor Analysis is a method that follows the idea of compressing a vast amount of variable data into a smaller chunk that captures the essence of the primary data.

Math behind PCA

Here a data matrix of n objects and m variables that represent these objects, which are correlated,

November-December 2019
ISSN: 0193-4120 Page No. 3907 - 3912

is considered and it is summarized into uncorrelated axis that is also known as principal axis and also popularly called as principal components that are obtained by calculating the linear combinations among the original p variables considered in the data collected about the n objects.

Each of the data points of the objects (n) is projected to focus like a cloud in the multi-dimensional space with a pivot for every one of the m factors. The centroid of these data points is characterized by the mean of all values of the variable. The difference of each factor is determined by executing the normal squared deviation of its n esteems around the mean of that variable calculated using the formula.

$$V_j = \frac{1}{t\text{-}1} \sum_{n=1}^{t} (Y_{jn} - \bar{Y}_j)^2$$

Where

V_j – Variance of j^{th} variable

t – Total no.of data objects

Y_{jn} –The value of variable j in object n

$\bar{Y}_j$ – Mean of variable j

How much the factors are directly related is spoken to by their covariance is given by:

$$C_{jk} = \frac{1}{t-1} \sum_{n=1}^{t} (Y_{jn} - \bar{Y}_j)(Y_{kn} - \bar{Y}_k)$$

C_{jk} – Covariance of variable j and variable k

Y_{jn} – The value of variable j in object n

$\bar{Y}_j$ – Mean of the variable j

Y_{kn} – The value of variable k in object n

$\bar{Y}_k$ – Mean of the variable k

The Principle Component Analysis represents a method that rotates the axis of the m-dimensional space formed by the cloud of data points to alternative positions (principal axis) such that the new axis secures the following properties:

- The axes are ordered such that the principal axis 1 has the highest variance, principal axis 2 has the following most elevated fluctuation, etc. and principal axis m has the least difference.
- The covariance among each pair of the primary principal axis is zero that is the essential principal axes are uncorrelated to one another.

PCA uses the Euclidean Distance formula calculated from the m variables as the proportion of uniqueness among the n objects considered. Principle Component Analysis determines the most ideal k dimensional where k value is not as high as m value and hence they portrayal of the Euclidean separations among items. The polynomial math for discovering the essential pivot is summed up to the acquired m factors:

- The course of PC 1 is of the most extreme difference in the m-dimensional cloud of data points considered.
- PC 2 is the axis toward the following most elevated fluctuation, subject to the imperative that it has zero covariance with the determined PC 1, etc. up to PC m.

Where each principal axis comprises a linear combination of the original two variables

$$\mathbf{PC_j = a_{i1}\, Y_1 + a_{i2}\, Y_2 + \ldots\, a_{in}\, Y_n,}$$

Where a_{ij} - the coefficients for factor i, multiplied by the measured value for variable j

Eigenvectors and Eigenvalues

Eigenvectors: Principal components (PC) reflect both common and unique variance of the variables and may be seen as a variance-focused approach

November-December 2019
ISSN: 0193-4120 Page No. 3907 - 3912

seeking to reproduce both the total variable variance with all components and to reproduce the correlations change with all segments and to repeat the connections. A factor with a low eigenvalue is contributing little to the explanation of variances in the variables and may be overlooked as redundant with respect to more crucial factors.

Eigenvalues: These values are commonly termed as characteristic roots. An eigenvalues of any considered factor measures the value of the variation in the total sample. The eigenvalue of a factor may be computed as the sum of its squared factor loadings for all the variables. If a factor has a low eigenvalue, then it contributes a little to the variance of the data, and hence the smaller eigenvalues can be considered redundant.

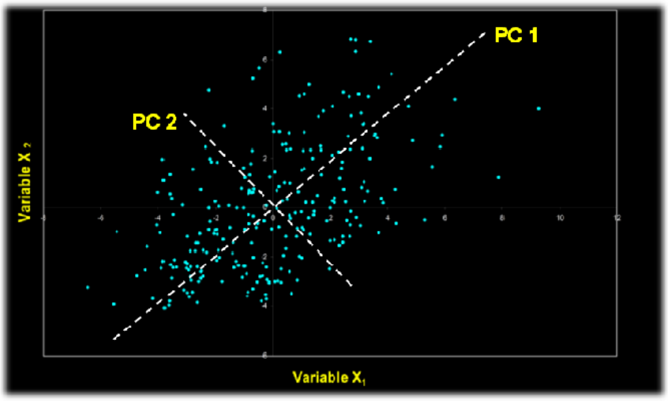

Figure 2: Plot representing the principal component axis

Consider the primary k central parts characterizing the k-measurements of the "hyper plane of best fit" to the point cloud of the all-out fluctuation of all p factors:

The central segments from 1 to k speak to the most extreme conceivable extent of that difference which can be shown in k measurements.

That is the squared Euclidean separations among points determined from their directions on PCs 1 to k obtain the most ideal portrayal of their squared Euclidean separations in the full p measurements. This principal is widely being applied in the field of medical science to find the DNA patterns in the genes and to find the gene sequence in the endangered species. This technique can also be applied in cricket game and also in fashion Designing.

Application of PRINCIPLE COMPONENT ANALYSIS in Fashion Designing

The data about the sales of clothes from various clothes shops from across the town is collected and plotted and this process is carry out for all the outlets of stores across the country. The thus obtained data will have the dimensions like colour, design, material, texture, price, brand, a season of sale, and model of design; size, etc. all this data is plotted in multidimensional graphs.

The data can be reduced by using the PRINCIPLE COMPONENT ANALYSIS method, and the hence obtained compressed data is plotted and the analysis and future prediction are done. The future prediction can be of the following types.

1. Which colour and design are more popular.
2. When are more people buying a specific material.
3. What should be the time of sale.
4. What are the products that need more discounts.
5. What are the profits by increasing the stock in one place.
6. What are the new places and materials of investment.
7. Which places new shops can be opened.

The concept of PRINCIPLE COMPONENT ANALYSIS can also be applied to many other fields like sports predictions, weather forecasting, and stock market analysis.

IV. Results and discussions

The proposed idea is applied to a synthetic data set which consists of 25 tuples and 5 attributes.

The tuples are sleeve length of the dress, wedge length in cm, collar type, waist size in cm and class the dress belongs to.

Table1: synthetic data set.

	sleve length	wedgelenght	collar type	waist size	class
20	30	100	2	36	C
21	45	120	1	28	C
22	15	100	2	38	C
23	45	130	1	36	C
24	15	130	3	40	C

The classes considered are Class A: the cotton material in colour red
Class B: the synthetic material in pink colour
Class c: the wool material in blue colour.

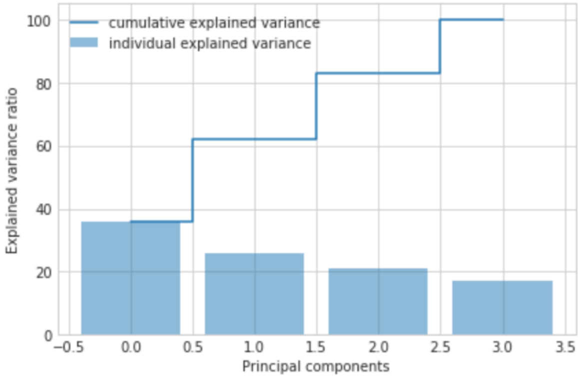

Figure 3: Variance of Calculated Principal Components

The above graph is obtained by plotting the variance of the four principal components calculated.

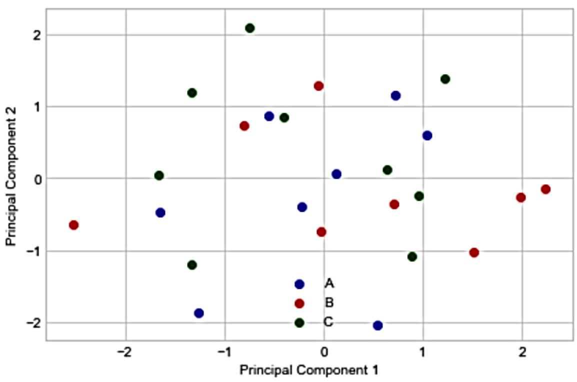

Figure4: Scatter Plot representation of the Reduced Data.

As it can be observed that the dataset considered here has 4 attributes and by following the method of Principle Component Analysis we have future reduced the dimensions to 2.

References

1. Abdulrahman Alkandari, Soha Jaber Aljaber.," Principle Component Analysis Algorithm (PCA) for Image Recognition", 2015 IEEE.
2. Luo Min, Liu Song., "Face Recognition Based on PCA and 2DPCA with Single Image Sample", IEEE 2012, Ninth Web Information Systems and Applications Conference.
3. Saritha kinkiri, Wim J.C. Melis., "Reducing data storage requirements for machine learning algorithm using Principle Component Analysis", 2016.
4. Sheng-Yen Chen, Chia-I Wei., "Hadoop-based design of Principle Component Analysis (PCA) on embedded heterogeneous platform", 2017 IEEE.
5. N.B. Chaphalkar, Sayali Sandbhor.," Sample Sufficiency for Principle Component Analysis in Real Property Valuation", SAI Computing Conference 2016 July 13-15, 2016 | London, UK, 2016 IEEE.
6. Alper Basturk, "INSPECTION OF DEFECTS IN FABRICS USING GABOR WAVELETS AND PRINCIPLE COMPONENT ANALYSIS", 2007 IEEE.
7. Nor'aini A.J, Rohilah S, Azilah S., "Classification of Iris Regions using Principle Component Analysis and Support Vector Machine", 2013 IEEE International Conference on Signal and Image Processing Applications (ICSIPA).
8. I.T.Jolliffe, "Principle Component Analysis", Second Edition,2002.
9. A. Mary Posonia, V.L.Jyothi (2016),” Extraction of perfect protein sequences with minimal processing cost using enhanced B+ tree algorithm”, Biomedical Research, special issue on S12345-S6789
10. A. Mary Posonia, Dr.V.L.Jyothi(2015), ”Improving Data Access Performance by Reverse Indexing”, International Journal of engineering and Technology(IJET),Vol 7 No 3,pp-1057- 1061

November-December 2019
ISSN: 0193-4120 Page No. 3907 - 3912

11. Mary Posonia, Dr. V. L. Jyothi, “XML Document Retrieval by Developing an Effective Indexing Technique”, in IEEE International Conference on IcoAC, MIT, Chennai, 2014, IEEE , **DOI:** 10.1109/ICoAC.2014.7229758, ISSN - 2377-6927
12. Vimal Kumar S., Vasudevan S. and MaryPosonia A, “Urban Mode of Dispatching Students from Hostel”, ARPN Journal of Engineering and Applied Science ,2017 ,Vol.12, No. 13.