2018 3rd International Electrical Engineering Conference (IEEC 2018)
Feb, 2018 at IEP Centre, Karachi, Pakistan

Tri-Stage Cascaded Data Compression Framework for Smart Distribution Systems

Syed Muhammad Atif[1*] and Anees Ahmed[2]

[1] Graduate School of Science and Engineering , PAF Karachi Institute of Economics and Technology, Karachi, 75190, Pakistan (syed.muhammad.atif@gmail.com)

[2] Graduate School of Science and Engineering , PAF Karachi Institute of Economics and Technology, Karachi, 75190, Pakistan (aneesemail@yahoo.com) * Corresponding author

Abstract: Modern smart distribution system requires storage, transmission and processing of big data generated by sensors installed in electric meters. On one hand, this data is essentially required for intelligent decision making by smart grid but on the other hand storage, transmission and processing of that huge amount of data is also a challenge. This paper proposes a data compression technique called Tri Compress that blends three different methods in order to achieve high compression rate for efficient storage and transmission. It is a lossy data compression technique. Our simulation results are excellent, i.e. data compression ratio is a low as 100:1, and shows that this technique is far better than contemporary techniques.

Keywords: Singular Value Decomposition, Normalization and Sparse Matrix Representation, smart grid, power system monitoring, lossy data compression, big data, compression ratio.

I. INTRODUCTION

We are currently living in the age of IoT. Like any other aspect of our life, power distribution systems are also taking advantage of it. Smart meters are rapidly replacing the conventional meters that allows power distributions companies to get insight of user power usage behavior and make their decision accordingly. Power grids are not only vital in distribution system but also very sensitive to fluctuations in demands. Sudden increase in demand lead to tripping of grid stations. Unfortunately, this local failure has cascading effects under certain circumstances leading to a catastrophe such as New York power outage 2003 [1-3]. Those catastrophic events urges us to design and deploy such power distribution system that are smart and intelligent enough to quickly take their decisions both reactively and proactively without any human intervention. As a result, a new era of smart power distribution systems begins. They are equipped with smart metering system that continuously provide power usage of their users. The system then utilizes this data for operations like monitoring, analysis and control. However, smart meter and other similar equipment continuously generate huge amount of data at constant rate that bring the challenges in terms of their transmission and storage. This paper focus the said problem and proposed a new solution named tri compress. As the name suggest this solution is the blend of three different methods precisely, Singular Value Decomposition (SVD), normalization and value-index sparse matrix representation. In nature, the proposed technique is loosy data compression technique.

This paper is organized as follows. Section I has presented the introduction. Section II will give related work. Section III, the core of this paper, will present the proposed idea and technique for data compression. Section IV will experimentally evaluate the newly proposed idea. Section V will provide the obtained results of simulation. Finally, section VI will conclude the paper.

II. RELATED WORK

With the advancements in smart distributions systems and integration of IoT, the amount of data available of storage, transmission and processing will become gigantic. This challenge draws attention of many researchers towards the development of data compression technique that specially cater needs of smart distribution systems [3-9]. Brindha and D. Sundararajan uses discrete wavelet techniques for compression of data. They uses a bi-orthogonal 5/3 spline filter for compression and able to achieve compression ratio of up to 8:1. Phasor measurement unit, a well-known image compression technique, is employed by Klump et al. [9] for compression of smart grid data that result in obtaining the best CR of 14.35:1. However, there is still a need of higher CR due to large amount of data which is specifically address in this paper.

The nature of data obtain from smart grid system allow us to store it in the form matrix as the data coming from different sensor belong to measurements taken for the same device at several different time instances. This form dataset representation is the most appropriate for SVD, a widely used technique in the field of image compression and many other [11-14]. This paper is utilize SVD for compression of data of smart grid systems. The reason is that it provide us a good tradeoff between information loss and degree of achieved compression. It is possible in case of smart grid data as it generally used by applications for monitoring and planning purposes that do not require high precision data.

III. PROPOSED SOLUTION

In this section, we explain our tri-stage cascaded data compression framework via singular value decomposition (SVD), normalization and sparse matrix representation (see Fig. 1.). The subsection A will explains how SVD will be utilized in our data compression framework model to achieve data compression. Subsection B will elaborates the second stage of our model i.e. normalization for efficient representation of data obtain after apply SVD in the first stage whereas subsection C provides detail regarding the third stage of our model that exploits sparse matrix representation for compressed data obtained from second stage so that data can be stored or transmitted in its most compressed form.

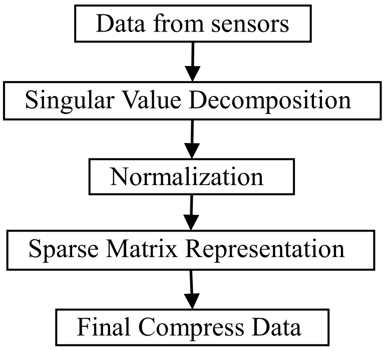

Fig. 1 Data Compression Framework.

A. Singular value decomposition (SVD)

Let **X** be the data collected from different sensors at regular time intervals. This data **X** can be considers as a matrix of dimension *m* by *t* where *m* is number of metering devices equipped with sensors and *t* is the time stamps as depicted in Fig. 2. Singular value decomposition (SVD) will degenerate this matrix **X** into three matrices **U**, **S** and **V** where **U** and **V** be the orthogonal unitary matrices of dimension *m* by *m* and *t* by *t* respectively where **S** be the diagonal matrix of dimension *m* by *t* that diagonal entries are arranged in descending order. Fig. 3 is giving a pictorial representation of this factorization. However, the effective dimension of **U**, **S** and **V** matrices are *m* by *r*, *r* by *r* and *r* by *t* where *r* is the rank of the matrix **X**. It is noteworthy that here we use matrix V with dimension *r* by *t* for convenient. It is equivalent to V^T of dimension *t* by *r* that is conventionally used in literature.

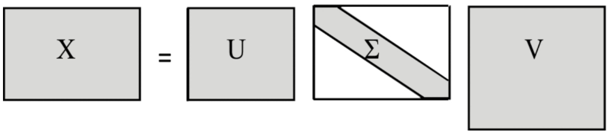

Fig. 2 SVD of matrix X.

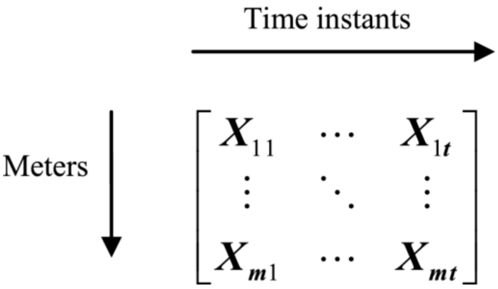

Fig. 3 Representation of data as matrix X.

In essence, SVD represents a matrix X as the sum of rank one matrices ordered in descending order of their respective frobenius norm.

$$X = \sum_{i=1}^{r} \sigma_i u_i v_i$$

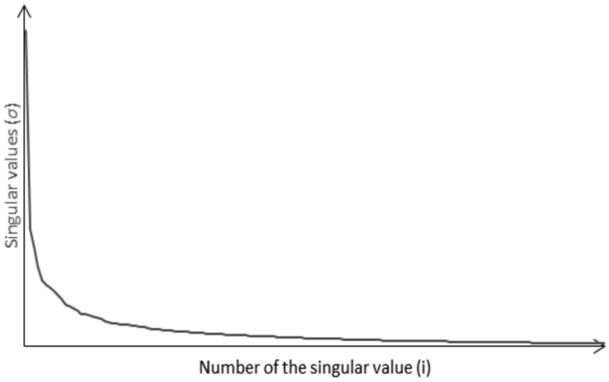

Fig. 4 Typical decrease in ordered singular values.

Since U and V are unitary matrices i.e. their column has unit L2 norm, so their corresponding singular value in the diagonal matrix S will represent the frobenius norm of respective sum term. Fig. 4 is illustrating the typical decay (decrease) in the singular values in SVD.

Hence, a matrix can be well approximated by the first few terms. Let **X** be approximated as **X_k** by the top k singular values then

$$X \approx X_k = \sum_{i=1}^{k} \sigma_i u_i v_i \tag{1}$$

or in matrix notation

$$X \approx \hat{X} = U_k \Sigma_k V_k \tag{2}$$

Where the dimension of U_k, Σ_k and V_k are *m* by *k*, *k* by *k* and *k* by *t* respectively and $k. \leq r$.

The storage or transmission capacity required for X without any manipulation is:

$$space(X) = m \times t \tag{3}$$

However, the storage or transmission capacity required for $\mathbf{X_k}$, the approximation of **X**, is:

$$\begin{aligned} space(X_k) &= space(U_k) \\ &+ space(\Sigma_k) + space(V_k) \end{aligned} \tag{4}$$

But,

$$space(U_k) = m \times k \tag{5}$$

$$space(\Sigma) = k \tag{6}$$

$$space(V_k) = k \times t \tag{7}$$

Hence,

$$\begin{aligned} space(X_k) &= m \times k + k + k \times t \\ &= (m+1+t)k \end{aligned} \tag{8}$$

Therefore, we have compressed data X_k is comparing to X i.e. *space*($\mathbf{X_k}$) < *space*(**X**) provided that m×t/(m+1+t)t < 1 or equivalently k << r ≤ m.

B. Normalization

Data coming from sensors in electrical devices as in our case is mainly wattage, voltage and current usage i.e. it is numerical in nature. Such a data is typically stored or transmitted by computing devices in floating point format or mantissa exponent format, a widely used standard for this format is IEEE 754. However, blindly using this format is not good in our case because of reason that people in vicinity usually have similar rage of electrical device as well as similar electrical usage behavior. It leads us to the conclusion that the variation in obtained data will be at most order of two or so. It is an ideal condition where technique like normalization can be readily use for data compression.

Normalization is way of representing decimal point numerical data. It transforms the entire data, by simply multiply them suitable powers of their system radix, so their exponents become identical. This transformations generally leads to precision loss. However, this precision loss is negligible when variation in data is vary low.

Normalization allows us to transmit or store the exponent part of data only once as it is now common. This leads to significant reduction in the size of that that need to be store or transmitted. For example there is 25% reduction in the size of data after normalization if the data is represented in single precision floating point IEEE 754 format as mantissa to exponent ratio in this format is 3:1.

C. Sparse Matrix Representation

Spare matrix is ones that vast number of entries are zeros. It is not a good idea to store (transmit) all the entries of such a matrix. Instead, one can store (transmit) a sparse matrix efficiently if it just store (transmit) only non-zero entries along with their indices (positions) in the matrix. It leads to reduction in size of data provided the amount of space required for zero entries is more than that required for index value representation of non-zero entries of the given sparse matrix. The matrix X of our data is well sparse in nature because most of the electronic (electrical) devices such as water pumping machine, iron, microwave oven, television etc. typically used only very small fraction of time. So, U_k and V_k matrices obtained after apply SVD of this data will also be sparse matrices i.e. most of the entries in most of the column of U_k (resp. most of the entries in most of the rows of V_k) will be zero. Thus, index value representation of U_k and V_k matrices will likely lead to significant data compression.

IV. SIMULATION AND EXPERIMENTAL RESULTS

In this section, we will present results obtained from simulation conducted to check the validity of our proposed tri-stage cascaded data compression framework specifically tailored for smart distribution (grid) system. We use TU Darnstadt tracebase data set. It is a freely available dataset that can be obtained from [15-16]. In our simulation study, we use data of five different devices namely DVD player, Subwoofer, TV and vacuum cleaner. There are total 80 devices precisely 28 DVD player, 12 Subwoofer, 16 TV and 24 vacuum cleaner. The wattage usage of these 80 devices is recorded for the entire day after about every 18s.

k	MAE (at all stages)	Compressed File Size after(in kB)			Compression Ratio after		
		SVD	Normalization	Sparse Matrix Representation	SVD	Normalization	Sparse Matrix Representation
1	8.747	15	14	8	100:1.89	100:1.77	100:1.01
2	7.280	33	26	11	100:4.12	100:3.28	100:1.39
3	6.328	48	37	14	100:6.06	100:4.67	100:1.77
4	4.932	64	51	21	100:8.08	100:6.44	100:2.65
5	4.406	79	61	23	100:9.97	100:7.70	100:2.90
8	2.248	127	101	41	100:16.03	100:12.75	100:5.18
15	0.180	271	224	152	100:34.22	100:28.28	100:19.19
19	0.040	367	307	241	100:46.33	100:38.76	100:30.43

Table 1 Summarized results of tri-stage cascaded data compression framework.

Thus, our test data matrix X has dimension of 80 by 4902. This data matrix requires 792KB when store as csv file format. We design a MatLab code to evaluate our methodology. As, our compression framework is loosy in nature so a matric is required to quantitatively measure the loss of information during compression. We use MAE as a matric to measure the loss of information during compression. It is defined as:

$$MAE = (\frac{1}{m \times t}) \sum_{i=1}^{m} \sum_{j=1}^{t} X(i,j) - X_k(i,j) \qquad (9)$$

Table 1 summaries the results of simulation study. The size in KB required to store compressed version of data matrix **X** in csv file format is recorded after each of the three stage in our model.

You may observe that compression ratio of decreases as *k* “number of rank one matrices in SVD summation use for approximation” increases. It is expected as increase in a will increase the dimension of U, S and V that in turn increase the size of file. However, higher value of k means better approximation of given data or less loss of information. Therefore, there is a steady decline in MAE as the value of k increases. It means that a tradeoff is required between compression ratio required to achieve and loss of information while using our framework. One have to compromise on the precision of data when require higher compression.

Normalization stage always able to further compress the data obtain from SVD stage. However, its compression ability is less than both stage one and two.

Spare matrix representation always dramatically compress the data due to high level of sparsity present in the data. You may conclude from table 1 that one that this stage always compress the data obtain from stage two by 100% or more.

V. CONCLUSION

This paper presents a lossy data compression framework that is specially tailored by keeping need high data compression needs of smart power distribution system. The framework compresses the data in three different stages using different techniques. In the first stage, it exploit redundancy in the data using SVD for compression. The second stage applies normalization on resulting data whereas the third stage transforms the compressed output of second stage into its equivalent index value sparse matrix representation. Our simulation results shows that this tri stage cascaded data compression framework is very efficient and may even produce promising compression ratio of 100:1.

REFERENCES

[1] Marx, Melissa A., Carla V. Rodriguez, Jane Greenko, Debjani Das, Richard Heffernan, Adam M. Karpati, Farzad Mostashari, Sharon Balter, Marcelle Layton, and Don Weiss. "Diarrheal illness detected through syndromic surveillance after a massive power outage: New York City, August 2003." *American Journal of Public Health*, vol. 96, no. 3, pp. 547-553, 2006.

[2] Anderson, G. Brooke, and Michelle L. Bell. "Lights out: impact of the August 2003 power outage on mortality in New York, NY." *Epidemiology (Cambridge, Mass.)* vol. 23, no. 2, pp. 189, 2012.

[3] Lin, Shao, Barbara A. Fletcher, Ming Luo, Robert Chinery, and Syni-An Hwang. "Health impact in New York City during the Northeastern blackout of 2003." *Public Health Reports* vol. 126, no. 3, pp. 384-393, 2011.

[4] M. Ringwelski, C. Renner, A. Reinhardt, A. Weigel, and V. Turau, “The Hitchhiker’s guide to choosing the compression algorithm for your smart meter data,” *Proceedings of IEEE International Energy Conference and Exhibition*,

pp. 935–940, 2012.

[5] A. Unterweger and D. Engel, “Resumable load data compression in smart grids,” IEEE Trans. Smart Grid, vol. 6, no. 2, pp. 919–929, Mar. 2015.

[6] F. Zhang et al., “Application of a real-time data compression and adapted protocol technique for WAMS,” *IEEE Transactions on Power System*, vol. 30, no. 2, pp. 653–662, 2015.

[7] S. Brindha and D. Sundararajan, “Power quality monitoring and compression using the discrete wavelet transform,” *Proceedings of International Conference on Advance Computer Communication Systems*, Coimbatore, pp. 1–6, 2013.

[8] J. Ning, J. Wang, W. Gao, and C. Liu, “A wavelet-based data compression technique for smart grid,” *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 212–218, 2011.

[9] R. Klump, P. Argawal, J. E. Tate, and H. Khurana, “Lossless compression of synchronized phasor measurements,” *Proceedings of IEEE Power Energy Society General Meeting*, pp. 1–7, 2010.

[10] J. Khan, S. Bhuiyan, G. Murphy, and M. Arline, “Embedded zerotree wavelet based data compression for smart grid,” *Proceedings of IEEE Industry Applications Society Annual Meeting*, 2013 IEEE, pp. 1-8, 2013.

[11] J.-Ayubi and M.-Rezaei, “Lossy color image compression based on singular value decomposition and GNU GZIP,” Advances in Computer Science: an International Journal, vol. 3, no. 3, pp. 16–21, 2014.

[12] J.-J. Wei, C.-C. Chang, N.-K. Chou, and G.-J. Jan, “ECG data compression using truncated singular value decomposition,” *IEEE Transactions on Information Technology and Biomedical*, vol. 5, no. 4, pp. 290–299, 2001.

[13] R. A. Sadek, “SVD based image processing applications: State of the art, contributions and research challenges,” *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 7, pp. 26–34, 2012.

[14] A. M. Rufai, G. Anbarjafari, and H. Demirel, “Lossy medical image compression using Huffman coding and singular value decomposition,” *Proceedings of Signal Processing Communication Applications Conference*, pp. 1-4, 2013.

[15] Unterweger, A. and Engel, D., “Resumable load data compression in smart grids”. *IEEE Transactions on Smart Grid*, vol. 6. no. 2, pp. 919-929.

[16] TU Darnstadt tracebase data set, https://github.com/areinhardt/tracebase [online]