Analysis on Big Data Frame Design and Key Technology Application of Veterinary Drug Supervision

Shuqing Han¹, Liwei Xing¹, Jing Zhang¹, Hongyu Zhang¹ and Jianzhai Wu¹*

¹Agricultural Information Institute, Chinese Academy of Agricultural Sciences/Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, 100081, China

*Corresponding author’s e-mail: wujianzhai@caas.cn

Abstract. This paper introduces the construction technology of veterinary drug big data framework, veterinary drug big data collection and processing technology, veterinary drug big data storage and calculation technology, veterinary drug big data sharing and exchange technology and veterinary drug big data multi-dimensional analysis technology, and expounds the significance of big data in veterinary drug supervision work, such as whole industry chain information aggregation of veterinary drug big data, visualization of flow and direction of veterinary drugs, traceability of veterinary drug safety and prediction of animal diseases. The application of big data in veterinary drug supervision promotes standardization, intellectualization, refinement and precision of veterinary drug production and management, promotes the development of agricultural modernization in China, and inserts the wings of science and technology into the construction of agricultural modernization.

1. Introduction

With the rapid development of the Internet, Internet of Things and cloud computing technology, the impact of massive data on human society is increasing. In the meantime, as a new natural resource [1], data need to be used reasonably, efficiently and fully in order to bring more benefits and value to people's life and work [2]. In order to transform data into knowledge, data warehouse, data mining and other technologies have been widely used to process and analyze data. They can not only provide reference for relevant organizations to make decisions, but also help to improve their decision-making ability, efficiency and accuracy. However, with the increase of data amount and the complexity of data structure, the commonly used data mining technology cannot meet the needs, which gives "big data" a deeper meaning. Compared with data mining, the application of big data is not only the difference of data magnitude, but also the improvement of data application level. Big data not only deals with the traditional regular data, but also analyses the disordered unstructured data, from which acquires the intrinsic knowledge of data, so as to provide support for prediction, planning and decision making in related fields.

With the continuous development of veterinary drug informationization, the construction of veterinary drug information systems, such as national veterinary drug product traceability system [3], national veterinary drug basic information inquiry system, provincial veterinary drug supervision platform, veterinary drug laboratory management system, veterinary drug enterprise management system, have provided vast amounts of data and resources for veterinary drug industry, and laid the foundation for the application of veterinary drug big data. The technology of veterinary drug database construction, veterinary drug mass data storage and parallel computing processing structure were
summarized in this paper, in a bid to provide ideas for further exploring the application and development of big data in veterinary drug supervision in China.

2. **Big data frame design of veterinary drug supervision**

2.1. **Analysis of big data characteristics of veterinary drug supervision**

Veterinary drug big data includes not only veterinary drug research and development, production, operation and supervision operation data, but also breeding, slaughter, quality inspection, epidemic prevention, feed, crops and other related data, as well as hot information on the network. There are not only structured data, semi-structured data, but also unstructured data; both static historical data and dynamic real-time data; both spatial distribution data and time series data. Establishment of data resource center throughout the whole process of veterinary drugs, application of distributed data storage method and parallel computing theory, design of data storage cluster technology and veterinary drug massive data processing technology architecture, research and develop veterinary drug big data construction technology with massive data, multi-source heterogeneous, whole industry chain characteristics, which is conducive to the realization of rapid extraction and calculation of dynamic information resources in the whole process of veterinary drug management.

2.2. **Big data organization and integration of veterinary drug supervision**

Based on the technology of big data storage and management, the whole process data of veterinary drug management are collected and integrated, and a comprehensive database of veterinary drug management facing big data is constructed. Massive spatio-temporal and attribute data from many complex operation processes, such as veterinary drug production environment, transportation process, storage environment, sales link and use process, are processed. It realizes the efficient management of structured, semi-structured and unstructured data of sensor data, map data, video image data, 2-D code data, location data and trajectory data with different collection and update frequencies. The database design mainly includes 6 aspects as follows:

- A database for monitoring the whole process of veterinary drugs. It mainly includes up to 100 sub-categories of over 10 categories of veterinary drug management monitoring data, including HACCP key control point monitoring data for veterinary drug production, GMP-based real-time dynamic sensing data for veterinary drug production key points, monitoring data for production operation specifications based on video analysis, veterinary drug 2-D code data, near infrared detection data and production number, production quantity monitoring data, coding data of veterinary drug monitoring, sensory data of veterinary drug circulation based on GSP, cold chain monitoring data of animal vaccine, monitoring data of veterinary drug circulation process, monitoring data of veterinary drug use process and data of web crawler, etc.
- Veterinary drug production and operation database; It mainly deals with the data of veterinary drug production and management enterprises, including the number, distribution, production scale, sales status and operation status of enterprises.
- Animal husbandry production and operation database; It mainly includes all kinds of data related to animal husbandry production, including the types, quantities, spatial distribution, production scale and production status of animal husbandry products.
- Basic geographic data bank; It mainly includes basic geographic data in the spatial distribution of veterinary drug production, transportation, storage, sale, circulation and use, including administrative divisions, POI and road network.
- Metadata; It mainly stores descriptions and codes of various tables and fields in the veterinary drug management database.
- Model and knowledge base; It mainly includes animal epidemic prediction, veterinary drug flow early warning, veterinary drug flow analysis and other models, production operation norms, production environment parameters, circulation process parameters and other expert knowledge.
3. Key technology analysis of big data framework for veterinary drug supervision

3.1. Veterinary drug big data collection and processing technology

The whole veterinary drug industry chain involves many data information sources, and the data has various formats and complex relationships. Therefore, it is necessary to sort out and aggregate the big data information of veterinary drug. The first is the information gathering technology for monitoring the whole process of veterinary drugs. In view of the whole process information links of veterinary drug production, management and use, the whole process monitoring information of veterinary drug is designed on the basis of data acquired by intelligent perception equipment in veterinary drug production, circulation and use process, taking the life cycle of veterinary drug as the main line, combining with supporting information such as animal husbandry, feed production, livestock slaughtering and network hotspots. Converging strategy forms a comprehensive data resource pool that runs through the whole chain of veterinary drugs. The second is the multi-source heterogeneous data processing technology of veterinary drugs. In view of the inconsistency, default, noise and high dimensionality of data resources, starting from the standardization and formatting of data, the rules of historical values of data from different sources are fully analyzed, and a multi-source heterogeneous data processing model of veterinary drugs is constructed based on the theory of cleaning, integration, transformation and protocol of big data. Standardization and normalization of data is realized in a bid to improve the quality of data. The third is the strategy of data different-location recovery for veterinary drugs. On the basis of Xinxiang base of Chinese Academy of Agricultural Sciences, in view of the consistency and synchronization of veterinary drug disaster preparedness data, this paper designs a database memory refresh mechanism, builds a network backup framework with good expandability based on agent confirmation mode, and realizes remote backup and timely recovery of veterinary drug data.

3.2. Veterinary drug big data storage and computing technology

The storage and calculation of information resources is an important support for data processing and analysis, and is also the key of veterinary drug management system. Its capability determines the operation effect of the whole system. (1) Distributed storage technology for mass data of veterinary drugs [5]; Combing and analyzing data types, data volume and data storage mode of veterinary drug and related industries, combining with data file management method of distributed storage, fully considering the expansibility and consistency of mass data storage, applying load balancing technology of metadata cluster and hot-spot file processing technology, designing veterinary drug information cloud storage source management framework, implements the whole process of veterinary drug administration and traceability of dynamic information resources management. (2) Massively parallel computing technology for veterinary drugs; analyzing and combing the basic data information and dynamic circulation information of veterinary drugs, calculating the storage capacity of distributed data of veterinary drugs, designing large-scale parallel computing function, combining with Spark parallel computing framework, constructing the parallel computing architecture of national veterinary drugs data, realizing the batch processing and flow processing capacity of massive data, and providing the underlying foundation for large-scale data analysis and processing.

3.3. Sharing and exchanging technology of veterinary drug big data

Veterinary drug data interaction and dynamic extraction is the core of effective monitoring of veterinary drug information. Supported by veterinary drug management information comprehensive service system structure based on the cloud platform, and based on Java EE platform, a B/S model veterinary drug data interaction and dynamic extraction system is developed. The first is data interaction and transformation mechanism. Based on XML technology, a multi-source heterogeneous big data conversion method for veterinary drugs is established. SDO data service object is used to develop a unified and standardized data interface, and a request-response mechanism is used to design a distributed data interaction mechanism to realize the dynamic interaction of big data for veterinary drugs. It realizes real-time docking with the national veterinary drug basic information query system and the national veterinary
drug product traceability system, the remote declaration system of veterinary drug approval number, etc. The third is a dynamic data extraction model based on permission classification. According to different users in the whole process of veterinary drug, combined with the purpose and responsibility of use, the criteria for delimitation of user rights are formulated. Based on the optimal search theory, a dynamic data extraction model based on authority classification is constructed. The hierarchical design is adopted for the state, provincial, municipal, county-level regulatory departments, production enterprises, business enterprises and ordinary users. The principle of stratified empowerment ensures that the quality of veterinary drugs can be monitored, the process can be traced back, and the government can supervise.

3.4. Multidimensional analysis technology of veterinary drug big data
The first is the multidimensional data model of veterinary drugs based on OLAP. On the basis of data warehouse, multi-dimensional data virtual space is formed by taking time dimension, space dimension and attribution dimension as subsets. Using OLAP multi-dimensional online processing theory, an object-oriented multi-dimensional data model of veterinary drugs is built to realize data display of high-dimensional aggregation matrix and low-dimensional details.

The second is veterinary drug multidimensional data analysis technology. Aiming at the massive data of multi-source and heterogeneous veterinary drugs stored in distributed database, supported by Microsoft SQL Analysis Services data mining service, the construction method of multi-dimensional data cube is studied. Combining with MapReduce parallel computing framework, a multi-dimensional data association index model is developed to realize comprehensive query, multi-dimensional retrieval, traceability tracking, process management and statistical analysis the whole process of veterinary drug production, circulation and use.

4. Application analysis of big data framework for veterinary drug supervision
Veterinary drug big data construction and intelligent management can be applied to veterinary drug big data information aggregation, veterinary drug flow direction information visualization, veterinary drug food safety traceability technology and animal disease prediction, etc. The specific applications are as follows:

4.1. Information aggregation of whole industry chain based on big data of veterinary drugs
The format types and logical relationships of multi-source heterogeneous information data of veterinary drugs are analyzed, and the big data information of veterinary drugs involving multiple data sources such as veterinary drug production, management, use, supervision, network hotspots are sorted out and aggregated; The construction of the National Veterinary Drug Data Center based on cloud platform, the aggregation of information related to the whole veterinary drug industry chain involved in the production, operation and use process, the construction of the network information intelligent perception acquisition model based on the network crawler technology and the real-time acquisition of the veterinary drug hot spot information in the network; The aggregation of veterinary drug related industry information, livestock and poultry breeding scale, quantity, feed, network hotspots, meteorological climate and other related aquaculture data information, and the use of national veterinary drug big data architecture lay the foundation for big data analysis and processing; The data types, data quantities and data storage methods of veterinary drug traceability are analyzed and sorted out, and the veterinary drug data are extracted and calculated rapidly based on the theory of parallel processing.

4.2. Spatial visualization display based on veterinary drug big data
The network topology map of veterinary drug circulation evolution in different time and space is preliminarily constructed, and the visualization display model of GIS map based on spatial analysis algorithm and data visualization display technology in different spatial and temporal dimensions is constructed; Visual display of dynamic change process of veterinary drug industry chain information; construction of vertical serial association model of veterinary drug identification information, basic data
information and dynamic circulation information; establishment of horizontal dynamic association model of veterinary drug data information, image information and video information;

4.3. Prediction and analysis of animal diseases based on big data of veterinary drugs
Establishment of veterinary drug quality and safety data traceability processing technology, establishment of veterinary drug product quality traceability technology system based on veterinary drug unique identification code; Establishment of Traceability Technology for veterinary drug residues exceeding the standard and traceability of veterinary drug use; rapid responsibility tracing of the issue of veterinary drug residues exceeding the standard caused by blind drug use, over-dose drug use and off-drug use; Establishment of risk assessment model for animal product quality and safety; construction of correlation model between veterinary drug dosage and animal epidemic disease and tracking of real-time development of animal epidemic disease; construction of prediction and analysis model of animal epidemic disease based on veterinary drug data mining.

5. Conclusion
Big data is another profound information technology revolution in human history after cloud computing and Internet of Things. It is not only an important wealth of the information society, but also a great change in work and thinking [6]. Understanding big data, sorting out big data analysis technology of veterinary drugs and the application of big data in veterinary drug supervision can provide strong support for veterinary drug enterprises, regulatory departments and inspection institutions to manage veterinary drugs, and also promote the progress of veterinary drug management concepts and governance models.

Aiming at the problems of dispersed data, insufficient real-time data and weak application ability of veterinary drug link, the key technologies of big data collection, storage, processing, analysis and application of veterinary drugs should be further improved. And the application of GIS spatial analysis technology to construct flow direction of veterinary drug and regional livestock and poultry feeding volume, analysis model of related relations off certain animal disease outbreak are also should be further explored, which can be used to more accurately identify the risk of veterinary drug residues in animal products and the trend of animal epidemic outbreaks, and to further explore the characteristics, rules and knowledge contained in veterinary drug data, so as to comprehensively enhance the innovative application ability of veterinary drug data.

Acknowledgments
This research was funded by the Fundamental Research Funds for Central Non-profit Scientific Institution (JBYW-AII-2017-06 and JBYW-AII-2017-33), the Fundamental Research Funds for the Central Research Institutes (Y2018PT82 and Y2018PT35) and The Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2016-AII). *Corresponding author: Jianzhai Wu (wujianzhai@caas.cn).

References
[1] Wu, H.Q. (2013) Opportunities and Challenges in the Era of Big data. J. Modern Management., 2013 (5): 33-34.
[2] Liu, Z.h., Zhang, Q.L. (2014) Summary of Research on Big data Technology. J. Journal of Zhejiang University (Engineering Edition)., 48(6): 957-972.
[3] Xiong Z. (2004) Study on the Technology and Application of Parallel Data Mining Facing Business Intelligence, Chongqing: Chongqing University.
[4] Gao, L.J., Liu, Y.B., Li, X.P., etc. (2014) Research on Information Collection and Processing Technology of National Veterinary Drug Traceability System. J. Chinese Veterinary Drug Journal., 48(9): 53-56.
[5] Cui X. (2010) Research on Distributed Storage Management and Parallel Processing Technology of Massive Spatial Data, Changsha: National University of Defense Technology.
[6] Victor, M.S., Kenneth, C. (2013) Big Data Age. Zhejiang People's Publishing House, Hangzhou.