International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-10 Issue-4, November 2021

Evaluation of Various DR Techniques in Massive Patient Datasets using HDFS

K. B. V. Brahma Rao, R Krishnam Raju Indukuri, P. Suresh Varma, M. V. Rama Sundari

Abstract*: The objective of comparing various dimensionality techniques is to reduce feature sets in order to group attributes effectively with less computational processing time and utilization of memory. The various reduction algorithms can decrease the dimensionality of dataset consisting of a huge number of interrelated variables, while retaining the dissimilarity present in the dataset as much as possible. In this paper we use, Standard Deviation, Variance, Principal Component Analysis, Linear Discriminant Analysis, Factor Analysis, Positive Region, Information Entropy and Independent Component Analysis reduction algorithms using Hadoop Distributed File System for massive patient datasets to achieve lossless data reduction and to acquire required knowledge. The experimental results demonstrate that the ICA technique can efficiently operate on massive datasets eliminates irrelevant data without loss of accuracy, reduces storage space for the data and also the computation time compared to other techniques.

Keywords: Dimensionality Reduction, Data Mining, Independent Component Analysis, Knowledge Reduction, HDFS

I. INTRODUCTION

The data is growing day by day in hospitals for the last ten years makes it difficult to store, manage and analyzing it either to make decisions of patients for right treatment. To deal with massive patient data detonation and knowledge reduction, we compared various dimensionality reduction techniques to acquire the required knowledge by eliminating irrelevant attributes without loss of accuracy. Dimensionality Reduction is a method to convert the given dataset of with more dimensions into fewer dimensions. In this method important information will not be lost, and redundant features will be eliminated along with unwanted data. Dimensionality reduction is important for making decision of a patient treatment because it leads to identify that set of features which alone shows most variability.

Manuscript received on September 26, 2021.
Revised Manuscript received on September 30, 2021.
Manuscript published on November 30, 2021.
* Correspondence Author
Dr. K. B. V. Brahma Rao*, Ph.D, Department of Computer Science and Engineering, Adikavi Nannaya University, Rajamahendravaram (A. P), India.
Dr. R Krishnam Raju Indukuri, Ph.D, Department of Computer Science and Engineering, Adikavi Nannaya University, Rajamahendravaram (A. P), India.
Dr. Suresh Varma Penumatsa, Professor & Dean of Academics Department of Computer Science & Engineering of Adikavi Nannaya University, Rajamahendravaram (A. P), India.
Dr. M. V. Rama Sundari, Ph.D, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam (A. P), India.

Doctors can use this information for using that feature sets for applying various analytical algorithms and thus would reduce the computational processing, memory and time. This may prove to be useful if we involve it for massive patient datasets. The dimensionality reduction is considered as the preprocessing mechanisms before grouping of the data. Without using dimensionality reduction, we need to work on all feature sets which are not influential much while grouping and there might be certain sets which has high influence in grouping. Hence there is no need to use less influential feature sets. The experiments are needed to identify which feature sets should be retained. The reduction techniques can reduce number of input variables and also number of covariants. The dependency one of the significant attributes is an important issue in data analysis. It is also required to find the partial dependency attributes because if some of the attributes are removed, we may loss the required data that effects on the accuracy of the result. For dimensionality reduction the techniques used are: Standard Deviation, Variance, PCA, LDA, Factor Analysis, Positive Region, Information Entropy and ICA.

1. Standard Deviation: Statistical standard deviation of each of the attributes is calculated and whichever attribute has the highest standard deviation. This attribute alone is used for grouping. Variance: Each of the attributes is calculated and whichever attribute has the highest value, that attribute alone is used for grouping.
2. PCA (Principal Component Analysis): This technique uses the concept of eigen value matrix and eigen vector. The PCA tries to reduce a dataset in bigger plane to a condensed plane possibly forming linear axis. The values obtained after PCA dimensionality reduction are used to determine covariance and the one which had higher values those attributes are chosen alone for grouping.
3. LDA (Linear Discriminant Analysis): This technique uses probability to find out LDA functions. The values in LDA functions of certain attributes can be used as rules for determining the groupings.
4. Factor Analysis: This technique can further be used with PCA for grouping. This identifies the field or features which is top most discriminant and using alone that features the experiment is conducted for grouping.
5. Positive Region: The P-lower approximation also known as Positive Region is a group of all attributes. The target set contains this group of attributes.
6. Information Entropy: Entropy is a function of attribute frequency of two attributes. The information gain is founded on the reduction in entropy after a dataset is split on an attribute.

The dataset is divided with the largest information gain attribute. This is repeated for all branches in the decision tree.

7. ICA (Independent Component Analysis): If two attributes are uncorrelated, it means there is no linear relation between them. If they are independent, it means they are not dependent on other attributes. These attributes are called independent components of the given dataset.

The massive datasets contain large number of attributes. We need to analyze and identify the important information that taken decision. The conversion of larger datasets into smaller one is a problem and some attributes further more may be highly correlated or highly similar to each other creating additional problems with their interpretation and modeling itself. To find the Knowledge has become a new challenge using big data. The dimensionality reduction techniques have been successfully used in data mining. The MapReduce technique has been using for big data analysis in the recent times. The dimensionality reduction techniques can implement in the MapReduce programming.

A parallel execution method is improving the performance of data mining for the effective computation of approximation. The parallel method provides flexible for executing large scale data. The cluster analysis is used to reduce the data. The dimensionality reduction algorithm will obtain multivariate data. The dimensionality reduction algorithm is used to obtain concise, accurate representation of the given data. The dimensionality reduction is necessary for processing the array of goals. Dimensionality reduction is a task specified method to deal with classification and regression problems. Mapping is required to visualize a high dimensional data as low dimension one. The transform coding typically involves dimension reduction. In various industries Big Data reduction is already initiated. In healthcare industry the Big Data analysis is on live wire. Though certain techniques like logistic regression are available, they can deal with interdependent elements is very much limited. These days meeting the healthcare expenses are very much discouraging and needs additional amount. More number of diagnostic tests to be done and unnecessary procedures are followed that becomes very lengthy. All this juncture data reduction is essential for effective diagnosis. The Big Data offers unlimited opportunities for healthcare researchers for analysis and it is estimated that developing and using prediction models in the healthcare industry could save money. The Big Data analysis finds place in retrieving required information from healthcare records in electronic mode. For Hadoop based applications the Hadoop Distribution File System (HDFS) is used. HDFS executes the NameNode and DataNode to execute the dataset in parallel. The distributed file system provides high-performance access to data across very scalable Hadoop clusters. The output of ICA algorithm is exposed to HDFS that executes using MapReduce. The MapReduce is a programming model and it consist two functions: Map and Reduce operations. This model operates on the massive datasets with a parallel, distributed algorithm on a cluster. A map function implements filtering and sorting, whereas reduce function makes a summary operation. The dimensionality reduction techniques can easily be applied for data mining that deals with huge datasets along with HDFS for parallel and distributed operations to acquire the required knowledge. And also eliminating redundancy, less memory space for storing data and reduce the execution time.

II. LITERATURE SURVEY

Rasendu Mishra [1] presented a survey of various dimensionality reduction techniques for reducing features sets in order to group documents effectively with less computational processing and time. They discussed the concept of dimensionality reduction can be used to respond to recommendations by collecting documents as per the query. C.O.S. Sorzano [2] presented the available one of the dimension reduction techniques plethora and discussed the mathematical understanding behind them. The mathematical procedures building possible this reduction are called dimensionality reduction techniques; they have extensively been established by fields like Statistics or Machine Learning. Yanyuam Ma [3] discussed different estimation and inference procedures at various levels with an intention of focusing the inherent ideas. They further discussed certain unresolved issues in the area that leads to future work. Milos Hauskrecht [4] discussed the basics of different approaches applied for selection of features and they also explained the effects on MS cancer proteomic dataset. It is helpful to perform analysis of high dimensional genomic data. Swati A Sonawale [5] presented a method of reducing the irrelevant features of the data to get a smaller set of features with more discriminative control for better performance. Nandakishore Kambhatla [6] described algorithms for voice and image data for comparing the performance with Principal Component Analysis (PCA) and neural networks of nonlinear PCA. In fact the local linear techniques outperform neural networks. Badrul M. Sarwar [7] explained two different methods based on Singular Value Decomposition (SVD) for reducing the dimensionality. The quality of the test system using SVD is compared with that of using collaborative filtering. The two experiments compare different perspectives of importance. Matthew Brand [8] presented an interactive graphical movie recommender which predicts as well as displays rankings of large number of movie tickets in the real time. The proposed system “learns” the user ratings on the SVD revised. Users can asynchronously join, add ratings, add movies, revise ratings, get recommendations, and delete themselves from the model. Fasong Wang [9] discussed the data mining problem in the light of ICA. The under-complete ICA model in data mining is given and then gives the most popular ICA algorithm-Natural Gradient Algorithm (NGA). Several applications such as latent variable decompositions, multivariate time series analysis and prediction, text document data analysis, extracting hidden signals in satellite images, weather data mining and so on are derived from ICA data mining. M. Usman Ali [10] proposed Principal Component Analysis (PCA) and Factor Analysis which are used for dimensionality reduction of Bioinformatics data. These techniques were applied on Leukaemia data set and the number of attributes was reduced very much. H.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Telgaonkar Archana [11] discussed the techniques of Dimensionality Reduction namely Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Performance analysis is carried out on high dimensional data set of UMIST, COIL and YALE which contain images of objects and human faces. KNN classifier and the Naive Bayes classifier classified the objects to compare performance of the techniques. Further these results differentiate the supervised learnings. K. Keerthi Vasan [12] highlighted the efficiency of PCA while detecting intrusion and determining the Reduction Ratio (RR). This work also focuses on the ideal number of principal components. Tonglin Zhang [13] proposed another PCA approach without the computation of principal components. This approach provides an exact solution to PCA for regression. Alireza Sarveniazi [14] reviewed the dimensionality reduction methods in detail that the last and the latest versions which are extensively developed in the past decade. Khaled Labib [15] projected a method for detecting attacks by Denial-of-Service and Network Probe attacks using Principal Component Analysis as a multivariate statistical tool. The paper discussed the nature of these attacks, and merits of using PCA for detecting intrusions. Sudeep Tanwar [16] underlined Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) techniques for performing DR over BD. Further the performance of both techniques in terms of correctness and mean square error (MSR). Lan Fu [17] described the Discrimination Analysis of Multivariate Statistical Analysis, Linear Dimensionality Reduction and Nonlinear Dimensionality Reduction Methods in the light of the wide range of applications of high-dimensional data. Zebin Wu [18] developed a parallel and distributed technique for hyperspectral dimensionality reduction and Principal Component Analysis (PCA), of cloud computing architectures. Marco Cavallo [19] proposed a different framework that interacts visually to improve dimensionality reduction based exploratory data analysis. Chaman Lal Sabharwal [20] followed PCA based algorithms in two different genres, Qualitative Spatial Reasoning (QSR) to achieve data reduction along with improved regression analysis. Nandakishore Kambhatla [21] developed a local linear approach for dimension reduction that provides accurate representation and is fast to compute. Laurens van der Maaten [22] presented a review and relative comparison of PCA and classical scaling techniques. The performance of the nonlinear techniques is investigated on artificial and natural tasks. The results reveal that nonlinear techniques perform well on selected artificial tasks, which cannot be necessarily prolonged to real-world tasks. Steven H. Berguin [23] proposed a method for dimensionality reduction that scales as plog(p), where p is the number of design variables. It works with the advantage of adjoin design methods to compute the covariance matrix of the gradient. This information is then used with PCA to develop a linear transformation which allows an aerodynamic optimization problem that reformulates in an equivalent coordinate system of lower dimensionality. G. N. Ramadevi [24] explored the real-time applications of the dimension reduction techniques. The PCA (Principal Component Analysis) is one of the dimensionality reduction techniques of feature reduction algorithm to reduce the dimensionality of the dataset without losing the data. PCA when applied on data before clustering will result more truthful and reduce the time substantially. PCA is used for data visualization and noise reduction. Jianqing Fan [25] presented an overview of methodological and theoretical developments of PCA over the last decade, with focus on its applications to big data analytics. They discussed relationship between PCA and factor analysis as well as its applications to large covariance estimation and multiple testing. Jiaying Weng [26] presented an overview of certain classic and modern dimension reduction methods, followed by a discussion of how to use the transformed variables in the context of analyzing survey data. Kerstin Bunte [27] reviewed the basic principles of dimensionality reduction and discussed some of the approaches that were published over the past years from the perspective of their application to big data. M. Song [28] discussed the effect of applying dimensionality reduction (preprocessing) techniques on the performance of trace clustering. They used three transformation techniques; singular value decomposition (SVD), random projection (RP), and Principal Components Analysis (PCA), and the state-of-the art trace clustering in process mining.

2.1 Materials and methods

In the review of literature very little work has been found towards comparison of dimensionality reduction algorithms using massive patient datasets and also not examined using HDFS. No comparison of data reduction using Independent Component Analysis (ICA) technique. The ICA technique is used on linear mixtures of some unknown latent attributes. The present paper compared various dimensionality reduction techniques including ICA technique. All dimensionality reduction techniques are examined with Hadoop platform which is a distributed platform and open source software is developed by Apache Software. The Hadoop platform needs only commodity hardware that is most sufficient in the clusters for processing massive datasets. The structure of processing of massive datasets various dimensionality reduction techniques and Hadoop Distributed File System (HDFS) is described in figure 1.

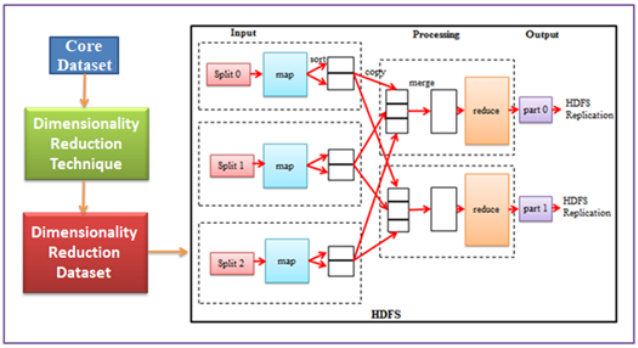

Fig. 1 Structure of HDFS with Dimensionality Reduction Technique

Hadoop Distributed File System (HDFS) is the very massive storage system for dig datasets used by Hadoop applications. HDFS creates multiple copies of data blocks and assigns them on data nodes, to enable reliable extremely rapid computations. Hadoop consists two most important modules: File Storage and Distributed Processing System.

The first module is File Storage is also known as "Hadoop Distributed File System (HDFS)". It is responsible for scalable, reliable, relatively low cost storage. The files are stored through a group of servers in HDFS and data availability is monitoring obstinately in a cluster servers. The second module of Hadoop is the similar data processing system is also known as "MapReduce". The Hadoop distributed file system and the MapReduce framework are continuously on the same set of nodes. The Hadoop MapReduce programming permits the execution of Java code and also uses software written in other languages.

All dimensionality reduction techniques will reduce the utilization of memory and processing time. The superfluous data is removed without important accuracy loss information. In this paper we have presented comparison of various dimensionality reduction techniques practically for dimensionality and knowledge reduction from big patient datasets. The comparison clearly shows that the Independent Component Analysis technique outperforms the latter one.

III. METHODOLOGY

In this section, we compared to examine the efficiency of various dimensionality reduction techniques using MapReduce for big patient datasets. We considered various dimensionality reduction techniques such as Standard Deviation (SD), Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Factor Analysis (FA), Positive Region (PR), Information Entropy (IE) and Independent Component Analysis (ICA). The following hardware and software used to do the experiments.

The experimentations have been carried out on six nodes in a cluster. The master node and five compute nodes. Each one of these computer nodes has the following features:

Processors: Intel Core i3 6th generation or above
Cores: 4 per processor (8 threads)
Network: One Gigabit Ethernet
Hard drive: 1 TB or above
RAM: 4 GB or above

The specific details of the software used are the following:

Python 3.0 or above
MapReduce implementation: Hadoop 2.8.0.
Apache Hadoop Distribution.
Maximum maps tasks: 33.
Maximum reducer tasks: 1.
Operating System: Windows 10 / Ubuntu 15 or above, Version 64 bits
Java SE Development Kit: JDK1.8 or above

IV. EXPERIMENTAL RESULTS

To calculate the performance of the dimensionality reduction techniques, we have considered measurements reduction of data size those effects processing speed and utilization of memory. A series of experimentations are conducted on the big patient dataset and compared the outcomes among dimensionality reduction techniques. To do the experiments we have taken minimum dataset size 70 megabytes to 200 megabytes. The output datasets sizes are compared with Independent Component Analysis reduction technique dataset size because in all experiments the ICA technique is performing well. The percentage of difference is computed for each reduction technique with ICA technique. The following table describes Dataset size in megabytes, Dimensionality Techniques, Reduced Dataset size in megabytes and the percentage of difference with respect to each dimensionality reduction technique with ICA technique.

Table-I: Performance metrics of Data Size and Reduction Data Size using different dimensionality reduction techniques

Data Size in MB	Dimensionality Reduction Technique	Reduction Data Size in MB	% of Difference w.r.t. SD/PCA/LDA/FA/PR/IE
70	SD	66	12.12
	PCA	64	9.38
	LDA	65	10.77
	FA	67	13.43
	PR	63	7.94
	IE	62	6.45
	ICA	58	5.9
80	SD	75	12
	PCA	72	8.33
	LDA	73	9.59
	FA	76	13.43
	PR	71	7.94
	IE	69	6.42
	ICA	66	5.92
90	SD	86	9.3
	PCA	83	6.02
	LDA	84	7.14
	FA	85	13.43
	PR	82	7.94
	IE	80	6.41
	ICA	78	5.93
100	SD	96	11.46
	PCA	93	8.6
	LDA	94	9.57
	FA	95	13.43
	PR	91	7.94
	IE	89	6
	ICA	85	5.91
110	SD	104	13.46
	PCA	100	10
	LDA	101	10.89
	FA	102	13.43
	PR	98	7.94
	IE	93	5.94
	ICA	90	5.4
120	SD	114	15.79
	PCA	110	12.73
	LDA	111	13.51
	FA	112	13.43
	PR	108	7.94
	IE	106	5.87
	ICA	96	5.2
130	SD	121	12.4
	PCA	118	10.17
	LDA	119	10.92
	FA	120	13.43
	PR	116	7.94
	IE	112	5.8
	ICA	106	5
140	SD	132	15.91
	PCA	127	12.6
	LDA	129	13.95
	FA	127	13.43
	PR	118	7.94
	IE	116	5.3
	ICA	111	4.8
150	SD	141	16.31
	PCA	138	14.49
	LDA	136	13.24
	FA	134	13.43
	PR	128	7.94
	IE	123	7
	ICA	118	6.5
160	SD	153	16.99
	PCA	146	13.01
	LDA	147	13.61
	FA	145	13.43
	PR	138	7.94
	IE	133	7.3
	ICA	127	7.1
170	SD	168	17.86
	PCA	163	15.34
	LDA	166	16.87
	FA	162	13.43
	PR	151	7.94
	IE	147	7.8
	ICA	138	7.3
180	SD	170	17.65
	PCA	167	16.17
	LDA	168	16.67
	FA	165	13.43
	PR	155	7.94
	IE	149	8
	ICA	140	7.5
190	SD	188	13.3
	PCA	181	9.94
	LDA	183	10.93
	FA	182	13.43
	PR	177	7.94
	IE	171	8.7
	ICA	163	7.6
200	SD	181	11.05
	PCA	176	8.52
	LDA	175	8
	FA	178	13.43
	PR	170	7.94
	IE	166	9.3
	ICA	161	7.8

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The names of the dimensionality reductions techniques are specified in acronyms. The following Fig. 2 describes comparison among seven dimensionality reduction techniques while the dataset size is 100 megabytes.

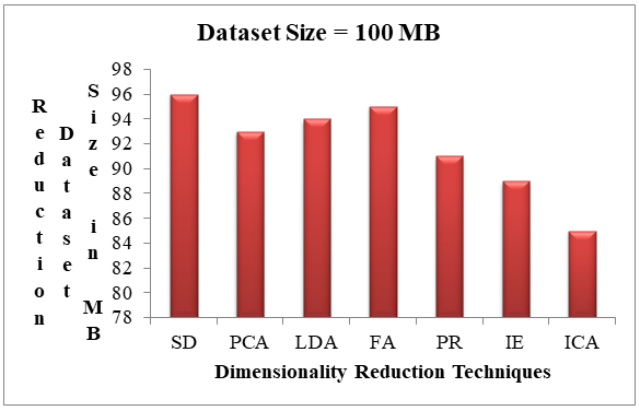

Fig. 2 Comparisons among Seven Different Dimensionality Reduction Techniques

The performance metrics shows that the ICA technique is performing well while comparing with the remaining dimensionality reduction techniques starting from the dataset size from 70 MB to 200 MB. While the dataset size is increasing, it shows that the dimensionality reduction using the Independent Component Analysis is producing better results rather than remaining techniques.

V. CONCLUSION

In the present paper, we compared various dimensionality reduction techniques using MapReduce that can handle massive datasets. The Hadoop MapReduce is an efficient distributed computational model for parallel processing with massive datasets. The dimensionality reduction techniques are compared from one to one. The experimental results demonstrate that the Independent Component Analysis technique is performing well compared to other dimensionality reduction techniques to reduce the unrequired dimensions without loss of accuracy of the result. Once the redundant data or dimensions are eliminated, the processing speed is increased and memory is utilized efficiently. The proposed future research work will focus on applications of the proposed distributed parallel method in dimensionality and knowledge reduction using semi-structural and unstructured data.

ACKNOWLEDGMENT

We would like to thanks the Director and management of B. V. Raju College, Vishnu Campus, Bhimavaram for give the assistance and support for this work. We are very much thankful to R&D Department of Adikavi Nannaya University, Andhra Pradesh, India for accepting and give their cooperation for this work. We are also thankful to Dr. Dr. Ch. V. Srinivas, Principal, B. V. Raju College for valuable suggestions towards this research work.

REFERENCES

1. C.O.S. Sorzano, J. Vargas, A. Pascual-Montano, "A Survey of Dimensionality Reduction Techniques, Cornell University", (2014), pp. 1-35.
2. Yanyuan Ma, Liping Zhu, "A Review on Dimension Reduction", International Statistical Review, (2013), pp. 134-150.
3. Milos Hauskrecht, Richard Pelikan, Michal Valko, James Lyons-Weiler, "Feature Selection and Dimensionality Reduction in Genomics and Proteomics", Fundamentals of Data Mining in Genomics and Proteomics, Springer, (2006), pp. 149-172.
4. Swati A Sonawale, Roshani Ade, "Dimensionality Reduction: An Effective Technique for Feature Selection", International Journal of Computer Applications, (2015), pp. 18-23.
5. Nandakishore Kambhatla, Todd K. Leen, "Dimension Reduction by Local Principal Component Analysis, Neural Computation", (1997), pp. 1493–1516.
6. Badrul M. Sarwar, George Karypis, Joseph A. Konstan, John T. Riedl, "Application of Dimensionality Reduction in Recommender System--A Case Study", Technical Report, (2000), pp. 1-15.
7. M. Usman Ali, Shahzad Ahmed, Javed Ferzund, "Using PCA and Factor Analysis for Dimensionality Reduction of Bio-informatics Data", IJACSA, (2017), pp. 415-426.
8. H.Telgaonkar Archana, Deshmukh Sachin, "Dimensionality Reduction and Classification through PCA and LDA", International Journal of Computer Applications, (2015), pp. 33-37.
9. K. Keerthi Vasan, B. Surendiran, "Dimensinality Reduction Using Principal Component Analysis for Network Intrusion Detection", Science Direct, Elsevier, (2016), pp. 510–512.
10. Alireza Sarveniazi, "An Actual Survey of Dimensionality Reduction", American Journal of Computational Mathematics, (2014), pp. 55-72.
11. Lan Fu, "The Discriminate Analysis and Dimension Reduction Methods of High Dimension", Open Journal of Social Sciences, Scientific Research, (2015), pp. 7-13.
12. Zebin Wu, Yonglong Li, David E. Goldberg, Jun Li, Fu Xiao, Zhihui Wei, "Parallel and Distributed Dimensionality Reduction of Hyperspectral Data on Cloud Computing Architectures", IEEE, (2016), pp. 2270-2278.
13. Chaman Lal Sabharwal, Bushra Anjum, "Data Reduction and Regression Using Principal Component Analysis in Qualitative Spatial Reasoning and Health Informatics", Scielo, (2016), pp. 31-42.
14. Nandakishore Kambhatla, "Dimension Reduction by Local Principal Component Analysis", ACM, (1997), pp. 1493-1516.
15. Laurens Vander Maaten, Eric Postma, Jaap vanden Herik, "Dimensionality Reduction: A Comparative Review", Tilburg centre for Creative Computing, Tilburg University, 2009, pp. 1-35.
16. Steven H. Berguin, Dimitri N. Mavris, "Dimensionality Reduction Using Principal Component Analysis Applied to the Gradient", AIAA Journal, (2015), pp. 1078-1090.
17. G.N.Ramadevi, K.Usharani, "Study on Dimensionality Reduction Techniques and Applications", International Journal Publications of Problems and Applications in Engineering Research (IJPAPER), (2013), pp. 136-140.
18. Jianqing Fan, Qiang Sun, Wen-Xin Zhou, Ziwei Zhu, "Principal Component Analysis for Big Data", Cornell University, (2018), pp. 1-20.
19. Jiaying Weng, Derek S. Young, "Some Dimension Reduction Strategies for the Analysis of Survey Data", Jouranl of Big Data, (2017), pp. 1-19.
20. M. Song, H. Yang, S.H. Siadat, M. Pechenizkiy, "A Comparative Study of Dimensionality Reduction Techniques to Enhance Trace Clustering Performances", Expert Systems with Applications, Elsevier, (2013), pp. 3722-3737.
21. Rasendu Mishra, Priti Sajja, "Experimental Survey of Various Dimensionality Reduction Techniques", Proceedings of International Conference on Inventive Computing Systems and Applications (ICICSA), (2018), pp. 12569-12574.
22. Matthew Brand, "Fast Online SVD Revisions for Lightweight Recommender Systems", Proceedings of International Conference on Data Mining, (2003), pp. 1-12.
23. Fasong Wang, Hongwei Li, Rui Li, "Data Mining with Independent Component Analysis", Proceedings of the IEEE International Conference on Intelligent Control and Automation, (2006), pp. 6043-6047.
24. Tonglin Zhang, Baijian Yang, "Big Data Dimension Reduction using PCA", IEEE International Conference on Smart Cloud, (2016), pp. 152-157.
25. Khaled Labib, V. Rao Vemuri, "An Application of Principal Component Analysis to the Detection and Visualization of Computer Network Attacks", Proceedings of SAR, (2004), pp. 1-10.

26. Sudeep Tanwar, Tilak Ramani, Sudhanshu Tyagi, "Dimensionality Reduction Using PCA and SVD in Big Data: A Comparative Case Study", Proceedings of the International Conference on Future Internet Technologies and Trends, (2018), 116-125.
27. Marco Cavallo, Cagatay Demiralp, "A Visual Interaction Framework for Dimensionality Reduction Based Data Exploration", Proceedings of International Conference on Human Factors in Computing Systems, (2018), pp. 1-13.
28. Kerstin Bunte, John Aldo Lee, "Unsupervised Dimensionality Reduction: The Challenges of Big Data Visualization", Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, (2015), pp. 487-494.

AUTHORS PROFILE

Dr. K. B. V. Brahma Rao has completed his Ph. D in Computer Science and Engineering from Adikavi Nannaya University, Rajamahendravaram, A. P., India. He has 25 years of teaching experience. He is currently working as a Professor in Department of MCA, B. V. Raju College, Vishnu Campus, A.P, India. He has published 4 papers in reputed international journals and 2 in conferences. His main research interests are Data Analytics.

Dr. R Krishnam Raju Indukuri has completed his Ph. D in Computer Science and Engineering from Adikavi Nannaya University, Rajamahendravaram, A. P., India. He has 20 years of teaching experience. He is currently working as a Principal of B. V. Raju College, Vishnu Campus, Bhimavaram, A.P, India. He has published 6 papers in reputed international journals and 4 in conferences. His main research interests are Cloud Computing and Data Analytics.

Dr. Suresh Varma Penumatsa is a Professor & Dean of Academics in the Department of Computer Science & Engineering of Adikavi Nannaya University, Rajamahendravaram, A. P., India. He is having 26 years of teaching experience. His areas of interest include Communication Networks, Image Processing, Speech Processing, Cloud Computing, Data Science and Machine Learning. He has published several research papers. He is a life member of ISTE, SMORSI, ISCA and IISA.

Dr. M. V. Rama Sundari has completed his Ph. D in Computer Science and Systems Engineering from Andhra University, Visakhapatnam, A. P., India. She has 18 years of teaching experience. Her areas of interest include Computer Networks, Data Bases and Data Analytics. She has published several research papers. She is a life member of IEEE, CSI, MISCA, AMIE, IAENG, IAOE, ACA and IARCP.