A Study on Traditional and Modern Database Management Systems

Dr.S.Srinivasu, D.Shailaja
Government Degree College(A), Khairatabad
Dr,A.Inna reddy
Dr.BRR Govt,Degree College, Jadcherla

INTRODUCTION

The importance of data management in smart grids and provides an extensive overview of data compression. It explores various data compression techniques, including both lossless and lossy methods, and highlights their relevance and applications within smart grid systems. The chapter also delves into signal processing techniques that enhance data compression efficiency, focusing on threshold optimization and intelligent techniques. Additionally, it reviews existing literature on data compression for smart grid applications. Finally, the chapter addresses the research gap in data compression, emphasizing areas where further exploration is needed to optimize smart grid performance and data management.

DATA MANAGEMENT IN SMART GRID SYSTEMS

The development of smart grid systems would require dynamic monitoring of the entire infrastructure, from generation to the consumer end. Hence, the deployment of monitoring and measuring devices at each grid network is necessary to acquire information in the smart grid systems.

A prevalent monitoring and measurement tool utilized in grid infrastructure is the smart meter. These devices are responsible for measuring energy consumption and export, recording parameters like power consumption, and transmitting the data to a central server via communication networks for management purposes. A typical smart metering system consists of metering and communication infrastructures, as depicted in Figure 1.1.

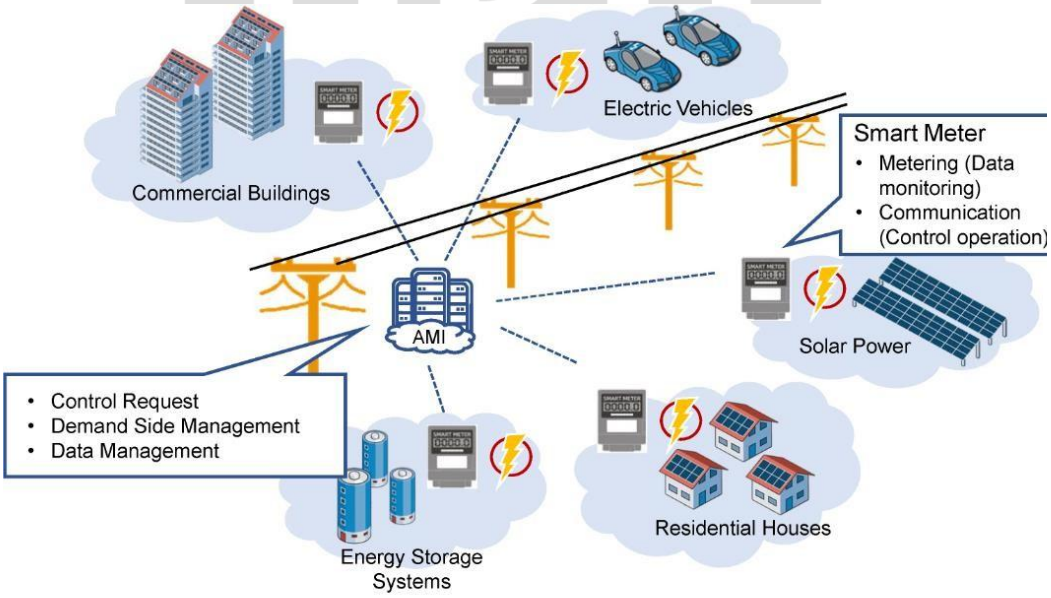

Figure 1.1 Structure of the smart metering and communication system

The metering system empowers the smart meter with capabilities such as recording voltage and current waveforms and storing data. Meanwhile, the communication infrastructure facilitates two-way communication between consumers and utilities. This infrastructure enables smart meters to establish connections with control centres for management and control purposes, thus establishing an advanced metering infrastructure (AMI).

These monitoring devices continuously capture data such as voltage, current, power, and electricity consumption, resulting in the need for effective management of a vast amount of data within the smart grid.

The high precision of monitoring devices and the vast amount of generated data also pose challenges for data communication, placing a significant burden on infrastructure. Moreover, these systems struggle with substantial expenses related to data storage. This scenario increases the complexity of managing data transmission and storage for smart grid systems.

OVERVIEW OF DATA COMPRESSION

Data compression is crucial for managing the large amount of data generated by smart grid systems, especially with the widespread use of smart meters and other grid devices. It aims to decrease data length while preserving important information, thereby making it easier to handle and enhancing the efficiency of data storage, communication, and processing within the smart grid systems.

Data compression offers several advantages for managing smart grid data effectively. At first, it helps reduce storage requirements by decreasing the length of historical data, which is crucial for long-term retention and analysis. Additionally, compressed data can be transmitted faster than uncompressed data, which is essential for smart grid applications where quick updates are necessary for decision-making and system stability. Moreover, by reducing the length of data to be processed, compression enhances the efficiency of data analytics, allowing for quicker insights. Furthermore, compressed data leads to reduced latency in data-driven operations, ensuring the responsiveness of control systems. Eventually, it promotes energy efficiency by consuming less energy for transmission and storage compared to uncompressed data, aligning with the goal of developing environmentally responsible smart grid systems. Consequently, this research proposes a methodology for effectively compressing large datasets in the context of smart grid systems.

In addition, with the rapid development of computing techniques, the data compression algorithm can be implemented for smart grid applications. Thus, it is feasible to embed the data compression algorithm into the monitoring devices so that the data can be compressed before it is sent out in order to reduce the data congestion. The compressed data is then sent to the receiving end, where the data can be reconstructed.

Therefore, employing data compression techniques is crucial to alleviating the strain on communication systems and storage resources. Moreover, while compressing data, it is important to retain as much valuable information as possible to accurately represent the original information. The fundamental requirements for data compression can be summarized as follows:

- Compress smart grid data to minimize the amount of data transmitted over communication systems.
- Ensure that compression preserves the valuable information within the data.
- Enable a nearly perfect reconstruction of the compressed data for analysis.

CLASSIFICATION OF DATA COMPRESSION TECHNIQUES

Data compression techniques are used in smart grid systems to reduce the length of data that needs to be stored, transmitted, and analysed. This can help improve the efficiency and effectiveness of smart grid operations.

There are two main types of data compression techniques: lossless and lossy compression techniques.

Lossless Compression

Lossless compression is a compression method that can preserve all the information in the original data. There are two common lossless compression algorithms, namely Huffman coding and the Lempel-Ziv (LZ) algorithm. However, lossless compression achieves a low compression ratio, that the compressed data is not significantly smaller, compared to the original uncompressed data. This is because lossless compression methods preserve all the information in the original data.

Lossy Compression

Lossy compression techniques compress data by losing some information while retaining the most valuable information of the original data. Common methods are wavelet transform (WT), principal component analysis (PCA), and singular value decomposition (SVD). Compared to lossless compression techniques, lossy compression achieves higher compression ratio by selectively discarding less important data. Due to this reason, researchers have focused on lossy compression techniques for data compression in smart grid applications.

SIGNAL PROCESSING TECHNIQUES FOR DATA COMPRESSION

Signal processing-based data compression techniques utilize mathematical transformations to analyze the data and exploit redundancies present in the data. These techniques are essential for both lossy and lossless compression, enabling the analysis of signals to achieve compression. They are particularly beneficial for smart grid applications, where saving storage space and bandwidth during transmission is vital. Examples of such techniques include using the discrete cosine transform (DCT), discrete fourier transform (DFT), wavelet transform (WT), principal component analysis (PCA), and singular value decomposition (SVD).

Review of Signal Processing Techniques

Dash *et al.* (2003) proposed an integrated approach combining spline wavelet and S-transform for addressing power quality disturbance data. The paper likely provides valuable insights into advanced signal processing techniques developed specifically for power quality monitoring and classification. By integrating the spline wavelet and S-transform, the authors aim to enhance the efficiency and accuracy of data compression, detection, and classification of power quality disturbances such as harmonics and transients. This integrated approach likely offers promising solutions for improving the reliability and performance of power systems by enabling timely detection and appropriate mitigation of disturbances.

However, despite its contributions, the paper may have limitations in terms of computational complexity, focus of findings on certain types of datasets, and scalability concerns, which could limit its applicability in realtime power quality monitoring systems with constrained processing capabilities.

Gharavi & Hu (2011) explored the use of communication technologies in smart grid systems, highlighting the role of wireless sensor networks (WSNs) and protocols such as ZigBee, Wi-Fi, and Power Line Communication (PLC). They emphasized that choosing the appropriate protocol depends on latency, bandwidth requirements, and network scalability, which are crucial for reliable data exchange across grid components.

Liu *et al*. (2014) proposed an error correction coding (ECC) mechanism using Reed-Solomon codes to enhance the reliability of data transmission in smart grids. Their work demonstrated that ECC helps mitigate transmission errors caused by noise and interference, ensuring that critical control data arrives intact.

Gunduz & Das (2020) examined cybersecurity threats in smart grid systems, focusing on vulnerabilities and potential solutions. They highlighted that the increased use of communication networks in smart grids exposes them to various cyber-attacks, including data violations and unauthorized control access. The authors discussed the importance of securing the data exchange between grid components and emphasized the role of encryption, authentication protocols, and detection systems in mitigating these risks. Additionally, they explored advanced methods for ensuring the confidentiality, integrity, and availability of smart grid data, suggesting a multi-layered security approach to enhance resilience against cyber threats.

Albu *et al.* (2010), presented the practical aspects of monitoring voltage and frequency in smart distribution grid. The paper likely provides insights into the challenges of data compression and accessibility, crucial for managing the vast amount of data generated by smart grid monitoring systems efficiently. By presenting a case study, the authors offer valuable real-world examples of implementing data compression techniques to improve accessibility and reduce the burden on communication networks. This approach contributes to enhancing the reliability and performance of smart distribution grid by enabling timely and effective monitoring of voltage and frequency variations. However, potential disadvantages of data compression techniques in this context could include the loss of fine-grained data resolution, which is necessary for certain analysis tasks. Additionally, the effectiveness of data compression methods may vary depending on the specific characteristics of the grid and the types of data being monitored, requiring careful consideration and adaptation to ensure optimal performance.

Qing *et al.* (2011) presented discrete cosine transform (DCT)-based data compression for power quality monitoring data. However, while DCT offers effective compression for signals with strong frequency localization, it may struggle with signals containing non-periodic or transient components, and those with power frequency changes. These limitations can lead to the loss of important information in the compressed data, particularly in scenarios where signal characteristics vary widely.

Kaushik *et al.* (2014) provided an overview of discrete fourier transform (DFT)-based data compression, highlighting its strong localization in the frequency domain. Despite their advantages, DFT-based compression techniques also face drawbacks. Similar to DCT, DFT struggles to effectively compress signals containing non-periodic or transient components, as well as those experiencing dynamic power frequency changes. Consequently, there is a risk of significant data loss when applying DFT-based compression on signals. These limitations hinder the effectiveness of DFT-based compression, particularly in scenarios where signal characteristics deviate from periodicity or experience dynamic frequency variations, potentially leading to the loss of important signal information.

Ning *et al.* (2011) introduced a compression algorithm based on wavelet transform (WT). This algorithm exploits WT's distinctive characteristics, particularly in how threshold selection directly impacts the efficient approximation of standard signal coefficients through thresholding. However, a significant hurdle arises from the variable length of compressed data associated with different threshold values. Hence, the key issue in WTbased data compression revolves around determining the most suitable threshold value.

Khan *et al.* (2016) presented wavelet packet transform (WPT)based data compression in smart grid applications. WPT offering a more detailed representation of signals compared to wavelet transform (WT), suffered from increased computational complexity, which could have hindered real-time processing requirements in smart grid systems. Moreover, the finer level of detail captured by WPT is not used for effective compression, leading to potential inefficiencies in the storage and transmission of data. Additionally, the increased complexity of WPT posed challenges in implementation and deployment within smart grid communication networks, where simplicity and efficiency are often prioritized.

Mehra *et al.* (2013) introduced a technique for compressing smart grid data employing principal component analysis (PCA). Under stable and fault conditions, the researchers have concluded that this method successfully filters out gaussian noise in signals and efficiently compresses both steadystate and transient signals. A critical obstacle in PCA-based data compression lies in accurately determining the optimal number of principal components, as it significantly influences compression efficacy and the degree of information loss after compression.

Notaristefano *et al.* (2013) proposed Symbolic Aggregation Approximation (SAX) to convert smart meter data into symbols, providing an innovative algorithm, for data dimension reduction and compression, particularly tailored for time series data. However, SAX's data recoverability is limited, which results in the loss of critical information after compression.

De Souza *et al.* (2017), proposed an SVD based lossy compression method. The method was applied to the data obtained from different measurement devices at different time points in a smart grid. It is simple to use and has high data compression efficiency. The computational complexity of SVD grows rapidly with the size of the data, making it less efficient for smart grid compression applications compared to WT. The experimental results in this study prove that SVD compression is not capable of compressing huge amount of data.

Wen *et al.* (2018), proposed a comprehensive examination of various compression techniques developed specifically for managing the vast amount of data generated by smart meters. The paper offers valuable insights into the challenges associated with handling smart meter big data, such as storage requirements, communication bandwidth constraints, and computational overhead. By surveying existing compression methods, including lossless and lossy techniques, the authors likely aim to highlight the potential benefits of data compression in terms of reducing storage costs, improving data transmission efficiency, and enabling faster processing of smart meter data. Additionally, the paper discussed emerging trends and future directions in smart meter data compression, such as the integration of advanced algorithms and the utilization of cloud computing resources for scalable data processing.

However, despite its contributions, the paper may have limitations in terms of comprehensively addressing the trade-off between compression efficiency and data accuracy. Some compression techniques may sacrifice data fidelity to achieve higher compression ratio, which could impact the reliability and usefulness of the compressed data for various smart grid applications.

Guzman *et al.* (2022) introduced a novel approach leveraging tensor decomposition for achieving higher compression ratio of time series data, particularly synchro phasor measurements from electrical systems and smart meter data. By establishing data into multidimensional arrays (tensors), correlations among heterogeneous data are effectively captured, reducing the length of the compressed signal without significant loss of information. However, the paper does exhibit limitations in terms of computational complexity, scalability issues with large datasets, and the need for adaptability to dynamic smart grid environment.

Overall, signal processing techniques, also known as traditional methods for data compression, have been used for many ye to reduce data size and facilitate efficient storage and transmission. Though they have their strengths, they also come with several disadvantages, compared to methods based on intelligent techniques. Limited compression ratio, lack of adaptability, limited learning capabilities and local parameter selection are some of the disadvantages of using signal processing techniques for data compression.

Importance of Threshold Optimization

Optimization is crucial in developing WT techniques for data compression in smart grid applications, as it ensures efficient compression while maintaining the integrity of essential information. In general, thresholding is critical for compressing data in smart grid applications. The selection of threshold values significantly impacts the length of compressed data. By fine-tuning the threshold value, researchers can optimize compression algorithms to effectively reduce the size of smart grid data without compromising its integrity. Moreover, compressed smart grid data maintains a high level of accuracy (low MSE), good signal quality (high SNR), and a significant reduction in data volume (CR). This optimization is crucial for smart grid applications, where large amount of data is generated from diverse sources such as sensors and meters, and efficient compression is necessary to manage bandwidth constraints, reduce transmission times and improve the effectiveness of data analytics process. Additionally, threshold optimization ensures the scalability and adaptability of compression techniques, allowing them to accommodate the variability of smart grid data and support the evolving needs of grid monitoring, control, and analysis. Overall, threshold optimization in the development of wavelet transform techniques for data compression in smart grid data is essential for optimizing data management, conserving bandwidth, supporting to reduce transmission losses, and ensuring the reliability and efficiency of smart grid operations.

INTELLIGENT TECHNIQUES FOR DATA COMPRESSION

Intelligent technique-based data compression involves the use of artificial intelligence (AI) and machine learning algorithms to adapt the specific characteristics of the data and optimize the compression process accordingly. Neural network-based algorithms and other machine learning algorithms are the examples of intelligent technique-based data compression.

Ibrahim & Morcos (2005) introduced a data compression technique leveraging adaptive fuzzy logic for power waveforms. This approach offered flexibility and adaptability in capturing complex waveform patterns, ensuring efficient compression while preserving critical information. However, challenges emerged related to computational complexity, especially for realtime applications, and the technique's effectiveness relied heavily on the accuracy and reliability of the fuzzy logic model. Additionally, ensuring adaptability to various waveform types required significant optimization efforts. Hence, while promising, careful consideration of computational demands and model accuracy was crucial for practical implementation in power system applications.

Tcheou *et al.* (2013) presented a comprehensive overview of data compression techniques customized for smart grid applications. It highlights the critical role of efficient data compression in managing the large amount of data generated by smart grid systems. By summarizing existing methods and discussing potential future trends, the paper provides valuable insights into the ongoing efforts to optimize data compression for smart grid. However, despite its thorough analysis, the paper may face limitations in fully addressing the scalability and adaptability challenges inherent in deploying compression techniques across diverse smart grid infrastructures. Additionally, while emphasizing the importance of compression efficiency, it cannot delve deeply into the trade-off between compression ratio and data fidelity, which are crucial considerations for ensuring the accuracy and reliability of smart grid operations.

Sun *et al.* (2016) have presented forward novel compression algorithms for data compression applications. By specifically using the electricity datasets within smart grid contexts, their methodology leverages the fundamental principles of compressed sensing theory. This suggests that the proposed compressed sensing theory algorithm offers a solution that surpasses conventional compression methods, potentially meeting distinct challenges or needs encountered in managing electricity datasets within smart grid systems. However, one potential limitation is the computational complexity involved in implementing compressed sensing algorithms, particularly for handling extensive datasets.

Wang *et al.* (2016) introduced a method for compressing smart meter data based on sparse and redundant representation. They proposed sparse coding (SC), a technique from artificial neural networks, to compress data from individual smart meter. Yu *et al.* (2016) also employed SC to compress smart meter data effectively and predict electricity demand based on the compressed data. Similarly, Prior *et al.* (2014) presented SC to enhance data transmission speed in the smart grid. SC achieved a higher compression ratio with minimal information loss compared to methods like PCA and DWT. Though this method achieves highest compression ratio, it is very difficult to find out the appropriate sparse coefficients.

A genetic algorithm (GA) and neural network (NN)-based compression algorithm for electrical power signals have been proposed by Barrosa *et al.* (2017). The GA is employed to select optimal signal samples, and the NN is employed for compressing the data and subsequently reconstructing the electric power signals. The paper investigates the collaboration between GA and NN, showcasing their complementary roles in achieving efficient compression. One potential drawback is the computational complexity, associated with the hybrid approach, particularly the parameter tuning and training requirements for the neural network.

Loia *et al.* (2017) proposed the utilization of fuzzy-based techniques to minimize memory and bandwidth requirements, thereby reducing the computational burden associated with processing smart grid data. This approach suggests an efficient way of managing data in a smart grid context. However, a potential drawback is the sensitivity of fuzzy logic to parameter settings and the need for careful tuning to achieve optimal results.

Li & Zheng (2019) introduced a neural network algorithm for compressing load data. The established procedure includes preprocessing the load data, training the neural network model, compressing the data using the trained model, and then reconstructing the compressed data. However, a drawback of this method might be the complexity and computational burden involved in training and deploying neural network models, especially in situations requiring processing large-scale datasets or real-time operations. Additionally, the effectiveness of the compression could heavily depend on the neural network's capacity to represent and generalize information, which cannot always be optimal, particularly in dynamic and evolving smart grid environments.

Das *et al.* (2020) introduced a machine learning-based strategy for compressing Dutch Residential Energy Datasets (DRED). The paper is expected to explore various machine learning techniques and their efficacy in compressing residential energy data. One possible limitation is the requirement for a substantial volume of data to train machine learning models, posing a challenge for compressing large-scale datasets.

Hashemipour *et al.* (2021) proposed an optimal singular value decomposition-based big data compression approach in a smart grid. In this paper, the thresholding optimization is primarily based on a single objective, typically aiming to minimize reconstruction error or maximize information retention. This single-objective optimization approach may overlook the potential trade-off between compression efficiency and other relevant performance metrics, such as CR, MSE, and SNR. Consequently, the drawback of this approach lies in its limited consideration of the broader optimization landscape, potentially leading to suboptimal solutions that do not fully align with the diverse objectives and constraints inherent in smart grid applications.

Subbarao *et al.* (2022) proposed hybrid method that merges a binary regression wavelet-surrogate tree with a hybrid thresholding technique. By combining this approach, they strengthen the advantages of each method, leading to improved data compression efficiency. Nevertheless, this hybrid method might face drawbacks such as increased implementation complexity and potential difficulties in parameter tuning, potentially affecting their practicality and scalability in smart grid environments.

Deterministic methods often encounter a significant limitation: they tend to become entrenched in local optima. Additionally, within a single objective optimization scenario, the quest for an optimal solution relies solely on one objective function. Consequently, deterministic methods prove unsuitable for data compression within smart grid applications.

The existing literature lacks sufficient emphasis on addressing the multi-objective optimization challenge associated with smart grid data compression. Consequently, this study endeavors to tackle the multi-objective optimization problem inherent in smart grid data compression. When approaching the data compression issue, various quality metrics, including CR, MSE, and SNR, serve as benchmarks for assessing the fidelity of compressed signals compared to the original data. Thus, this research introduces a novel multi-objective optimization technique to address these challenges. In a multi-objective optimization scenario, achieving an optimal solution involves optimizing multiple objective functions simultaneously.

In the literature, various versions of multi-objective evolutionary algorithms have been proposed (Beume *et al.* 2007). However, traditional evolutionary multi-objective optimization algorithms (EMOA) encounter obstacles such as restricted exploration, premature convergence, dependence on decomposition techniques, and absence of explicit memory (Deb *et al.* (2008). In response to these challenges Campos Jr. *et al.* (2019) proposed the MO-PSO algorithm. MO-PSO which stands out as a promising method that combines the strengths of particle swarm optimization with multi-objective optimization techniques.

Inspired by the behavior of horses, the Horse Herd Optimization (HHO) algorithm has emerged as a potent tool for tackling intricate optimization problems. In a study by Naeimi *et al.* (2021), HHO showcased superior performance compared to several other optimization algorithms, including grasshopper optimization, multiple verse optimizers, the sine-cosine method, the dragon-fly algorithm, moth flame optimization, and grey wolf optimization, particularly in scenarios with high dimensions. HHO's strengths lie in its simplicity, efficient memory utilization, and ability to sustain diverse solutions effectively. Building on the advancements of HHO, the MO-HHO method has been successfully applied to determine the optimal threshold for data compression in smart grid applications.

This research leverages a diverse array of datasets to demonstrate the efficacy of multi-objective HHO in data compression, with the aim of validating the proposed methodology. The suggested approach surpasses other data compression techniques in terms of signal-to-noise ratio (SNR) and mean square error (MSE), while also achieving significant reductions in data length.

SUMMARY

In smart grid applications, compression involves representing information in a condensed form, resulting in a reduction in the length of the data after the compression process. Depending on the type of compression algorithm, the original version of the data can be retrieved via the reconstruction process. WT-based compression is the most widely used and well-known technique for power systems and smart grid data compression. In general, WT-based data compression is carried out by using a universal threshold for ignoring particular wavelet coefficients. But the performance of data compression varies for different threshold values. Hence, selecting an optimal threshold is a challenging task for data compression. Therefore, to solve this issue, an effective optimization algorithm is needed. In this work, a multi-objective optimization algorithm such as the MO-PSO, and the MOHHO algorithm are developed to find out the optimal threshold. The suggested multi-objective algorithm accurately determines the global optimum threshold. Hence, it maintains a good compromise between SNR, MSE and CR. The effectiveness of the proposed algorithm is examined using four different datasets. Various datasets from the IEEE power quality wave data, household electric power consumption data, and appliance-level energy consumption from Dutch Residential Energy Datasets (DRED) and an online dataset from an experimental set-up are used to test the proposed algorithm. The outcome demonstrates that the suggested multi-objective algorithm performs better than the other conventional algorithms.