International Journal of Psychosocial Rehabilitation,
Vol. 24, Issue 06, 2020 ISSN: 1475-7192

Dimensionality Reduction using Deep Learning Techniques

[1]B. ChandraSekhar,[2]Abboju Niranjan,[3]Dr G.VenkataRamiReddy

ABSTRACT*— Dimensionality reduction is a compelling way to deal with scaling back the information. It is a strategy that endeavors to extend a lot of high dimensional vectors to a lower dimensionality space while holding measurements among them. The Artificial Intelligence and information mining systems may not be successful for high-dimensional information due to the scourge of dimensionality and question precision and efficiency will debase quickly as the measurement increments. It is a typical preprocessing step for feature extraction, classification and different undertakings. Learning a classifier on low-dimensional data sources is quick. All the more critically, Dimensionality reduction can help to get familiar with a superior classifier, especially when the information has a low-dimensional structure, and with little datasets, where it has a regularizing impact that can help abstain from over fitting. This paper deals with the working of deep learning technique for dimensionality reduction on various datasets. It is also compared with the traditional machine learning method i.e. Principal Component Analysis (PCA).*

Keywords-- *Autoencoder, Dimensionality Reduction, Principal Component Analysis (PCA), Reconstruction of data.*

I. INTRODUCTION

Dimensionality reduction Produce a smaller low-dimensional encoding of a given high-dimensional data index. Complex learning is a significant issue over a wide assortment of data handling fields including image recognition, video analysis, AI, and database route. In numerous issues, the deliberate information vectors are high-dimensional yet we may have motivation to accept that the information lie close to a lower-dimensional complex. In other words, the high-dimensional information are numerous, backhanded estimations of a basic source, which regularly can't be straightforwardly estimated. Learning a reasonable low-dimensional complex from high-dimensional information is basically equivalent to learning this hidden source. The research centers around huge information representation that depends on dimensionality decrease strategies. This paper represents a staggered technique for information binding and perception. It partitions the entire data processing process into discrete steps and applies specific dimensionality decrease strategy considering for investigating data volume and type.

The real procedures of dimensionality reduction are 1) Feature determination. 2) Feature reduction (decrease). Feature determination is a procedure that picks an ideal subset of highlights as indicated by target work. Its targets are 1)To lessen dimensionality and expel noise 2)To improve mining execution 3)Speed of learning 4)Predictive precision 5)Simplicity and comprehensibility of mined outcomes. Feature determination or Feature reduction alludes to the mapping of the first high-dimensional information onto a lower-dimensional space. For a given

[1]*Assistant Professor of of CSE, Malla Reddy Institute of Technology Hyderabad, 500075, India,chandrasekhar.batti@gmail.com*
[2] *Research scholar of JNTUH,Hyderabad, 500085, India,abboju.niranjan@gmail.com.*
[3] *Professor of CSE, SIT, JNTUH Hyderabad, 500085, India, gvr_reddi@yahoo.co.in.*

arrangement of information purposes of p factors register their low-dimensional portrayal. The real contrasts between feature reduction and feature choice are: In feature reduction every single unique element are utilized and the changed highlights are straight mixes of the first includes and in feature determination, just a subset of the first highlights are chosen and they are Continuous versus discrete. The real objectives of Techniques of dimensionality decrease are: 1) High viability, 2) Able to deal with both immaterial and excess features, 3) individual feature assessment cannot be done purely, 4) High proficiency, 5) Less expensive than existing subset assessment strategies, 6) Not heuristic search methods. Many of the researchers are focusing on the dimensionality reduction of the large datasets by enhancing the speed and accuracy without the data loss. Lei Yu.et.al. made a research on dimensionality reduction for data mining techniques [1].Mizuta.M proposed a dimension reduction methods[2]. Sorzano.et.al made an A.P. A survey on dimensionality reduction techniques [3]. Zubova .et.al. elaborated the dimensionality reduction using parallel computing[4]. Sugiyama used the local fisher discriminant analysis for dimensionality reduction of multimodal labeled data [5]. Vander and Hinton visualized the data using t-SNE [6]. Wang and Carreira made a nonlinear low dimensional regressing using auxiliary coordinates [7]. Wang and Carreira conducted research on optimization of deeply nested systems [8]. Davis .et.al conducted experiment on information theoretic metric learning [9]. Globerson, A., and Roweis represented the metric learning by collapsing classes [10]. Goldberger, J.et.al. have done experiment on neighborhood component analysis [11]. Rish.et.al proposed a Closed-form supervised dimensionality reduction with generalized linear models [12]. The motivation behind the proposed system is because of its dimensionality reduction and achieving the high accuracy without loss of data [13]. Therefore, in this paper, reconstructable feature vector of the respective data sets with gaining the low dimensionality. The rest of the paper is organized as follows. Section 2 represents the proposed methodology. Section 3 represents the set of experiments carried out on datasets of face and flower finally, Section 4 concludes the paper.

II. METHODOLOGY

Multimodal biometrics captures different traits as a single trait to perform the accurate recognition of an individual. The proposed system illustrates the feature level fusion of iris and retina images. The process carried out in the following steps.

Step 1: Image Acquisition

The acquired color images of flowers, cars and facial recognition system from various datasets JNTU UCEV, STANFORD CAR and FLOWER data setss respectively are used for dimensionality reduction. These are transformed to grayscale images.

Step 2: Preprocessing

Preprocessing expects to feature the fundamental highlights by invalidating the picture distortion. Ordinary Daughman's calculation is utilized to wipe out uproarious information and protect the structure of image data. This administrator works successfully for nonlinear edge identification and furthermore finds the most extreme edge quality that reproduces and upgrade the nature of picture information.

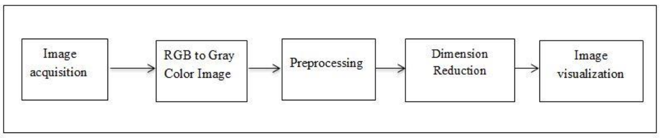

Figure1: Process flow of proposed system

Step 2: Dimensionality reduction using Autoencoder

An indistinguishable structure is framed at whatever point light goes into the image. Autoencoder is utilized on the standardized picture for hearty highlights extraction, upgrades the precision and mitigates the reconstruction error. Autoencoder is adequate for more profound portrayal and reproduction of the picture information. It diminishes the dimensionality of the information picture via preparing a lot of information into littler portrayals at shrouded layers. Recreation of the information picture as yield is picked up by interpreting the encoded highlight vector. Autoencoder neural system uses back engendering by leveling objective incentive to the information esteem f^i=x^i. It distinguishes the capacity and sets the objective worth equivalent to the information esteem through learning on the estimate. Autoencoders give the better outcomes even the quantity of concealed units are enormous. Concealed(Hidden) unit 'I' gets actuated to 'qi'. This portrayal remotely can't play out the activity on explicit Input 'f', qi(2)(f) helps in actuation of the concealed unit. Averaging the preparation information for example enactment of concealed unit 'I' is communicated in "numerical equation (1)".

$$\hat{E}_i = \frac{1}{x} \sum_{j=1}^{x} [q_i^{(2)}(f^{(j)})] \qquad (1)$$

$$\hat{E}_i = E \qquad (2)$$

Where E is called sparse parameter which small value closely to zero given in "Equation (2)"

III. EXPERIMENTAL RESULTS

The proposed system deals with the dimensionality reduction of the images that helps to save the time and reduces the cost. The feature of reconstruction of the images helps to decode the original image with minimal error. Dimensionality reduction and Representation system accuracy is calculated by assuming JNTU UCEV, STANFORD CAR and FLOWER data sets with 500 images of samples of jntu ucev (faces),car,flower with .jpg format. Acquired images are of various sizes with different dimensions. The proposed system embellished from "Figure 2" to "Figure 12". "Figure 2" depicts the enrolled templates of facial recognition system of JNTUK UCEV students and "Figure 3" illustrates the preprocessed images of jntukucev . An autoencoder is applied on Jntukucev extraction for feature extraction and dimensionality reduction are shown in "Figure 4". "Figure 5 and 6" illustrates the enrolled images and preprocessed images of flower dataset." "Figure 7" shows encoded images using the autoencoder. "Figure 8 , Figure 9 and Figure10" represents enrolled , preprocessed images and reduced feature extracted images of car dataset. Various comparisons is done on different datasets to achieve the dimensionality reduction and finally compared with the traditional method i.e. Principal Component Analysis(PCA) .Thereby, Experiment is conducted on various dimensional images for achieving the dimensionality reduction and to enhance the reconstruction of the original data with minimal error and accuracy of the data is shown in table1. "Figures 11, 12 and 13" shows the enhancement of the accuracy and efficiency of the proposed system i.e. autoencoder is compared with PCA on various datasets of flower, JNTUKUCEV and car datasets.

Figure 2 : Enrolled images of students of JNTUKUCEV

Figure 3: Pre-processed images of students JNTUKUCEV.

Figure 4: Autoencoder employed on JNTUKUCEV for dimensionality reduction.

Figure 5: Enrolled images of students of car dataset.

Figure 6: Pre-processed images of car dataset.

Figure 7: Autoencoder employed on car dataset for dimensionality reduction.

Figure 8: Enrolled images of students of flower dataset.

Figure 9: Pre-processed images of flower dataset.

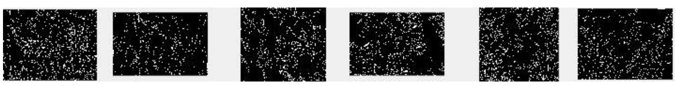

Figure 10: Autoencoder employed on flower dataset for dimensionality reduction.

Table1: Autoencoder is compared with PCA on different dimensional images

S.no	Original Image Dimension	Dimension Reduction after applying PCA	Dimension Reduction after applying autoencoder with no of hidden layers
1	250x300	250X250	100X250
2	591X500	591X499	190X591
3	754X500	754X499	150X754
4	284X177	284X176	170X 284
5	272X186	272X185	125X272
6	341X148	341X147	110X341
7	300X168	300X167	125X300
8	120X160	120X120	90X120
9	492X702	492X 492	10X492
10	320X280	320X279	200X 320

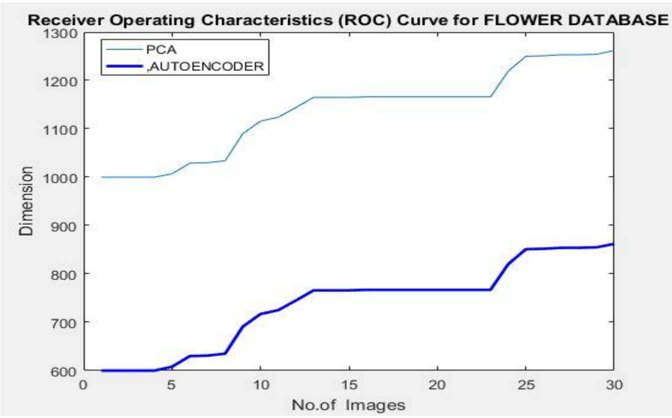

Figure 11: ROC curve depicting the high performance of proposed system compared to PCA on Flower dataset.

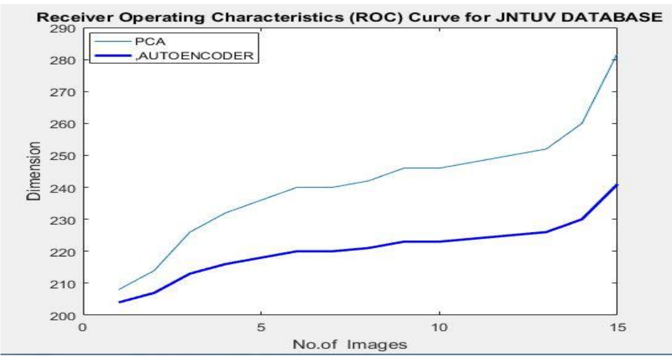

Figure 12: ROC curve shows the high performance of proposed system compared to PCA on JNTUKUCEV dataset.

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

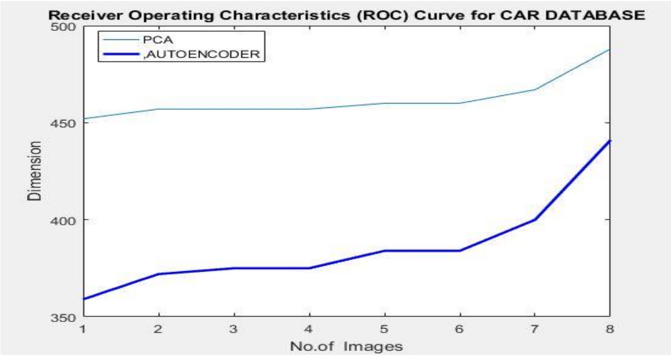

Figure 13: ROC curve depicting the high performance of proposed system compared to PCA on CAR dataset

IV. CONCLUSION

In this paper, we have examined the dimensionality reduction using the deep learning technique i.e. Autoencoder. The unsupervised technique helps to safeguard the image quality and reconstructs the original image with a minimal error. The deeper representation of the image helps to extracts the robust features that enhances the recognition and accuracy. The proposed system results the low dimensionality images with high quality when it is compared with the traditional technique i.e. Principal Component Analysis Image safeguarding is a very significant utilization of dimensionality reduction as it empowers us to use less memory and to improve the speed of execution of any projects which utilize these images. In the future work Convolution Neural Network can be employed in proposed system for acquiring low dimensionality images with high accurate and quality images.

REFERENCES

1. Lei Yu Binghamton University, Jieping Ye,Huan Liu ,Arizona State University, “Dimensionality Reduction for datamining-Techniques, Applications and Trends”.
2. Mizuta, M. Dimension Reduction Methods. Hum¬boldt-Universität Berlin, Center for Applied Statistics and Economics (CASE), 2007, 15.
3. Sorzano, C. O. S., Vargas, J., Montano, A. P. A Survey of Dimensionality Reduction Techniques. 2014. https:// arxiv.org/abs/1403.2877 (last accessed 2 November 2017).
4. Zubova, J., Kurasova, O., Liutvinavicius, M. Parallel Computing for Dimensionality Reduction. Information and Software Technologies, Springer-Verlag, 230-241, 2016. ISBN 978-3-319-46254-7.
5. Sugiyama, M. 2007. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. J. Machine Learning Research 8:1027–1061.
6. vander Maaten, L. J. P., and Hinton, G. E. 2008. Visualizing data using t-SNE. J. Machine Learning Research 9:2579– 2605.
7. Wang, W., and Carreira-Perpin˜an, M. ´ A. 2012. Nonlinear ´ low-dimensional regression using auxiliary coordinates. In Lawrence, N., and Girolami, M., eds., Proc. of the 15th Int. Workshop on Artificial Intelligence and Statistics (AISTATS 2012), 1295–1304.

8. Carreira-Perpin˜an, M. ´ A., and Wang, W. 2014. Distributed ´ optimization of deeply nested systems. In Kaski, S., and Corander, J., eds., Proc. of the 17th Int. Workshop on Artificial Intelligence and Statistics (AISTATS 2014), 10–19.
9. Davis, J. V.; Kulis, B.; Jain, P.; Sra, S.; and Dhillon, I. S. 2007. Information-theoretic metric learning. In Ghahramani, Z., ed., Proc. of the 24th Int. Conf. Machine Learning (ICML'07). [10].Globerson, A., and Roweis, S. 2006. Metric learning by collapsing classes. In Weiss, Y.; Scholkopf, B.; and Platt, J., ¨ eds., Advances in Neural Information Processing Systems (NIPS), volume 18, 451–458. MIT Press, Cambridge, MA.
10. Goldberger, J.; Roweis, S.; Hinton, G.; and Salakhutdinov, R. 2005. Neighbourhood components analysis. In Saul, L. K.; Weiss, Y.; and Bottou, L., eds., Advances in Neural Information Processing Systems (NIPS), volume 17, 513– 520. MIT Press, Cambridge, MA.
11. Rish, I.; Grabarnilk, G.; Cecchi, G.; Pereira, F.; and Gordon, G. 2008. Closed-form supervised dimensionality reduction with generalized linear models. In McCallum, A., and Roweis, S., eds., Proc. of the 25th Int. Conf. Machine Learning (ICML'08), 832–839.
12. ViswanadhaRaju, S., Vidyasree, P., Gudavalli, M., & Sekhar, B. C. (2016, November). Minimum cost fused feature representation and reconstruction with autoencoder in bimodal recognition system. In Proceedings of the International Conference on Big Data and Advanced Wireless Technologies (p. 17). ACM.