Research Report: Optimal Singular Value Decomposition Based Big Data Compression Approach In Smart Grids

Reported By: Alec Mabhiza Chirawu
Reporter Chinese Name: (亚历克上)
Reporter Student Number: M202161029

Advance Communication Coding , Information And Engineering Department, USTB

Professor: 杜利平(Liping Du)

29 October 2021

Summary

This report discusses the optimal singular value decomposition based big data compression approach in smart grids. A review of some of the available methods provides insights into the changing and improvement of automatic delivery data for electricity power with a two-way reliable flow of electricity and information among different equipment on the grids. .

Key findings include:

- There has been marked an increase in the transferring of data especially in the modern days.
- While the increase of big data also triggered the electricity power uses as the data being transmitted increases.
- The number of small electricity applicants have increased the amount of electricity usage on different times requiring monitoring to save electricity power.
- The same method of compressing data on electricity power also played an important row in image processing benchmarks.

The information presented in this report has been gathered from secondary sources and optimal singular value decomposition based big data compression approach in smart grids article.

The report has been prepared for submission as Advanced Communication Coding end of Course at University of Science and Technology in Beijing , Information and Communication Department.

TABLE OF CONTENTS

1 Introduction

The application of optimal singular value decomposition based big data compression approach in smart grids because smart grids distribute more data daily so due to that big data compression is required for easy transferring and data storage in small grids. If exploited properly , this data can reveal a great deal about the customers and generating units. So data compression techniques can bring great benefit to the smart grid by making it easier to handle large amounts of data.

1.1 Methodology

Information for this report was sourced from various secondary sources, all listed in the Reference List. The Data from publications by the IEEE transactions on industry applications also proved valuable. This report is not a comprehensive review of the available researches , but provides a broad overview of the topic.

1.2 Scope of the report

Where ever the term SVD on its own is used, it is reference Singular Value Decomposition. Different lossy compression schemes have been developed for smart grid applications based on wavelet decomposition, discrete cosine transform and fuzzy based approaches. There is also ongoing research on spatial domain methods like neural network - based methods and deep stacked auto encoders. The main idea is to save required memory and bandwidth. At the end providing a good performance for both compression and decompression of the considered data from the results.

2. Findings

2.1 Compression in smart grids

The importance of data compression in the smart grid will be more highlighted as a result of an increase in data exchange and the required memory is likely to occur. In noise reduction, some of the singular values with greater energy have been kept and the other ones are considered as noisy. The scree was employed to deal with this problem. Determining the proper cutoff number, that is the threshold to data division was the main challenge. This method can be employed effectively in different points of the grid where the volume of the sent and received data is high.

2.2 SVD Decomposition

An SVD of matrix is a factorization, as can be seen below, which decomposes the matrix A into U, V, and Σ. U is a m × m real or complex unitary matrix, and V is an n × n real or complex unitary matrix. The diagonal entries σi of Σ are known as the singular values of A. A decomposition of a diagonal matrix into elements that are on its original diameter are singular values that are placed in descending order. Each singular value is involved in the retrieving process of the original matrix.

2.3 Optimization Framework

The general form of an optimization problem that can be seen in the following equation is reviewed. Determining the number of singular values to retain can significantly affect the performance of the SVD-based data compression. The more singular values would lead to more accuracy, and of course, less compression ratio.

2.4 Experiments(Results)

A data matrix is employed to investigate the performance of the existing heuristic algorithms in solving the defined optimization problem. Various data types such as energy consumption and renewable production of houses in the U.K. for nine months with half an hour time-step and the electrical data over a single 24-h period from 443 unique houses on February 4, 2011 are selected. The presented scheme has been implemented using MATLAB 2018a on the market data. Differential evolution , simulated annealing, teaching learning based optimization, particle swarm optimization, as well as the GA with and without the mutation (GA-M and GA) are employed to investigate the ability of the existing evolutionary algorithms.

The compression ratio and percentage of redundant data for the obtained solution by each algorithm. Ofcourse, the algorithms with the same objective value propose the same compression ratio. But, the goal was to check the solvability of the proposed optimization problem. The quality of the compression scheme was evaluated in the next cases through comparisons with the other rank reduction methods. It is assumed that the maximum tolerable error must be lower than 0.5% of the Euclidean norm of the original data. This percentage is equal to 0.48 in the proposed method. The second method of the Guttman–Kaiser and the method by de Souza satisfy the accuracy constraint with a lower compression ratio.

2.4 ***Proposed method***

In the proposed method, the original data matrix is decomposed into U, V, and Σ by the SVD decomposition. Then, the optimal rank reduction is specified and three rank-reduced matrices will be ready to be sent. Since the main goal of the problem is data compression, the inverse of the compression ratio can be considered as the objective function in a minimization problem.

The objective function can be replaced by minus x, the compression ratio will be maximized. As an important consideration, the compression is valuable if the information can be retrieved with an acceptable accuracy. The constraint reinforces the proximity of two matrices with Euclidean norm criteria.

The lower bound expression can be modified based on the communication network requirements. In some cases, the communication network bottlenecks play the main role in determining the compression ratio. So the value can change between the lower bound based on the equation and the upper bound that is the total number of the singular values.

The proposed optimization framework contains one decision variable (x) regardless of the dimension of the original data. It is applied to the output of the SVD decomposition, which means that before the algorithm starts, the initial data should be applied to. Then, with $\alpha = 5\%$, the proposed method increases the compression ratio.

The performance of the SVD-based data compression can be affected by the data type. Two matrices of the same size may require different number of SVD's to meet the same error threshold, depending on the dependence of the rows or columns of the matrices. When the rows of the matrix are dependent, or there is a correlation between them, decision making gets more complicated. In these cases, most of the SVD's have small values,and suddenly after a knee point, their value increases sharply. The performance of various algorithms, as well as the proposed method on each image quality of the obtained points with various methods can be analyzed similarly to the previous cases. The knee point does not provide an accurate compression level in this case.

3. Conclusion

The SVD-based data compression is basically a tradeoff between the data accuracy and the compression level. Various methods decide for the number of the singular values in rank reduction problems and some of them reveal good performance. An important feature of the proposed method is that the objective function can easily be changed to find the optimal rank of the matrix for other applications. The computational complexity of the heuristic methods for solving the optimal rank reduction problem is proportional to the number of iterations. For instance, the SVD decomposition in MATLAB is in the order of O(max(m,n)2). In terms of computational complexity, the proposed method is similar to the other SVD rank reduction algorithms. An optimization method that finds the optimum point regarding the accuracy constraint is required in data compression, especially for applications that are sensitive to the error like energy market or smart grid data.
An optimization method that finds the optimum point regarding the accuracy constraint is required in data compression. The proposed framework achieves the higher compression ratio as well as the satisfaction of accuracy constraint simultaneously and the redundant section is cut down. It is simple and efficient and could be utilized by market operators, load aggregators, electricity retailers, and Combined Heat and Power(CHP) and micro grid operation.

4. Recommendation

The information collected for this report provides a broad overview of key changes in the SVD-based methods.Further analysis would be possible if the relevant data from year 2020-2021 are taken into test and comparison. The reliance on secondary has resulted in some patchy data. For example, it is not possible to identify for any data loss during data compression by the following categories:

- High population
- Real time

There for greater access to primary data would enable a more thorough analysis to be made.

5. Reference List

J. C. S. de Souza, T. M. L. Assis., and B. Pal, “Data compression in smart distribution systems via singular value decomposition,” *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 275–284, Jan. 2017.

V. Loia, S. Tomasiello, and A. Vaccaro, “Fuzzy transform based compres sion of electric signal waveforms for smart grid,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 1, pp. 121–132, Jan. 2017.

M. M. Dixit and C. Vijaya, “Effects of hybrid SVD–DCT based image compression scheme using variable rank matrix and modified vector quan tization,” in *Innovations in Computer Science and Engineering (Lecture Notes in Networks and Systems)*, vol. 32, H. Saini, R. Sayal, A. Govardhan, and R. Buyya, Eds. Singapore: Springer, 2019.

C. Yu, H. Li, and X. Wang, “SVD-based image compression, encryption, and identity authentication algorithm on cloud,” *IET Image Process.*, vol. 13, no. 12, pp. 2224–2232, 2019.

S. K. Mukhopadhyay, M. O. Ahmad, and M. N. S. Swamy, “SVD and ASCII character encoding-based compression of multiple biosignals for remote healthcare systems,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 1, pp. 137–150, Feb. 2018.

C. Yu, H. Li, and X. Wang, “SVD-based image compression, encryption, and identity authentication algorithm on cloud,” *IET Image Process.*, vol. 13, no. 12, pp. 2224–2232, 2019.

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.
[18] X. Huang, T. Hu, C. Ye, G. Xu, X. Wang, and L. Chen, “Electric load data compression and classifification based on deep stacked auto-encoders,” *Energies*, vol. 12, no. 4, pp. 653. Feb. 2019.