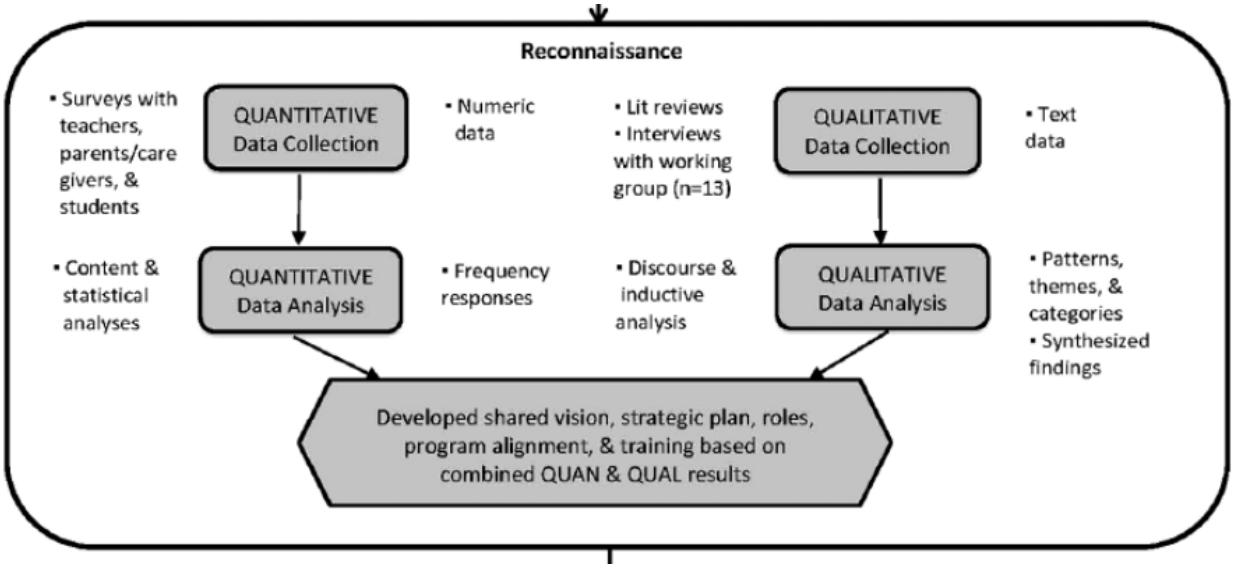
Research Approach Identification

Based on Rahman (2017) studies, the nuance definition of qualitative research varies on literature review. Corbin & Strauss (2015) states that any finding which has not concluded with any statistical procedure & quantitative. Moreover, as stated by Flick (2014) '*analyzing subjective meaning or the social production of issue, events, or practices by collecting non-standardized data and analyzing text and images rather than number and statistic*'. Despite the nuance definition, the overarching concept requires the dual perspective of positive and negative. Based on the research approach implemented on the studies, it was found that the study was conducting a qualitative analysis regarding Deep Learning in relationship with Big Data Analytics. In view of developing the emergence of Deep Learning & Big Data Analytics, systematic literature review implemented as the main method of data collection (Ivankova & Wingo, 2018).



With the industry expanded exponentially within the globe, big data analytics could undergo numerical transformation, leading to different challenges depending on the maturity of the market (Gupta and Rani, 2018). To overcome the dynamic environment, systematic literature review would be essential in qualitative research, to preserve relevance of the current research approach

to the future. The research could have implemented grounded theory, allowing continuous generation of new insights and theories grounded in the data (systematic literature review). The cycle creates continuous feedback as it is a close loop, starting with collection of theoretical sampling, analysis of literature review and development of concept. The methodology falls under three stages, Open coding → Axial Coding → Selective Coding (Noble & Mitchell, 2016). Although the current paper does explicitly implement the overarching concept, the underlying framework could be implied on the current research paper.

Critical Analysis of Research Paradigm

Ontological Perspective

Ontology derives from the exploration of concepts such as existence, identity, causality, substance, and the nature of reality. Within the current context of computer sciences, it is referred as an engineer artifact, '*constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary*' as stated by Maedche (2002). Essentially, serving as a standardized structure for representing knowledge, clarity, consistency within the domain.

Realist

1. *Reality exists independently.*

As mentioned by Sobh and Perry (2006), '*searching towards an understanding of the common reality of an economic system in which many people operate inter-dependently. That is, realists believe that there is a "real" world "out there" to discover*'. The scope of the research was extended into the scientific field, with the hope of expecting future research to discover optimized solution to the research gap as one could argue.

Next, the current research recognized complex concepts with the intention of reviewing the identifying gaps within the field of Big Data Analytics. It posits that these patterns and structures exist independently of human perception or interpretation within deep learning (Wooi Keong & Md Husin, 2023).

Epistemological Perspective

'*Epistemology is an area of philosophy concerned with the nature and justification of human knowledge'* as stated by Hofer and Pintrich (1997), exploring the theories and beliefs about knowledge does one acquires.

Constructivism

1. *Construct individual experience onto the research approach.*

The research approach provides clarity regarding the complex concept of deep learning and big data analytics along with its challenges within. The research approach could have taken a different approach, such as interview, survey etc. However, the literature framework was specifically developed in this format therefore it could be argued that the followed closely with the engineer artifact ontological perspective. Hence, the research approach follows a constructivist because it allows flexibility for the researcher to construct any information onto the research approach (Vogel-Walcutt et al, 2010; Mir & Watson, 2000).

Strength of Research Approach

1. *In-dept evaluation.*

By conducting qualitative research, the approach facilitates research understanding on a deeper level. Referring to the nuance definition qualitative research, interpreting the feeling, opinion and experiences uncover the action intent of the sample (Rahman, 2017). Referring to the research approach, the strength relies on the in-dept elaboration of the topic of big data analytics and deep

learning. Covering an overview of deep learning data analysis in data mining and machine learning, key characteristic of big data & challenges in big data analytics, application of deep learning in big data analytics and challenges of deep learning in big data analytics.

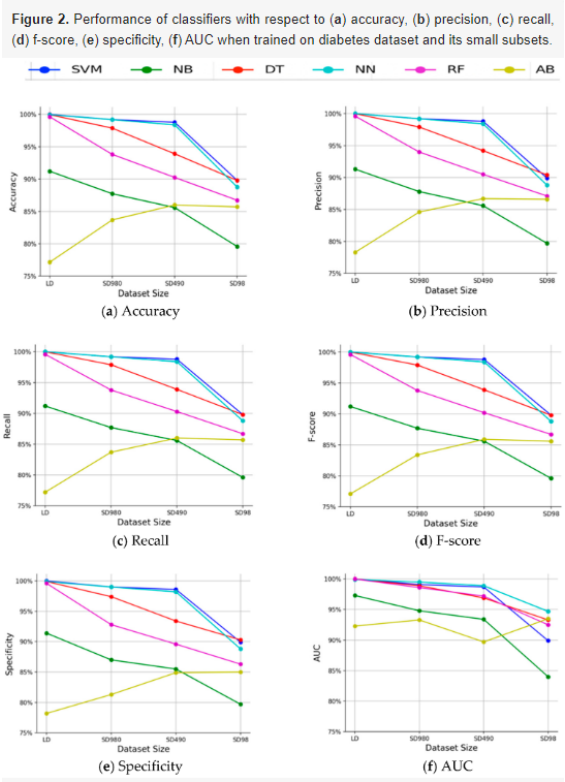
2. *Epistemological Viewpoints.*

The environment and social interaction shaped the perspective of an individual. Although the overarching perspective share similarity, each perspective carries its own comprehension of the concept, creating a nuance within the same perspective (Rahman, 2017; Ivankova & Wingo, 2018). Stating an objective reality that is independent despite with or without the perception of an individual, suggesting a gap of knowledge to be discovered (Healy & Perry, 2000; Wooi Keong & Md Husin, 2023). The gap of the research was focused on elaborating the challenges of deep learning in big data analytics. Suggesting a pattern existing independently, restricted by the limitation.

Weakness of Research Approach

1. *Limited Generalizability.*

As stated by Rahman (2017), '*smaller sample size raises the issue of generalizability to the whole population of the research*'. Despite the quantity of research not explicitly mentioned, it raises concern regarding generalizability (Johnson and Onwuegbuzie, 2004; Rahman, 2017; Firestone, 1993). Althnian et al (2021) research found that the algorithm's performance drops when the sample changes from small to large. Therefore, it indicates that the research currently conducted may lack thereof enough supporting evidence to generalize the study into the population.



2. *Lower Credibility.*

Based on Johnson and Onwuegbuzie (2004) and Rahman (2017) research, both studies suggest that qualitative research holds lower credibility compared to quantitative research. Although not implicitly mentioned within, the impact on attracting potential stakeholders influences people with decision making capability.

Recommendations

1. *Incorporate quantitative research to cross validate the findings.*

   Exploring the objective reality despite with or without the observation of an individual, it provides an overview regarding the current limitation. To strengthen the research, it would be recommended to incorporate quantitative research, developing a robust framework to specifically quantify the limitation (Johnson and Onwuegbuzie, 2004).

2. *Introduce a variety of research approaches.*

Despite the current research approach, collected data by systematically reviewing literature review, leading to the formation of a realism viewpoints. The current research could incorporate a variety of research approaches such as interview, survey etc., exploring the social nuance regarding the limitation. By extracting knowledge from experts within the field, it could have developed a pragmatic research paradigm, taking a value-oriented approach to the research (Johnson and Onwuegbuzie, 2004).

References

1. Althnian, A., AlSaeed, D., Al-Baity, H. H., Samha, A. K., Dris, A. B., Alzakari, N., … & Kurdi, H. (2021). Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Applied Sciences, 11(2), 796. https://doi.org/10.3390/app11020796
2. Corbin, J., & Strauss, A. (2015). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (4th ed.). Sage.
3. FIRESTONE, W. A. (1993). Alternative Arguments for Generalizing From Data as Applied to Qualitative Research. Educational Researcher, 22(4), 16–23.
4. Flick, U. (2014). An Introduction to Qualitative Research. SAGE.
5. Gupta, D. and Rani, R. (2018). A study of big data evolution and research challenges. Journal of Information Science, 45(3), 322-340. https://doi.org/10.1177/0165551518789880
6. Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. Qualitative Market Research: An International Journal, 3(3), 118–126. https://doi.org/10.1108/13522750010333861
7. Hofer, B. K. and Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88-140. https://doi.org/10.3102/00346543067001088
8. Ivankova, N., & Wingo, N. (2018). Applying Mixed Methods in Action Research: Methodological Potentials and Advantages. American Behavioral Scientist, 62(7), 978–997. https://doi.org/10.1177/0002764218772673
9. Johnson, R. B. and Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. Educational Researcher, 33(7), 14-26. https://doi.org/10.3102/0013189x033007014

10. Maedche, A. (2002). ONTOLOGY-DEFINITION & OVERVIEW. https://doi.org/https://doi.org/10.1007/978-1-4615-0925-7_2
11. Mir, R., & Watson, A. (2000). STRATEGIC MANAGEMENT AND THE PHILOSOPHY OF SCIENCE: THE CASE FOR A CONSTRUCTIVIST METHODOLOGY. In Strategic Management Journal Strat. Mgmt. J (Vol. 21).
12. Noble, H., & Mitchell, G. (2016). What is grounded theory? Evidence-Based Nursing, 19(2), 34–35. https://doi.org/10.1136/eb-2016-102306
13. Rahman, S. (2017). The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language "Testing and Assessment" Research: a Literature Review. Journal of Education and Learning, 6(1), 102–112.
14. Sobh, R. and Perry, C. (2006). Research design and data analysis in realism research. European Journal of Marketing, 40(11/12), 1194-1209. https://doi.org/10.1108/03090560610702777
15. Vogel-Walcutt, J. J., Gebrim, J. B., Bowers, C. A., Carper, T. M., & Nicholson, D. (2010). Cognitive load theory vs. constructivist approaches: which best leads to efficient, deep learning?. Journal of Computer Assisted Learning, 27(2), 133-145. https://doi.org/10.1111/j.1365-2729.2010.00381.x
16. Wooi Keong, Y., & Md Husin, M. (2023). Understanding Research Paradigms: A Scientific Guide. Journal of Contemporary Issues in Business and Government, 27(2). https://doi.org/10.47750/cibg.2021.27.02.588

Appendix

Chen et al. Journal of Big Data (2024) 11:34
https://doi.org/10.1186/s40537-023-00874-6

Journal of Big Data

METHODOLOGY Open Access

Data-driven multinomial random forest: a new random forest variant with strong consistency

JunHao Chen[1*], XueLi Wang[1*] and Fei Lei[2]

*Correspondence:
chenjunhao@emails.bjut.edu.cn;
xlwang@bjut.edu.cn

[1] School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing, China
[2] Faculty of Information Technology, Beijing University of Technology, Beijing, China

Abstract

In this paper, we modify the proof methods of some previously weakly consistent variants of random forest into strongly consistent proof methods, and improve the data utilization of these variants in order to obtain better theoretical properties and experimental performance. In addition, we propose the Data-driven Multinomial Random Forest (DMRF) algorithm, which has the same complexity with BreimanRF (proposed by Breiman) while satisfying strong consistency with probability 1. It has better performance in classification and regression tasks than previous RF variants that only satisfy weak consistency, and in most cases even surpasses BreimanRF in classification tasks. To the best of our knowledge, DMRF is currently a low-complexity and high-performing variation of random forest that achieves strong consistency with probability 1.

Keywords: Random forest, Strong consistency, Classification, Regression, Machine learning

Introduction

Random Forest (RF, also called standard RF or BreimanRF) [1] is an ensemble learning algorithm that makes classification or regression predictions by taking the majority vote or average of the results of multiple decision trees. Due to its simple and easy-to-understand nature, rapid training, and good performance, it is widely used in many fields, such as data mining [2–4], computer vision [5–7], ecology [8, 9], and bioinformatics [10].

Although the RF has excellent performance in practical problems, analyzing its theoretical properties is quite difficult due to its highly data-dependent tree-building process. These theoretical properties include consistency, which can be weak or strong. Weak consistency refers to the expectation of the algorithm's loss function converges to the minimum value as the data size tends to infinity, while strong consistency refers to the algorithm's loss function converges to the minimum value as the data size tends to infinity [11]. Consistency is an important criterion for evaluating whether an algorithm is excellent, especially in the era of big data.

Many researchers have made important contributions to the discussion of consistency-related issues in RF, proposing many variants of RF with weak consistency, such