Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 1, 2020

Review Article

BIGDATA AND MACHINE LEARNING MODELS FOR DIMENTIONALITY REDUCTION PLATFORM

K.S. PRAHARSHITA, SAI SUCHARITHA ARAVABHUMI[2], SASHANK ATTALURI[3], SWATHI MANDAVA[4], S. RAGHAVENDRAN[5], S.K. HASANE AHAMMAD[6]

[1]Student, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, A.P, India.
[2]Student, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, A.P, India.
[3]Student, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, A.P, India.
[4]Student, VRSEC, Vijayawada, Krishna District, A P, India.
[5]Associate Professor, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai.
[6]Research Scholar, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, A.P, India.

Received: 21.11.2019 *Revised: 02.12.2019* *Accepted: 14.01.2020*

ABSTRACT

Various machine learning, cloud computing and big data models are most utilized control parameters in software applications. Therefore these systems are most demand at current trending decades. But these models are needs very low data access time, speed for process. Day to day life data storage servers and devices are costly and hardware complex, so dimensionality is increases with rapid manner. Any type of optimization techniques access takes more time consumption for high dimensional data. So, more applications related problems are occurs only at high dimensional data space does not at low space dimensional data. In this research proposes a dimensional reduction technique with Logistic regression (LR) model. This LR model is most helpful for dimensional reduction and clustering problems. LRML method has reduced the dimensional data size and achieved thee efficiency by 95.3% and ratio of reduction by 35.76%.

Keywords: Machine Learning, Dimensional Adjustment, HDF, Byte Duplication.

DOI: http://dx.doi.org/10.31838/jcr.07.01.88

INTRODUCTION

Internet of Things (IoT) carries numerous irrelevant gadgets related together to form a communitarian figuring condition. IoT dazzles sporadic boundaries as far as network, computational energy and power belongings, which make it certainly particular from the ones, remember by means of methods for the sanctioned guarantee in disseminating structures. So as to stay far away from the tense safety in the IoT area, systems and gadgets need to be validated. In this paper, we don't forget the inserted gadget health only, accepting that machine security is pleasantly in the vicinity. It tends to be visible that the methods of lifestyles of little registering devices inside the IoT territory are specifically level to exceptional wellness assaults. In this work, we provide the conditions of implanted safety, the responses to opposing diverse attacks and the time for opposing satire sealing of the setup gadgets by means of methods for trusted figuring.

Objectives of Dimensionality Reduction

1. *Low data storage accuracy improvement.*
2. *Fats and accurate calculations for reduction of duplication data.*
3. *Storage room requirement.*
4. *Unwanted countless measurement for data reduction.*
5. *Delete the repeated data and highlights.*

The recent technologies related to datasets and number of records, attributes has been analyzed and developed with the help of big data platform. Along, this machine learning tools also useful for dimensionality reduction techniques. In this technique parallel processing data analysis algorithms useful for time confinement applications. As well as the time, speed and efficiency one of the major task with respect to duplicate data dimensionality reduction applications. Moreover various applications are presented but needs improvements. Existed methods like genetic algorithm, differential enaling, conventional models and decision trees are useful for duplicate data reduction models but, these have low performance metrics. When the dimensionality reduction applications performed, the access time automatically decreased. One of the most important data reduction mechanism implemented by using 2009kdd, this is the large dataset consist of 15k data columns. In data mining optimization techniques various algorithms are performed with column wise data reduction models. But, these are very slower than the conventional models. The first significant project has implemented [5] this decreases the data columns with efficient manner. Using this datasets loose a low amount of information this is disadvantages of methods [6].

LITERATURE SURVEY

Dimensionality decrease is a powerful manner to cope with scaling restored the statistics [1]. It is a methodology that tries to amplify diverse over the pinnacle dimensional vectors to a decrease dimensionality area while maintaining estimations among them [2]. The AI and certainties mining strategies may not be compelling for high-dimensional realities in attitude at the scourge of dimensionality and inquiry accuracy and adequacy will degrade quick in view of the estimation increases [3]. The Dimension decline is used for 1) Visualization: To projection of high-dimensional measurements onto 2D or 3D. 2) Data Compression: Efficient collecting and healing. Three) Noise removal: Positive impact on request accuracy. Dimensionality techniques are carried out for the remedy of excessive dimensional statistics as in exquisite clarification microarray assessment, content material characterization, with burdens to a primary extensive assortment of capabilities, with diverse unessential and dreary functions and late examinations results, sparkle off overabundance primarily based detail willpower. The notion for measurement lower can be referenced as seeks after [4-5] 1) the unmistakable evidence of a diminished recreation plan of features which can be farsighted of influences can be useful from a statistics revelation aspect of view. 2) For some getting to know computations, the education, however portrayal time will increase surely with the quantity of functions. 3) Noisy or unimportant functions may additionally comparably affect request as perceptive capabilities so they may have an effect on conversely on precision [6-7].

RELATED METHODS

Post Process Dimensionality Reduction

In this model capacity of the device and cloud storage has updated with post process manner. In this technique all determination datasets are replicated with later hour. The significant function is that is fitness function performed rows and columns have data with a diagonal manner [8]. Before this process characters and datasets, frames have automatically arranged in a association manner. These applications are very useful for static and dynamic document verification framework.

Duplication with Inline Process

In this technique inline centric duplication has performed on datasets. In this hash analysis is main goal of the system; the dimensionality reduction has been performed with the help of rectangular outstanding column analysis. These frameworks correct the error dimensional reduction techniques [11-21]. This is the modern technique utilized for PCA, MDS linear and nonlinear applications respectively.

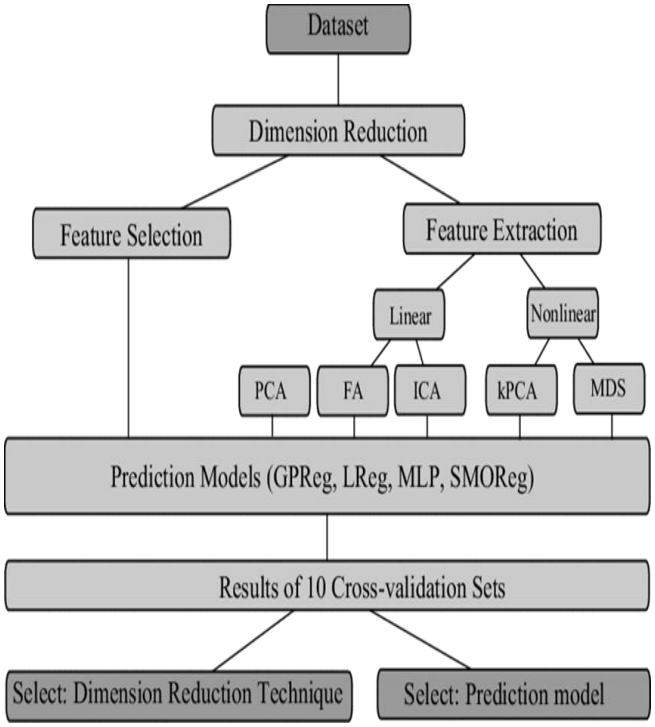

Figure 1: 2009KDD Training Model

Figure 1 explains about dataset dimension reduction procedure. In this technique future selection and extraction are the major steps, in future extraction linear and nonlinear are the again classified techniques. In linear technology principle component analysis, FA and ICA are major optimization dimensionality reduction steps. Coming to nonlinear model KPCA and MDS are the significant methods. These above steps are unified at prediction model. Coming to result section, dimensionality reduction and prediction validate the datasets columns.

LRML-HDFF PROPOSED METHOD

In this section proposed a logistic regression, machine learning, handoops v(HDF) mechanism. This technique has implemented by using python software with the help of JUPYTER notebook. This mechanism is explained in the below figure 2.

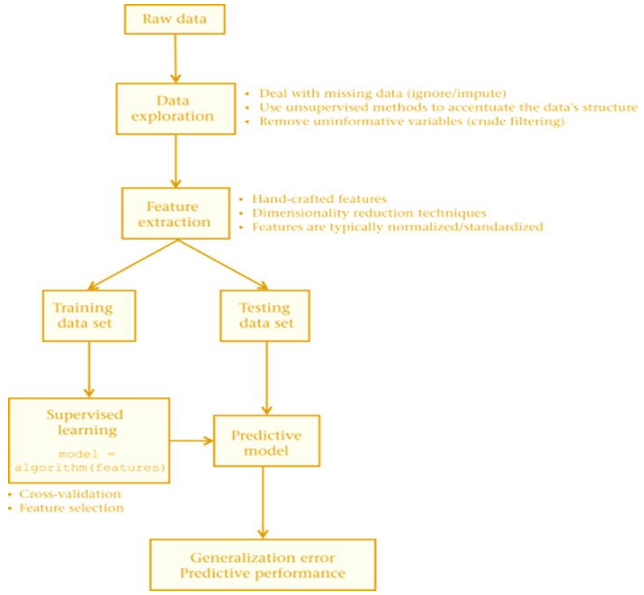

Figure 2: LRML Model Flow

In the first step raw data has been taken has input, coming to second step data exploration is performed. This data deals with ignore missing data, unsupervised methods did not access the structure. This trouble has been removed by unique variable filtering method, coming to next stage future extraction has been performed. In this manual adjustment done by users, also dimensionality reduction technique selected here. The entire data is normalized into typical manner. This section further classified into training and testing data, by using training data cross validation supervised learning with LRML takes place. At prediction HDF model is incorporated.

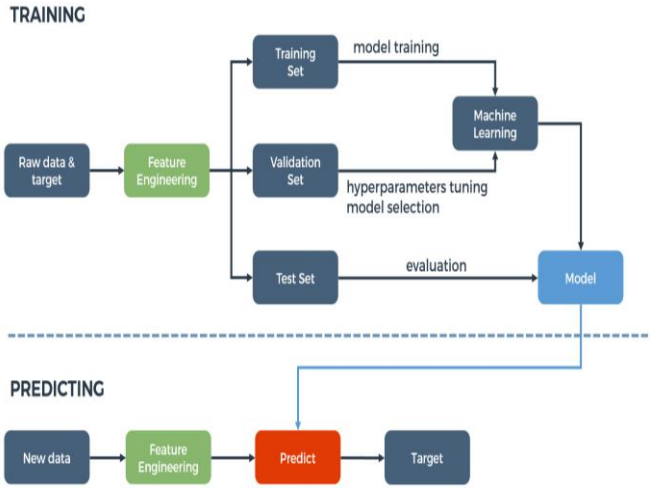

Figure 3: LRML Block Diagram

Figure 3 demonstrate that logistic regression based machine learning data dimensionality reduction and prediction model. In this hyper parameters are tuned with test data, at training new data samples are utilized to predict the data. At final target model and validation model compared with various assessment parameters and conclude that LRML is best.

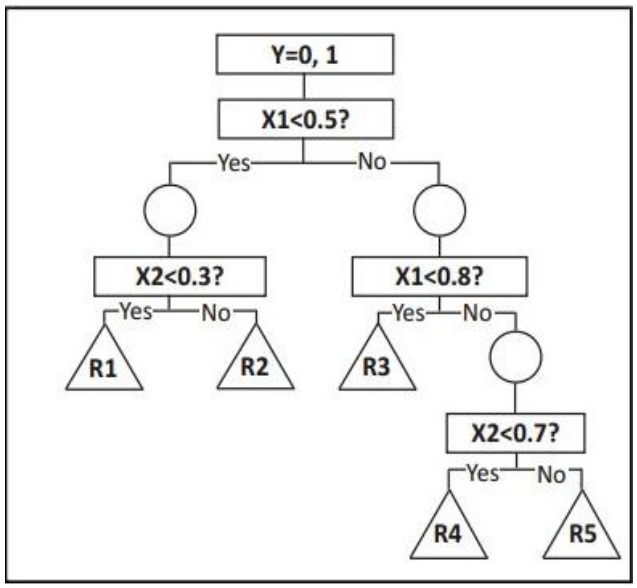

Figure 4: LR Weight Model

Figure 4 demonstrate that LR mechanism with weights, the output weights are decided by impulse function weights. In this section yes or no conditions are decided the outputs. If condition is true registers are capture the weight, else registers are place this information at unconfined registry. At this situation the classification is done by using reduction weight based instruction at dataset. The usage of proposed instruction used for LR tree version and the approval is chosen in appropriate column data.

RESULTS

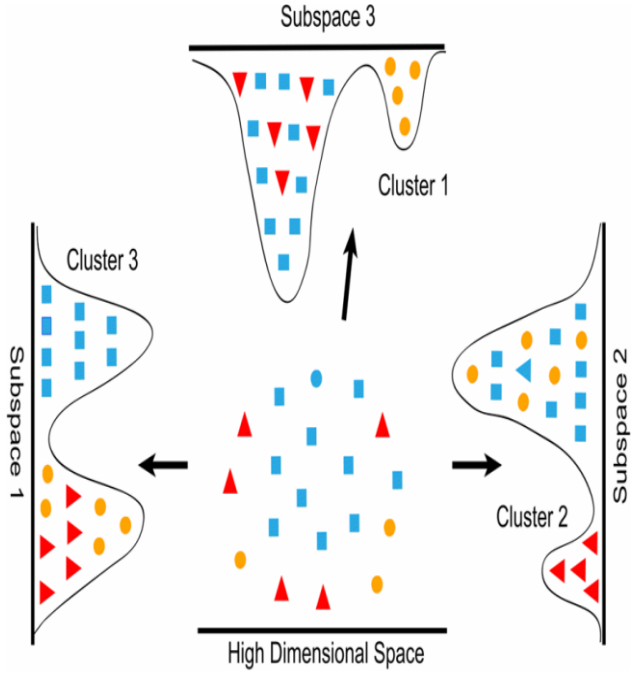

Figure 5: High Dimensional Data

Fig. 5 demonstrates that cluster based high dimensionality reduction techniques. Here, four clusters are used to decrease the subspace column data. The above explanation is useful for better dimensionality reduction.

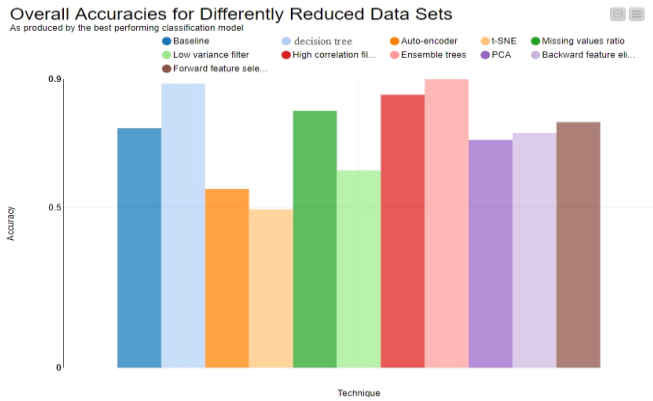

Figure 6: Accuracy Analysis

Fig 6 explains about various models comparison at accuracy point of view. In this proposed LRML-HDF got more accuracy compared to conventional methods. In this physical and regular dimensional reduction has been proposed on scientific datasets. This prediction gives the better throughput and more dimensionality reduction ratio.

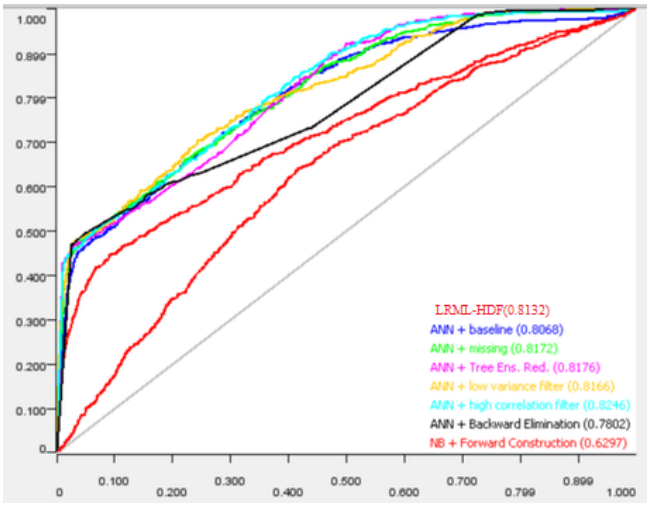

Figure 7: Reduction Ratio

Figure 7 Demonstrate that LRML-HDF gives the more dimensionality reduction ration which is more accurate and efficient.

CONCLUSION

This research explains about 2009kdd dataset dimensionality reduction process, the results explain that LRML method gives the more accuracy compared to conventional models. In this various models like PCA, non-direct, KPCA, DTML achieves less accuracy and low compression ratio. But, by using LRML-HDF can achieve more compression ratio and accuracy in this investigation achieved the efficiency by 95.3% and ratio of reduction by 35.76%.

REFERENCES

1. The experimental results, prototype system, source code, and preprocessed datasets. https://takanori-fujiwara.github.io/s/dr-cl/.
2. The original cPCA implementation. https://github.com/abidlabs/contrastive. Accessed: 2019-3-6. [3] H. Abdi and D. Valentin. Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics,* pp. 651–657, 2007.

3. Marella, S.T., Karthikeya, K., Kushwanth, V.S., & Bezawada, A. (2018). Enhancement of Performance and Economy of Data Centers by Virtualization. *International Journal of Simulation--Systems, Science & Technology*, 19(6).
4. Abid and J. Zou. Contrastive variational autoencoder enhances salient features. arXiv preprint arXiv:1902.04601, 2019.
5. Marella, S.T., Karthikeya, K., Myla, S., Sai, M.M., & Allam, V. Detecting Fraudulent Credit Card Transactions Using Outlier Detection.
6. Abid, M.J. Zhang, V.K. Bagaria, and J. Zou. Contrastive principal component analysis. arXiv preprint arXiv:1709.06716, 2017.
7. Ahammad, S.H., Rajesh, V., Saikumar, K., Jalakam, S., & Kumar, G.N.S. (2019). Statistical analysis of spinal cord injury severity detection on high dimensional MRI data. *International Journal of Electrical and Computer Engineering,* 9(5), 3457-3464. doi:10.11591/ijece.v9i5.pp3457-3464
8. Ahammad, S.H., Rajesh, V., Neetha, A., Sai Jeesmitha, B., & Srikanth, A. (2019). Automatic segmentation of spinal cord diffusion MR images for disease location finding. *Indonesian Journal of Electrical Engineering and Computer Science,* 15(3), 1313-1321. doi:10.11591/ijeecs.v15.i3.pp1313-1321
9. Ahammad, S.K Hasane, V. Rajesh, and MD Zia Ur Rahman. & quot; Fast and Accurate Feature Extraction-Based Segmentation Framework for Spinal Cord Injury Severity Classification. & quot; *IEEE Access* 7 (2019): 46092-46103
10. Abid, M. J. Zhang, V. K. Bagaria, and J. Zou. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nature Communications,* 9(1):2134, 2018.
11. Myla, S., Marella, S.T., Goud, A.S., Ahammad, S.H., Kumar, G. N. S., & Inthiyaz, S. Design Decision Taking System for Student Career Selection For Accurate Academic System.
12. E. Acuna and C. Rodriguez. The treatment of missing values and its effect on classifier accuracy. *In Classification, Clustering, and Data Mining Applications,* pp. 639–647. Springer, 2004.
13. M. Ankerst, M.M. Breunig, H.P. Kriegel, and J. Sander. OPTICS: Ordering points to identify the clustering structure. *In Proceedings of ACM SIGMOD International Conference on Management of Data,* pp. 49–60, 1999.
14. Bezawada, A., Marella, S.T., & Gunasekhar, T. (2018). A Systematic Analysis of Load Balancing in Cloud Computing. *International Journal of Simulation--Systems, Science & Technology,* 19(6).
15. Z. Bar-Joseph, D.K. Gifford, and T.S. Jaakkola. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics,* 17(1): S22–S29, 2001.
16. M. Behrisch, B. Bach, N. Henry Riche, T. Schreck, and J.D. Fekete. Matrix reordering methods for table and network visualization. Computer Graphics Forum, 35(3):693–716, 2016.
17. Hasane Ahammad, S., Rajesh, V., Hanumatsai, N., Venumadhav, A., Sasank, N.S.S., Bhargav Gupta, K.K., & Inithiyaz, S. (2019). MRI image training and finding acute spine injury with the help of hemorrhagic and non-hemorrhagic rope wounds method. *Indian Journal of Public Health Research and Development,* 10(7), 404-408.
18. Saikumar, K., Rajesh, V., Ramya, N., Ahammad, S.H., & Kumar, G.N. S. (2019). A deep learning process for spine and heart segmentation using pixel-based convolutional networks. *Journal of International Pharmaceutical Research,* 46, 278-282. Retrieved from www.scopus.com
19. Hasane Ahammad, S. K., & Rajesh, V. (2018). Image processing based segmentation techniques for spinal cord in MRI. *Indian Journal of Public Health Research and Development,* 9(6), 317-323. doi:10.5958/0976-5506.2018.00571.5
20. Vijaykumar, G., Gantala, A., Gade, M.S.L., Anjaneyulu, P., & Ahammad, S.H. (2017). Microcontroller based heartbeat monitoring and display on PC. *Journal of Advanced Research in Dynamical and Control Systems,* 9(4), 250-260. Retrieved from www.scopus.com
21. Raj Kumar, A., Kumar, G.N.S., Chithanoori, J.K., Mallik, K.S.K., Srinivas, P., & Hasane Ahammad, S. (2019). Design and analysis of a heavy vehicle chassis by using E-glass epoxy & S-2 glass materials. *International Journal of Recent Technology and Engineering,* 7(6), 903-905. Retrieved from www.scopus.com