Research on User Data Analysis of Agricultural Product E-commerce Based on Big Data Technology

XIE Chao¹, HE Defu²*

¹School of Management, Wuhan Donghu University, Wuhan, China
²College of Computer, Hubei University of Education, Wuhan, China
*E-mail: hedefu1023@sina.com (Corresponding author)

Abstract. With the rapid development of rural e-commerce, users will produce a large amount of disordered original log data on the platform. In order to give full play to the user data value of agricultural e-commerce, this paper combs the concepts of big data and user data of agricultural e-commerce, and presents the user data analysis service and application system architecture of agricultural e-commerce based on big data technology, so that big data can be used in agriculture. The application value in the field of Commodity E-commerce has been fully embodied, which has played a role in promoting the further development of e-commerce.

1. Introduction
In recent years, the government has attached great importance to the development of e-commerce in rural areas. Governments at all levels have issued a series of guiding principles on accelerating the development of e-commerce in rural areas, such as "Internet plus" three year action plan for the implementation of modern agriculture, and a series of policies to guide the development of agricultural e-commerce. Enterprises are enthusiastic about developing rural e-commerce, and many enterprises, such as Alibaba, Suning Easy-to-buy and Jingdong, are contributing to rural e-commerce. With the support of the government and the efforts of the industry, rural e-commerce has developed rapidly. Taking the rural e-commerce market as an example, China E-commerce Research Center released the "Data Monitoring Report of China's Online Retail Market in 2018", which shows that in 2018, the total retail sales of rural e-commerce reached 137 trillion, an increase of 30.4% compared with the previous year, and the total retail sales of agricultural products reached 230.5 billion, an increase of 33.8% compared with the previous year Rural e-commerce is developing rapidly. In addition, in 2017, 832 poverty-stricken counties nationwide achieved a net retail sales volume of 120.79 billion, an increase of 52.1% over the previous year, 13.0% higher than the rural growth rate. However, for a long time, compared with the "wind-driven water" and "rush to the countryside" of urban industrial products, the upstream of agricultural products appears to be "lack of stamina" and "enthusiasm lags behind". The most effective way to solve this obstacle is to learn from the latest sales technology, with the rapid development of mobile Internet and the explosion of network information, make the demand of buyers and sellers form seamless docking, and promote the rapid development of agricultural products with the times.

Nowadays, with the vigorous development of e-commerce, Internet Finance and other business, a large number of e-commerce data, transaction data and payment data of agricultural products are growing day by day. How to collect, store, analyze, process, retrieve and effectively utilize the value of these data, create benefits, and build core competitiveness are the challenges and opportunities...
faced by current agricultural products e-commerce. When users use the agricultural products e-commerce platform to complete their business, the platform will record their trajectory of action, and the agricultural products e-commerce platform has comprehensive data information. Agricultural products e-commerce data includes browsing, purchasing and consumption records of all registered users, users' evaluation of commodities, and trading records of platform merchants, product trading volume, inventory, logistics information, and credit information of merchants, etc. It is necessary to apply big data technology to the in-depth analysis and utilization of these massive data.

2. Overview of Big Data

Big data, or huge data, refers to information that is too large to be captured, managed, processed and collated in a reasonable time through the current mainstream software tools to help enterprises make more positive business decisions. The earliest reference to the term "big data" can be traced back to Apache org's open source project Nutch. At that time, big data was used to describe a large number of data sets that needed batch processing or analysis simultaneously to update the network search index. With the release of Google MapReduce and Google File System, big data is no longer only used to describe a large amount of data, but also to cover the speed of data processing. The industry usually uses four V (Volume, Variety, Value, Velocity) to summarize the characteristics of big data.

Volume, data volume is huge. If we only consider the storage capacity, we should start with the smallest data storage unit bit and go up Byte, KB, MB, GB, TB, PB, EB in sequence. So far, the amount of data on all printed materials produced by human beings can reach hundreds of PB, and the amount of data recorded in history about all the words that human beings said at the beginning should be measured by EB. Up to now, the data volume of all printed materials produced by human beings is 200 PB, while the data volume of all words spoken by human beings in history is about 5 EB (1EB=210PB).

Variety, there are many types of data. The diversity of types divides data into structured data and unstructured data. Compared with the structured data which is easy to store in the past, there are more and more unstructured data, including network logs, audio, video, pictures, geographic location information and so on. These multi-type data put forward higher requirements for data processing ability.

Value, low value density. The value density is inversely proportional to the total amount of data. The larger the total amount of data is, the more invalid and redundant data are. How to quickly purify the value of data through powerful machine algorithms is an urgent problem to be solved under the background of big data.

Velocity, fast processing speed. This is the most prominent feature of big data distinguishing from traditional data mining. According to IDC's "Digital Universe" report, it is estimated that by 2020, global data usage will reach 35.2ZB. In the face of such a huge amount of data, the efficiency of data processing is the life of enterprises.

Big data requires special technology to organize and access huge amounts of data using special data structures in order to effectively process data stored across multiple servers and discrete data. The technologies suitable for big data include large-scale parallel processing database and data mining, data visualization, distributed file system, distributed database and cloud computing platform. Big data includes three dimensions: theoretical dimension, technical dimension and practical dimension. The core of big data lies in mining hidden value in data for users.

3. User Data of Agricultural Electronic Commerce

In order to analyze the user data of agricultural e-commerce, it is necessary to collect the behavior log data based on the user on the platform of agricultural e-commerce.

User's behavior data is stored in the background database in the form of log in the agricultural e-commerce industry. In the process of running the agricultural e-commerce platform, a large number of original logs are generated. These original logs take the user operation and service provided by the platform for a given period of time as a session. A large number of session logs constitute the user's
original behavior log. Because of the huge amount of logs, the original behavior logs are usually stored in the data warehouse with distributed technology. User behavior data of agricultural products e-commerce industry records almost all users' behavior on the platform of agricultural products e-commerce, mainly including: agricultural products browsing, agricultural products search, agricultural products click, agricultural products purchase, agricultural products scoring and service evaluation.

User behavior data of agricultural e-commerce can be analyzed from three basic dimensions: user attribute dimension, agricultural product attribute dimension and user-agricultural product interaction dimension. Users of e-commerce platform include potential users and value users.

Firstly, the dimensionality of user attributes is analyzed. When the purchasing behavior occurs in the agricultural e-commerce platform, users change from potential users to value users. Value users can be subdivided into potential value users and high value users according to their loyalty, purchase amount and shopping trend. Usually from the demographic characteristics of users, recent purchase time, shopping frequency, shopping amount, shopping categories, scoring tendency and other characteristics, mining user-specific attribute categories, better understanding of users, personalized recommendation of corresponding agricultural products and targeted services.

Then, from the dimension analysis of agricultural product attributes. Agricultural products e-commerce platform has inherent characteristics because the threshold of opening stores and shelf cost are far lower than traditional retail industry. E-commerce sales of long-tailed agricultural products are more abundant, there is a "can not buy, only unexpected" prosperity situation, but long-tailed agricultural products due to its cold door nature, destined to be unable to home page. Then, because of the large base of users and the great differences of interests among users, after deeply mining the users'interest points, we can personalize the long tail agricultural products to the corresponding users and improve the sales volume of cold-end agricultural products.

Finally, the interaction between users and agricultural products is analyzed. The interaction between users and agricultural products is the most important information in e-commerce behavior log, which reflects the user's interest in agricultural products. A successful case in personalized recommendation system is the use of such information. According to the initiative of user interaction, it can be divided into explicit feedback behavior and implicit feedback behavior. Explicit feedback behavior is the user's active preference for the goods. In the agricultural e-commerce platform, users can generally score the purchase of agricultural products by 0-5 points, and express their satisfaction with agricultural products by score. The other side of explicit feedback behavior is implicit feedback behavior. Implicit feedback behavior needs to be analyzed and mined by data mining and other technical means in order to discover the hidden user interest information, but can not clearly reflect the user's preferences and attitudes. In the agricultural products e-commerce platform, users click on and browse the characteristics of agricultural products. Users click on and browse a certain agricultural product, which does not necessarily mean that users like this agricultural product. It is also possible that the agricultural product belongs to the front page of the platform, which belongs to the main agricultural products, and users are more likely to be clicked on it; and the number of clicks on a certain agricultural product and the length of time spent browsing a certain agricultural product by users. It can also implicitly reflect the user's preferences. The more clicks on agricultural products, the longer time it takes to browse agricultural products, the more interested it is. Implicit feedback can not be used explicitly, but the cost of acquisition is lower and the amount of data is larger. The research on the mining method and application of implicit feedback data can better meet the needs of users. Especially in the recommendation system technology of agricultural products e-commerce, implicit feedback behavior usually hides the potential shopping trends and needs of users, so now it is combined with explicit feedback behavior and implicit feedback behavior to mine interest.

4. User Data Analysis Service and Application of Agricultural Product E-commerce Based on Big Data

With the rapid development of big data technology, 5G mobile network, Internet of Things and artificial intelligence technology, and the popularity of mobile Internet and smart phones, data
information overload has appeared. How to effectively implement intelligent recommendation of data information has become a hot issue in the field of artificial intelligence and big data. User data analysis system for agricultural e-commerce based on big data is mainly composed of big data collection module, user interaction module, intelligent recommendation processing module, evaluation module and system management module. There will be massive data interaction between the module and the module. The big data collection module is responsible for collecting massive data of platform users. After collecting massive data, it can be stored in the core database. Through data mining technology, user data information can be analyzed and provided to the platform. It can also be recommended to other users for reference. The system management module is responsible for the management of the whole system, and the evaluation module is responsible for the effective evaluation of the recommended information. The evaluation results are fed back to the system in order to achieve efficient management and provide an effective data base for the intelligent push of data information, so as to realize the efficient intelligent user data analysis service. The user data analysis system of agricultural e-commerce based on big data is shown in Figure 1.

The user data analysis system of agricultural e-commerce based on big data is mainly divided into five parts: data layer, preprocessing layer, algorithm layer, service layer and application layer. Among them, the data layer mainly includes user center, Commodity Center and log engine; the preprocessing layer mainly includes content features, user portraits and group segmentation, etc. Preprocessing refers to feature extraction and feature construction of data from multiple data sources; the algorithm layer mainly includes user-based collaborative filtering, commodity-based collaborative filtering, knowledge-based recommendation, content-based recommendation and complementation. The service layer mainly includes intelligent recommendation engine, A/B and transformation monitoring, and the application layer mainly includes client. User data analysis system of agricultural products e-commerce applies big data technology and artificial intelligence technology in the process of e-commerce, and makes intelligent recommendation for products and services for different users. Providing intelligent services to users can effectively satisfy users'data and information enhancement needs of products, and provide guarantee for improving service satisfaction of agricultural products e-commerce.
5. Conclusion
Under the background of big data, in order to improve the economic benefits of agricultural e-commerce, it is necessary to grasp the user's application demand for products and obtain the most accurate data information support. Based on the user's data of agricultural e-commerce, using data mining, artificial intelligence, in-depth learning and other technologies, we can analyze the key information in a large number of disordered logs and know the user's shopping interest. Users provide personalized shopping services, which can deepen the contact with consumers and bring the advantages of e-commerce platform into full play, thus promoting their own development.

Acknowledgments
This work was supported by the grants from Hubei Provincial Collaborative Innovation Centre of Agricultural E-Commerce (Wuhan Donghu university research [2019] No.10 Document) and Youth Foundation WuHan Donghu University in 2019 "Research on Traceability System Construction of Quality and Safety of E-commerce Agricultural Products from the Perspective of Blockchain".

References
[1] Li Chao. Analysis of Personalized Recommendation System for E-commerce Based on Big Data [J]. Journal of Commercial Economics, 2019(02):69-72.
[2] Yu Tao. Research on Key Technologies of Electronic Commerce Based on Big Data [J]. Science & Technology Information, 2018, 16(36):35-36.
[3] Zhang Pengpeng. User Data Mining and Application of Mother-to-Child E-commerce Based on Big Data [J]. Enterprise Reform and Management, 2018(21):80-85.
[4] Song Hao, Jiang Fuwen. User Data Mining of Online and Offline E-commerce Based on Big Data [J]. Electronic Technology & Software Engineering, 2017(16):19-23.