

Impact of big data congestion in IT: An adaptive knowledge-based Bayesian network

Soheli Farhana, Adidah Lajis, Zalizah Awang Long, Haidawati Nasir

MIIT, Universiti Kuala Lumpur (UniKL), Malaysia

Article Info

Article history:

Received May 6, 2019

Revised Oct 17, 2019

Accepted Oct 25, 2019

Keywords:

Algorithm

Bayesian network

Big data

ABSTRACT

Recent progress on real-time systems are growing high in information technology which is showing importance in every single innovative field. Different applications in IT simultaneously produce the enormous measure of information that should be taken care of. In this paper, a novel algorithm of adaptive knowledge-based Bayesian network is proposed to deal with the impact of big data congestion in decision processing. A Bayesian system show is utilized to oversee learning arrangement toward all path for the basic leadership process. Information of Bayesian systems is routinely discharged as an ideal arrangement, where the examination work is to find a development that misuses a measurably inspired score. By and large, available information apparatuses manage this ideal arrangement by methods for normal hunt strategies. As it required enormous measure of information space, along these lines it is a tedious method that ought to be stayed away from. The circumstance ends up unequivocal once huge information include in hunting down ideal arrangement. A calculation is acquainted with achieve quicker preparing of ideal arrangement by constraining the pursuit information space. The proposed algorithm consists of recursive calculation in the inquiry space. The outcome demonstrates that the ideal component of the proposed algorithm can deal with enormous information by processing time, and a higher level of expectation rates.

*Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Zalizah Awang Long

Malaysian Institute of Information technology (MIIT),

Universiti Kuala Lumpur (UniKL),

1016 Jalan Sultan Ismail, 50250 Kuala Lumpur, Malaysia.

Email: zalizah@unikl.edu.my

1. INTRODUCTION

Bayesian system is utilized to anticipate ideal arrangement in this decade. There has been a development for undertaking research on Bayesian systems in the field of information technology [1-5]. By summing up the two methodologies for discovering the complex structure of Bayesian system. The principal approach presents learning as a requirement fulfillment issue. In this methodology, we survey the properties of restrictive freedom among the traits in the information. Normally this is finished utilizing a measurable speculation test for optimal solution. We at that point assemble a system that shows the required conditions and independencies. This should be possible utilizing fruitful methods. Instances of this methodology incorporate with some Bayesian methods [6-8]. The second methodology presents learning as a Big-data issue. We begin by characterizing a measurably stimulated score that depicts the wellness of every possible structure to the observed information. These scores incorporate Bayesian scores [9-12].

A vital advancement of the optimal solution for handling big data is a challenge while IT industry is introduced by providing real-time data [13-15]. Real-time data is huge as data update to the storage in every millisecond. Therefore, an optimal solution using Bayesian network absolutely required to handle the big data congestion in IT industry [16-20]. Such 'conventional' look strategies don't make a difference any

information about the normal structure of the system to be scholarly. For instance, eager slope climbing seeks techniques look at all conceivable neighbourhood changes in each progression and apply the one that prompts the greatest enhancement in score. The standard decision for neighborhood changes is edge expansion, edge cancellation, and edge inversion. The expense of these assessments ends up intense when we gain from enormous informational indexes. Since the assessment of new applicants requires gathering different insights about the information, it turns out to be progressively costly as the quantity of occurrences develops. To gather these measurements, we for the most part need to play out a disregard the information. Albeit, ongoing strategies might decrease the expense of this accumulation action, regardless we expect non-minor calculation time for each new arrangement of measurements we require [21-24]. Also, on the off chance that we consider areas with substantial number of properties, the quantity of conceivable applicants develops exponentially. In this research, we propose and implement an adaptive knowledge-based bayesian network algorithm to classify the impact of big data congestion in information technology and establish the novel method by comparing the processing time with existing method.

2. PHYSICS IMPACT OF BIG DATA IN IT APPLICATION

The application in IT are nearly acquiring a huge insurgency the manner in which we interface with "jobs". This will happen utilizing a lot of information that many keen gadgets will deliver and that will change the manner in which Big Data is taken care of. Give us a chance to see how. As long as the data getting from the different sources mostly considering the real-time data that are huge in size is known as big data. It has some impact in IT application process can be identified in following sections.

2.1. Security of Big Data

Expanded information requires expanded security. In IT application, the systems will make a system of shifted gadgets that will likewise prompt a pooling of changed sorts of information. IT Data Security will be another test put crosswise over with Big Data Security experts. In the event that any security escape clause occurs, the whole system of associated gadgets will be put in danger of control. The check and verification of gadgets that are added to the IT system will wind up essential. It will end up being the undertaking of Big Data associations to make a checkpoint to review the gadgets that are added to the system.

2.2. Storage of Big Data

The principal thing that strikes a chord with an expansion in the selection of IT is the measure of information that will be produced. Going with such a lot of information comes the duty of putting away it from numerous channels. Enormous Data associations are preparing up to move to the PaaS (Platform as a Service) show with the goal that they approach an adaptable and versatile technique to deal with IT information.

2.3. Analysis of Big Data

Big Data and information technology are interconnected with one another. IT will produce gigantic measures of information that must be broke down if the systems work precisely. The systems may create some excess information and that is the reason it ends up vital for Big Data associations to spend their investigation control on the information that is essential. In this way, another component of information order will be included so the Big Data Analytics instruments convey better execution. Different sources send in truckloads of information for examination to Big Data associations. At the present time, Big Data organizations are just barely getting to be fit for taking care of this colossal measure of information in an exceedingly secure way. The change we are expecting on the Big Data front would be the selection of adaptable and versatile answers for improve security, information putting away, and information investigation abilities.

3. MATERIALS AND METHODS

For adapting little scale data-base system, Bayesian Network (BN) is as of now all around investigated. However, learning in high dimensional areas e.g. informal communities and IT areas are confronting issue of high dimensionality. These spaces deliver datasets with a few hundred or a huge number of factors. It is really a serious issue for this kind of high dimension data for a Bayesian Network system. In this section, a brief overview is carried out for Bayesian Network (BN). By considering the limitation of general Bayesian Network on Big data, an algorithm is proposed to resolve the issues of BN.

3.1. Bayesian Network

There is only a solitary recommendation from Nielsen and Nielsen watching out for consistent Bayesian Network modification [25]. The procedure screens of adaptable model. The circumstance suggests to movement happened k observations back. Thusly, if we relearn show from last k cases, we will misfortune data learned before k cases. The change recognition part consistently forms the models that gets the information. This processes contention evaluates in every precedent v , information D in the hubs X . Strife evaluate watches the policy of v fits the nearby B with the factor of X_i utilizing method [26-27]. The current method will indicate the next assumption on the data handler that the individual components of a perception v are emphatically corresponded that can be written as,

$$\log \frac{P_{\beta}(X_i)}{P_{\beta}(X_i|X_j(\forall j \neq i))} < 0 \tag{1}$$

The (1) tracks the historical backdrop for contention evaluation as c_{X_i} for node X_i . Researchers initially ascertain change the cosine part for the last $2k$ measures c_1, c_2, \dots, c_{2k} in c_{X_i} ,

$$C_2 \equiv \sum_{j=1}^{2k} c_j \cos\left(\frac{\pi(j+\frac{1}{2})}{2k}\right) \tag{2}$$

The part in (2) determined after gathering of every perception indicate the propensity of $2k$ struggle esteems in orchestrating the inclined line. The heavy esteem shows the parameter k to handle with the perception of next arrangement of suitable data from randomly coming from different sources of hubs. Researchers tended to have unique instance for gradual Bayesian Network algorithm by discovering the basic adjustment. It concluded with the past information by storing data using this method.

3.2. Proposed adaptive knowledge-based Bayesian Network algorithm

To develop adaptive knowledge-based Bayesian Network, we derive the Bayesian Network firsr. From the Bayesian network, B can be expressed by the joint matrix $E=(C_{ij})$ of its attributes and the conditional probability table $P(x_j|\pi_j), j=1,2,\dots,m$ of each node. The conditional probability table (CPT) $P(x_j|\pi_j)$ of each node can be equivalent to $P(x_j, \pi_j)$, and $P(x_j, \pi_j)$ satisfies the relationship below:

$$\sum_{\forall x_j \in v(j), \forall \pi_j \in \varphi(j)} P(x_j, \pi_j) = 1 \tag{3}$$

Where $v(j)$ are value space of x_j and $\varphi(j)$ are value space of π_j .

$P(x_j, \pi_j)$ is more flexible than $P(x_j|\pi_j)$ because $P(x_j, \pi_j)$ fulfilled the relationship. So, the encoding below is used to descript a Bayesian network:

$$(P(x_1, \pi_1), P(x_2, \pi_2) \dots \dots \dots, P(x_n, \pi_n), E) \tag{4}$$

Where E is the joint matrix. To adjust recently erudite parameter E from different hubs. To recycle past data programmed to E from the past stage. In this way, E will refresh for conveyance of new information D_1 dependent on the suspicion that just hubs in C require refreshing of their probabilistic ties. In learn piece stage it takes in a somewhat coordinated sections G_{X_i} for every hub in C . At that point comparing pieces in E additionally removed from every hub located at C . Every one of the parts at that point converged into a solitary diagram G_1 , utilizing four bearing guidelines that save however much of B 's structure as could reasonably be expected and without damaging the newfound data. The factors that are restrictively free of hub X_i molding on some arrangement of factors, are non-nearby X_i . So also, the factors discovered ward of hub X_i are adjoining X_i in the chart sections. Generally imperative based learning techniques find just the course of the curves of v utilizing free evaluation in bearing residual bends. Further, rest of the connections in conclusive chart coordinates utilizing conditions bellow,

- In the event that the difference between A and E is a connection, C tends to A is a circular segment, and the nearby point coordinate the connection the differences between A and E as A tend to E .
- On the off chance that the differences between A and E is a connection and there is a coordinated way from A to B , at that point coordinate the connection $A - E$ as $A \rightarrow E$.
- In the event that Rules 1 and 2 can't be connected, picked a connecti on $A - E$ at arbitrary with the end goal that the hubs discovered unaltered at main period calculation in β state.
- In the event that Rules 1 and 3 can't be connected, picked a connection indiscriminately and guide it haphazardly.

4. RESULTS AND DISCUSSIONS

4.1. Experimental setup

In this area, a few trials are completed by contrasting proposed calculation and the ordinary Bayesian Network calculation. The objective of this work is to assess the proposed calculation's capacity to manage high dimensional information and portraying the circumstance where it beats alternate calculations. The outcomes will demonstrate the adaptability of the calculation and observational assessment of the proposed calculation is to legitimize its adequacy.

4.2. Results

Here we examine the conduct of client characterized parameter K . The impact of various introductions for hunt period of the proposed calculation. As a matter of first importance, we examine the adequacy of recurrence counter reserve. It is utilized in the various investigations displayed in this examination. Here, we demonstrate how recurrence counter store proposed to lessen the learning time over steady process over enormous number of employments. Figure 1 shows about the execution time for colossal number of employments by utilizing proposed algorithm and regular Bayesian Network.

Client characterized of k which indicate quantity for occupations that put away at each progression of executing the calculation. Along these lines, it reserves all connected factors with the objective k information. The conduct of the proposed calculation with various k esteems is appeared in Figure 2. We can see that the estimation of k could be among normal and most extreme level of the hypothetical chart. Sensibly, the multifaceted nature straightly increments regarding k . This expansion in the multifaceted nature is immaterial when contrasted with the aggregate capacity calls, though it is important that the proposed calculation took in a decent estimation from k . The information has been tested from non-stationary area, and the calculation don't experience any float.

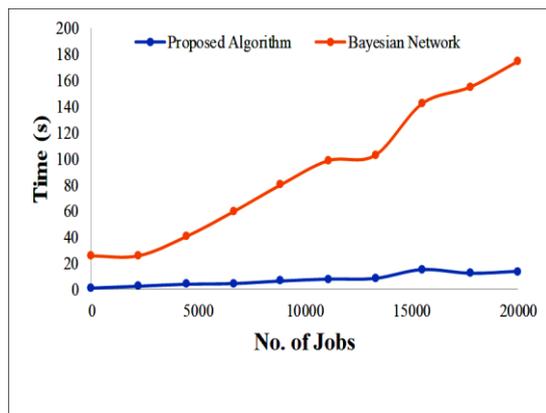


Figure 1. Comparison of big data processing time

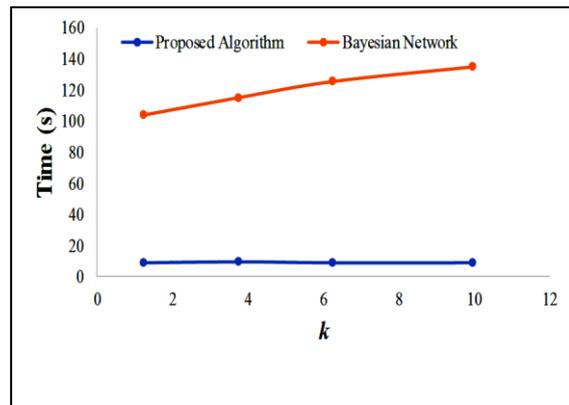


Figure 2. Comparison on different k value

5. CONCLUSION

This examination demonstrates an experimental assessment of the proposed calculation on Bayesian Network to take care the effect of Bigdata in IT industry application. The proposed adaptive knowledge-based Bayesian Network was developed and examined. From the analysis, it was found that the proposed algorithm processing time is better than the existing algorithm. Diverse reenactment has been taken care of assess the flawlessness and the time utilization of the proposed calculation. The reproduction led dependent on the diverse information sources applications. Examination were led with the customary and proposed Bayesian Network calculation for big information control on IT industry application.

REFERENCES

- [1] Mahler, R. P., "Statistical multisource-multitarget information fusion," *Artech House, Inc*, 2007.
- [2] Li C., Zhang R., Li J. & Stoica P., "Model order determination for signed measurements via the Bayesian information criterion," *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 366-370, 2018.
- [3] Liu X., Lu R., Ma J., Chen L. & Qin B., "Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification," *IEEE journal of biomedical and health informatics*, vol. 20(2), pp. 655-668, 2016.

- [4] Koudas N. & Bansal N. *U.S. Patent No. 9,256,667*. Washington, DC: U.S. Patent and Trademark Office, 2016.
- [5] Liu D., Liang D. & Wang C., "A novel three-way decision model based on incomplete information system," *Knowledge-Based Systems*, vol. 91, pp. 32-45, 2016.
- [6] Vandekerckhove J., Rouder J. N. & Kruschke J. K., "Bayesian methods for advancing psychological science," 2018
- [7] Linkov I., Massey O., Keisler J., Rusyn I. & Hartung T., "From weight of evidence" to quantitative data integration using multicriteria decision analysis and Bayesian methods," *Altex*, vol. 32(1), p. 3, 2015.
- [8] Dorazio R. M., "Bayesian data analysis in population ecology: motivations, methods, and benefits," *Population ecology*, vol. 58(1), pp. 31-44, 2016.
- [9] Perryman A. L., Patel J. S., Russo R., Singleton E., Connell N., Ekins S. & Freundlich J. S., "Naïve bayesian models for vero cell cytotoxicity," *Pharmaceutical research*, vol. 35(9), p. 170, 2018.
- [10] Christinaki E., Poli R. & Citi L., "Bayesian transfer learning for the prediction of self-reported well-being scores," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 41-44, 2018.
- [11] Contaldi C., Vafaee F. & Nelson P. C., "Bayesian network hybrid learning using an elite-guided genetic algorithm," *Artificial Intelligence Review*, pp. 1-28, 2018.
- [12] Tang Y., Zhang Q., Liu H. & Wang W., "Adaptive bayesian network structure learning from big datasets," *International Conference on Database Systems for Advanced Applications*, pp. 158-168, 2017.
- [13] De Francisci Morales G., Bifet A., Khan L., Gama J. & Fan W., "IoT big data stream mining," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2119-2120, 2016.
- [14] Mishra N., Lin C. C. & Chang H. T., "A cognitive adopted framework for IoT big-data management and knowledge discovery prospective," *International Journal of Distributed Sensor Networks*, vol. 11(10), 718390, 2015.
- [15] Cai H., Xu B., Jiang L. & Vasilakos A. V., "IoT-based big data storage systems in cloud computing: Perspectives and challenges," *IEEE Internet of Things Journal*, vol. 4(1), pp. 75-87, 2017.
- [16] Sagioglu S., Sinanc D., "Big data: A review," *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pp. 42-47, 2013.
- [17] Raghupathi W, Raghupathi V., "Big data analytics in healthcare: promise and potential," *Health information science and systems*, vol. 2(1), p. 3, 2014.
- [18] Lee J, Kao HA, Yang S., "Service innovation and smart analytics for industry 4.0 and big data environment," *Procedia Cirp*, vol. 16, pp. 3-8, 2014.
- [19] Tao F., Cheng J., Qi Q., Zhang M., Zhang H., Sui F., "Digital twin-driven product design, manufacturing and service with big data," *The International Journal of Advanced Manufacturing Technology*, vol. 94(9-12), pp. 3563-76, 2018.
- [20] Flyverbom M, Deibert R, Matten D., "The governance of digital technology, big data, and the internet: New roles and responsibilities for business," *Business & Society*, vol. 58(1), pp. 3-19, 2019.
- [21] Suci G., Vulpe A., Fratu O. & Suci V., "M2M remote telemetry and cloud IoT big data processing in viticulture," *Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International*, pp. 1117-1121, 2015.
- [22] Chawla S., Malec D. & Sivan B., "The power of randomness in bayesian optimal mechanism design," *Games and Economic Behavior*, vol. 91, pp. 297-317, 2015.
- [23] Dong G. & Chen Z., "Data driven energy management in a home microgrid based on bayesian optimal algorithm," *IEEE Transactions on Industrial Informatics*, 2018.
- [24] Ryan E. G., Drovandi C. C., McGree J. M., & Pettitt A. N., "A review of modern computational algorithms for bayesian optimal design," *International Statistical Review*, vol. 84(1), pp. 128-154, 2016.
- [25] Nielsen S. H. & Nielsen T. D., "Adapting bayes network structures to non-stationary domains," *International Journal of Approximate Reasoning*, vol. 49(2), pp. 379-397, 2008.
- [26] Jensen F. V., Chamberlain B., Nordahl T. & Jensen F., "Analysis in HUGIN of data conflict," *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, Elsevier Science Inc.*, pp. 519-528, 1990.
- [27] Crimins F., "Numerical recipes in C++: The art of scientific computing," *Applied Biochemistry and Biotechnology*, vol. 104(1), pp. 95-96, 2003.

BIOGRAPHIES OF AUTHORS



Dr. Soheli Farhana joined University of Kuala Lumpur, Malaysia in December 2018. She received her PhD in Electrical Engineering from the International Islamic University Malaysia (IIUM) in 2016 where she was developed a new model of Carbon Nanotube. She was the Post-Doctoral Fellow at IIUM in 2016 where her research focused on the Development of Nano-Soft Tools. Dr. Farhana was the Visiting Scholar at MIT, MA, USA in 2017. Dr. Farhana's research focuses on the Carbon Nanotube, Computer Graphics, Software Engineering, and Artificial Intelligence. She is currently doing research on Big Data Analytics and Real-time Distributed Systems. Dr. Farhana received the IIUM Faculty of Engineering Best Doctoral Dissertation Award at 32nd IIUM Convocation in 2016. She was the recipient of IIUM President Award, Gold Medal from International Invention, Innovation & Technology Exhibition (ITEX), and the Best Paper Award from the engineering congress in 2015 at Imperial College, United Kingdom. She was also the recipient of the Professional of the Year Award from Worldwide Who's Who in 2018.



Associate Prof Dr Adidah Lajis is an established lecturer and researcher in artificial intelligence and cognitive computing. Her research interests include natural language processing, text mining, cognitive assessment and intelligent system. She actively participated in short terms research projects and received several research grants include Fundamental Research Grant Scheme by Malaysia Ministry of Education. She is a reviewer for Cognitive Computation Journal and some of conference proceedings paper.



Assoc. Prof. Dr. Zalizah Awang Long's passion is for the improvement in student character building, can be traced back to 2015 where she spent significant time working on designing and developing frameworks, models and various activities focusing on student development. As a Director of the Center for Student Development then, she led the way by introducing a new paradigm in student development. Assoc. Prof. Zalizah is the individual responsible for the new student transcripts model known as GHOCs - Graduate Higher Order Critical Thinking Skills. In addition to that, she also initiated the establishment of Ulul Albab and Wakaf at UniKL. As a dedicated educator for more than 20 years, Assoc. Prof. Zalizah's life evolves very much around teaching activities, as well as research. To add on, she has contributed to a number of publications related to her passion on student character building and also her interest in data mining. In 2012, her PhD thesis received the Excellence Thesis Award. Her passion in research drives her to be constantly active in research matters, as she participates in both local and international conferences as part of the reviewer team, co-chair and board member. Assoc. Prof. Dr. Zalizah Awang Long is now serving as the Dean of Universiti Kuala Lumpur - Malaysian Institute of Information Technology (UniKL MIIT).



Assoc. Prof. Dr. Haidawati Mohamad Nasir is a senior lecturer at the Universiti Kuala Lumpur, Malaysian Institute of Information Technology. She obtained her PhD degree from the department of Electrical and Electronic, University of Strathclyde, Glasgow in 2012. Her research interest is in the signal and image processing with application to super-resolution image reconstruction, machine learning, and image enhancement and computer vision.