January-February 2020
ISSN: 0193-4120Page No. 10654 -10658

Survey on Different Dimensionality Reduction Techniques using Machine Learning Framework

V. Jagadeeswar Reddy[1], R. Sheeja[2]

UG student[1], Assistant professor[2]
Department of Computer Science and Engineering
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,
Chennai, India
v.nanireddy@gmail.com[1], cjabbn@gmail.com[2]

Article Info
Volume 82
Page Number: 10654 -10658
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 19 February 2020*

Abstract
The data are available from the cloud storage. For the particular method or algorithm all the data are get uploaded in the server. When we need from the upload data only the essential part is get fetched from the server. This essential data can be fetched using the data reduction process. This process can use the method of Fisher data analysis. It can analysis the data in the cloud path can check the requirement what they need, based upon the need the intrinsic information is fetched. The dimensionality reduction can put forward the exact localization of the data in the server path. This method is said to be Locality projection method. In this paper they proposes the use of the both LPP and the FDA which is the together to form LFDA. This LFDA can easily fetch the essential data from the server among common data with the help of the Eigen value method. The data classification and the visualization can takes the special step in this method to locate the required data. The simulation has been made using the LFDA method for the study.

Keywords: *LFDA, Eigen values, LPP, FDA, Data Reduction.*

1. Introduction

The data in the sets can be reduced by using the dimension reduction technique. The Dimension reduction can involves a pre processing stage in which the required data can be localized from the overall data. The dimension reduction can involves in two steps first one is the data extraction and the second one is the data localization. The data localization can locate the particular need of the data from the common data. After the localization of the data, the localized data is get extracted from it and use it for the future use. During this process the two types of noise is get takes place the DR can removes the two sets of noise one is the independent random noise and the other one is the un-wanted degrees of signal. This noise can create a loss of data or the required data is not meets the result. But in this DR they can eliminate the noise. The data in the sheet are localized based upon the dimension if the data are in the 3*3 matrix from that the required data is localized by using the row and the column of the data. The Eigen value method is followed to fetches the data. This paper can proposes the method LFDA this can localized data in the matrix set after the localization the matrix Eigen value can be noted from that Eigen value data is extracted.

The Eigen value has both rows and the column value, when we provide the Eigen value the data in the Eigen value is extracted. This model can be explained by and example when the cloud as the all the data of the particular user. When the server or the other user needs the DOB and age of the particular user they fetches only the DOB data and find out the both results. The Dimension reduction can involves a pre processing stage in which the required data can be localized from the overall data. The dimension reduction can involves in two steps first one is the data extraction and the second one is the data localization. The data localization can locate the particular need of the data from the common data in this the required data is localized from the common data and extract the data result without any loss. For fetching particular data the matrix method can be used this can gives the exact result without any change and loss. The low dimension SVM method is employed to localize and

January-February 2020
ISSN: 0193-4120Page No. 10654 -10658

the extract the data much faster and easier. The classification of matrix set can be employed to predict the data from the sets. Based upon the dimension of the matrix the reduction can be made according to the Eigen values. This method is faster and effective compared other algorithm LFDA which localized the particular Eigen value of the matrix and extract the data from the sets.

2. Literature Survey

Chen-fuchianget., al., proposed about the dimensionality reduction in the graphs. The lancos algorithm has been used in the reduction technique in which the compute graph has get involved. The CG includes the compute multi partied the compute bipartite. This system has extended the limit towards the compute bipartite in which the broken set of matrix gets removed. The CBG method is get implemented in which the broken of matrix set can be involves in the graphical manner. Consider a 3*3 in which the broken set is resulted into 3 sets. The edges of the matrix set is get broken for the 5 sets of nodes the matrix is get resulted as the 5*5 matrix from the N*N matrix. The CBG model is get implemented as the broken symmetry is evolved. Based upon the number of broken sets the matrix size is get determined. From this sets the data inside is get extracted. The data reduction can be made using the broken symmetry in the matrix model. The k means edge system is implemented at which set is want to extract form the entire matrix system. The past Lanczos algorithm is expanded to provide the accurate set of data. This model is much efficient and time consuming [1].

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}$$

CédricTraizetet, al., proposed the dimensionality reduction is not only involved in the data extraction it also gives the extent towards the image system. The image can be get localized using the particular part, the intense data at the particular area is get extracted. The boundaries have been made at the overall image. At each set of part the dimensionality reduction takes place but the efficient of the image is not obtained as per our expected needs. This system had extent the hand towards the PCA system, it creates the matrix set throughout the image and the sections are get separated with the boundaries. By using the PCA the k system can locate the specific set of matrix part and localized the data, by using the PCA the intense high clarity image is get obtained. Compared to the three methods Auto decoder and Koherns and PCA. The PCA system can provides the exact result in which the resolution of the image is also high with less time consumption. It can provide more accuracy compared to other algorithm [2].

WenboYu ; Miao Zhang et., al., proposed the dimensionality reduction in the hyperspectal method which requires a large storage space. The dimensionality reduction is high because of the exact localization of the data from the sets. So the space requirement and the accurate extraction of the result this paper proposes the method manifold learning. This can provides the exact data to get extracted. The manifold learning can classified the data based upon the class similarity, class size class diversity. Based on these stages the data reduction has been made. Here the feature extraction and the feature localization has been made. The localization at the matrix set is based upon the values of the sets. The particular set data is get extracted and shown as the result. The pre processing is involved in the process to provide the accurate value. The input value which we will provided is compared with the data in the sets from the entire data. Data is get matched the matrix set address is encountered and the data inside the set is extracted [3].

Wei Wang ; Wei-guoShen ; Ya-xin Sun ; Bin et., al., proposed the less density of data is low impact in considering or to extract. To avoid this

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2/2\sigma_i^2\right)}$$

dimensionality is made to high to make to data to be intense. As the high dimensionality is implemented the space in the server gets consumed large area. To avoid this low dimensionality technique is used with image intensity should be high so the data localization is made easier. As the data fetches from the localized part is projected in the resultant area. The goal of this paper to use the low dimensionality reduction system with intense image. The data are get cluster at the initial stage based upon the formation of boundary the data at each set can be extracted with high clarity. Noise in the sets can be removed with the help of the DR the intense noise and the unsupervised random detecting noise. After the removal of the noise the data which is obtained is in pure form. The comparison of the data is made in the each sets with the help of the rows and the columns. The result obtained is more accurate [4].

HongleiZhang; MancefGabbouj et., al., proposed the two methods PCA and the LDA are the main in the dimensionality reduction. They can reduce the dimensions of the vector in the matrix set. The edge predicting is not provides the better result. The PCA and the LDA can face several problems. The PCA is not get expanded in the sets label. The LDA can used in the label sets but the data extraction form the label set is not much effective. To make the system much better this can proposes the novel based dimensionality reduction. They can find the data in the each sets and locate Eigen area. The data in the multi labeled are getting separated from the various distances. To locate the distance and the localized area this approach can pays a way. The distance between the two sets of the data can be monitored using this method. So that the extraction of the data is more easy. The distance can be calculated with the help of the graph. The graph shows a linear movement and the extraction of the data is also takes place in the linear. The class can give the labeled set of data in each dimension [5].

January-February 2020
ISSN: 0193-4120Page No. 10654 -10658

LinaXun ; Qing Yan et., al., proposed the hyper spectra cal images due to the huge amount of information. This method can uses the high dimensionality to predict the data in the labeled samples. So the need for the space is high. In this paper they proposes the use of the class probability semi supervised learning. This method can be labeled to the limited number of samples. The samples are get plotted in the geometrical area, the data values are within the boundary. The extension has to extended for the various labeled samples. The samples are in the area away from the boundary, the extraction method has get implemented from the system manner. The multi labeled samples are get provided by the graph which shows the amount of data gets extracted. The size of the labeled samples is to get extended in the range of N*N matrix. The matrix whose data can be fetched from the labeled samples. The samples are provided with the sufficient data and the extraction can be made by the class probability semi supervised learning. The DR can plot the data in the graph so the data distance can be identified and the clustering of the sets has been made. From the Eigen value data can be extracted [6].

Christian Uhlet., al., proposed the DR technique has been implemented in the signal processing stage in which it can't provide the effective data. So in this paper they propose the method of dynamic component analysis the classification of the matrix set is made by the boundaries. The Eigen values can provide the data from labeled samples. The localization of the specified sets can be observed. The input data is get compared with the data in the sets. The data in the specified set is localized and the exact distance between the one to another data is measured based upon the measured values. The labeled is get marked in the graph. The graph shows the type of data is to be extracted. This system is more effective and accurate [7].

Wei Wanget., al., proposed the web service in which the SOAP system is used as the protocols for the communication system. The dimensionality reduction can locate the set in the matrix using the Eigen values. The data in the multi labeled are getting separated from the various distance. To locate the distance and the localized area this approach can pays a way. The distance between the two sets of the data can be monitored using this method. The matrix whose data can be fetched from the labeled samples. The samples are provided with the sufficient data and the extraction can be made by the class probability semi supervised learning. The localized data is get extracted from the set and it gets transmitted to the user by using QOS method [8].

Aina Suiet., al., proposed about the future work in which the video retrieval in the dimensionality reduction. In present work for extracting the data the high dimensional reduction has been needed. For the video retrieval the dimensionality should be high. This paper proposes the low dimensional reduction with the intense data in the matrix set. So the data is more effective and the quality of the data in the each specific part is more accurate. The use of the AlexNet it can built a framework using the deep learning technique. Data can be accessed from the frame work set. Gathered information are get collected in the data set in which the localization of the data in the matrix can be predicted using the Eigen [9].

Xiaoxiao Maet., al., proposed the projection of the dimensional reduction. The entire data is clustered in the set. From that clustered data the essential data is extracted is done by the localization and the extraction method. The data are accumulated in the form labeled samples. The classification is carried out in the entire set of data. The matrix is get separated in a individual pattern. The graph has been created and the data are get plotted in the sets of the matrix. From the preferred rows and the column the data has been extracted. The SLGE system is proposed in which it uses the low dimension technique system is get implemented. It can save the storage memory and the reduction path. The hyperspectral system is involved in which the data is localized in the specific set of boundaries. The distance between the data in the labeled samples is predicted by the method and the data featuring has been made with reduced noise signal in the dimension matrix [10].

3. Various Dimensionality Reduction Algorithms

3.1:-Principal component analysis:-PCA is one of the most popular dimensionality reduction techniques used to reduce the dimensions of the data from 3d to 2d. It is used for noise filtering and data visualization and much more. PCA reduces no of features between the data sets by detecting the correlations between the data sets present. PCA aims to find the directions of maximum variance in high-dimensional data and projects it onto a new subspace with equal or fewer dimensions than the original one.[5]

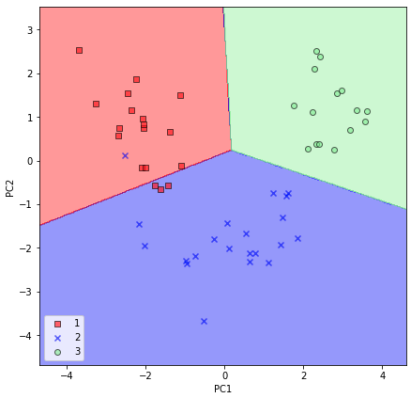

Figure 1: PCA reduction rate

3.2:-Linear discriminant analysis:- similar to PCA LDA is a dimensionality reduction technique used to reduce the dimensions. LDA will also find the feature that maximize the separation between multiple classes.
PCA is considered as an unsupervised algorithm where as LDA is a supervised algorithm.

The general LDA approach is very similar to a Principal Component Analysis (for more information about the PCA, see the previous article Implementing a

January-February 2020
ISSN: 0193-4120Page No. 10654 -10658

Principal Component Analysis (PCA) in Python step by step), here we have used iris data set to extract the result of LDA.

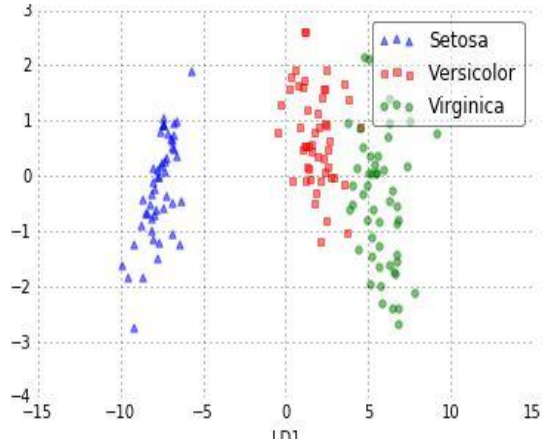

Figure 2: LDA reduction rate for iris data set

3.3:-Independent component analysis: ICA is based on information theory and is also one of the most widely used dimensionality reduction technique. The main difference between PCA and ICA is PCA always looks for uncorrelated factors whereas ICA always looks for independent factors and classes. It decomposes the mixed signal into its independent sources' signals.

3.4:-T-distributed stochastic neighbor embed:-

T-SNE is one of the few algorithms which is capable of retaining both local and global structure of data at the same time.

It calculates the probability similarity of points in high dimensional space as well as in low dimensional space.

Table 1: Comparison of various dimensionality reduction techniques

Parameters	PCA	LDA	ICA	T-SNE	PACK
Reduction rate	62%	73%	73%	81%	89%
Accuracy on validation set	74%	82%	81%	85%	87%
Best threshold	-	0.03	-	0.2	-
Avg precision	0.176	0.239	0.143	0.254	0.147
AuC	72%	81%	76%	82%	81%

4. Proposed System

Pack algorithm:- Pack algorithm is a dimensionality reduction algorithm. It is a usual classification and regression model with a closed form step that coordinates both. The resulting nonlinear low dimensional classifier achieves classification errors competitive with state of art and it is fast at training and testing when compared to other techniques or algorithms. Detection of point anomalies is a very important issue in a large scale of fields from Astronomy and Biology to network intrusions. Clustering has been employed by many researchers to solve such problems and LFDA seems like the most efficient technique. Due to its high computational complexity, this work focused on decreasing it by decreasing the dimensionality of the data points. For this reason, pack algorithm and clustering algorithm applied on the new data sets provided by LFDA. Our analysis shows that a speedup of 25% was achieved while the quality was 80% reducing the dimensionality of data set to half**.**

5. Result

The dimensions have been reduced and we can visualize the different transformed components. There is very less correlation between the transformed variables. We can see that the correlation between the components obtained from PACK algorithm is quite less as compared to the correlation between the components obtained from others. Hence, PACK algorithm tends to give better results.

6. Conclusion

We have surveyed different dimensionality reduction techniques in machine learning framework and it is compared with pack algorithm. And the main of the paper is to extract the particular set of data from the entire data. For this dimensional reduction method is used. Here for the data reduction two methods has been carried feature localization and the feature extraction. The feature localization is to locate the data from the entire matrix set by using the Eigen value. From the located set the data from the samples is get extracted. The extraction can be made using LFDA method. This method shows the reduction data as accurate and the system efficiency is maximum.

References

[1] Dimensionality Reduction of the Complete Bipartite Graph with K Edges Removed for Quantum Walks Viktoria Koscinski ; Chen-Fu Chiang IEEE 2018.

[2] Temporal Dimensionality Reduction for Land Cover Map Production Using High Resolution Image Time Series Jordi Inglanda; CédricTraizet IEEE 2018.

[3] Learning a Stable Local Manifold Representation for Hyperspectral Linear Dimensionality Reduction WenboYu ; Miao Zhang ; Yi Shen IEEE 2018.

[4] Dimensionality reduction via adjusting data distribution density Wei Wang ; Wei-guo Shen ; Ya-xin Sun ; Bin Chen ; Rong Zhu IEEE 2018.

[5] Feature Dimensionality Reduction with Graph Embedding and Generalized Hamming Distance Honglei Zhang ; Mancef Gabbouj IEEE 2018.

[6] Class-Probability Based Semi-Supervised Dimensionality Reduction for Hyperspectral Images Lu Liang ; Yi Xia ; LinaXun ; Qing Yan; Dexiang Zhang IEEE 2018.

[7] Dynamical Component Analysis (DYCA): Dimensionality Reduction for High-Dimensional

January-February 2020
ISSN: 0193-4120Page No. 10654 -10658

Deterministic Time-Series Bastian Seifert ; KatharinaKorn ; SteffenHartmann ; Chri stian Uhl IEEE 2018.

[8] New Dimensionality Reduction Method of Wind Power Curve Based on Deep Learning YujingZhang ; Ying Qiao ; Zongxiang Lu ; Wei Wang IEEE 2018.

[9] Research on Feature Dimensionality Reduction in Content Based Public Cultural Video Retrieval YumengLiu ; Aina Sui IEEE 2018.

[10] patial-Spectral Graph-Based Nonlinear Embedding Dimensionality Reduction for Hyperspectral Image ClassificaitonXiangrong Zhang ; Yaru Han ; NingHuyan ; Chen Li ; JieFeng ; Li Gao ; Xiaoxiao Ma IEEE 2018.

[11] V. Chandola, A. Banerjee, V. Kumar, "Anomaly Detection for Discrete Sequences: A Survey", In: IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 5, 2012. pp. 823- 839.

[12] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space", Philosophical Magazine IEEE2018

[13] J. Beran, Y. Feng, S. Ghosh, R. Kulik, "Long-Memory Processes. Probabilistic Properties and Statistical Methods"IEEE2018

[14] I. Syarif, A. Prugel-Benett, and G. Wills, "Unsupervised Clustering Approach for Network Anomaly Detection", In: Benlamri R. (eds) Networked Digital Technologies. NDT 2012.Communications in Computer and Information Science, vol 293.Springer, Berlin, Heidelberg. IEEE2018

[15] A. V. Chernov, I. K. Savvas, and M. A. Butakova, "Detection of Point Anomalies in Railway Intelligent Control System Using Fast Clustering Techniques", 3rd International Scientific Conference "Intelligent Information Technologies for Industry, 2018, Springer

[16] J. Gan and Y. Tao, "Dbscan revisited: Mis-claim, un-fixability, and approximation"IEEE2018

[17] Cunningham, P. (2018) Dimension Reduction University College Dublin, Technical Report

[18] Samet, H. (2018) Foundations of Multidimensional and Metric Data Structures. Morgan Kaufmann.

[19] Zhang, Zhenyue; Zha, Hongyuan (2018). "Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment".

[20] van der Maaten, L. J. P., Postma, E. O., and van den Herik, H. J. (2018). Dimensionality reduction: a comparative review (TiCC-TR 2018-005), Tilburg University Technical Report