ITM Web of Conferences **73**, 03023 (2025) https://doi.org/10.1051/itmconf/20257303023
IWADI 2024

Advancements in Coded Computation: Integrating Encoding Matrices with Data Shuffling for Enhanced Data Transmission Efficiency

Shijie Yuan*

School of Telecommunications Engineering and Intelligence, Dongguan University of Technology,
Dongguan, 523808, China

Abstract. In the interconnected age of big data, cloud computing, and the Internet of Things, the demand for robust data processing and transmission systems is critical. This study delves into the fundamental principles, technological advantages, and applications of coded computation, emphasizing the integration of encoding matrices and data shuffling techniques. Encoding matrices enhance data reliability, fault tolerance, and security, reducing transmission and storage costs. Data shuffling techniques, by reordering data, decrease communication overhead and computational burden, thereby optimizing the coding computation process. This paper analyzes various data shuffling methods, their integration with encoding matrices, and their impact on computational efficiency and data transmission. The application of these technologies promises substantial improvements in the efficiency of data systems, offering vital advancements for modern computing environments. By refining the design of encoding matrices and data shuffling strategies, the potential to elevate the performance of coded computations is explored, with implications for the progressive development of information technology.

1 Introduction

In the current landscape of digital transformation, the interconnection of devices and platforms through the Internet of Things, big data, and cloud computing has escalated the complexities of data processing and transmission. The burgeoning volume and velocity of data across global networks necessitate robust, fault-tolerant, and efficient computational methodologies. Coded computation has emerged as a pivotal technique, enhancing the reliability and efficiency of data systems by introducing redundancy to mitigate transmission errors and optimize storage [1].

Despite the advancements in coded computation, there remains a significant challenge in balancing computational load and minimizing communication overhead, particularly in large-scale systems. The traditional approaches, while effective, often fall short in addressing the dynamic requirements of modern computing environments, which demand

---

* Corresponding author: ayujiebo@fsu.edu.pa

higher flexibility and scalability. To address these challenges, innovative methods such as encoding matrices and data shuffling techniques are being explored. These methods aim to optimize the coding and decoding processes, thus improving fault tolerance and data security while reducing the overall system complexity [2].

This paper delves into the integration of encoding matrices with data shuffling techniques, a strategy that promises to enhance the computational efficiency and reliability of data transmission. By systematically reviewing and analyzing various encoding matrices and their application in data shuffling, this study assesses their impact on the performance of coded computation systems. Furthermore, the paper explores the potential of these integrated techniques in various computing scenarios, highlighting their advantages in reducing energy consumption, improving processing speed, and ensuring data integrity in distributed networks [1,2]. Through this comprehensive approach, the paper aims to provide substantial advancements in the field of coded computation, contributing to the development of more efficient and secure data transmission methodologies.

2 Concept of concept of coded computations coding calculation

2.1 Fundamental principles

Coded computing is a branch of mathematics and computer science, this paper focuses on how to effectively design, analyze and apply coding systems for data transmission and storage. The basic principles of coded computing include the integration of error-correcting codes and coding theory into key concepts and techniques for computing tasks, Using these basic principles and taking advantage of error correction, redundancy, efficient coding and decoding, coding theory, data shuffling techniques, and parallelism Concepts to improve data reliability, processing efficiency, and fault tolerance in a variety of computing applications, These principles form the basis for designing and implementing accurate and efficient coded computing systems.

2.2 Technological advantages and applications

Technical advantages: Coding computing has excellent fault-tolerant ability, good data security and efficient data processing ability. And a high degree of scalability. In terms of fault tolerance, coded computation enhances fault tolerance by combining with error correction codes. This helps to detect and correct errors in the data. This improves the reliability and integrity of the data, especially in systems prone to noise or interruptions. Data security: By introducing redundancy and error correction functions, Encoded computing can enhance data security by preventing data corruption and unauthorized access. Efficient data processing: Encoded computing reduces data access latency and increases data transfer speed by encoding data in a structured wayDegree to optimize data processing efficiency. Scalability: Encoded computing is highly scalable, It can be applied to large-scale distributed computing environment, cloud computing and data-intensive applications. Thereby realizing efficient data processing and storage.

Application: The technical advantages of coded computing make it possible to improve the data reliability, security, efficiency and efficiency of modern computing systems in various industries and sectors. Fault tolerance is widely used to improve the data reliability, security and efficiency of modern computing systems in various industries and departments. And fault tolerance.

Cloud computing: Coded computing is widely used in cloud computing systems. O as to improve the data reliability, safety and efficiency of the distributed server for processing a large amount of data. Such as:

Data storage systems: In data storage systems, encoded computation is used to improve data integrity and reduce storage overhead. Enhance fault tolerance and ensure the integrity of stored data.

Wireless communication: Coded computation is used in wireless communication systems to reduce transmission errors, improve signal quality, and enhance that reliability of data transmission over the wireless channel.

Distributed computing: a combination of coding theory techniques and distributed computing, coding computing can solve the problem of mitigating the effects of stragglers. And the fault tolerance and the security of the computing network are improved [2].

3 Overview of encoding matrices

Definition of matrix:

Matrix properties: Definition：An encoding matrix is a matrix that converts an original data vector into an encoded or coded form by performing matrix multiplication [3]. This transformation introduces redundancy into the data, enabling error detection and correction during decoding

3.1 Characteristics of the coding matrix:

Linear transformation: The encoding matrix represents the linear transformation applied to the input data vector. This transformation is usually defined by a coding scheme to introduce redundancy for error correction.

Size and structure: The coding matrix may be a square or rectangular matrix, the size of which is determined by the requirements of the coding scheme. The size and structure of the matrix are crucial for proper encoding and decoding.

Error correction capability: The properties of the coding matrix, such as its rank and specific properties derived from coding theory, determine the error correction capability of the code. Higher rank matrices generally provide better error correction capability [4].

Efficiency and complexity: The design of the encoding matrix affects the efficiency of the encoding and decoding process. Efficient coding matrix can reduce the computational complexity and improve the speed of error correction.

Application specific: The encoding matrix is tailored to the requirements of the specific application, Such as communication systems, storage systems, network coding, and distributed computing. Different matrices may be appropriate for different use cases based on factors such as error probability, data transfer rate, and computational resources.

3.2 Classification and applications

The classification of coding matrix is diverse, and the coding matrix under each classification has unique characteristics and application scenarios.

But it is mainly classified by the construction mode and the application field. In practical application, it is necessary to select the appropriate type of coding matrix according to the specific needs and scenarios.

3.2.1 Classification by Structure

According to the classification of structure, it can be divided into randomness and confirmation.

Random coding matrix. Features: Including Bernoulli encoding matrix, Gaussian encoding matrix, local Fourier encoding matrix, etc. These coding matrices are usually able to satisfy the incoherence with the vector basis, so that the sensing matrix satisfies the RIP criterion. However, due to the strong uncertainty of the random coding matrix, they are easy to cause waste of resources in hardware and are not easy to implement. When a random encoding matrix is used for encoding measurements, it is often necessary to perform multiple experiments to reduce the uncertainty of the results.

Deterministic coding matrix: Features: including Toeplitz coding matrix, Hadamard coding matrix, etc. The computational complexity of this kind of coding matrix is low, and it is easy to develop the coding matrix template. And the stability of the image reconstruction quality can be improved.

3.2.2 Classification by Application Field

According to the application field, it is mainly divided into digital remote control system field and image system field.

Coding Matrix in Digital Remote Control System. Definition: a component of a decoder in a digital remote control system, consisting of a bistable trigger and a coding switch. Each bistable state has two States, "1" and "2", and there are 2 n possible combinations of n bistable cascades. Each combination is a binary code group, and the coding switches are connected according to the binary code group. Purpose: To convert commands (positions of operating keys) into electrical signals.

Coding matrix for image processing. Features: In the field of image processing, coding matrix is usually used for image compression, reconstruction and feature extraction tasks. For example, in millimeter-wave single-pixel imaging, the design of coding matrix is very important to improve the quality and stability of image reconstruction [5]. According to the characteristics of millimeter-wave images, researchers have proposed a method to construct the encoding matrix based on two-dimensional Gabor filters. The matrix has good self-adaptability and can realize random sampling of a properly selected part of the feature space, the local feature information in the image is obtained.

4 Survey of published encoding matrices

4.1 Introduction to published matrices

Spark-based parallel matrix: a Spark-based parallel matrix computation algorithm, The algorithm can handle the multiplication of large dense matrices, Based on this, a parallel computing algorithm of Smith canonical form is proposed. The algorithm can realize the distributed parallel computing of the Smith canonical form of large-scale matrices on the Spark platform [6].

Smith's normal form algorithm for large-scale matrices: This algorithm can support the parallel calculation of block matrices. O that the Smith canonical form can be computed more efficiently. The algorithm can support the parallel computation of the block matrix, so it can compute the Smith canonical form more efficiently.

Network coding based on sparse matrix: For distributed secure storage systems, the concepts of sparse matrix and irregular matrix in LDPC codes are combined [7].

4.2 Comparison of performance and applicability

Spark based parallel matrix: able to handle larger scale matrix operations. And network transmission can be effectively reduce. And has expandability [8].

Large-scale matrix Smith standard calculation algorithm: It greatly reduces the amount of data transmitted by the network and improves the performance of the algorithm.

The field of mathematics has been widely used, and it has also been applied to many scenarios in the computer field.

Network coding based on sparse matrix: Through coding optimization, a sparse check matrix is constructed to improve the efficiency of coding and decoding. It reduces the complexity and has a good application prospect in distributed storage systems [9]. It mainly uses random shuffling, layered shuffling, circular shuffling, block shuffling and other technologies to achieve data shuffling. It is mainly used in artificial intelligence, statistical analysis, data privacy protection and so on.

5 Review of data shuffling methods

The first proposed shuffle algorithm is Fisher-Yates shuffle algorithm proposed in 1938. Although this method can effectively realize data exchange, However, it takes extra time to check the conflict of data, and then Knuth et al. Proposed the KDS algorithm on this basis. The disadvantage is solved, but the new disadvantage is that the original sequence cannot be saved. The hash algorithm generates a sequence that does not conflict with the hash table, but the disadvantage is that it consumes a lot of time. Guid algorithm can not achieve the desired effect because of the internal data conversion of random function. The folding and shuffling algorithm is used in large-scale sampling algorithm, and the scrambled data set has good randomness of data distribution. Compared with some traditional shuffling methods, the folding shuffling algorithm also has better time efficiency [10].

6 Discussion

This paper has examined the integration of encoding matrices with data shuffling techniques within the framework of coded computation. The discussion has highlighted the pivotal role these technologies play in addressing the challenges associated with the increasing complexities of data processing and transmission in modern digital infrastructures. The exploration of encoding matrices underscores their utility in enhancing data security, reliability, and fault tolerance, critical aspects in the efficient management of data across diverse computing platforms. Data shuffling techniques, on the other hand, serve as a complementary strategy by optimizing the arrangement of data to minimize communication overhead and improve overall computational efficiency. The interplay between encoding matrices and data shuffling has been shown to significantly bolster the performance of data transmission systems, particularly in environments characterized by high data traffic and susceptibility to errors, such as cloud computing and IoT systems. Furthermore, the evaluation of different encoding matrices and data shuffling methods has revealed a variety of application-specific benefits. For instance, random and deterministic matrices each provide unique advantages depending on the operational context. Random matrices, typically favored in scenarios requiring robust error handling and security measures, introduce unpredictability in data encoding, which is crucial for thwarting potential data breaches. Deterministic matrices, however, are noted for their lower computational overhead and stability, making them suitable for applications where computational resources are a limiting factor [3,4].

The analysis also extends to the practical implementation challenges and opportunities these technologies present. While the theoretical benefits of integrating encoding matrices with data shuffling are considerable, practical deployment must consider factors such as computational resource allocation, system architecture, and the specific needs of the application environment. As such, future research should focus on developing more adaptive coding strategies that can dynamically adjust to the varying demands of real-world computing environments.

7 Conclusion

This paper has explored the integration of encoding matrices and data shuffling techniques within the context of coded computation, highlighting their significant contributions to enhancing the efficiency and reliability of data transmission. The analysis demonstrated that encoding matrices improve fault tolerance, data security, and reduce operational costs, while data shuffling optimizes the overall computational load by efficiently managing communication overhead. Together, these technologies address the evolving demands of modern digital infrastructures, particularly in large-scale and complex data environments such as cloud computing and the Internet of Things.

The investigation into various encoding matrices and data shuffling methods revealed that their strategic application could significantly influence the performance and robustness of data systems. For example, random encoding matrices, suitable for high-security needs, and deterministic matrices, ideal for systems with limited computational resources, each cater to specific operational scenarios. The study also underscored the critical role of these technologies in mitigating the impact of stragglers in distributed computing settings, thereby enhancing the overall system throughput and resilience.

References

1. Z. Zhao, Y. Zhang, X. Zhu, J. Zuo, "Research on time series anomaly detection algorithm and application," in 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), vol. 1, pp. 16-20, IEEE (2019).
2. G. Ling, G. Sun, On the definition of determinant and matrix. J. Jiaozuo Norm. Coll. 35(1), 66–67 (2019) (in Chinese).
3. H. Xu, X. Zhu, Z. Zhao, X. Wei, X. Wang, J. Zuo, Research of pipeline leak detection technology and application prospect of petrochemical wharf. 2020 IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC), Chongqing, China, 263–271 (2020).
4. Y. Xiong, Application of matrix in information coding. J. People's Public Secur. Univ. China (Nat. Sci. Ed.) 23(1), 75–78 (2017).
5. X. Zhu, Z. Zhao, X. Wei, X. Wang, J. Zuo, Action recognition method based on wavelet transform and neural network in wireless network. In Proc. 2021 5th Int. Conf. Digit. Signal Process., 60–65 (2021, February).
6. Z. Zhao, X. Zhu, X. Wei, X. Wang, J. Zuo, "Application of Workflow Technology in the Integrated Management Platform of Smart Park," in 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), vol. 4, pp. 1433-1437, IEEE (2021).
7. Y. Huang, R. Wang, M. Ju, X. Zhu, Y. Xie, "Reconstructing global daily XCO2 at 1°× 1° spatial resolution from 2016 to 2019 with multisource satellite observation data," Journal of Applied Remote Sensing, 18(2), 028502-028502 (2024).

8. R. Wang, J. Zhu, S. Wang, T. Wang, J. Huang, X. Zhu, Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking. Int. J. Multimed. Inf. Retrieval 13(4), 39 (2024).
9. H. Wang, X. Li, D. Zheng, Network coding secure storage system based on sparse matrix. Inf. Secur. Commun. Secur. 6, 80–82 (2011).
10. Y. Zhang, X. Wang, J. Wen, X. Zhu, "WiFi-based non-contact human presence detection technology," Scientific Reports, 14(1), 3605 (2024).