

Soft-Computing: A Fundamental Approach in Analyzing Big Data

Ejiofor C. I¹ & Mgbefuluike .I .J²

Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University ^{1,2}

Abstract

This research paper has designed a novel model: soft-computing model in analyzing big data. It focuses on voluminous data while addressing data velocity. The model comprises of mediator, data filter, collector, predictor and acceptor, all model components. The enhancement of data volume is handled using data filters while data velocity is handled using predictor. Unified Modeling Language (UML) portrays the behavioral functionalities of the model. The proffered benefit of the model will be explored on full implemented.

Keywords—Model, soft-computing, Big data

1.0 Introduction

The advent of computer and electronic devices has resulted in tremendously growth of data globally (Boyd and Crawford, 2011). This growth could be associated with proliferation of novel and existing technologies, the internet boom, the integration of electronic devices and the quest for human dependent information (Segaran and Hammerbacher, 2011). This facilitator to data has resulted in the development and preservation of Big Data. The concept of big data refers to data, data structure or data sets which are voluminous and complex enough preventing traditional or conventional processing with application software interacting with computer base platform: Operating Systems (OS) (Dedić, and Stanier, 2017). Big data involves data storage, analysis, searching, querying, and updating and information privacy. These processes of big data can be constrained under; Big data velocity, big data volume, big data variety and big data velocity (Boyd and Crawford, 2011). The tremendous investment of companies in attaining and constraining the influences of big data has been seen annually with seventy-five (75%) of companies pledging considerable amount of their revenue of investment in addressing issues with big data (<https://www.simplilearn.com/big-data-applications-in-industries-article>).

The banking sectors has experienced it fair share of big data with considerable increase in banking data recorded over the years which perhaps is gradually hindering traditional processes (Ibrahim et al.,

2015). This data have been generated in response to threat or security issues, increase in customer base, the fright for loan policies and other banking facilitating issues (Wu et al, 2015). The healthcare sector has also experienced huge amount of big data due to the integration of healthcare service with Information and Communication Technology (ICT), facilitating prompt diagnoses, treatment and the appropriate dissemination of information across various department and units within medical centers and hospitals. Huge investment in analytical tool has helped in managing big data. This integration of big data has indeed provided need resolution for data and information processing (Lee, 2013).

Technical experts are gradually exposing big data proliferating the educational sector, with big analytic tools playing a large role in resolving these constrained issues. Lack of analyses of big data will indeed constrain the optimal functionalities with this sector which may affect staff recruitment, prompt record keeping, and student registration. These issues are best resolved using several simulation and validation tools, not readily available with conventional software (Preis, 2012).

Although several tools has been used in analyzing and simulating big data processes, model base implementation tool are indeed scare, this research paper provide a soft-computing model in addressing Big data.

2.0 Overview of Soft-Computing and Related Researches

The sections cover several research works on Big Data and a comprehensive discussion on the concept of soft-computing. Table 2.1 provide this literatures on big data.

Table 2.1: Review of related works on Big Data

SN	Author (Year)/Title	Goal	Strength/Finding	Limitation/Weakness	Further Research
1.	Gali, (2012) <i>Research Trends: Special Issue on Big Data</i>	Big Data and various Processing Paradigm	Expose comprehensive different techniques for analyzing big data	No proposed technique or Model	No Application of reviewed technique to Model development
2.	Samiddha and Ravi, (2016) Big Data Concepts, Applications, Challenges and Future Scope	A descriptive discussion on big data	Highlight the application of Big Data in different sector ii. Highlight the challenges hindering Big data in different sector	No proposed technique or Model	No Application of reviewed technique to Model development
3.	Raghavendra et al., (2015) The anatomy of big data computing	The anatomy of big data computing	Comprehensive discussion on Big data	No proposed technique or Model	No Application of reviewed technique to Model development
4.	Marcos et al, (2015) Big Data computing and clouds: Trends and future directions	Big Data computing and clouds: Trends and future directions	i. data management and supporting architectures. ii. model development and scoring iii. visualization and user interaction iv. Business models.	No proposed technique or Model	No Application of reviewed technique to Model development

Table 2.1 provides a brief description of previous research works on big data. These researches portray clearly the area of explorations. It shows clearly the neglect from previous works in proposing and implementing a model in addressing the area of concern under big data which includes velocity, variety and volume. This research paper proposes a design soft-computing model in address “data volume and velocity” as concept under big data.

Soft-Computing is a broad approach under artificial

intelligent, a sub-domain under Artificial Intelligence (AI). Soft-computing focuses on fusing methodologies with the aim of creating models to solve real problem which cannot be solved mathematically (Charurvedi, 2008). These models are usually evaluated, analyzed and supported using mathematical tool adapted to meet current challenges. The solutions provided by solutions are usually uncertain and unpredictable, simply because precision are not tied to soft-computing problem.

Soft-computing problems are modeled using imprecise boundaries: 0 to 1 (Zadeh, 1994). Therefore a problem can exist within these boundaries unlike conventional computational processing where a problem can only exist within one class. Most complex system deals either in biology, medicine, humanities or management science deal largely with none polynomial problems which are difficulty to resolve using conventional processing. Soft computing can determine imprecision and uncertainty by determining the closeness of object in relative boundaries: partial truth, and approximation to achieve practicability, robustness and low solution cost. As such it forms the basis of a considerable amount of machine learning techniques (Zadeh, 1994; Charurvedi, 2008).

Machine learning is a sub-domain of Artificial Intelligence and subsequently a sub domain of computer science which gives computer system the ability to learn without explicitly programmed (Wernick et al., 2010). This concept can be traced to pattern matching which is linked to soft computing, computational theory, and computational processing. This learn paradigm construct algorithm which learn either through supervised, unsupervised or reinforced learning. Soft-computing covers sub-domains such as machine learning, fuzzy logic, neural network, computational reasoning, features selection, data mining and even data preprocessing. It is the intent of this research to feature selection (preprocessing components: Genetic Algorithm: GA) for data filtering and Artificial Neural Network: Specially Back Propagation Gradient Descent (BPGD) for machine learning. These components will be integrated within the novel model in addressing Voluminous (GA) and Velocity (BPDG) data in term preprocessing or optimization and prediction

3.0 Soft-Computing Model for Big Data Analysis

The designed soft computing for Big data analysis and evaluation was conceptualized with the aim of addressing the pronounced limitation of existing researches on big data: the neglect for a précised

model in addressing the three pinnacles of big data: volume, variety and velocity. The design model was visualized with the main aim of addressing data volume and velocity. This soft-computing model comprises of five main components: Mediator, Data Filter, Collector, Predictor and Acceptor. Figure 3.1 portray the Soft-computing for Big Data Analysis

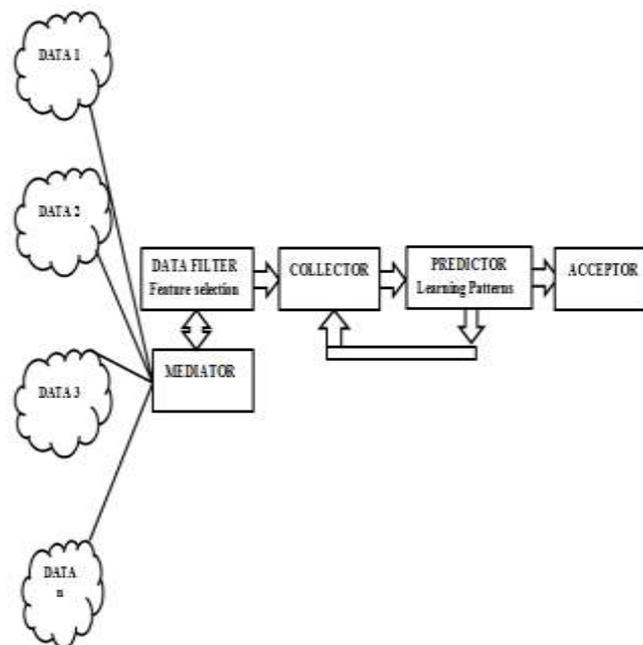


Figure 3.1: A Soft-Computing for Big Data Analysis (SC-BDA)

The Soft-Computing for Big Data Analysis (SC-BDA) comprises of:

- a. **Mediator:** The voluminous data received from different arena serve as input to the mediator. The mediator is the intermediary or mediator between the huge voluminous data received from various data collector centers and the data filters. The mediator accepts a wide range of unstructured, semi-structured and structured dataset. This data include symbols, numerical values, and alphabetical constructs. Upon receiving this string of uniformed patterns, the mediator transmits it to the data filters whose prerogative is to reduce the volume of these data using predefined inbuilt constructs. The mediator also acts as a repository for unwanted data returned by the data filter. The retained data sent to the data filter serves as output to the mediator
- b. **Data Filter:** The mediated received data serve as input to the data filter. The data filter achieves big data reduction using the predefined approach of feature selection. The feature selection identify appropriate data; useful and relevant data in line with the predefined policy and

procedures of the set organization in used. These techniques of feature selection actually constrain the voluminous data existing in big data and constraining the inability of conventional software from processing them in considerable amount upon which this software can process and produce useable results. The diminutive data serves as output for the filter data.

- c. **Collector:** The refined and diminutive data received by the collector from the data filter serve as input to the collector, which an intermediary between the predictor and data filter. The serves as a high speed repository while training, pattern matching and machine learning is taken place. The predictor receives its input data from the collector, these data are usually recycled back to the collector during training until equilibrium is achieved or training target has been obtained.
- d. **Predictor:** The predictor obtains training data from the collector which has been subsequently filtered by the data filter. This data has been resolved from ambiguities, and poor features eliminated. This less voluminous data is learnt from by the predictor in turning out a predictive output depending on the configuration or the set of the organization. This predictive output serves as the output of predictor.
- e. **Acceptor:** The acceptor accepts the final model output. This output has been constrained through the reduction of voluminous data and has been enhance providing velocity (speed0 in prediction. This output is the user acceptable output.

4.0 Design of the Soft-Computing Model for Analysis Big Data

The implementation of this model will be been enhanced using object components. Therefore object oriented programming tools like Java programming language, C++ will be appropriate for implementation. This implementation will be better complemented using object oriented approach in design the integral components of these models. Therefore Unified Modeling Language (UML) as an object tools was proposed and implemented for such design. UML Sequence diagram showcase the integral model functionalities, exposing th exchange of message between model objects. Figure 4.1 portray the Soft-Computing for Big Data Analysis (SC-BDA). The sequence diagram comprises of five main objects: mediator, data filter, collector, predictor and acceptor. These objects interact in a synchronous manner, with the *mediator* as an object

providing the first input in the form of diminutive data which is passed on to the next object: datafilter. The datafilter interact with the mediator in collecting the mediated data, filter this data based on predefined rules and pass to the collector. These details are accepted by the object: *collector*, serving as the intermediary for housing training data. The predictor receives synchronous message from collector and perform predictive training on these data. The acceptor as an object accepts the final output which subsequently used.

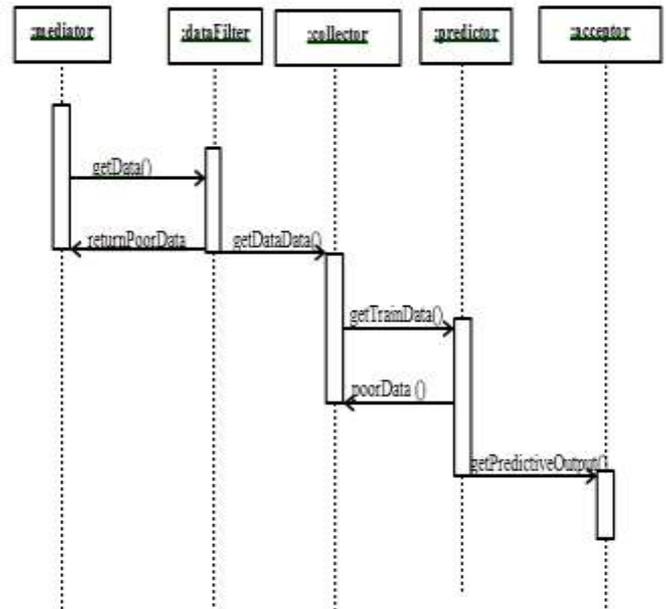


Figure 4.1: Soft-Computing for Big Data Analysis (SC-BDA).

4.0 Discussion

The Soft-Computing for Big Data Analysis (SC-BDA) is a universal model with the ability in addressing big data volume and velocity using the integral components of data filters through feature selection and prediction through learning patterns. The model on full implementation will obtains the following benefits:

- a. Eliminate voluminous irrelevant data.
- b. Enhance prompt decision making with preprocessed data
- c. Handle data uncertainty and imprecision
- d. Efficiently manage organizational resources
- e. Forecasting efficiently with available resources

The significance of this model in addressing voluminous data and handling data velocity was the inspiration behind this research paper.

5.0 Conclusion

This research paper has designed a soft-computing model for analysis big data; specifically it was streamlined toward addressing big data volume and velocity. The model comprises of five main components; mediator, data filter, collector, predictor and acceptor. The functionalities of the model were exposed using object oriented design methodologies specifically unified modeling language which portrays the behavior action of the model. The prompt decision making and management of organizational resource are fundamental benefits of this model. This model is indeed a plus in addressing Big data issues.

References

- [1] Boyd, D; Crawford, K. (2011). "Six Provocations for Big Data". *Social Science Research Network: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. doi:10.2139/ssrn.1926431
- [2] Charurvedi (2008), "Soft Computing: Techniques and Its Applications in Electrical Engineering", Springer,
- [3] Dedić, N.; Stanier, C. (2017). "Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery". 285. Berlin ; Heidelberg: Springer International Publishing. ISSN 1865-1356. OCLC 909580101..
- [4] Gali H. (2012), *Research Trends: Special issue on Big Data*, retrieved online from https://www.researchtrends.com/wpcontent/uploads/2012/09/Research_Trends_Issue30.pdf
- [5] Ibrahim; T. H., Abaker; Y., Ibrar; B. A., Nor; M., Salimah; G., Abdullah; U. K., S. (2015). "big data" on cloud computing: Review and open research issues". *Information Systems*. 47: 98–115.
- [6] Lee, J.; Lapira, E.; Bagheri, B.; Kao, H.. "Recent advances and trends in predictive manufacturing systems in big data environment (2013)". *Manufacturing Letters*. 1 (1): 38–41. doi:10.1016/j.mfglet.2013.09.005.
- [7] Marcos D. A., Rodrigo N. C., Silvia B., Marco A.S. N.. And R. B. (2015), Big Data computing and clouds: Trends and future direction (2015), Big Data computing and clouds: Trends and future directions, *Journal of . Parallel Distributed Computing*. 79–80 (2015) 3–15
- [8] Preis, T; M., Helen Susannah; Stanley, H. Eugene; Bishop, Steven R. (2012). "Quantifying the Advantage of Looking Forward". *Scientific Reports*. 2: 350. doi:10.1038/srep00350. PMC 3320057 PMID 22482034.
- [9] Raghavendra K. Pramod K. K., Arun A.,Raghavendra R. C.and Rajkumar B.Raghavendra R. C. and Rajkumar B (2015), The anatomy of big data computing, retrieved online from cloud-bus.org/papers/BigDataAnatomy.pdf
- [10]Samiddha M. and Ravi S., (2016), Big Data Concepts, Applications, Challenges and Future Scope, *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 5, Issue 2, Pp. 66 - 72
- [11] Segaran, T.; Hammerbacher, J. (2009). *Beautiful Data: The Stories Behind Elegant Data Solutions*. O'Reilly Media. p. 257. ISBN 978-0-596-15711-1.
- [12] Wernick, Y., Brankov, Y. and Strother (2010), Machine Learning in Medical Imaging, *IEEE Signal Processing Magazine*, vol. 27, no. 4, July 2010, pp. 25–38
- [13] Wu, D.; Rosen, D.W.; Wang, L.; Schaefer, D. (2015). "Cloud-Based Design and Manufacturing: A New Paradigm in Digital Manufacturing and Design Innovation". *Computer-Aided Design*. 59 (1): 1–14. doi:10.1016/j.cad.2014.07.006.
- [14]Zadeh L. A. (1994), "Fuzzy Logic, Neural Networks, and Soft computing," *Communications of the ACM*, March 1994, Vol. 37 No. 3, pages 77-84.
- [15] Zadeh L. A. (1994), "Fuzzy Logic, Neural Networks, and Soft computing," *Communications of the ACM*, March 1994, Vol. 37 No. 3, pages 77-84. <https://www.simplilearn.com/big-data-applications-in-industries-article>