18th IMEKO TC10 Conference
"Measurement for Diagnostics, Optimisation and Control to Support Sustainability and Resilience"
Warsaw, Poland, September 26–27, 2022

Big Data Analysis in Smart Grid Systems

Yu Jun[1], Olena Hordiichuk-Bublivska[2], Yan Lingyu[3], Marian Kyryk[4], Mykola Beshley[5], Hu Jiwei[6]

[1] *Wuhan Fiberhome Technical Services Co., Ltd., Wuhan, China, yujun@fiberhome.com*
[2] *Department of telecommunications, Lviv Polytechnic National University, Lviv, Ukraine, olena.v.hordiichuk-bublivska@lpnu.ua*
[3] *School of Computer Science, Hubei University of Technology, Wuhan, China, yanlingyu@hbut.edu.cn*
[4] *Department of telecommunications Lviv Polytechnic National University, Lviv, Ukraine, marian.i.kyryk@lpnu.ua*
[5] *Department of telecommunications Lviv Polytechnic National University, Lviv, Ukraine Department of Information Systems, Faculty of Management, Comenius University in Bratislava, Bratislava, Slovakia, mykola.i.beshlei@lpnu.ua*
[6] *Wuhan Fiberhome Technical Services Co., Ltd., Wuhan, China, hujiwei@fiberhome.com*

***Abstract* – The problem of Big Data processing in large industrial systems requires the use of machine learning methods. A smart grid system is an example of improving the efficiency of traditional data processing systems, which allows much more efficient and flexible distribution of electricity to end-users. However, for a smart grid to work properly, it needs to constantly monitor data from sensors and meters. The Singular Value Decomposition (SVD) algorithm is used to improve the efficiency of Big Data processing and reduce its dimensionality. The paper proposes the use of advanced SVD, which can work in distributed industrial systems and ensure the reliability and speed of data processing.**

Keywords* – *Smart grid, Industrial Internet of Things, Federated Machine Learning, Affordable and Clean Energy, Big Data, Singular Value Decomposition.

I. INTRODUCTION

The United Nations adopted in 2015 the roadmap to cope with poverty, inequality and ensure sustainable development till 2030. This road map is entitled The Global Goals [1]. One of the core directions is the direction of affordable and clean energy. Between 2000 and 2018, there is a considerable raise from 78% to 90% in the amount of people who have access to electricity. Despite the considerable progress, a lot of problems should be addressed in the energy sector. The problems of this sector are very versatile. They range from the sensors [2], to instrumentation [3] to metrology [4], and to genuine problems with the energy production and transfer [5]. In the epoch of the internet of things [6] appeared the concept of the smart grid. This paper considers this domain. The concept of Industry 4.0 involves the digitalization of production processes and the presentation of the results of industrial systems as products available to end-users. Smart grim is an example of an optimized distribution of electricity in a large-scale system between end-users. The analysis of the received information allows revealing places of losses of electricity, inefficient use of the subscriber equipment, the unsatisfactory condition of network infrastructure, etc. All these data allow quickly making decisions based on identified problems and eliminate them with minimal human intervention.

It is proposed to use the SVD algorithm to process large data on end devices of intelligent systems. This algorithm is often used in reference systems to determine the relationships between different data. The paper proposes the use of the SVD algorithm not only to search for anomalous data but also to pre-reduce the dimension of information arrays by discarding unimportant.

II. FEATURES OF MANUFACTURE DIGITALIZATION

A. Industrial Internet of Things

If earlier all industrial productions were managed mainly by service personnel, now the possibility of more flexible management has developed. The use of a large number of end devices that have the functions of collecting data from the environment, analyzing them, and the ability to make decisions (in part or in full) opens new perspectives for the development of industrial systems.

The decentralized architecture of large-scale industrial systems determines the interaction between many devices to achieve a common goal. The use of distributed data processing algorithms allows taking into account all the features of the system. Distributed computing involves working together on a single task of multiple devices at the same time [7, 8]. The examples of IIoT systems are shown in Fig.1.

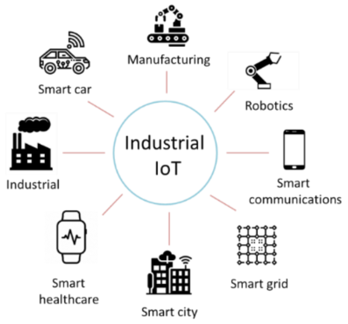

Fig.1. Examples of IIoT systems

B. Big Data processing in smart grid

Smart grid is an important component of industrial systems. Thanks to the smart grid, it is possible to distribute electricity according to the needs of end-users and with the greatest efficiency [9-12]. As in the Industrial Internet of Things, sensors are used to collect and process data, which then transmits data to central devices.

In the Industrial Internet of Things and the smart grid, Big Data comes from a variety of end devices that collect data. The information must then be processed on the central device and returned to the end-users. The duration of calculations should be such that it satisfies the requirements for the quality of service provision.

Big Data processing requires the use of the latest methods of information analysis. Machine learning and artificial intelligence are some of the effective tools for analyzing large amounts of information. Because there is always plenty of data collected from industrial devices, this allows to always effectively teach machine learning algorithms. However, the excessive amount of information also causes difficulties, as the duration of the analysis and the load on the data processing device increases. Optimization methods are used to determine the most important data for further processing and discard redundant ones. Determining the required information you to perform further calculations faster and more efficiently [13, 14]. The global model is then sent to all local nodes that are correcting local machine learning models. This approach significantly increases the reliability of user data, as it is not transmitted through transmission channels and is not at risk of being intercepted or lost [15-18].

III. SINGULAR DECOMPOSITION OF DATA IN SMART GRID

The SVD algorithm can be used in smart grid systems to reduce data dimensions and search for dependencies [19]. According to the algorithm, the schedule of the data matrix has the form:

$$M = U \times \sum \times V^T . \tag{1}$$

where $\sum(m \times n)$ – diagonal matrix, $U(m \times m)$ and $V(n \times n)$ - identity matrixes.

Data transformation can be represented as the deformation of a two-dimensional geometric figure. A rectangular shape can be transformed linearly (without changing the diagonal line) and nonlinearly. The nonlinear transformation of the figure is much harder to describe, so it is more convenient to bring it to a lily look. For preventing nonlinear transformation, we can first rotate the figure to a certain angle. Thus, we can represent the transformed figure as a result of multiplying the initial by certain coefficients.

So, we can write:

$$v_1,\ v_2 . = Mv_1,\ Mv_2 \tag{2}$$

Redefine the parameters of the figure v as vectors:

$$M \times v_1 = \sigma_1 \times u_1,$$

$$M \times v_2 = \sigma_2 \times u_2 . \tag{3}$$

Based on the properties of vectors:

$$x = (xv_1)v_1 + (xv_2)v_2 , \tag{4}$$

let's define:

$$M = \sigma_1 u_1 v_1 + \sigma_2 u_2 v_2 . \tag{5}$$

Then write the result:

$$[M] \times [v_1 \quad \dots \quad v_n] = [u_1 \quad \dots \quad u_m] \times \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \sigma_n \end{bmatrix} . \tag{6}$$

Singular numbers $\sigma_1,\dots,\sigma_n$ are arranged in descending order and determine the degree of relationships between matrices U and V . Based on the matrix $\sum$ it can be rejected the number k of diagonal elements n . Also rejected is the number k of rows in the matrices U and V . Thus, by discarding smaller values of diagonal elements, it can reduce the amount of data for further calculations, but keep the properties of the original matrix. Based on preliminary calculations:

$$M \times V = U \times \sum , \tag{7}$$

$$M \times V \times V^T = U \times \sum \times V^T . \tag{8}$$

$$M = U \times \sum \times V^T . \tag{9}$$

FedSVD algorithm is an advanced method of SVD that can be used in distributed systems, and data can be presented as:

$$M = \{m_1, m_2, \dots, m_k\}, m \tag{10}$$

$$m_i = P \times m \times Q^R \quad (11)$$

where m is the data from the end users; P, Q is the data are generated by a special masking server and transmitted to end devices to encrypt local data; R is the data is generated on terminal devices to calculate the masking of private information before transmission over the communication channel [20]. And it can be write:

$$M \in R^{mxn}, \quad (12)$$

$$M = \{m_1, m_2, \ldots, m_i\}, \quad (13)$$

$$m_i = U * \Sigma * V_i^T. \quad (14)$$

$$M = U * \Sigma * V_{\Sigma i}^T. \quad (15)$$

Based on the features of smart grid systems described above, we propose a modified method of SVD. This method works similarly to the FedSVD, but all the data needed to mask private information is generated directly on the host site. (Fig.2).

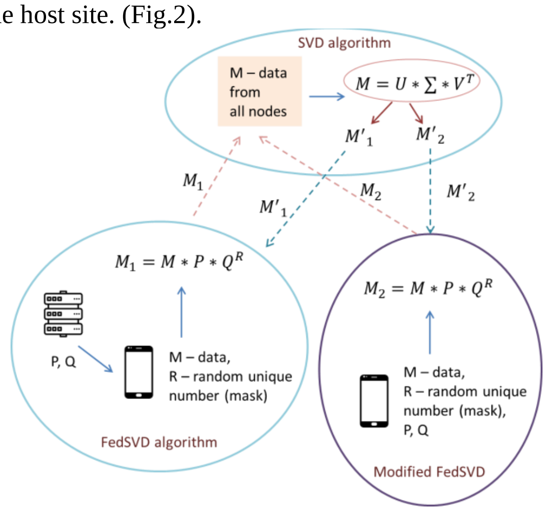

Fig.2. Implementation of FedSVD and improved FedSVD methods

Modification of the federated SVD avoids the transfer of values required for data masking between different devices. This reduces the risk of interception or damage. Because all calculations for pre-processing and masking before sending to the central node are performed locally, this requires the use of end devices with fairly high power. Some of the devices in the system may not have the necessary parameters of computing power. It is also permissible to perform data calculating on edge devices that are close to the local ones. The results of the data processing efficiency using different SVD algorithms are presented in the next section.

IV. RESEARCH OF BIG DATA PROCESSING EFFICIENCY USING FEDSVD ALGORITHM

In this paper, the operation of SVD algorithms, as well as FedSVD and its improved version, following Fig.5 are researched. The simulation was performed using the Python programming language and built-in libraries for processing matrices and calculating singular schedules. First of all, we compare the duration of calculations of each algorithm when performing the task of finding anomalous deviations for different volumes of data sets (Fig. 3).

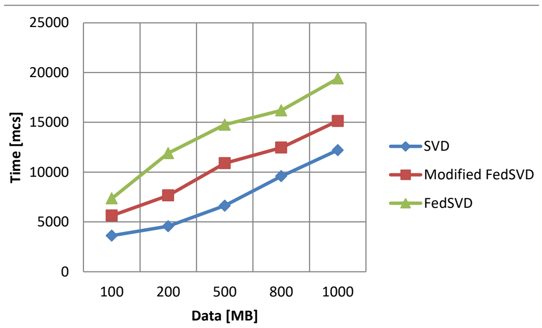

Fig.3. The duration of calculations for SVD, FedSVD and modified FedSVD algorithms depends on the data size

From Fig.3 we have seen that because the FedSVD algorithm uses additional operations to protect local data, it works slower than the classic SVD. However, the proposed modification and generation of all data necessary to mask user information directly on local nodes reduces the duration of calculations for all studied volumes of data. The results obtained are especially important for industrial systems in which automatic control and troubleshooting take place. As the amount of data coming from different end nodes increases, there is a risk of overloading data processing devices and slowing down calculations. Now let's examine the effect of data optimization on the accuracy of calculations. As noted in the previous section, the change in the number of diagonal elements of the matrix Σ (singular numbers) in SVD allows for the reduction of the rank of matrices. Therefore, we changed the singular numbers used for further calculations and compared the accuracy of the results of the calculations to find anomalous deviations in the data set of 1 GB. The results of the calculations are shown in Fig.4.

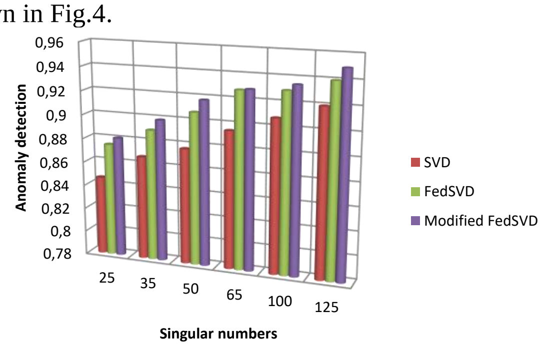

Fig.4. The accuracy of data processing for different SVD algorithms depends on the singular numbers

From Fig.4 we can conclude that the proposed algorithm of the modified FedSVD allows for the best preserving the properties of the original data. The results of the research show that the modified FedSVD also allows processing data quickly and reliably. High privacy of users' private information is also ensured.

V. CONCLUSIONS AND OUTLOOK

Smart grids deliver electricity based on the needs of end-users, who can also influence the management of the system. Resource exchange and flexible management can minimize losses and increase the efficiency of electricity distribution systems. The large amounts of information that need to be processed during the smart grid management process cause inconvenience to accurate and fast data analysis. The paper proposes an advanced SVD that can work in distributed systems and has high reliability and confidentiality of data. The results of the study showed that the proposed method accurately processes the data of users of the smart grid and protects them from outside interference.

REFERENCES

[1] **https://www.undp.org/sustainable-development-goals**

[2] **Ma, N., Kochan, O., Jun, S., & Kochan, V.** (2014). Decreasing of thermocouple inhomogeneity impact on temperature measurement error. In 13th IMEKO TC10 Workshop on Technical Diagnostics 2014: Advanced Measurement Tools in Technical Diagnostics for Systems' Reliability and Safety (pp. 85-90).

[3] **Winiecki, W., & Bilski,** P. (2008, May). Multi-core programming approach in the real-time virtual instrumentation. In 2008 IEEE Instrumentation and Measurement Technology Conference (pp. 1031-1036).

[4] **Hu, Z., Kochan, R., Klym, H., Kochan, O, & Jun, S.** (2014). Method of integral nonlinearity testing and correction of multi-range ADC by direct measurement of output voltages of multi-resistors divider. In 13th IMEKO TC10 Workshop on Technical Diagnostics 2014: Advanced Measurement Tools in Technical Diagnostics for Systems' Reliability and Safety (pp. 56-60).

[5] **Mizusawa, T., Miura, S., Taketa, T., & Hiranaka, Y.** Distributed Power Usage Control and Estimation of Total Demand. 15th IMEKO TC10 Workshop on Technical Diagnostics 2017 - "Technical Diagnostics in Cyber-Physical Era" P. 96-99

[6] **Hu, Z., Odarchenko, R., Gnatyuk, S., Zaliskyi, M., Chaplits, A., Bondar, S., & Borovik, V.** (2020). Statistical Techniques for Detecting Cyberattacks on Computer Networks Based on an Analysis of Abnormal Traffic Behavior. International Journal of Computer Network & Information Security, 12(6), 1-13

[7] **Xu, H., Przystupa, K., Fang, C., Marciniak, A., Kochan, O., & Beshley, M.** (2020). A combination strategy of feature selection based on an integrated optimization algorithm and weighted k-nearest neighbor to improve the performance of network intrusion detection. *Electronics*, *9*(8), 1206.

[8] **Shah, S. A., Seker, D. Z., Rathore, M. M., Hameed, S., Ben, S., DraheimY. D.**: Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?, *IEEE Access*, Vol. 7, 2019, pp. 91885-91903, doi: 10.1109/ACCESS.2019.2928233.

[9] **Zhang, P., Sun, H., Situ, J., Jiang,C., Xie, D.**: Federated Transfer Learning for IIoT Devices With Low Computing Power Based on Blockchain and Edge Computing, I*EEE Access*, 2021. doi:10.1109/access.2021.3095078

[10] **Saha, H. N. et al.**: Disaster management using Internet of Things, *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, 2017, pp. 81-85, doi: 10.1109/IEMECON.2017.8079566.

[11] **Xu, X., Zhang, L., Sotiriadis, S., Asimakopoulou, E., Li, M., Bessis,N.**: CLOTHO: A Large-Scale Internet of Things-Based Crowd Evacuation Planning System for Disaster Management, in I*EEE Internet of Things Journal*, Vol. 5, No. 5, 2018, pp. 3559-3568. doi: 10.1109/JIOT.2018.2818885.

[12] **Klymash, M., Savchuk, R., Pozdnyakov, P., Beshley, M.**: The researching and modeling of structures of mobile networks for providing of multiservice radio access, *Proceedings of International Conference on Modern Problem of Radio Engineering, Telecommunications and Computer Science*, 2012, pp. 281-282.

[13] **Porambage, P., Okwuibe, J., Liyanage, M., Ylianttila,M., Taleb, T.**: Survey on Multi-Access Edge Computing for Internet of Things Realization, in *IEEE Communications Surveys & Tutorials*, Vol. 20, No. 4, 2018, pp. 2961-2991, doi: 10.1109/COMST.2018.2849509.

[14] **Klymash, M., Seliuchenko, M., Beshley, M., Redchuk, S.**: Increasing wavelengths utilization efficiency in OTNoDWDM network based on local resource distribution method, *2015 Second International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T)*, 2015, pp. 157-160, doi: 10.1109/INFOCOMMST.2015.735730.

[15] **Beshley, M., Pryslupskyi, A., Panchenko,O., Beshley, H.**: SDN/Cloud Solutions for Intent-Based Networking, *2019 3rd International Conference on Advanced Information and Communications Technologies (AICT)*, 2019, pp. 22-25, doi: 10.1109/AIACT.2019.8847731.

[16] **Jun, S., Kochan, O., & Kochan, R.** (2016). Thermocouples with built-in self-testing. *International journal of thermophysics*, *37*(4), 1-9.

[17] **Karimipour, H., Dinavahi, V.**: On false data injection attack against dynamic state estimation on smart power grids, *2017 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, 2017, pp. 388-393, doi: 10.1109/SEGE.2017.8052831.

[18] **Fan, L., Li, J., Pan, Y., Wang, S., Yan, C., Yao, D.**: Research and Application of Smart Grid Early Warning Decision Platform Based on Big Data Analysis, *2019 4th International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, 2019, pp. 645-648, doi: 10.1109/IGBSG.2019.8886291.

[19] **Banafa, A.**: *2 The Industrial Internet of Things (IIoT): Challenges, Requirements and Benefits," in Secure and Smart Internet of Things (IoT): Using Blockchain and AI* , River Publishers, 2018, pp.7-12.

[20] **Zhou, Z., Song, Y., Xiang, P., Fang, S.**: Research on Improving Intelligent Inspection Efficiency of Substation Based on Big Data Analysis, *2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, 2020, pp. 99-102, doi: 10.1109/ACIRS49895.2020.9162602.

[21] **Di, C., Leye, W., Lianzhi F., Junxue, Z., Kai, C.,Qiang, Y.**: Federated Singular Vector Decomposition, 2021.