Research and optimisation of deep learning algorithms for big data

Xuye Wang, Jixu Xia, Jiahao Zheng*

Shenyang Aerospace University, Liaoning, China

ABSTRACT

This research is dedicated to the in-depth study and optimisation of deep learning algorithms for big data environments. With the continuous development of big data technology, deep learning algorithms have achieved remarkable results in processing large-scale, high-dimensional and diverse data. However, in the context of big data, there are still problems such as low training efficiency, insufficient model generalisation performance, and excessive resource consumption, which urgently require in-depth research to enhance the applicability of deep learning algorithms in big data applications. First, we review the basic principles and architectures of deep learning algorithms and explore the challenges they face in the big data environment. In the review of related work, we summarise the current status of big data and deep learning research, review existing optimisation algorithms, and point out the problems and research gaps. By analysing the characteristics of big data, including data size, diversity, real-time requirements, and storage and computation constraints, we delve into the impact of these characteristics on deep learning, providing theoretical support for subsequent research. In the research methodology and algorithm design section, we propose a series of deep learning algorithm optimisation strategies for big data environments, including distributed and parallel computing, big data sampling and preprocessing, parameter optimisation and tuning, and model architecture tuning and innovation. Through the comprehensive application of these strategies, we aim to improve the efficiency and performance of deep learning algorithms in big data environments. Through experimental design and evaluation, we select representative big data sets for validation and use multiple evaluation metrics to comprehensively assess the optimised algorithms. Through comparative experiments and performance analyses with traditional methods, we demonstrate the superior performance of the proposed algorithm. Finally, by interpreting and discussing the experimental results, we summarise the main research findings and propose directions for future improvement and outlooks for in-depth research. The results of this study are not only of guiding significance for the application of deep learning algorithms in big data environments, but also provide useful theoretical support and practical experience for research in related fields.

Keywords: Big Data; Deep Learning Algorithms; Distributed Computing; Parallel Computing; Data Sampling

1 INTRODUCTION

With the rapid development of information technology and the popularity of the Internet, the era of big data has arrived. The emergence of big data brings new requirements for data processing and analysis capabilities, and also provides a broad application scenario for deep learning algorithms. Deep learning algorithms have made remarkable achievements in the fields of image recognition, natural language processing, and speech recognition with their excellent feature learning and pattern recognition capabilities [1]. However, with the increasing data size and data diversity, traditional deep learning algorithms face a series of challenges in the big data environment, such as low training efficiency, insufficient model generalisation performance, and excessive resource consumption.

This study aims to explore deep learning algorithms for big data in depth and propose corresponding optimisation strategies to improve their applicability and performance in big data applications [2]. By reviewing the fundamentals of deep learning algorithms, the characteristics of big data, and the existing research, we will reveal the current problems and challenges faced and provide a theoretical foundation for subsequent research [3].

In the big data environment, the dramatic increase in data size puts the traditional deep learning algorithms under severe pressure on computational and storage resources. In addition, the diversity of data and real-time requirements also pose new challenges to the design and training of deep learning models [4]. Therefore, we need to think deeply about how deep learning algorithms can be better adapted to the demands of the big data environment through innovative methods and optimisation strategies.

2 RELEATED WORK

In the era of big data, deep learning algorithms have achieved remarkable success in handling complex data tasks, however, deep learning algorithms in big data environments still face a series of challenges. In order to better understand these challenges and propose solutions, many researchers have conducted related work [5]. This section will review previous research relevant to this study, including the current state of deep learning algorithms in big data environments, existing optimisation algorithms, and problems and challenges.

The achievements of deep learning algorithms in big data applications have attracted widespread attention. Researchers have successfully applied deep learning algorithms in the fields of image recognition, natural language processing, and speech recognition with impressive results [6]. For example, deep convolutional neural networks (CNNs) have achieved state-of-the-art performance in image classification tasks, and recurrent neural networks (RNNs) have made significant progress in natural language processing. These applications have driven the widespread use of deep learning in the big data domain, but have also exposed some problems in the context of large-scale data [7].

In order to improve the performance of deep learning algorithms, many researchers have worked on developing various optimisation algorithms. With the emergence of big data, several studies have focused on the optimisation of deep learning algorithms in distributed and parallel computing environments [8]. Strategies such as distributed training and model parallelisation have been proposed to accelerate the training process of deep learning models.

In addition, some adaptive learning rate tuning and parameter optimisation methods have been introduced to improve the convergence speed and generalisation performance of the models [9].

Despite the excellent performance of deep learning in big data applications, there are still some challenges that need to be addressed. One of them is inefficient training. Due to the dramatic increase in data size, the traditional deep learning training process becomes more and more time-consuming [10]. In addition, the diversity and real-time requirements of big data put new demands on the design and training of deep learning models. Models also face many difficulties in terms of generalisation performance, resource consumption and interpretability.

Although there have been many studies on deep learning in big data environments, there are still some unanswered questions and research gaps [11]. For example, the problems of interpretability of deep learning models in the context of big data and how to better adapt to heterogeneous and dynamically changing big data environments still require in-depth research.

3 RESEARCH METHODOLOGY AND ALGORITHM DESIGN

In order to solve the problems faced by deep learning algorithms in the big data environment, this study adopts a series of research methods and designs corresponding algorithms to optimise the performance of deep learning algorithms in big data applications. The following are the main directions of the research methods and algorithm design:

3.1 Distributed and Parallel Computing Methods

Distributed computing frameworks, such as TensorFlow, PyTorch, etc., are used to divide the training task of the deep learning model into multiple sub-tasks and perform parallel computing on multiple computing nodes [12]. Introduce parametric server architecture using model parallelism and data parallelism strategies to make full use of computational resources and improve training efficiency. Distributed synchronous and asynchronous training algorithms are designed to balance the model update speed and global consistency.

3.2 Big Data Sampling and Preprocessing Methods

Design sampling and preprocessing methods to better adapt to the characteristics of big data by deeply analysing the diversity and real-time requirements of big data. Propose efficient data sampling algorithms to dynamically adjust the sampling strategy according to the data distribution to ensure adequate learning of various types of data during the training process [13]. Design real-time data preprocessing methods to deal with streaming data, reduce preprocessing time, and maintain the real-time nature of the model.

3.3 Parameter optimisation and adjustment methods

Through in-depth research on the parameter optimisation of deep learning models, methods such as adaptive learning rate and sparse parameter technology are proposed to improve the generalisation performance and training efficiency of the model. Innovative design of parameter adaptive learning rate adjustment algorithm, according to different levels and time steps to adjust the learning rate, balancing the convergence speed and model performance. Sparse parameter technology is introduced to reduce the redundancy of model parameters and reduce model complexity.

3.4 Model Architecture Adjustment and Innovation

Conduct an in-depth analysis of the current application of deep learning models in big data environments and identify the key points of the model architecture that need to be adjusted. Improved model architectures are proposed to introduce innovative designs such as attention mechanisms, deeper network hierarchies, and cross-modal information fusion in order to improve the model's ability to express large-scale data.

3.5 Experimental Design and Evaluation

Select representative large data sets and consider multiple performance evaluation metrics to ensure that the experimental results are convincing and interpretable. Design detailed experimental protocols, including comparative experiments and performance analysis, to comprehensively assess the effectiveness of the proposed method [14]. Provide guidance for further improvement through in-depth interpretation of the experimental results. With the above research methodology and algorithm design, this study aims to provide innovative solutions for deep learning algorithms in big data environments to cope with the problems of low training efficiency and insufficient model generalisation performance.

4 CONCLUSION

Through this study, we have conducted in-depth research and optimisation of deep learning algorithms for big data to address the many challenges faced in big data environments. By introducing distributed and parallel computing strategies, we successfully improve the training efficiency of deep learning algorithms in a big data environment. The use of distributed synchronous and asynchronous training methods enables the model to better adapt to the processing demands of large-scale data. We designed efficient data sampling and real-time preprocessing methods to adapt to the diversity and real-time requirements of big data. These strategies not only improve the generalisation performance of the model, but also reduce the overhead of computational and storage resources during the training process. By introducing adaptive learning rate tuning and sparse parameter techniques, we successfully improve the training efficiency and generalisation performance of deep learning models. These optimisation strategies perform well on different data and tasks, reducing the risk of model overfitting. We propose improved model architectures that introduce innovative designs such as attention mechanisms, deeper network hierarchies, and cross-modal information fusion. These improvements enable deep learning models to better represent the complex features of large-scale data. Through comparative experiments and performance analyses, we verify the superior performance of the proposed method. However, there are still some unanswered questions and room for improvement, such as the study of model interpretability and better adaptation to heterogeneous data, which will be the direction of future research.

In summary, this research provides a comprehensive and innovative solution for deep learning algorithms for big data, aiming to provide more powerful and efficient deep learning tools in big data application scenarios in the future. We are satisfied with the results of the current work and look forward to further in-depth research to address the challenges in more complex big data environments in the future.

5 DISCUSSION

During the research and optimisation of deep learning algorithms for big data, we have conducted an in-depth analysis of the research problem and addressed the challenges faced by adopting a series of research methods and optimisation strategies. By introducing distributed and parallel computing strategies, we have successfully improved the training efficiency of deep learning algorithms. However, in a big data environment, different types of data may have different distribution characteristics, so more in-depth research is still needed for the robustness and adaptability of the optimisation algorithms. Future work can explore more flexible distribution strategies to cope with the challenges of different data distributions. Better adaptability to heterogeneous data is also a direction of interest.

The data sampling and preprocessing methods we devised improve the model generalisation performance while reducing the computational and storage resource overheads during training. However, real-time data preprocessing may introduce additional computational burden in some cases, which needs to be carefully weighed in practical applications. Future research can further optimise the real-time data preprocessing method to improve its efficiency while considering the practicality in different application scenarios.

The introduction of adaptive learning rate tuning and sparse parameter techniques has a significant effect on the performance improvement of deep learning models. However, adaptive learning rate tuning is still a challenge when dealing with non-smooth data and requires more sophisticated strategies. Future research could explore smarter and adaptive parameter optimisation methods to adapt to changes under different data distributions. In addition, parameter optimisation strategies for different deep learning model architectures are also a worthwhile research direction.

The improved model architecture introduces a series of innovative designs, including attention mechanisms, deeper network hierarchies, and cross-modal information fusion. These designs have made significant progress in improving model representation for large-scale data. Future research could explore additional model architecture improvements, especially customised designs for different types of data and tasks. Research on lightweight models is also a challenging direction.

Through detailed experimental design and performance analysis, we verified the superior performance of the proposed method. However, there are still some unanswered questions and directions for further research, such as the interpretability of the model and better adaptation to heterogeneous data. Future research can enhance the interpretation of the experimental results and delve into the inner workings of the model. Meanwhile, more research on heterogeneous data processing can promote the application of deep learning in a wider range of fields.

Overall, this study has achieved remarkable results, but there are still many potential research directions that need to be explored in depth. With the continuous development of technology and the expansion of big data application scenarios, we expect more breakthroughs in the future.

REFERENCES

[1] Navazi, F., Yuan, Y., & Archer, N. (2023). An examination of the hybrid meta-heuristic machine learning algorithms for early diagnosis of type II diabetes using big data feature selection. *Healthcare Analytics*, 4, 100227.

[2] Khalaf, F. S., & Essa, E. I. (2023). Optimization-based feature selection and hybrid machine learning algorithms for big data classification: An investigation. *Journal of Advanced Sciences and Nanotechnology*, *2*(2), 226-237.

[3] Li, X., Huang, Z., & Ning, W. (2023). Intelligent manufacturing quality prediction model and evaluation system based on big data machine learning. *Computers and Electrical Engineering*, *111*, 108904.

[4] Gao, H., & Zhang, X. (2023, August). Distributed Optimization for Big Data Analytics: Beyond Minimization. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 5800-5801).

[5] Keerthi, D. S., Maithri, C., Dayanidhi, M., Kavitha, M. V. G., & Ananda Babu, J. (2023). Segmentation and classification of medical big data on brain tumor using bacteria foraging optimization algorithm along with learning automata. *Internet Technology Letters*, e468.

[6] Kaveh, M., & Mesgari, M. S. (2023). Application of meta-heuristic algorithms for training neural networks and deep learning architectures: A comprehensive review. *Neural Processing Letters*, *55*(4), 4519-4622.

[7] Pandey, B. K., Veeramanickam, M. R. M., Ahmad, S., Rodriguez, C., & Esenarro, D. (2023). ExpSSOA-Deep maxout: Exponential Shuffled shepherd optimization based Deep maxout network for intrusion detection using big data in cloud computing framework. *Computers & Security*, *124*, 102975.

[8] Alqahtani, T. M. (2023). Big Data Analytics with Optimal Deep Learning Model for Medical Image Classification. *Comput. Syst. Sci. Eng.*, *44*(2), 1433-1449.

[9] Mbamba, C. K., & Batstone, D. J. (2023). Optimization of deep learning models for forecasting performance in the water industry using genetic algorithms. *Computers & Chemical Engineering*, *175*, 108276.

[10] Shadmani, A., Nikoo, M. R., Gandomi, A. H., Wang, R. Q., & Golparvar, B. (2023). A review of machine learning and deep learning applications in wave energy forecasting and WEC optimization. *Energy Strategy Reviews*, *49*, 101180.

[11] Deng, A. (2023). Database task processing optimization based on performance evaluation and machine learning algorithm. *Soft Computing*, 1-11.

[12] Akhtar, M. M., Ahamad, D., Shatat, A. S. A., & Shatat, A. S. A. (2023). Big Data Classification in IOT Healthcare Application Using Optimal Deep Learning. *International Journal of Semantic Computing*, *17*(01), 33-58.

[13] Yazdinejad, A., Dehghantanha, A., Parizi, R. M., & Epiphaniou, G. (2023). An optimized fuzzy deep learning model for data classification based on nsga-ii. *Neurocomputing*, *522*, 116-128.

[14] Ali, I. M. S., & Hariprasad, D. (2023). Hyper-heuristic salp swarm optimization of multi-kernel support vector machines for big data classification. *International Journal of Information Technology*, *15*(2), 651-663.