Operating an I-GMM-based data preparation approach on a cloud platform

Dr. A. Avani

Assistant Professor, Department of CSE, Anu Bose Institute of Technology, Paloncha, India

avanialla@gmail.com

ABSTRACT

The amount of information that is kept in the sustainable energy cloud platform may be significantly improved via the use of pre-processing, which in turn makes the process of data analysis more effective. In this research, an updated GMM model is used to build a technique for the preparation of data pertaining to power grids. The solution is based upon the distributed architecture of cloud computing. This method starts with building a smart grid cloud platform on top of the cloud storage architecture. After that, you should execute feature extraction on the grid data that has been saved on the cloud platform. In order to finish the preparation of the data, the upgraded GMM model is employed to both categorise and preserve raw grid data. This brings the whole process full circle. Experiments have been conducted to prove, in the end, that this strategy is effective.

I. INTRODUCTION

The energy demand in a number of different places is expanding at a quick rate as a direct result of the rapid expansion of economic development. During the same time period, the growth and maturation of the Internet of Things technology has also led to its widespread use in the electrical grid. Because of all of these factors, the data included in the smart grid indicates a meteoric rise [1][2]. However, these data types are diverse [3,] the data sources are complex [4,] and there are also issues such as missing data, abnormal data, and inconsistent formats [5,] which cause significant disruption to power grid companies' data gathering and analysis. [As an example:] [As an example:] [As an example:] [As an example:] [Therefore, preparing the data in the smart grid, and then categorising and storing it, may significantly increase the data's overall quality while also reducing the amount of computation time and storage space that is necessary for data analysis. These have the potential to lower the costs associated with running the grid.

According to various processing aims, data preparation may be split into data cleaning, data integration, and data categorization, et cetera. The primary goal of data cleaning is to get rid of any errors, such as missing data, noise pollution, inconsistent formatting, and so on. For instance, Lin et al. [6] find a solution to the issue of anomalous converter status data by using a data cleaning method that is based on cluster analysis analysis and neural networks; and Lv et al. [7] propose a smart grid incomplete data verifying algorithm that is based on machine learning. Both of these methods are able to solve the problem. The primary goal of data integration is to transform information obtained from a wide variety of data sources into a form that can be stored consistently. For instance, Li et al. [8] presented a wide-area distributed power quality data integration architecture for power grids. This design would make use of distributed data integration technology as well as wide-area distributed backup and recovery technology. [Citation needed] The process of distinguishing data based on its origin, format, and application

is known as data classification. In order to categorise the power load, for instance, Xiao et al. [9] used the LIBSVM classifier that was enhanced by grid search. Each of these strategies deals with the data contained inside the grid in order to accomplish a certain goal. Smart grids that have already made a lot of use of the Internet of Things and cloud computing, on the other hand, need a strategy for preparing data that includes cleaning, summarizing, and categorising the data and is compatible with smart grid cloud platforms.

In light of this, the study presents a technique for the pre-processing of grid data that is based on the distributed architecture of cloud computing and makes use of an upgraded version of the Gaussian Mixture Model. This approach begins by constructing a smart grid cloud platform using the cloud computing architecture. After that, the data that is saved on the smart grid cloud platform is subjected to data cleaning procedures. Finally, the upgraded GMM model is used to categorise and store the grid data, which significantly enhances the quality of the data and makes future grid data mining and analysis in the virtualized environment much easier to do.

II. PREPARATION

A. Computing in the Cloud

The power grid has long made use of computer technology for the purposes of operation monitoring and simulation computation. This practise has been going on for quite some time. A straightforward parallel computing platform may once have been adequate for meeting the requirements of smart grids, but this is no longer the case as a result of ongoing growth in the size of the power grid, the growth of interconnections between regions, and the improvement in the production level of power grid equipment. The super-large size, great scalability, cheap cost, and automation that come with cloud computing make it an excellent choice for meeting the development requirements of the power grid at present [10]. The use of cloud computing has become widespread across a variety of sectors. The creation of a cloud computing centre has the potential to provide smart grids with all facets and degrees of service. Through the organisation of a unified software platform for the power grid, it is possible to manage and maintain the model data of something like the overall infrastructure in a uniform manner. Furthermore, through the distributed computing function of the cloud platform, it is possible to simulate and analyse large amounts of data.

The architecture of cloud computing is primarily composed of two components known as service and management. The purpose of the service component is primarily to provide customers with a variety of cloud-based services, which may be broken down into three levels: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) [11]. [Citation needed] Cloud management is the primary focus of the management section. It serves as the basis for the three-tiered cloud service, which is composed of a user layer, a mechanism layer, and a detection layer.

The purpose of this study is to provide a cloud platform for smart grids that is based on the architecture of cloud computing. In addition, there are four alternative cloud models available in order to cater to the requirements of various users: the public cloud, the private cloud, the hybrid cloud, and the industrial cloud [12]. By using the private cloud, which works behind the power company's firewall and is not connected to the internet, to build a cloud platform within the power grid, the power sector is protected from possible threats.

B. GMMA's

time-honoured approach to data analysis is called the Gaussian mixture model, or GMM for short. The Gaussian model postulates that the same category of data will be roughly distributed throughout a space with a lot of dimensions. As a consequence of this, a single Gaussian density distribution function is all that is required to adequately characterise the distribution of a certain category of data. A class may be described using the model. The Gaussian mixture model evolved as a solution for the data distribution showing a no ellipsoidal distribution. This model mixes numerous Gaussian density distribution functions, as illustrated in the following formula:

$$P^X = \sum_{i=1}^{n} w_i P_i^X \qquad (1)$$

While wi is really the weighting of the i-th Gaussian model, ,, X P F X x I is the i-th Gaussian density distribution function, and n is the number of Gaussian models. n refers to the number of Gaussian models. Finding X Pi requires first locating the mean x and then the covariance matrix. This is the first step in the procedure. As a consequence, in order to resolve the Gaussian mixture model, one must first solve the equations wi, x, and. The answer that one gets from doing so is the probability that the test data belongs to each category.

The expected maximum algorithm, often known as the EM algorithm, is the approach that is most frequently used for the solution of the parameters described above. These are the steps that make up the algorithm.

For the data set X x x x 1 2,,..., m, the number of data types is n

1) For the data set I x X, initialise wj, x, and m, where j n may be any number from 1 to m. m is the total number of data types.

2) Proceed to Step E: The posterior probability that corresponds to j is calculated as follows:

$$\wp_i^j = \frac{w_j P_i^X}{\sum_{j=1}^{n} w_j P_i^X} \qquad (2)$$

3) Proceed to Step M and update wj, x, and :

$$w_j = \frac{\sum_{i=1}^{n} \wp_i^j}{n} \qquad (3)$$

$$\bar{x} = \frac{\sum_{i=1}^{n} \bar{x} \wp_i^j}{\sum_{i=1}^{n} \wp_i^j} \qquad (4)$$

$$\Re = \frac{\sum_{i=1}^{n} \wp_i^j \left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T}{\sum_{i=1}^{n} \wp_i^j} \qquad (5)$$

Continue to iterate through steps E and M until you reach the best possible option.

III. EXPERIMENT ANALYSIS

This study verifies the effectiveness of the GMM method by using it as a benchmark in a series of experiments that compare it against a more conventional version. The experimental data is obtained from the data that is held on the software platform of a province power grid. After the data has been gathered, it is partitioned into data sets of 100, 500, 1000, and 2000 separately for the purpose of experimental tests. As can be seen in the table that follows, the I-GMM method that was introduced in this research has superior

classifier performance than that of the conventional GMM technique.

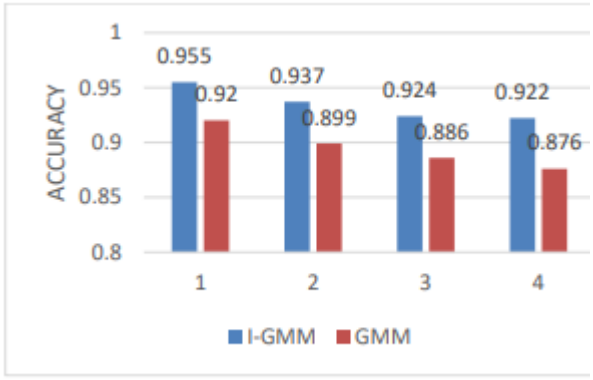

Figure 3 is a chart that compares the accuracy rates of the I-GMM with the GMM.

IV. CONCLUSIONS

The data that is stored in the smart grid needs to be pre-processed in order to improve both the quality of the data that is collected and stored by the smart grid and the speed with which data analysis is done. platform in the cloud on the basis of the decentralised infrastructure of given the context of cloud computing, this work makes use of the enhanced GMM paradigm for the development of a data pre-processing approach for power grids. This approach will begin by constructing a cloud-based smart grid platform. about the infrastructure of cloud computing. After that, conduct some data cleaning up on the grid data that has been saved on the cloud platform. In conclusion, the enhanced GMM model is used in order to categorise and storing the grid data is necessary in order to finish the pre-processing of the data. In conclusion, the usefulness of this approach is shown via experiment.

REFERENCES

[1] A. Barua, D. Muthirayan, P. P. Khargonekar and M. A. Al Faruque, "Hierarchical Temporal Memory Based Machine Learning for Real-Time, Unsupervised Anomaly Detection in Smart Grid: WiP Abstract," 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS), Sydney, Australia, 2020, pp. 188-189.

[2] B. Zhao, K. Fan, W. You, K. Yang, Z. Wang and H. Li, "A Weightbased k-prototypes Algorithm for Anomaly Detection in Smart Grid," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020, pp. 1-6.

[3] W. Mao, X. Cao, Q. zhou, T. Yan and Y. Zhang, "Anomaly Detection for Power Consumption Data based on Isolated Forest," 2018 International Conference on Power System Technology (POWERCON), Guangzhou, 2018, pp. 4169-4174.

[4] C. Li, H. Jiang and Q. Ge, "Power Data Cleaning in Micro Grid," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 3776-3781.

[5] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu and L. Zhu, "A missing power data filling method based on improved random forest algorithm," in Chinese Journal of Electrical Engineering, vol. 5, no. 4, pp. 33-39, Dec. 2019.

[6] J. Lin, G. Sheng, Y. Yan, Q. Zhang and X. Jiang, "Online Monitoring Data Cleaning of Transformer Considering Time Series Correlation," 2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Denver, CO, 2018, pp. 1-9, doi: 10.1109/TDC.2018.8440521.

[7] Z. Lv, W. Deng, Z. Zhang, N. Guo and G. Yan, "A Data Fusion and Data Cleaning System for Smart Grids Big Data," 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking

(ISPA/BDCloud/SocialCom/SustainCom), Xiamen, China, 2019, pp. 802-807.

[8] L. Jin et al., "Research on Wide-area Distributed Power Quality Data Fusion Technology of Power Grid," 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 2019, pp. 185-188.

[9] X. Qi, X. Wu, Y. Ji, X. Wang and H. Li, "Research on Classification of Power Load Data Based on LIBSVM," 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 2019, pp. 158-162.

[10] J. Shen, T. Zhou, D. He, Y. Zhang, X. Sun and Y. Xiang, "Block Design-Based Key Agreement for Group Data Sharing in Cloud Computing," in IEEE Transactions on Dependable and Secure Computing, vol. 16, no. 6, pp. 996-1010, 1 Nov.-Dec. 2019.

[11] Y. Jiugen and X. Ruonan, "Cloud Computing-based Big Data mining Connotation and Solution," 2020 15th International Conference on Computer Science & Education (ICCSE), Delft, Netherlands, 2020, pp. 245-248.

[12] A. Sun, G. Gao, T. Ji and X. Tu, "One Quantifiable Security Evaluation Model for Cloud Computing Platform," 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD), Lanzhou, 2018, pp. 197-201.

About Author

Dr. A. Avani obtained M. Tech degree from JNTU, Hyderabad, India. She is at present working as professor in Department of CSE of Anu Bose Institute of Technology, Paloncha, Telangana, India. Her area of interest is Data mining.