Exploration on the Construction of Sharing Platform of Power Big Data Application

Yang Jing[1], Zhang Quan, Gong Lihua, Zhu Longzhu, Liu Kunpeng
Customer Service Center, State Grid Corporation of China

Abstract. Big data application sharing supermarket is a big data application sharing portal. The first, one or more enterprise users can share and use the same big data application product. This product comes from the contribution of enterprise users of a certain sharing supermarket. In this way, we can solve the problem of unbalanced development of big data capabilities, and avoid repeated construction and repeated investment in the same kind of big data applications; the second. Through such a sharing platform, many enterprises carry out cooperation such as federal learning to jointly create big data application products.

Keywords. big data, platform, sharing, block-chain, federated learning

1. Introduction

At present, the application of big data has made great development and is widely used in various industries, Kuo, Y. H., & Kusiak, A.[1], The review of the literature suggests that production research enabled by data has shifted from that based on analytical models to data-driven. Manufacturing and data envelopment analysis have been the most popular application areas of data-driven methodologies. The research published to date indicates that data mining is becoming a dominant methodology in production research. Future trends and opportunities for data-driven production research are presented. Lu, Y., & Xu, X.[2],focus on how to ensure efficient synchronization of the mobile health record in unreliable mobile environments. The proposed middleware is a software-as-a-service (SaaS) layer which enables the mobile health record. Zhang, Y., Zhang, R., Wang, Y., Guo, H., Zhong, R. Y., Qu, T., & Li, Z.[3],the composition of big data service layer is described in detail, and a sales predicting method based on neural network is introduced. The salability of products is divided, and the qualitative economic production volume mechanism is finally given. Based on the framework, an intelligent service system for enterprises with the characteristics of mass production is implemented. Experimental results show that the big data service framework can support chemical enterprises to make decisions to reduce costs, and provides an effective method for Smart Product Service System (PSS).

Some experts introduce the application of big data technology and framework in manufacturing industry. A. Papacharalampopoulos, J. Stavridis, P. Stavropoulos, G. Chryssolouris[4],The current study addresses the challenges, associated with the framework of the thermal oriented processes, having holistic (digital) modelling as a

---

main objective. Herein two different case studies are performed; numerical examples regarding big data impact on manufacturing and simulation-based paradigms of control design taking into account communications. Implementation of the aforementioned takes into account the controller's complexity.D. Mourtzis, E. Vlachou, N. Milas, N. Tapoglou, J. Mehnen[5], proposes a Cloud-based, knowledge-enriched framework for machining efficiency based on machine tool monitoring. More precisely, it focuses on the optimization of the machining parameters and moves through an event-driven optimization algorithm, utilizing the existing machining knowledge captured by the monitoring system.

In China, there are many provincial companies under the State Grid Corporation of China. In recent years, they have produced a large number of big data application results based on power business scenarios. These big data application results belong to different provincial companies of State Grid Corporation, and have the promotional value in other internal units, such as distribution transformer overload influence factor mining and prediction service[6]-[7], optimized layout of electric vehicle charging facilities [8]-[9], electricity charge risk forecast and other big data analysis models [10]-[14].

An attempt is made in this paper to explore different issues such as, how to share big data applications faster, avoid repeated investment from the perspective of State Grid Corporation, shorten the development cycle from the perspective of power companies, and in turn save human and material resources. For example, if a provincial company A has the product of electricity stealing analysis then other provincial companies such as B-E can use the products of the provincial company A by mere sharing the platform of big data application. Therefore, the provincial companies B-E need not invest separately in power stealing analysis and thus, can save both manpower and precious material resources.
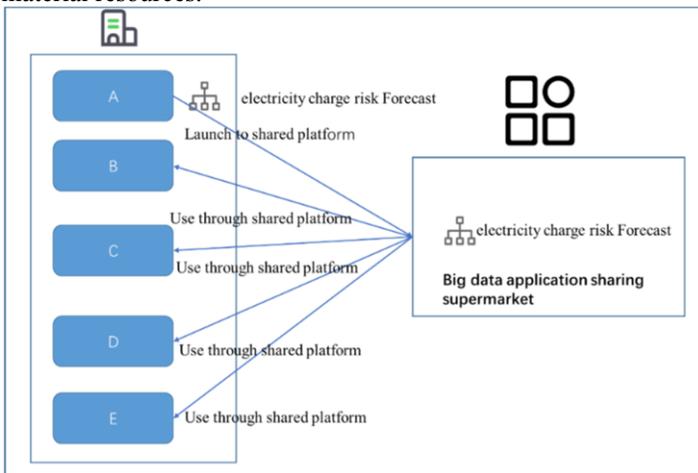

Figure 1 Big data application product sharing

At present, the power industry only realizes data level achievement sharing, but the sharing of model, algorithm, tool and product achievements have not yet been realized. The power big data platform is positioned as the energy big data sharing and operation service platform, focusing on the data sharing service function, data mining analysis function, data product development function and data product operation function [15]. The present paper focuses on the data sharing function of the platform where users can

develop new data products and excavate data value based on data sharing and platform capability. The data sharing is an important means to solve the unbalanced distribution of big data resources and realize its reasonable allocation [16]. We introduce a data sharing platform OpenEI created by non-governmental organizations in the United States, share their own energy related data and data modeling tools for different organizations and institutions on this website for other users to download, analyze and model. The data, on this data sharing platform, is mostly in the form of small examples, such as 24-hour power load data. In addition, the paper mentioned the design of power big data sharing platform, mainly from the management, service, exchange and monitoring of power data. The key point is the sharing of power data.

This paper explores the construction of big data application sharing supermarket to realize the sharing of big data application results. Big data application achievements include models, algorithms, tools, etc., but they are not just limited to themselves. But they are the application of big data technology in a certain business scenario, such as electricity tariff risk prediction, electricity theft analysis, optimized layout of electric vehicles, heavy overload of station area, etc.

2. Design of data application sharing platform

2.1 Platform positioning

The platform of the big data application sharing supermarket is positioned from the following four aspects. First, the big data product trading platform, which provides three types of big data products, namely, data services, tool models and applications. Second, the mutual benefit cooperation platform, which integrates and shares the data resources, technical resources and business resources of relevant parties between ecosystems and within the big data commercialization service ecosystems so as to promote the transaction matchmaking between supply and demand sides and multiple parties to achieve profit sharing and mutual benefits. Third, the innovation driven platform of big data application, which realizes online selection and assignment of big data topics, realizes objective evaluation of the value of big data results through the real feedback of the market and users, and conducts targeted promotion of excellent big data results according to key work points in relevant stage. And finally, the big data technology exchange platform which encompasses online technology exchange forum and offline training, special competition, achievement exchange, technology salon, etc., in which personnel training can be promoted and innovation ability is enhanced.

2.2 Division of cooperation relationship of big data application sharing supermarket

From the perspective of cooperation, the platform is divided into the supplier and demander of on-shelf products and the operator of the sharing supermarket. Suppliers of on-shelf products mainly refer to the enterprises or units that provide big data on-shelf products to the sharing supermarket, which can be subdivided into data service, tool model and product application suppliers according to the supply function. The demand side of the on-shelf products mainly refers to the enterprises or units that have actual business requirements for big data on-shelf products. The operator of the sharing supermarket is the operating unit entrusted by the company. Whereas, the internal units

mainly include directly affiliated units, provincial (municipal) companies, industrial units and scientific research institutes.

From the users' point of view, it can be divided into subscribers, publishers and operators as follows:

The subscriber can log in the big data application sharing supermarket as a member, order and use the required big data products through the sharing supermarket.

The publisher can log in the big data application sharing supermarket as a member, release its own products through the sharing supermarket, and promote the products.

The operator engages in the operation and management side of the big data application sharing supermarket.

Members of the big data application sharing supermarket can have multiple identities i. e., they can be both subscribers as well as publishers and can publish their own excellent products to other members to meet their own business needs.

2.3 On-shelf product design of big data application sharing supermarket

Based on the application characteristics, the supermarket products can be divided into data products, application products and tool products. The data products refer to the on-shelf products that provide data resources or data processing and desensitization result sets directly for users under the premise of compliance with laws and regulations. The application products refer to the big data applications with UI interfaces that can directly provide analysis and prediction results to the end users. The tool products refer to the big data model and algorithm, which can be used only when integrated into the program code or the training platform.

2.4 Internal business model design

Mainly the internal partners of the State Grid Corporation trade in the virtual mode of points. The initial score of each unit is 500 points. The on-shelf products are divided into 10—50 points according to the classification and scale. The suppliers of the on-shelf products can set such points by themselves with different charging standards by times, weeks, months and years. The demanders of the on-shelf products can use the points to purchase the products according to the different charging standards of the products. On successful completion of the transaction, corresponding fee of the buyer is deducted, the operating party of the sharing supermarket withdraws the service fee from each transaction at a rate of 6%, and the supplier of the on-shelf product gets the remaining 94% transaction fee.

2.5 Design of product sharing mode

Data products can be shared through downloading, internal data interface of the company, and federated learning, specifically as follows:

(1) The documents that do not involve sensitive detailed data (like the industry analysis report, general standard training set, etc.) can be directly shared through downloading and internal interface of the company.

(2) The data involving sensitive information is shared in the way of model training for the demander. The specific implementation method is that when the demander, who needs the data training model, uploads his model to the sharing platform, the data side downloads the model and uses its own data for the training. After completion of the

training, only the trained model is provided to the data demander, which can realize data sharing thereby, avoiding the sensitive data leakage.

Application products are remotely shared and used, where users directly use big data application products by uploading data or processing data in the middle platform. The application products involve specific business scenarios, including data processing, integration, calculation, analysis and display, with strong business context. These are shared and used remotely. The user needs to process the data according to the application product instructions.

Tool products are shared and used by downloading or remote calling, and users can directly use big data application products by uploading data or processing data in the middle platform. The demand side of tool products can download the model for direct use, perform further training based on the current model, and remotely use the product after processing according to the tool product instructions.

3. Design of big data application sharing platform

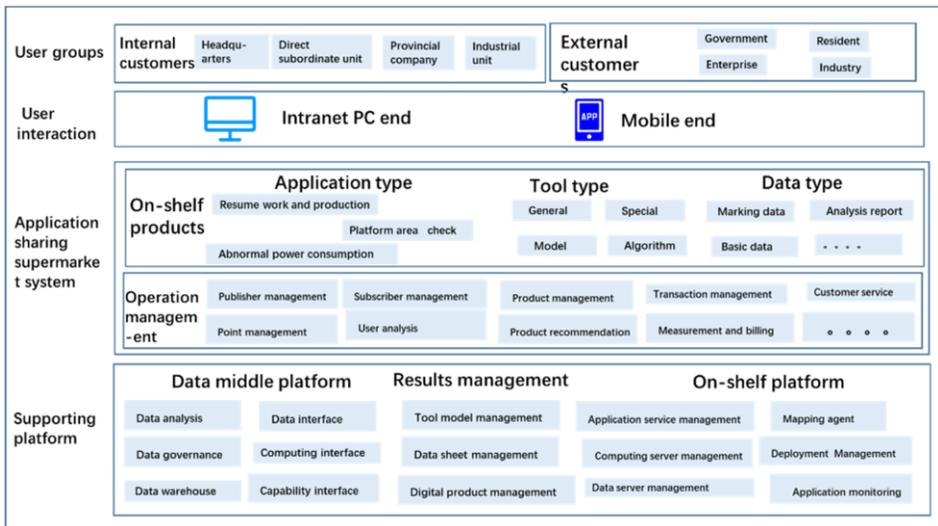### 3.1 Business framework of big data application sharing supermarket

Figure 2 Business framework of big data application sharing supermarket

The business architecture of shared supermarket is divided into four parts: supporting platform, application sharing supermarket system, user interaction and user group.

The supporting platform includes data middle platform, achievement management function and on-shelf platform, which provides source data, data processing ability and product shelf management for the products on the shelves of shared supermarkets.
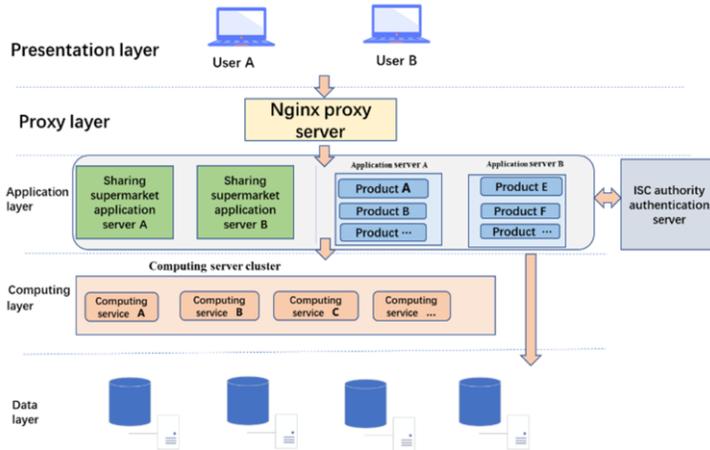
The application sharing supermarket system includes products on the shelf and operation management of the shared supermarket. Similar to online shopping mall, users can select and purchase big data application products, including application class, tool class and data class. Through the operation function, the management of on-shelf and off-shelf approval, product iterative update management, transaction management

and shared supermarket operation management include: user operation, product operation and transaction operation.

User interaction can be done by provincial company users through the web PC terminal and mobile app end for sharing supermarket.

The current users are the company headquarters, directly affiliated institutions and provincial companies, and the future users can further be expanded to the government, enterprises and residents.

3.2 Framework of arrangement of big data application sharing supermarket

Figure 3 Arrangement framework of big data application sharing supermarket

The deployment architecture of shared supermarket is divided into data layer, computing layer, application layer, agent layer and presentation layer.

The data layer includes business source data, index processing and result data. Source data and result data are stored in Oracle, and indicators are processed in hive Library of big data platform.

The calculation layer supports the model operation of the application products on the shelf, and outputs the results based on the intermediate result table of the data layer. In order to reduce the impact of data computing model on the application server, the application service is separated from the computing model as far as possible, and a unified computing server is planned. Virtualenv or Anaconda is installed on the computing server, and the python environment for different application products is established, and the application service makes remote calls.

The application layer includes the application of shared supermarket and the application service of products on the shelf, including web application service, middleware service and message service.

The agent layer uses nginx proxy to map the servers and port addresses of shared supermarkets and products on shelves to nginx servers, so as to realize unified IP port access and reduce the security risks brought by the opening permissions of IP ports of all servers.

The presentation layer is open to users through the web page scheme. Users log in to the sharing supermarket through the web page to order and use the products.

3.3 Property rights protection of big data application products

The sharing supermarket protects the property rights of big data application products through blockchain technology. JI Lusheng [17] proposed an anonymous address technology scheme based on block chain to use resource services to process encrypted addresses for personal data rather than sending them directly to the third-party services. WANG Jingwei [18] proposed a blockchain based data marketplace framework demonstration system to verify its feasibility and security, and providing reference solutions for building practical data marketplaces in the future. YAN Shu [19] introduced the characteristics of blockchain technology to expound the main ideas of transforming the authorized storage link, data traceability and the realization of smart contract by using blockchain to comb the overall framework of data circulation. In view of this, an attempt is made in this work to give the application examples of data circulation using blockchain in foreign countries, and introduce some other numbers according to the circulation of new technology.

In the blockchain mode, the complete authorization and authentication process of big data application sharing supermarket is shown in Figure 4. The put-on-shelf side signs the electronic agreement and authorizes corresponding authority to share supermarket. Firstly, the shared supermarket, stores the evidence locally through the application system, and then uploads the authorization information to the authorization information chain. The application system executes the code on the chain, initiates the query on the chain, and records the authorization information to the block. When the demand side submits the product requirements, the authentication transaction is initiated on the chain to confirm the validation of the authorized user. Then, the verification node on the chain returns the authorization information. If it is authorized, corresponding big data product will be returned. The blockchain model avoids the defects of the traditional model. Any node can record the authorization information which cannot be changed. Multiple sides can share the authorization records in real time, so the query efficiency is high. In addition, the authorization is decoupled from the business, and can be added and exited at any time.
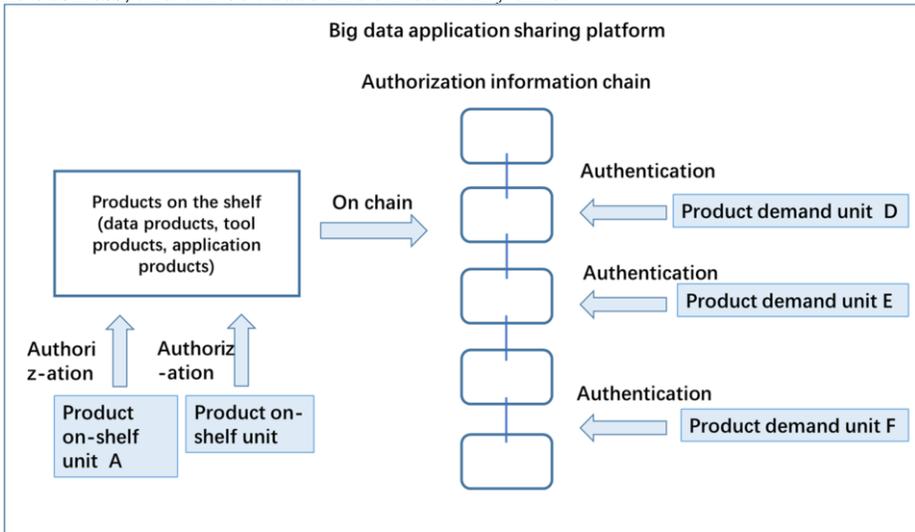

Figure 4 Big data application sharing supported by blockchain

Following three aspects are explored in this paper:

1) Blockchain-based data product certification: The blockchain characteristics like distributed storage, multi-node consensus synchronization and hash encryption storage of files are used to provide certification for data products released by the product publisher to prevent tampering, subjective change, malicious deletion and other operations brought by the network security threats, and effectively guarantee the reliability and authenticity of network security data.

2) Blockchain-based tool and model certification: The blockchain characteristics like distributed storage, multi-node consensus synchronization and hash encryption storage of the files are used to provide certification for the models and tools released by the product publisher. For other users, certification is provided for the results of continuing training based on the current model. The intellectual property rights and sharing of partners in federated learning are recorded.

3) Certification for process behavior involved in the purchase and use of products: The process and behavior involved in the order and use are recorded in the chain, so as to ensure the authenticity and traceability of detailed information such as the time and content of the behavior of each side, and prompt handling of abnormal behavior, if any, and other information.

The work is carried out in the following order:

1) Data, tool, application and other product certification: The platform is a blockchain-based electronic data certification system. The system makes full use of the characteristics of the blockchain like decentralization and non-tampering, anchoring the product information on the blockchain, controlling the access of different users to the electronic data, and effectively solving the security problem of certification storage of the electronic data. It provides users with services like data upload, download, query, use and authorization. At the same time, distributed storage technology and access control mechanism are introduced to increase the reliability of the system.

2) Certification for ordering and using process of sharing supermarket products: For products on the shelf in this system, the enterprise users or individual users can register as system users of the platform. After successful registration, users must first conduct real name authentication, and then they can provide certification for the electronic data in cloud forensics system. Users can put all kinds of big data application products on the shelves. In order to protect the privacy of data and tool products, users can choose whether to encrypt or not.

3.4 Joint construction of big data application products based on Federated learning

Federated learning is a new basic technology of artificial intelligence. Google first proposed the concept of federated learning in 2016 [20]. Its main idea is to build a machine learning model based on data sets on multiple devices and preventing data leakage at the same time. Subsequently, scholars have further studied more secured and personalized federated learning mechanism, and optimized the data distribution imbalance, user management mechanism and other directions.

The big data application sharing supermarket brings together the internal units to update the model through federated learning. The model publisher publishes the model that needs further training to the sharing supermarket for collecting federated learning partners. After downloading the model, the partner uses the data of the unit to optimize and improve the model, and then uploads it. The updated model is immediately aggregated with the updates of other users to improve the sharing model.

The federated learning zone is built for users based on the following three aspects:

(1) Secure multi-party computation: The security model involves multiple participants, and each participant needs to provide the security evidence of the simulation platform to ensure that each participant cannot obtain additional information except input and output. However, the ideal model needs a complex computing framework, and the effectiveness of its implementation is limited. In some scenarios, some knowledge disclosure can be considered and data disclosure with security guarantee can be accepted.

(2) Differential privacy: The differential privacy or anonymous technology is used to protect data privacy and conceals some sensitive attributes by adding noise to the data or using generalization method, so that the third party cannot distinguish the subjects. However, the specific implementation still requires data transmission to other places, which involves data authenticity and privacy issues.

(3) Homomorphic encryption: Privacy protection is realized by parameter exchange of machine learning under the encryption mechanism. The data and models themselves are not transmitted.
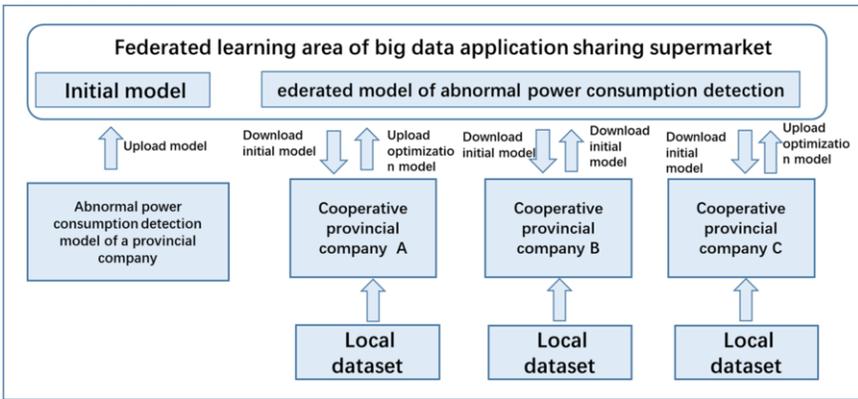

Figure 5 Federated learning zone of big data application sharing supermarket

Taking the forward propagation method of convolutional neural network as an example,
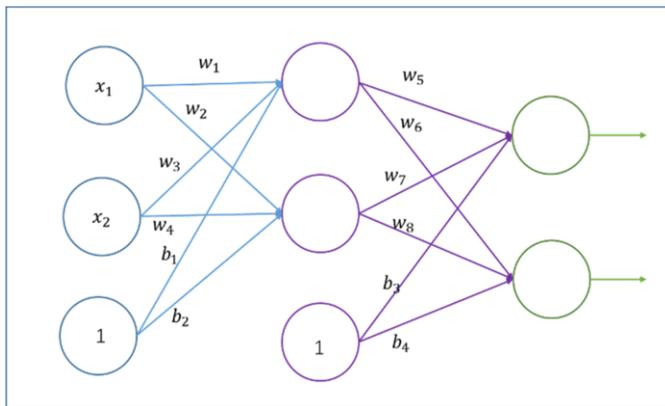

Figure 6 Schematic diagram of three-layer neural network

The communication scheme of federal learning privacy protection proposed in reference [21] is adopted; different TLS / SSL security channels are established between trainers and the parameter servers. The classical Advanced Encryption Standard (AES) encryption algorithm is used among the trainees to encrypt the public key PK and the private key sk of homomorphic encryption algorithm, the structural parameters of deep neural network model, the selected optimization algorithm and its learning rate.

If t is the number of iteration steps and t=1, the provincial companies get the provincial model gradient based on the local data, $J(\omega)$ is loss function, and ω is the parameter, then Gradient encryption is given by,

$$g(u_i, t) = \frac{\partial \ (J(\omega))}{\partial \omega} \tag{1}$$

H is a homomorphic encryption function

$$G(u_i, t) = H(g(u_i, t)) \tag{2}$$

Upload to the server of federal learning management system

The server carries out the weighted average:

$$\overline{G}_t = \sum_{i=1}^{n} p_i G(u_i, t) \tag{3}$$

The server sends it to the provincial company, and the provincial company gets the Decryption gradient, and D is the decryption function:

$$g(u_i, t + 1) = D(\overline{G}_t) \tag{4}$$

Proceed to the next iteration

$$t = t + 1 \tag{5}$$

Effect of federal learning

Table 1 Model effect

Model effect			
Model	**Sample size**	**Prediction accuracy (%)**	**Recall (%)**
Provincial companies independently identify the electricity theft (Neural network)	10000 (avg)	55.5(avg)	57.8 (avg)
Provincial companies use the federal learning model (Neural network)	10000 (avg) 50000 (avg)	80.6 (avg)	64.3 (avg)

4. Application effect

Table 2 shows the application of products and provincial usage on the shelves in the sharing supermarket.

Table 2 Big data products and usage

Big data products and usage			
ID	**Product name**	**Number of provincial companies using products**	**Saving investment**
1	Electricity stealing identification	20	19
2	Risk prediction of electricity charge	20	19
3	Layout optimization of electric vehicles	10	9
4	Customer service panoramic view	5	4
5	Efficiency evaluation of business hall	16	15
6	Customer portrait	27	26
7	Identification of abnormal power consumption	29	28
8	Overload prediction of distribution transformer	10	9
9	Customer service efficiency analysis	5	4
10	Power index monitoring for resumption of work and production	27	26
11	Power marketing data service sharing	20	19
12	Intelligent identification of abnormal electricity price	15	14
13	Traffic prediction analysis	6	5

5. Conclusions

Big data application sharing supermarket is an effective means of big data application sharing. In power application, it can achieve the purpose of providing services to many provincial companies by matured products of any provincial company. At the same time, as a sharing platform, it can help provincial companies to create new big data application products through federal learning.

6. Future work

At present, the big data application sharing supermarket is used by the internal users of the State Grid of China. How to cooperate with the external units and build the big data application sharing ecosystem are the key topics to be researched in the future.

At present, the planning for the external units is as follows: Firstly, leading Internet enterprises such as Alibaba, Huawei, Baidu and Tencent mainly rely on their advanced Internet thinking, creation, technical capabilities, large ecosystem, and reliable service resource capacity; secondly, the external research institutions mostly rely on their extensive research on theoretical basis, scientific research strength and talent advantages in their professional fields; thirdly, the government agencies, enterprises and institutions have actual demand for power data and application scenarios that can be applied to the industrial data.

The external key business planning of big data application sharing supermarket includes: first to carry out the transaction of big data on-shelf products with the support of supermarket construction; second, to understand the value-added realization of data through transaction and settlement; third, to gather partners, suppliers and demanders through sharing supermarket, and make use of platform advantages to facilitate multilateral transactions and product crowdfunding to enable transaction growth and enhance value-added realization of the data; fourth, to carry out the operation of sharing supermarket, expand the data commercialization service ecosystem, attract more ecological partners to participate, enrich the on-shelf products, and promote big data innovation by trading; and fifth, to carry out optimization iteration of sharing supermarket, constantly expand functions, consolidate data service capacity, and improve data security.

References

[1] Kuo, Y. H., & Kusiak, A. From data to big data in production research: the past and future trends. International Journal of Production Research.2019, 57(15-16):4828-4853.
[2] Lu, Y., & Xu, X. Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services. Robotics and Computer-Integrated Manufacturing.2019,57:92-102.
[3] Zhang, Y., Zhang, R., Wang, Y., Guo, H., Zhong, R. Y., Qu, T., & Li, Z. Big data driven decision-making for batch-based production systems. Procedia CIRP.2019,83:814-818.
[4] A. Papacharalampopoulos, J. Stavridis, P. Stavropoulos, G. Chryssolouris, Cloud-based Control of Thermal Based Manufacturing Processes. Procedia CIRP, 5th CIRP Global Web Conference-Research and Innovation for Future Production .2016, 55:254-259.
[5] D. Mourtzis, E. Vlachou, N. Milas, N. Tapoglou, J. Mehnen, A cloud-based knowledge enriched framework for machining efficiency based on machine tool monitoring, Journal of Engineering Manufacturing, IMECHE Part B, Accepted for publication.2017,233:278-292.

[6]   Qiu xiaogeng, Dong Xiangyu, Zhang Peng, Liang Wei, Guo Jun, Bai Kaifeng, Liang Zhixian, Feng Chao. Analysis of distribution transformer overload warning based on big data [J]. POWER SYSTEMS AND BIG DATA 2018, 21(10):38-42.

[7]   He Jianzhang, Wang Haibo, Ji Zhixiang, Meng Xiangjun, Zhang Tao. Heavy overload prediction of distribution transformer based on Stochastic Forest theory [J]．Power System Technology, 2018, 21(10):38-42.

[8]   Xie Yuande, Zhang Lin, Deng Shali, he jiemeng. Study on optimization network layout of electric vehicle charging facilities [J]. Practice and understanding of Mathematics, 2020,50(10):168-176.

[9]   Sun Yue, Jiang Cheng, Wang Zhizhi, Tang chunsen. Optimal layout of dynamic wireless power supply system for electric vehicles based on PSGA [J]. Power system automation, 2019, 43(9):125-131.

[10] Li Jianbin, Wu Binbin, Zhu Yakui, Wang Yue, and Yagang, Zhao Shasha. Risk prediction and control of customer electricity tariff based on big data analysis [J]. Power Systems and Big Data ,2019, 22 (2):1-6.

[11] He Rong, Zhang Xiangdong, Qiu Lin, Chen libing, Zhou Qian, Zhang Yan. Electricity tariff risk prevention and control based on clustering and logistic regression [J]. Power Systems and Big Data, 2019, 22(12): 42-49.

[12] Zhao Hong, Shen Jianzhong, Wang Jun, Zhang Cheng, Qu Qing. Risk prediction model of electricity charge recovery based on customer portrait and machine learning algorithm and its application [J]. Microcomputer Applications, 2020, 36(2):93-96.

[13] Wang Zongwei, Zhao Guoqi, Jin Peng, Yang Jing, Wang Hailong, Zhang Zuobin. Research on the risk identification model of electricity tariff based on customer credit system [J].Machine Design and Manufacturing Engineering , 2019, 48(10): 99-104.

[14] Xie Ying, Wang Zheng, Zhao Yongliang. Research on risk identification model of customer electricity consumption behavior [J]. Zhejiang Electric Power, 2017,36(12):53-56.

[15] Li J, Xu Z P, Zhou C L, Kang C T. Application architecture design of energy big data sharing and operation service platform [J]. Power and Energy, 2019, 40 (5): 568-571.

[16]   Qu H N, Pang X W, et al. Application value of power big data and analysis and design of its sharing platform. Management and Administration[J], 2017 (7):104-108.

[17] JI Lusheng , ZHANG Guiling ,YANG Jiarun. The protection scheme of personal data under the chain based on block chain, Computer Engineering[J], 2020,https://doi.org/10.19678/j.issn.1000-3428.00.EC0057024.

[18] WANG Jingwei, ZHENG Zhenzhe, WU Fan, CHEN Guihai. Block chain based data marketplaces[J]. Big Data Research. http://kns.cnki.net/kcms/detail/10.1321.g2.20200311.1050.004.html.

[19] YAN Shu, QING Sude, WEI Kai. Application of block chain in data circulation[J]. BIG DATA RESEARCH.2018 (1):3-12.

[20] McMahan H B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) .2017. JMLR: W&CP volume 54:1273-1282.

[21] Phong L T, Aono Y, Hayashi T, Wang L H.Privacy-preserving deep learning via additively homomorphic encryption. IEEE Transactions on Information Forensics and Security. 2017, 13(5):1333-1345.