ISSN 2455 – 9091

IJCSET JUNE Volume 8 Issue 6

International Journal of Computer Science Engineering and Technology (IJCSET)

https://www.doi.org/10.5281/zenodo.6793656

INDUSTRIAL BIG DATA ANALYTICS FOR DEEP DENOISING AUTOENCODER

[1] Dr. G. A. Mylavathi, [2] Dr. B. Srinivasan
[1, 2] Assistant Professor
[1, 2] Department Of Computer Science
[1, 2] Gobi Arts & Science College, Gobichettipalayam, 638453
[1, 2] Tamilnadu, India.

ABSTRACT - The new improvement of cyber-physical systems, big data, and cloud computing and industrial wireless networks, another period of industrial big data is presented. Deep learning has critical potential for arrangements giving in refined industrial applications. In this paper, an idea of device electrocardiogram (DECG) is introduced, and an algorithm in light of deep denoising autoencoder (DDA) is proposed for the expectation of the leftover valuable existence of industrial hardware. The proposed algorithm called coordinated DDA and the algorithm are furnished and DECG is contrasted and conventional processing plant data framework, and the plausibility and adequacy of the proposed algorithm are approved tentatively. The proposed idea and algorithm consolidate regular industrial situation and advance artificial intelligence, which can possibly speed up the execution of industry.

Keywords: [Big data, Cyber-physical systems, Deep learning, Industrial big data.]

1. INTRODUCTION

Big data refers to particularly colossal datasets with complex plans that are hard to manage utilizing standard techniques and devices. The term cooperation incorporates, catch, stockpiling, formatting, extraction, duration, integration, analysis, and visualization With the quick progress of PCs and the Internet, worldwide data size is growing, and limit, analysis, and processing of big data have acquired essential thought. An extending number of appraisal foundations and attempts are suitably looking at the improvement of big data analysis and processing frameworks, and planning application research on big data. Big data processing models consistently embrace a dissipated figuring structure due to huge data sets and computational costs. After a short time, the most overall utilized big data processing programming model is Map Reduce, which was proposed by Google. As the most all around utilized orbited big data processing stage, Hadoop is the open source execution of Map Reduce and can give reliable, useful, and adaptable data processing models. Hadoop maintains building groups with run of the mill unpretentious hosts, and each limit node has its own free processing unit and certain enrolling power.

Big Data is the term portraying enormous arrangements of different data organized, unstructured, and semi organized that is constantly produced at a high velocity and in high volumes. A developing number of organizations presently utilize this data to uncover significant experiences and further develop their independent direction, yet they can't store and handle it through conventional data stockpiling and processing units.

The improvement of Internet of Things (IoT) and cloud advances caused arrangement of sensors and line design in assembling climate, which expanded the expense of assembling upkeep. The prescient support, as a significant part in assembling upkeep, assumes a crucial part in planning, upkeep the board, and quality improvement. By and large, the prescient upkeep can be classified into experience-based models, physical science based models, and data driven models. The device electrocardiogram (DECG) guideline is presented, and another strategy in light of deep learning and DECG is proposed for prediction of RUL of gear as well as production line. DECG is like screen the soundness of the human body records devices process duration with all its sub-processes. Because of considerably more data gathered from DECG, it's feasible to present deep learning and completely upgrade the exhibition of deep learning for explicit application. In light of deep learning and countless rush to-disappointment tests, the proposed algorithm can give an exact prediction of device RUL.

2. EXISTING METHODOLOGIES

1. Dao Zhou et.al proposed Distributed Data Analytics Platform for Wide-Area Synchrophasor Measurement Systems. s synchrophasor data begin to assume a huge part in power system activity and dynamic review, data processing and data analysis capacity are basic to Wide area measurement systems (WAMS). The Frequency Monitoring Network (FNET/GridEye) is a WAMS network that gathers data from many Frequency Disturbance Recorders (FDRs) at the conveyance level. The past FNET/GridEye data focus is restricted by its data stockpiling ability and calculation power. Focusing on scalability, extensibility, concurrency and robustness, a distributed data analytics platform is proposed in this paper to handle huge volume, high speed dataset. An assortment of ongoing and non-continuous synchrophasor data analytics applications are facilitated by this platform. The calculation load is imparted to adjust by different nodes of the analytics cluster, and big data analytics apparatuses, for example, Apache Spark are taken on to oversee huge volume data and to help the data processing speed. Future data analytics applications can be effectively formed and connected to the system with basic configuration.

2. M. U. Bokhari et.al proposed a three layered architecture model for storing and analyzing big data. The three layers are data gathering layer, data storing layer and data analysis and report age layer. To assemble and deal with the colossal volume of big data coming from fast sources, for example, sensors or social media, a cluster of high velocity nodes or cuts off are kept in the data gathering layer. The data stockpiling layer is answerable for storing the big data. The Hadoop Distributed File System (HDFS) can be utilized for data capacity. In the data analysis layer, AI techniques, for example, ANN, Naive Bayes, SVM and Principal Component Analysis and so on are utilized to stir information from the immense complex data pieces.

3. Anand Paul et.al proposed Smartbuddy: Defining Human Behaviors Using Big Data Analytics in Social Internet of Things. As we dig into the Internet of Things, we are seeing escalated connection and heterogeneous correspondence among various devices over the Internet. Thus, these devices create a monstrous volume of big data. The capability of these data has been dissected by complex network hypothesis, portraying a particular branch, known as "human dynamics." In this expansion, the objective is to depict human conduct in the social area progressively. These goals are beginning to be practicable through the amount of data gave by smartphones, social networks, and smart cities. These make the climate more astute and offer a canny space to detect our exercises or activities and the development of the ecosystem. To address the previously mentioned needs, this work presents another idea of SmartBuddy that spotlights on the analysis, the ecosystem given by smart cities, wearable devices (e.g., body area network), and big data to decide human behaviors as well as human dynamics.

3. PROPOSED METHODOLOGY

Big Data analytics encompasses the processes of collecting, processing, filtering/cleansing, and analyzing extensive datasets so that organizations can use them to develop, grow, and produce better products. Let's take a closer look at these procedures.

3.1 Industrial Big Data Analytics

In this paper, the algorithm for industrial big data analytics, data acquisition and pre-processing, is presented. As it was a lot of comparative devices or a gathering of devices creates a lot of data. Data acquisition and prediction ought to be achieved before data analysis. To accomplish higher prediction precision and the best utilization of cleaned data, the capacity of deep figuring out how to achieve understanding and information from big data is utilized.

Deep learning is performed with referenced records, which are utilized as info and output data. In prediction stage, in the event that at least one comparative devices are in a sharp wear stage, the state data of these devices are treated as contribution of deep neural network, while network output is the predicted RUL.

3.2 Rul Prediction Based On Deep Learning

In recent years, the data-driven models play a vital part in dynamic support and RUL prediction. Nonetheless, one of the main model errands is highlight designing, which includes a few expert data tasks, like dimensionality decrease, and depends significantly on unambiguous production situation. In assembling climate, various sorts of processing units for the most part have their own functioning components and support rationales. Subsequently, it very well may be difficult to accomplish countless relating progressed data-driven models for upkeep, which is superfluous and causes an asset squandering.

In this paper, propose deep learning for prediction of the excess valuable life. One of the clearest benefits of deep learning is its capacity to remove the highlights consequently, for example, convolutional neural network (CNN) and recurrent neural network (RNN). Also, as the layer gets deeper, the quantity of highlights normally reduces and acquired highlights become more unique. In application situation, the verifiable data of each cycle working time are utilized for preparing. In light of the production foundation deciphered previously, and to accomplish the best element extraction utilizing deep learning, the coordinated deep denoising auto-encoder (IDDA) is brought into assembling climate.

The proposed IDDA comprises of DDA and a direct regression analysis. As referenced before objective is to foresee the equipment RUL in view of present status, which is on the off chance that, the functioning season of equipment processes. Taking into account that gathered data comprise of time series, to accomplish a superior prediction, at each time second during the preparation phase the dataset is parted into far off dataset and recent datasets, and afterward, they are utilized as contributions of two unique DDA. The far off datasets denote the datasets that are far away from current time second, while the recent datasets denote datasets that are near current time second. According to viewpoint, a precise prediction requires sensible combination of harm inclination and present statuses. In this way, the far off datasets are utilized to mimic the harm pattern, while the recent datasets are utilized to recreate the smoothing system of recent change. With expect to keep away from the overfitting, a renowned stunt called dropout is applied to data stream in two deep models. A while later two outputs and inevitable straight regression is performed to change the discrete datasets to equipment RUL. The subtleties of proposed algorithm (see Algorithm 1), including foreseeing phase, are introduced in the following.

Algorithm 1 Algorithm of Integrated Deep Denoising Auto-Encoder

Training Input: $D = \{(R_i, Y_i)\}_{i=1}^{N}$ where R_i and Y_i denote the i^{th} equipment's records and the remaining useful life, respectively

Prediction input: distant records $I_d^p = \{P_i\}_{i=1}^{S}$ and recent records $I_r^p = \{P_i\}_{i=S}^{K}$

1. //training phase
2. Split D into $I_d \leftarrow \{R_{ij}\}_{i=1}^{N}{}_{j=1}^{S}, I_r \leftarrow \{R_{ij}\}_{i=1}^{N}{}_{j=1}^{K}$ and $Y \leftarrow \{\{Y_i\}_{i=1}^{N}$
3. For i ← 1 to N //for each equipment
4. $I_d^i \leftarrow \{R_{ij}\}_{j=1}^{S}$
5. $I_r^i \leftarrow \{R_{ij}\}_{j=S}^{K}$
6. Output1←DeepDenoisingAuto-encoder1 (input← I_d^i)

7. Output2←DeepDenoisingAuto-encoder2 (input←I_r^i)
8. //integrate the outputs of two deep neural networks into one output
9. Output ← Fusion (Output1, Output2)
10. //perform regression to transform the discrete output into prediction
11. Y_i^{Δ} ← Regression (Output)
12. // calculate Euclidean loss between predictions and records
13. $Loss \leftarrow \frac{1}{2N}\sum_{i=1}^{N}\|Y_i - Y_i^{\Delta}\|_2^2$
14. //use stochastic optimization to update the parameters of the IDDA
15. // α is the learning rate
16. Stochastic Optimization (Loss, α)
17. End for
18. // prediction phase
19. Poutput1 ← DeepDenoisingAuto-encoder1 (input ←I_d^p)
20. Poutput2 ← DeepDenoisingAuto-encoder2 (input ←I_r^p)
21. Poutput ← Fusion (POutput1, POutput2)
22. Prediction ← Regression (POutput)

At the point when the authentic data from K time minutes are assembled and cleaned, the first step is to divided the data into far off datasets and recent datasets. In this paper, datasets gathered at K time minutes don't have a uniform conveyance constantly. As displayed in line 2 of proposed algorithm, time minutes from NO.1 to NO.S have uniform conveyance, which gives an enormous scope to harm tendency reproduction. Then again, the remainder of time minutes, which structure the time span from NO.S to NO.N, psychologist to completely typify the recent trademark. Moreover, since the datasets have similar data structure, DDAs has a similar construction. In the preparation phase, after data fusion and regression, the distinction of prediction and recorded values is characterized as Euclidean loss. In addition, assuming Adam stochastic enhancement is taken on; quicker integrators can be accomplished contrasted and conventional SGD or batch processing.

4. EXPERIMENT RESULT

4.1 Accuracy

Datasets	WAMS	HDFS	Proposed IDDA
100	78	85	97
200	74	87	96
300	71	83	94
400	69	81	92
500	65	78	90

Table 1.Comparison Table of Accuracy

The Comparison table 1 of Accuracy Values explains the different values of existing algorithms (WAMS, HDFS) and proposed IDDA. While comparing the Existing algorithm (WAMS, HDFS) and proposed IDDA, provides the better results. The existing algorithm values start from 65 to 78, 78 to 85 and proposed IDDA values start from 90 to 97. The proposed IDDA gives the great results.

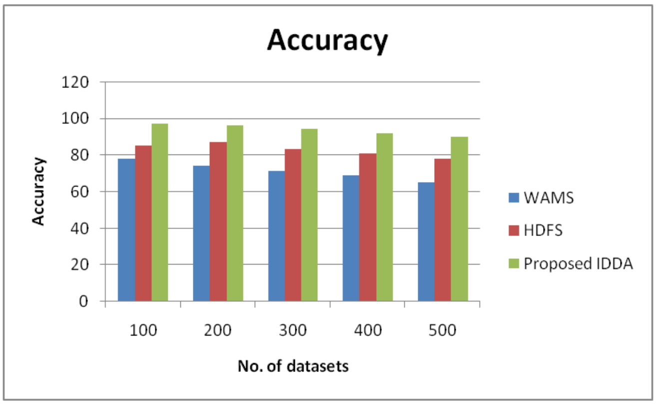

Figure 1.Comparison chart of Accuracy

The Figure 1 Shows the comparison chart of Accuracy demonstrates the existing1, existing 2 (WAMS, HDFS) and proposed IDDA. X axis denote the No. of datasets and y axis denotes the Accuracy in percentage. The proposed IDDA values are better than the existing algorithm. The existing algorithm values start from 65 to 78, 78 to 85 and proposed IDDA values start from 90 to 97. The proposed IDDA gives the great results.

4.2 Precision

Datasets	WAMS	HDFS	Proposed IDDA
100	0.65	0.72	0.98
200	0.69	0.71	0.95
300	0.64	0.68	0.93
400	0.63	0.65	0.89
500	0.61	0.62	0.88

Table 2.Comparison table of Precision

The Comparison table 2 of Precision Values explains the different values of existing algorithms (WAMS, HDFS) and proposed IDDA. While comparing the Existing algorithm (WAMS, HDFS) and proposed IDDA, provides the better results. The existing algorithm values start from 0.61 to 0.65, 0.62 to 0.72 and proposed IDDA values start from 0.88 to 0.98. The proposed IDDA gives the great results.

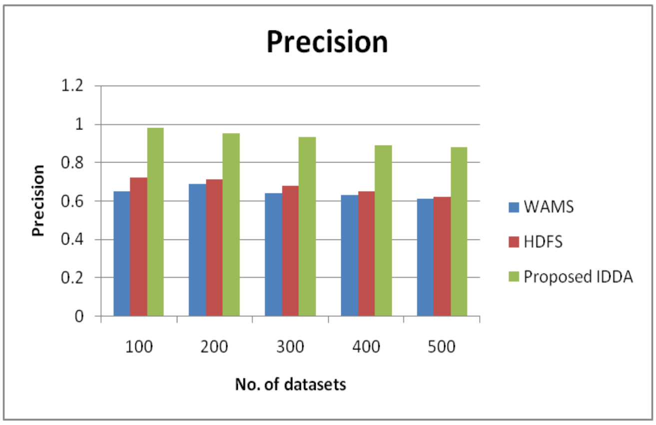

Figure 2.Comparison chart of Precision

The Figure 2 Shows the comparison chart of Precision demonstrates the existing1, existing 2 (WAMS, HDFS) and proposed IDDA. X axis denote the No. of datasets and y axis denotes the Precision in percentage. The proposed

IDDA values are better than the existing algorithm. The existing algorithm values start from 0.61 to 0.65, 0.62 to 0.72 and proposed IDDA values start from 0.88 to 0.98. The proposed IDDA gives the great results.

CONCLUSION

In this paper, a new algorithm for RUL prediction in view of DECG and deep learning is introduced. The idea of DECG was presented. Then reduce the effect of specialists' insight and human choice on prediction, a deep learning philosophy proposed IDDA algorithm, which embraces IDDA and regression activity was utilized. They got results have shown a high viability of proposed algorithm. By and by, the correlation results have shown prevalence of proposed algorithm and its plausibility to speed up the execution of Industry.

REFERENCES

[1]. K. Park, M. C. Nguyen and H. Won, "Web-based collaborative big data analytics on big data as a service platform," 2015 17th International Conference on Advanced Communication Technology (ICACT), 2015, pp. 564-567, doi: 10.1109/ICACT.2015.7224859.

[2]. O. Gogliano Sobrinho et al., "Big data analytics in support of the under-rail maintenance management at Vitória – Minas Railway," 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 6026-6028, doi: 10.1109/BigData52589.2021.9671739.

[3]. I. K. Nti, J. A. Quarcoo, J. Aning and G. K. Fosu, "A mini-review of machine learning in big data analytics: Applications, challenges, and prospects," in Big Data Mining and Analytics, vol. 5, no. 2, pp. 81-97, June 2022, doi: 10.26599/BDMA.2021.9020028.

[4]. N. W. Grady, J. A. Payne and H. Parker, "Agile big data analytics: AnalyticsOps for data science," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 2331-2339, doi: 10.1109/BigData.2017.8258187.

[5]. A. Hennig et al., "Big social data analytics of changes in consumer behaviour and opinion of a TV broadcaster," 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 3839-3848, doi: 10.1109/BigData.2016.7841057.

[6]. D. Wu, "A big data analytics framework for forecasting rare customer complaints: A use case of predicting MA members' complaints to CMS," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 3965-3967, doi: 10.1109/BigData.2017.8258406.

[7]. O. Baker and C. N. Thien, "A New Approach to Use Big Data Tools to Substitute Unstructured Data Warehouse," 2020 IEEE Conference on Big Data and Analytics (ICBDA), 2020, pp. 26-31, doi: 10.1109/ICBDA50157.2020.9289757.

[8]. M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara and K. Sadatdiynov, "A survey of data partitioning and sampling methods to support big data analysis," in Big Data Mining and Analytics, vol. 3, no. 2, pp. 85-101, June 2020, doi: 10.26599/BDMA.2019.9020015.

[9]. R. Norman, J. Bolin, E. T. Powell, S. Amin and J. Nacker, "Using Big Data Analytics to Create a Predictive Model for Joint Strike Fighter," 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 3590-3596, doi: 10.1109/BigData.2018.8622388.

[10]. H. H. Larsen, J. M. Forsberg, S. V. Hemstad, R. R. Mukkamala, A. Hussain and R. Vatrapu, "TV ratings vs. social media engagement: Big social data analytics of the Scandinavian TV talk show Skavlan," 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 3849-3858, doi: 10.1109/BigData.2016.7841058.