

Big Data Analytics Applying the Fusion Approach of Multicriteria Decision Making with Deep Learning Algorithms

Swarajya Lakshmi V Papineni¹, Snigdha Yarlalagadda², Harita Akkineni³, A. Mallikarjuna Reddy⁴

¹Professor, Department of Information Technology, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, AP, India

²Technology Analyst, Infosys Limited, Bangalore, Karnataka, India

³Associate Professor, Department of Information Technology, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, AP, India

⁴Assistant Professor, Department of CSE, Anurag University, Hyderabad, AP, India

¹papinenivsl@gmail.com, ²snigdha.y928@gmail.com, ³akkinenih@gmail.com, ⁴mallikarjunreddycse@cvsr.ac.in

Abstract: Data is evolving with the rapid progress of population and communication for various types of devices such as networks, cloud computing, Internet of Things (IoT), actuators, and sensors. The increment of data and communication content goes with the equivalence of velocity, speed, size, and value to provide the useful and meaningful knowledge that helps to solve the future challenging tasks and latest issues. Besides, multicriteria based decision making is one of the key issues to solve for various issues related to the alternative effects in big data analysis. It tends to find a solution based on the latest machine learning techniques that include algorithms like decision making and deep learning mechanism based on multicriteria in providing insights to big data. On the other hand, the derivations are made for it to go with the approximations to increase the duality of runtime and improve the entire system's potentiality and efficacy. In essence, several fields, including business, agriculture, information technology, and computer science, use deep learning and multicriteria-based decision-making problems. This paper aims to provide various applications that involve the concepts of deep learning techniques and exploiting the multicriteria approaches for issues that are facing in big data analytics by proposing new studies with the fusion approaches of data-driven techniques.

Keywords: Big data, deep learning algorithm, data-driven approach, machine learning, multiple criteria decision-making.

I. INTRODUCTION

From the analysis of big data, the major challenge is to manage the unrelated set of data to be converted on a single basis [1]. The conventional methods were failed to do so the process for the approach to provide the correct insight for the data ingestion that includes the data division based on the velocity, size, and in different shapes or formats which are to be further intended to be responsive enough for it to be applicable for the number of optional applications, with this progress of population and the

communication field of different kinds of things like networking, soft computing, IoT, and variety of knowledge shared for a meaningful understanding of data contents on the related fields in computer science, information technology, and agriculture. According to the studies, almost locally, the 1.5 ZB of data [2] has been created that made the challenges in case of information for insight knowledge on the researchers' issues. In the global view, it has been decremented to 1.3 ZB for it to prepare easy as on the data created and for the utilization of applications beyond the real-time environment. There are various tools and algorithms to solve and handle the large set of data called to be the big data [3] in an effective manner for the electronic data available from various sources. By the end of 2020, the data that can be created can be approximating at the rate of 50 ZB. These developed tools are capable of achieving the process of the acquired data, later, applying various [4] filtering technique for the process to be done for the data to convert in an understandable meaning and to find proper insights to communicate the output with external tools under the ongoing process for data to be exchangeable.

For the process of making the data to be understandable, various machine learning (ML) algorithms were applicable, such as neural networks, gravitational search, support vector machine, tabu search, particle swarm optimization (PSO), long short term memory network (LSTM), genetic algorithm, fuzzy logic, and deep learning techniques. It is all related to the data-driven approach for it to be provided with a large sequence of data to understand the process of data for training the network. Combining the decision-making analysis makes the process entirely multiple criteria under technical issues for providing the researchers with a concrete idea for future development in the analysis of the big data sets. Besides, the processing involves the advantageous out of making it possible for application in different fields for better choice-making, effective and efficiency overall in enhancing the potentiality for the practitioner of the support [5],[6] making system in finding the novelty in the application of various domains involved.



Various decision-making support systems exist in the real-world environment that improve potentiality in achieving this concerning issue [7]. Although, decision seeking issue plays a crucial step in the successful outcome of an institution or organization. Depending on the multiple criteria, the provided difficulties are further available to handle the proper data acquisition aspect. In recent times, the problem solving on big data is up to the benchmark [8] by processing the output. Still, they did not facilitate the deep learning technique to attain more intellectual decision support on the entire system based integration. This paper studies the combined analysis of data collected [9], and processing is based on the proposed novel fusion approach by integrating deep learning with the multi-criterion decision-making problem.

II. Existing Methods

In the existing methods, it is found that the recent findings on the approaches to solve the issues and challenges behind the big data analysis to achieve precision on the processing of data and acquiring it to further applications for the other algorithms to the exploitation of support-oriented systems. A single based deep learning (DL) model with the integration of shallow kind model has been established to use it in hierarchical structural tree format. Though it is developed for the sample discretization into various levels to form the clusters, every deep learning model is trained by the related discrete model [10]. Five samples have been developed. Thus, proposed a big data hierarchical deep learning, known to be the BDHDLs, to organize the multiple criterion models. This unique study makes the distribution of cluster adapted [11], [12] for the model that has been trained. There exists five different phased. In the foremost phase, the features of behavior and contents are chosen using the tools and techniques that involve big data's objective.

Later, the second phase compares the overall data sequence on a single cluster format by doing it in a parallel improvement by the K-means approach of the cluster [13]. The hierarchical based on the clustering procedure will be added for every concentration of single cluster included with minimum excellence simultaneously in analogous with the produces a cluster of the obtained subtree. This will be fed for the final generalized approximation to get the information of achieving the greater data pattern from the cluster analysis. The estimation error can be reduced as the computing rate occurs by organizing the multiple criterion models. This unique study makes the distribution [14] of cluster adapted for deep learning schematics. Combining the decision-making analysis makes the process entirely multiple criteria under technical issues for providing the researchers with a concrete idea for future development in the analysis of the big data sets.

In Multivariate regression, we can easily find the output from the optimized theta coefficients. It is fast compared to the neural network approach since it requires less training data to model the regression analysis. K-Nearest Neighbor (KNN) is one of the supervised learning techniques in

machine learning, which analyzed that for both classification basis and regression problems, we use KNN for the regression case to predict the continuous values. Our work suggests many new techniques in solving the machine algorithms oriented that introduce the deep insights of implementing the collective influences that involve the multiple existences of variables in finding a solution for the processing data found at the data collected in the system's electronic fields.

III. Proposed Method

Large sequence data employs the deep learning-based with a fusion of multiple criteria-based decision making the hierarchical clustering algorithm for it to discover various features of which increased products hard obtaining precise characteristics for the remoteness ascertaining likeness among tests, uproarious for inclusive of unessential data at present joined group staggered progressive based on decision tree named structure. On the other hand, the unessential data significantly diminish the viability interruption discovery framework.

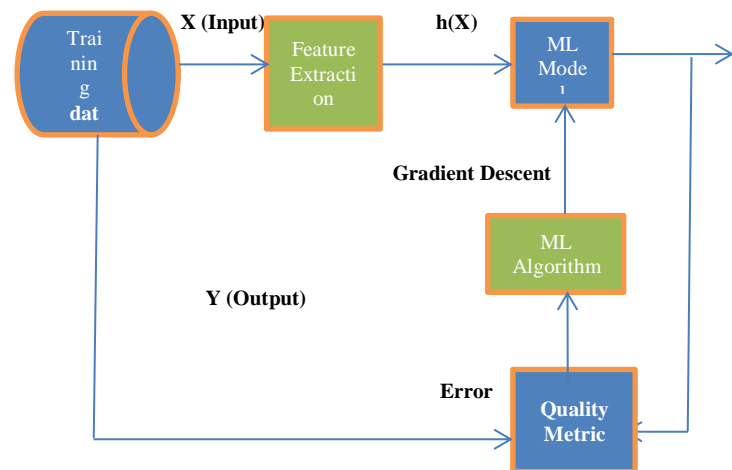


Fig 1: Block diagram of Multiple Variate Linear Regression based deep learning algorithm

To improve the interruption discovery execution of the progressive effects for the profoundly prepared bunch staggered group dependent mixed on conducting highlights for the substance highlights [13-14]. Various decision-making support systems exist in the real-world environment that improve potentiality in achieving this concerning issue. Since design doesn't take the structure of information into thought, it might confront trouble to deal with information with uncommon qualities. Figure 1 represents the square outline of Multiple Variate Linear Regression based ML-DL calculation. Compared to the normal, profoundly associated network, it has two significant qualities: (1) sparseness of association and (2) sharing of weight. The algorithm steps discussed below give how the end to end process is implemented.

Algorithm steps to implement MC-DL for being used in the analytics of big data:

1. The input of the model considered: a, b, c d model parameters defined from the datasets collected from the running records.
2. Output: Is to be observed based on the ordered list of the problem solved
3. Initializations with a batch of the data stored
4. For each dataset

$$B(a_i, k) = b(a_i) + b(N_i) \text{ for the entire dataset followed}$$

The overall effectiveness of ai's coefficient is calculated at the influencing relation between the agents of one and other initial values if they exist at the global set of problems.

$b(N_i)$ = Neighboring set of values

5. The initial value that was fed to the system is expressed as,

$$\mu_{ij} = b(a_i) + b(a_j) / K$$

The constant normalization parameters are the beneficial agent factors, $b(N) = \sum_{j \in N} \mu_{ij} b(a_j)$

6. The mapping function defined for the entire hidden layers for the DL is

$a_{i+1} = f_k(W_i * a_i + b_i)$ with its activation, i.e., the sigmoid function may be cumulatively added with the weights and hidden layers connecting the networking of DL.

The decision value that is normalized for the variable of x is

$$\text{Decision value}(x) = \frac{f_k(x) - \text{mean value}}{\delta_k}$$

7. End for
8. Return

Computing benefit factors of the neighboring agents, Mutual coefficients under the common agents can be calculated based on the weighting index.

$$\mu_{ij} = \frac{b(a_i) + b(a_j)}{k} * (C_{a_j} + 1)$$

Therefore, the overall assessment of the benefit plan will be investigated.

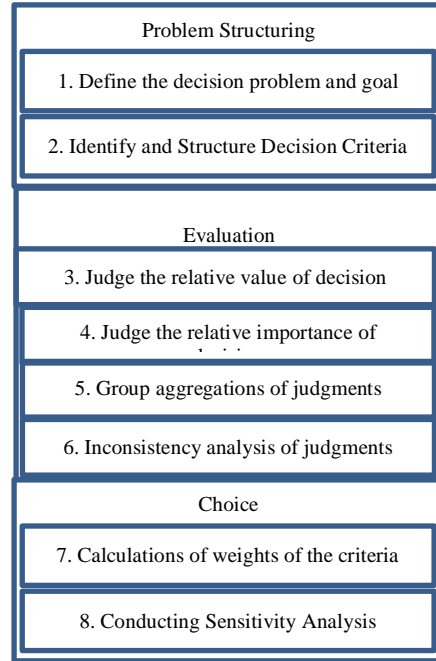


Fig 2: Entries based on ranking for multiple criterion decision making

Entries based on ranking for multiple criterion decision making are depicted in figure 2. Though it is developed for the sample discretization into various levels from the clusters, every model of deep learning is trained by the model on related discrete up to which five samples have been developed. Thus, the proposed big data exploits the key feature of the hierarchical oriented [15, 21-22] deep learning technique, known to be the BDHDLs, to organize the multiple criterion models [16-20]. This unique study makes the distribution of cluster adapted for the model that has been trained.

IV. Results and Discussion

Every model of deep learning is trained by the model on related discrete up to which five samples. The real time-based end to end process of the data that has been processed is depicted in figure basis. It attains for the start process of which it has the historical relationship for long term. Also, for real-time datasets, it will be developed on a streaming basis of a memory unit that can store the large set and finally attain into the storage part of the processing unit. For each data point p, the data points whose input variables (X) are most near to point (p) are selected.

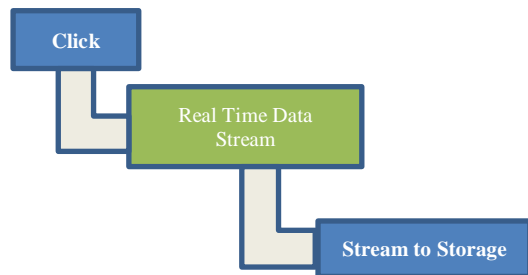


Figure 3: End to end data processing

Besides, it is a nonparametric supervised learning technique in machine learning applicable to regressive analysis and classification-based problems. Here we use the DL for the regression case to predict the continuous values. Here suppose we select many neighbors. For each data point p , the data points whose input variables (X) are most near to point (p) are selected. Then, distances are calculated using Euclidean distance, Manhattan, or Minkowski distances method. Then the mean of these 10 points is used as the predicted output value for p .

Table 1: Numerical Prediction accuracy for the fields estimated

Techniques applied	Statistical analysis	Estimation error
Decision tree	-0.9560	1015.56
KNN	-5.432	1867.55
Ridge	-11.2645	986.324
Linear Regression	-9.5399	3512.369
Proposed MLDM	-0.0123	586.369

Table 2: Categorical Prediction accuracy for the fields estimated

Techniques applied	Precision value	Accuracy
Gaussiandistribution	0.21	0.13
Bernoulis approximation	0.23	0.49
Decision tree	0.43	0.98
Support vector machine (SVM)	0.68	0.59

The results obtained for the accuracy values and prediction of different fields, such as numerical and categorical, are done as per the utilization of data processed. This will solve the challenges that are faced by the system on which the data is imposed.

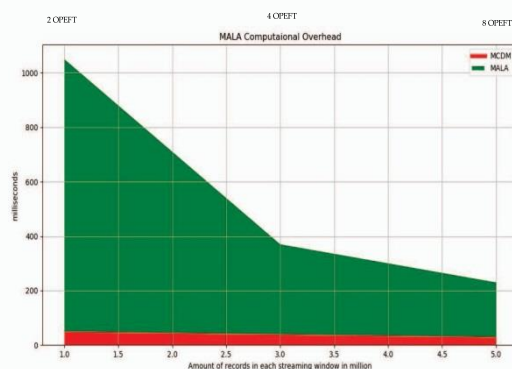


Fig: 4 Multiagent Lambda architecture for computational overhead

In online detection, the structural data provided will be in attaining the data acquisition termed in the interlink of property and runs through the query kind of

language for better result analysis. In figure 4, the green area indicates the time of processing by the MALA. The red area indicates the latency obtained because of multicriteria decision-making. In the response, the time obtained is at the rate of average. The latency of MCDM reduced linearly with the node number of cluster increments.

Further, obtaining the window for the length of data supposed to be in the selection aspect. However, this will not enhance the faster estimation time because of the metadata availability and an extensive sequence of data with the processing data acquired in the overall factor's minimal impishness. The latency test occurred overhead to process the delay time, 4.9% rate in the average responsiveness.

V. CONCLUSION

In this study, the present scenario of machine learning with multiple criteria decision-based problem solving has played a challenging role in increasing the population to deal with big data problems. The challenges facing the growth of volume in the data applications in various domains have been highlighted. With the framework of networking that is oriented with the machine learning algorithms, the online mode system with the equivalence rate has been attained worthy for the implementation of analytical solving. There is a trade-off between the data velocity and the processing time. Thus, attaining the cost-effective based standalone algorithms for existing benchmark results with the proposed fusion approach. The proposed study provides improved knowledge for the research on this particular stream. It supports achieving better insights in attaining the effectiveness, robustness, and accuracy in predicting the values for multiple criteria based supportive decision modeled system and their framework. Thus, providing significant improvement in the performance of advanced techniques compared to existing techniques for big data analysis through mixed data processing resources.

REFERENCES

- [1] Z. Miao and L. Fan, A novel multi-agent decision making architecture based on dual's dual problem formulation, IEEE Transactions on Smart Grid, 99(2016) 1–1.
- [2] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, Network anomaly detection: Methods, systems and tools, IEEE Commun. Surv. Tutor., 16(1)(2014) 303–336.
- [3] M. Weiten, Ontostudio® as an ontology engineering environment, in Semantic knowledge management. Springer, (2009) 51–60.
- [4] N. M. Meenachi and M. S. Baba, Web ontology language editors for the semantic web-a survey, International Journal of Computer Applications, 53(2012) 12.
- [5] R. Klapsing, G. Neumann, and W. Conen, Semantics in web engineering: applying the resource description framework, IEEE MultiMedia, 8(2)(2001) 62–68.
- [6] K. R. Kurte, S. S. Durbha, R. L. King, N. H. Younan, and R. Vatsavai, Semantics-enabled framework for spatial image information mining of linked earth observation data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(1)(2017) 29–44.
- [7] S. BOU, H. KITAGAWA, and T. AMAGASA, Cbix: Incremental sliding-window aggregation for real-time analytics over out-of-order data streams, DEIM Forum, F7-5, (2018)
- [8] U. Fiore, F. Palmieri, A. Castiglione, and A. De Santis, Network anomaly detection with the restricted boltzmann machine, Neurocomputing, 122(2013) 13–23.

- [9] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks*, 61(2015) 85–117.
- [10] R. Vijayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, Deep learning approach for an intelligent intrusion detection system, *IEEE Access*, 7,(2019) 41525–41550.
- [11] S. M. Kasongo and Y. X. Sun, A deep learning method with filter-based feature engineering for a wireless intrusion detection system, *IEEE Access*, 7(2019) 38597–38607.
- [12] W. Shu and H. Shen, Multi-criteria feature selection on cost-sensitive data with missing values, *Pattern Recognition*, 51(2016) 268–280.
- [13] K. Vougas, T. Sakellariopoulos, A. Kotsinas, G.-R. P. Foukas, and V. G. Gorgoulis, Machine learning and data mining frameworks for predicting drug response in cancer: an overview and a novel in silico screening process based on association rule mining, *Pharmacology & Therapeutics*, 203, Article ID 107395, (2019).
- [14] Mallikarjuna Reddy, A., Rupa Kinnera, G., Chandrasekhara Reddy, T., Vishnu Murthy, G., et al., (2019), Generating cancelable fingerprint template using triangular structures, *Journal of Computational and Theoretical Nanoscience*, 16(5) 5-6 1951-1955, doi: <https://doi.org/10.1166/jctn.2019.7830>.
- [15] A. Mallikarjuna, B. Karuna Sree, et al, Security towards Flooding Attacks in Inter-Domain Routing Object using Ad hoc Network *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(3)(2019) 545-547.
- [16] A Mallikarjuna Reddy, Vakulabharanam Venkata Krishna, Lingamgunta Sumalatha and Avuku Obulesh, Age Classification Using Motif and Statistical Feature Derived On Gradient Facial Images, *Recent Advances in Computer Science and Communications* (2020) 13:965. <https://doi.org/10.2174/2213275912666190417151247>
- [17] Srinivasa Reddy, K., Suneela, B., Inthiyaz, S., Kumar, G.N.S., Mallikarjuna Reddy, A. Texture filtration module under stabilization via random forest optimization methodology, *International Journal of Advanced Trends in Computer Science and Engineering*, 8(3)(2019).
- [18] A. Diez-Olivan, J. Del Ser, D. Galar, and B. Sierra, Data fusion and machine learning for industrial prognosis: trends and perspectives towards industry 4.0,” *Information Fusion*, 50(2019) 92–111.
- [19] Mallikarjuna Reddy, A., Venkata Krishna, V. and Sumalatha, L. Face recognition approaches: A survey. *International Journal of Engineering and Technology (UAE)*, 4.6,6(7)(2018) 117-121. doi: 10.14419/ijet.v7i4.6.20446.
- [20] C. R. T. G. Sirisha and A. M. Reddy, Smart Healthcare Analysis and Therapy for Voice Disorder using Cloud and Edge Computing, 2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Mangalore, India, (2018) 103-106. doi: 10.1109/iCATccT44854.2018.9001280.
- [21] I. M. Cavalcante, E. M. Frazzon, F. A. Forcellini, and D. Ivanov, A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing, *International Journal of Information Management*, 49(2019) 86–97.
- [22] Y. Liu, Y. Qian, Y. Jiang, and J. Shang, Using favorite data to analyze asymmetric competition: machine learning models, *European Journal of Operational Research*, 287(2)(2020) 600–615.