Versatile Multivariate Data Pruning in Smart Grid IoT Networks

Sharda Tripathi, Mayukh Roy Chowdhury, and Swades De
Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India

***Abstract*—With wide scale sensor deployments in smart grid IoT networks, there has been a manyfold increase in the variety and quantity of data generated in the network. In this work, the problem data reduction in smart grid IoT network is addressed to enhance the resource utilization without hampering the required quality of service. A novel versatile algorithm for multivariate data pruning at the edge devices in smart grid IoT networks is presented. This is achieved via a two stage data reduction mechanism which first exploits the inter-variable correlation to cut down on the number of transmitted variables, followed by adaptive data compression in temporal domain using adaptive compressive sampling. It is shown that, with the application of the proposed algorithm at the edge nodes, around 23% savings in bandwidth requirement can be achieved with minimum loss of information.**

***Index Terms*—Multivariate data compression, edge processing, adaptive compressive sampling, singular value decomposition, PMU data, smart grid communication, IoT networks**

I. Introduction

The emerging 5G communication technologies facilitate pervasive monitoring and control of power grid through smart Internet-of-Things (IoT) networks. Here, the Phasor Measurement Units (PMUs) act as IoT sensor nodes or edge devices for measuring the system state in terms of time-synchronized phasors of voltage and current, frequency, and rate of change of frequency and communicating it to the control center for decision making purposes. Owing to complex architecture of power grid, widespread deployment of PMUs, and rapid increase in their data sampling rate, massive data comprising of multiple variables which characterize the health of power system are required to be transmitted. Thus, efficient data handling at the source node itself is vital for optimal resource utilization in the communication network. To this end, adaptive pruning of multivariate data based on its temporal variation and inter-variable correlation without affecting the quality of power grid health monitoring is of our research interest.

A. Related Works and Motivation

Several lossy and lossless algorithms for the compression of PMU data have been proposed in the literature. These include dimensionality reduction techniques such as PCA [1], ICA, non-negative matrix factorization [2], transform based signal processing algorithms using DWT, DCT [3], JPEG based compression [4], and statistical learning techniques [5], [6]. It may be noted here that, most of the pruning techniques proposed for PMU data reduction are targeted to operate at the data aggregation points where extensive data from different PMUs are available, rather than at the edge nodes. Some of the works [1], [2] proposed a PMU data compression mechanism for spatial as well as temporal dimension to account for redundant information from PMUs in the close proximity of each other. Another important case is of pruning multiple variables measured at a single PMU. In state-of-the-art, multivariate data pruning is addressed by their independent compression, or if inter-variable correlation is considered, data buffering is required which makes this technique suitable for aggregated data over a certain time frame. It is observed that pruning algorithms proposed so far for PMU data are good for offline data compression during the archival stage, or for the transmission of aggregated data to the control center. Online multivariate data compression at the edge nodes in smart grid is challenging particularly due to the limited scope of batch formation for data compression owing to stringent latency constraints in data delivery.

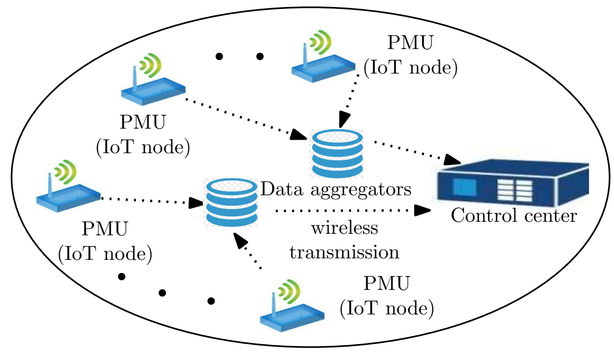

Fig. 1: IoT network for smart grid communication.

B. Contribution and Significance

To this end, in this work we propose a versatile multivariate data pruning algorithm for delay sensitive data at the edge node in smart grid communication, where the trade-off between latency and compression is addressed by adaptively choosing the batchsize based on number of variables to be transmitted from each PMU, and the inherent sparsity in each of the variable streams. Further, the proposed algorithm performs a two-fold data compression, first by exploiting the inter-variable correlation using principal component analysis to reduce the number of transmitted variables, and subsequently pruning down the number of transmitted samples corresponding to each variable in the temporal domain using adaptive compressive sampling. The proposed adaptive multivariate

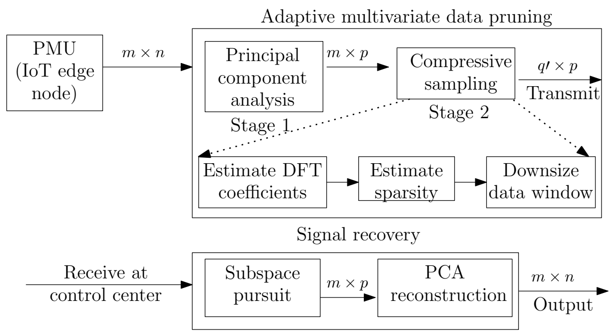

Fig. 2: Adaptive multivariate data pruning for PMU data.

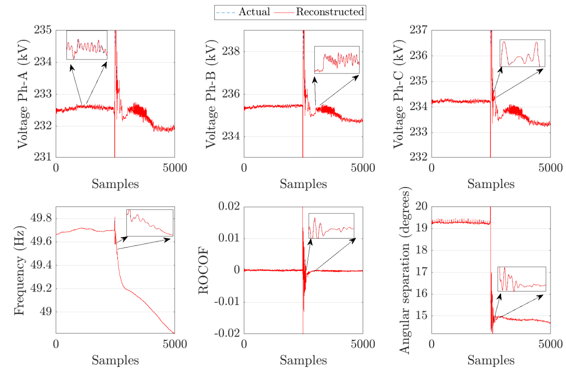

Fig. 3: Reconstruction performance of proposed multivariate data pruning algorithm for different variables.

pruning at the edge nodes signifies first level data reduction in the communication network. We anticipate that, with the proposed algorithm working in unison with the data reduction techniques at the aggregation points, overall communication resource requirement from PMUs to the control center will be remarkably curtailed.

II. Proposed Multivariate Data Pruning Algorithm for PMUs

A smart grid monitoring IoT system model is shown in Fig.1. It is observed that due to high sampling rate at the PMU, much of the information contained in successive data samples is redundant. Though reduction of PMU data at the edge node is of our primary interest, it should be ensured that the proposed data pruning does not hamper the quality of power grid health monitoring. Thus, the key idea here is to vary the compressibility of data in accordance with the occurrence of grid transients to bring down the resource requirements as well as preserve the underlying process dynamics.

The proposed multivariate data pruning for PMU data at the edge nodes comprises of two compression stages. In the former stage the variables measured by the PMU are decorrelated using principal component analysis which involves the application of orthogonal transformations to yield principal components that are linearly uncorrelated. Let $X \in \mathbb{R}^{m\times n}$ be the input data matrix, where n is the number of variables, and m is the number of samples taken for each variable. Depending upon the number of variables measured by the PMU, m is chosen to be $n+1$, which is typically the minimum batch size required to perform the principal component analysis. The principal components are obtained in vector $Y \in \mathbb{R}^{m\times n}$, such that $Y = XZ$, where $Z \in \mathbb{R}^{n\times n}$ denotes the matrix consisting of orthogonal basis vectors. The number of principal components $p, p \ll n$ are then identified as the dimensions which preserve 99% variance of the input signal. Further, the projection of input data X along p principal components is obtained as $\hat{Y} = X\hat{V}$, such that $\hat{Y} \in \mathbb{R}^{m\times p}$, and $\hat{V} \in \mathbb{R}^{n\times p}$.

In the latter step, data reduction in temporal domain is performed on all the selected principal components using adaptive compressive sampling, wherein the sparsity in each uncorrelated component is exploited such that the compressed signal can be reconstructed from far fewer samples than required by the Nyquist criteria. Let $\{y_k = y_{k1}, y_{k2}, \cdots, y_{kq}\}$ be the k^{th} column of $\hat{Y}$, then y_k can be expressed as $y_k = \psi f$, where ψ is the sparse basis matrix of size $q \times q$ and f is the column vector of coefficients corresponding to ψ. Depending upon the sparsity in vector y_k, only $q\prime$ ($q\prime \ll q$) random samples out of q are transmitted. The transmitted samples are denoted by $\widetilde{y_k} = \phi y_k = \phi\psi f$ where ϕ is a $q\prime \times q$ sensing matrix. In this work, subspace pursuit algorithm [7] has been used to recover the temporally compressed principal components. Further, the reconstruction of all the measured variables from the principal components is performed through the knowledge of $\hat{V}$ using inverse matrix operations at the control center. Discrete Fourier transform and identity matrix are chosen as sparse basis and sensing matrix, respectively, such that incoherence and restricted isometry property are satisfied for successful signal reconstruction. In Fig. 2, a detailed block diagram of the proposed adaptive multivariate data pruning algorithm for PMU data reduction is presented.

III. Results

In this section, performance of the proposed multivariate data pruning algorithm for PMU data is presented. Real dataset from the tripping incident of Rihand thermal power station unit in India on June 1st, 2010 has been used for the analysis in this work. It comprises of 6 variables including three phase voltages, frequency, rate-of-change-of-frequency (ROCOF), and angular separation between PMUs located at 4 different locations, namely, Vindhyachal, Dadri, Kanpur, and Moga.

A. Performance of Proposed Adaptive Multivariate Data Pruning Algorithm

As discussed in Section II, signal recovery in the temporal domain takes place through subspace pursuit algorithm, followed by reconstruction of all measured variables from the principal components using inverse matrix operations at the control center. In Fig. 3, reconstructed data is plotted against the actual values of all measured variables: 3 phase voltages, powerline frequency, ROCOF, and the angular separation for the PMU at Vindhyahal location. From the plots it is observed that the reconstructed data follows the actual values very closely. Normalized root mean square (nRMSE) is used here

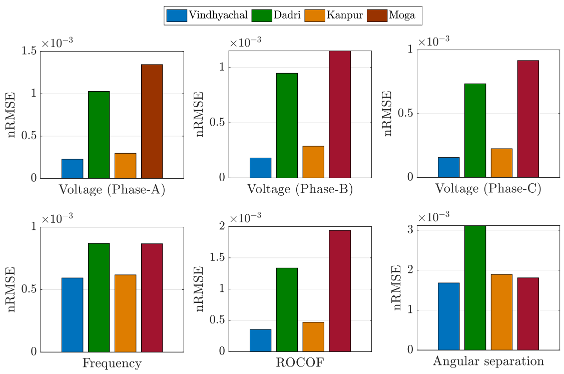

Fig. 4: Comparison of nRMSE for different variables across different locations.

to quantify the reconstruction accuracy. For required quality of service, nRMSE below 0.2 is considered acceptable [8]. Mean nRMSE for 3 phase voltages is obtained as 1.88×10^{-4}, for frequency, ROCOF, and angular separation it is observed to be 5.9×10^{-4}, 3.5×10^{-4}, and 1.68×10^{-3}, respectively, each of which is well below the acceptable limit.

In Fig. 4, nRMSE comparison at 4 different PMU locations: Vindhyachal, Dadri, Kanpur, and Moga is presented for 3 phase voltages, powerline frequency, ROCOF and angular separation. It is observed that for every variable at each PMU location, nRMSE is on the order of 10^{-3}, which is within the permissible error threshold. *Thus, multivariate compression of PMU data is achieved at the edge node with minimum information loss.*

B. Bandwidth Saving with Proposed Adaptive Multivariate Data Pruning Algorithm

Bandwidth saving in transmission of compressed data is measured by the relative difference between the number of uncompressed data points and the number of data points transmitted after adaptive multivariate data pruning. Mathematically, it is evaluated as $(m.n - q\prime.p)/m.n$. It may be recalled from Section II that n is the number of variables measured by the PMU, m corresponds to the number of samples for each variable, p denotes the number of principal components, and $q\prime$ is the number of samples in each of the temporally compressed principal component stream. A comparison of bandwidth saving across 4 given PMU locations: Vindhyachal, Dadri, Kanpur, and Moga is presented in Table I. Mean bandwidth saving observed across these locations is 23.6%. *Thus, by applying adaptive multivariate data pruning at the PMUs, around 23% reduction can be attained in the amount of data content transmitted from a single IoT edge node.*

IV. Conclusion

In this work, an adaptive pruning algorithm has been developed for the compression of multivariate data sampled at the PMUs in smart grid IoT networks. In the proposed algorithm, by exploiting the inter-variable correlation using principal component analysis, the number of variables required to be transmitted are reduced, and subsequently redundancy in temporal domain is eliminated through adaptive compressive sampling. Through extensive simulations on a real dataset it has been demonstrated that around 23% bandwidth saving can be achieved at the edge nodes with nRMSE for data reconstruction on the order of 10^{-3}, thereby signifying negligible loss of information in the data pruning and recovery process. Future works will be aimed at enhancing the scalability and reducing the computational complexity of the proposed algorithm for its wide range adoption in smart grid IoT networks.

TABLE I: Comparison of bandwidth saving obtained from proposed multivariate data pruning across different locations

Location	Bandwidth saving
Vindhyachal	29.35%
Dadri	19.68%
Kanpur	27.24%
Moga	18.07%

References

[1] G. Lee and Y. Shin, “Multiscale PMU data compression based on wide-area event detection,” in *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct 2017, pp. 437–442.

[2] A. Kummerow, S. Nicolai, and P. Bretschneider, “Spatial and temporal PMU data compression for efficient data archiving in modern control centres,” in *2018 IEEE International Energy Conference (ENERGYCON)*, June 2018, pp. 1–6.

[3] J. Khan, S. Bhuiyan, G. Murphy, and J. Williams, “PMU data analysis in smart grid using WPD,” in *IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society*, Oct 2014, pp. 2438–2444.

[4] S. Acharya and C. L. DeMarco, “Exploiting network-induced correlation for efficient compression of PMU data,” in *2018 North American Power Symposium (NAPS)*, Sep. 2018, pp. 1–6.

[5] S. Tripathi and S. De, “Assessment of power system stability using reduced-rate synchrophasor data,” in *2015 IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS)*, Dec 2015, pp. 1–5.

[6] ——, “Dynamic prediction of powerline frequency for wide area monitoring and control,” *IEEE Trans Ind. Informat.*, vol. 14, no. 7, pp. 2837–2846, July 2018.

[7] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[8] J. Qin, Q. Zhao, H. Yin, Y. Jin, and C. Liu, “Numerical simulation and experiment on optical packet header recognition utilizing reservoir computing based on optoelectronic feedback,” *IEEE Photon. J.*, vol. 9, no. 1, pp. 1–11, Feb 2017.