

Development Problems and Countermeasures of Rural E-Commerce Logistics in the Context of Big Data and Internet of Things

Xianfeng Zhu*

Abstract

As the Internet has expanded and the continuous expansion of online shopping in China, many rural areas also have sales outlets. Due to the impact of economic conditions, rural locations have inadequate e-commerce logistical infrastructure, the number of outlets is small, and each other is in a decentralized state. For various reasons, the advancement of rural e-commerce logistics lags far behind that in urban areas. As the Internet of Things with big data grow in popularity, we can create and enhance the assurance system for the booming e-commerce in rural areas by building the support system of rural online shopping platform, and strengthening the joint distribution of logistics terminals based on data mining, so as to encourage the quick and healthy growth of rural online shopping.

Keywords

Big Data, E-Commerce, Internet of Things, Logistics, Rural Areas

1. Introduction

As the Internet has expanded and the maturation of China's online shopping business, countryside also witnesses the fast growth of rural online shopping [1]. However, rural growth is relatively backward, and there is a shortage of rural logistics enterprises, resulting in many problems in rural logistics [2]. However, logistics is a vital basic condition for the growth of e-commerce. The backwardness of rural logistics will hinder increasing e-commerce. Therefore, improving the development of logistics is the prerequisite to improve the expansion of online shopping. To promote the expansion of online shopping, many academics have conducted extensive study on how to better improve rural online shopping. Based on the development of logistics supply chain under big data, Zhao et al. [3] used M Company as an example, thoroughly examined its supply chain management cost control method, and proposed pertinent countermeasures and recommendations. Sun and Gu [4] suggested creating an e-commerce entity monitoring Internet of Things (IoT)-based system and established the assessment index system. As stated by the test results, the regulatory system proposed in the study has certain security.

Huang [5] used an IoT-based system to streamline the supply chain management process to enhance

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received June 16, 2022; first revision September 15, 2022; second revision October 25, 2022; third revision December 14, 2022; accepted January 25, 2023.

* Corresponding Author: Xianfeng Zhu (zhu_xianfeng@outlook.com)

School of Economics and Management, Jiaozuo University, Jiaozuo, China (zhu_xianfeng@outlook.com)

it. From the above research and literature, it is evident that the IoT and large data can be used to establish the corresponding logistics system and improve the rural logistics distribution functions and services. Currently, rural e-commerce logistics in China still has a lot of issues, such as the uncoordinated rural industrial structure. In order to further study and address these issues, this study will adopt the "Internet plus" thinking, use big data and the Internet, establish a rural logistics information platform, and encourage the expansion of logistics for rural online shopping as a whole.

The first part of the paper is the introduction. In this part, the background and related research of the growth of rural online shopping logistics are described, and the purpose and method of this study are explained. Section 2 describes the current scenario of rural logistics development in China. Section 3 uses data mining technology to create a shared distribution system for rural logistics terminals. Section 4 discusses how large data and IoT are influencing rural e-commerce. The main innovation of this research is to use data mining technology to integrate the terminal distribution of various express companies and build a rural logistics terminal common distribution system.

2. The Development Status of Rural E-commerce Logistics in China

2.1 Status Quo of Logistics Service System

Based on the principle of accurate service, China is committed to developing rural areas, strengthening circulation and talent transfer, so as to build an accurate, timely and efficient e-commerce service system [6]. China's rural logistics security has made great achievements. The following are the primary strategies for advancing China's rural logistics system [7,8].

First, establish a multi-level service model in rural areas. At present, China's rural regions now have 616,851 village-level service stations and 24,057 county-level e-commerce development service stations. Second, improve the multi-level logistics system at county and township levels. The third is to carry out multi-level training disciplines in schools, universities, and enterprises, and nearly 200,000 college students have received e-commerce training. The fourth is to train multi-disciplinary talents in cities, towns, and other rural areas. Up to now, about 3 million rural e-commerce talents have been trained.

2.2 Development Status of Transportation System

2.2.1 Railway construction

Since the program for rural revitalization's implementation by the State Council, the Chinese nation has increased its investment in rural railway construction, and China's railway fixed asset investment and railway operating mileage have increased year by year. The national railway mileage increased by nearly 14.88% from 2015 to 2019. The development of railways has greatly increased the mileage of rural areas and created favorable conditions for an increase in rural e-commerce.

2.2.2 Road construction

In 2019, the length of China's rural roads reached 4.25 million kilometers, a 4.0% year-over-year growth, and the mileage of roads in 2018 was 0.8% higher than that in 2017. The geographical conditions in rural areas are complex, and road transportation can better meet local material transportation needs.

2.3 Development Status of Logistics Mode

Increasing urban-rural integration has promoted the development of rural family farms. Most farmers cooperate with processing companies and large suppliers and are located at the upstream of the agricultural product logistics chain. At present, the scale of third-party logistics enterprises in rural areas is relatively small, and they are still in the early stage of logistics distribution. Although postal logistics is a relatively mature logistics enterprise in rural areas, it still needs to improve efficiency and professionalism in order to further expand the sales market of agricultural products. Although China Post has a lot of service locations in remote regions, the growth rate of service outlets is very slow, and the problem that hinders the separation of logistics and transportation of rural family farms remains unsolved.

3. Joint Distribution of Logistics Terminal Based on Data Mining

Rural e-commerce has grown quickly, and the logistics express sector has followed suit, but there are many problems in its development process, the most prominent of which is the rural express terminal distribution problem. In view of this, this study will design a logistics express terminal distribution mode to realize joint distribution, and use data mining technology to plan the terminal distribution system. Data mining technology will enhance the distribution efficiency and superior service of the logistics express terminal, build a joint distribution system suitable for China's rural logistics terminals, and help to enhance the overall level of rural logistics services.

3.1 Data Mining and Logistics Terminal Distribution Status

Data mining means the process of using algorithms to search for information hidden in a significant quantity of data. Data mining modeling generally includes business analysis, data collection, data cleaning, model construction, evaluation model, and application. The logistics that are provided to customers are referred to as "terminal logistics," and they are a logistics activity that has as its primary goal the terminal of the distribution connection. At present, terminal logistics mainly has problems such as distribution difficulty, distribution safety, vehicle traffic difficulty, and fierce competition, while rural terminal logistics distribution is difficult and time-consuming. Due to the scattered and distant recipients in rural areas, rural terminal distribution has always been a problem.

Joint delivery refers to the delivery of multiple express companies in a certain region to improve logistics efficiency. Its essence is to reduce costs and enhance the use of logistical resources in extensive activities. In the context of e-commerce, the logistics express terminal joint distribution is mainly divided into the terminal joint distribution led by the third party and the terminal joint distribution led by the express alliance. Both methods can reduce the distribution cost. Under the joint distribution, it will also reduce the urban traffic pressure, encourage resource pooling in logistics, and increase the quality of logistics services overall.

3.2 End Joint Distribution based on Data Mining

Due to the growth of online shopping, the quantity of rural express delivery is also growing rapidly every year. The issues of high distribution cost and low distribution efficiency in terminal logistics are

becoming more and more obvious. Select a suitable distribution mode to meet the basic requirements of rural logistics, improve logistics efficiency, and reduce distribution costs. At the same time, under the premise of joint distribution at the end, the supervised learning method in data mining is used to select and model the joint distribution mode, so as to increase the final joint distribution's effectiveness.

Among the classification algorithms of data mining, there are mainly decision tree, artificial neural network, and other algorithms. Among them, decision tree is the most commonly used and main classification and regression algorithm in data mining, and it is easier to extract the final rules. Random forest is an improved algorithm of decision tree. Compared with decision tree, random forest has better stability and can use C4.5 algorithm to generate decision tree. Suppose that there are samples in the set of S data samples, $C_i (i = 1, \dots, m)$ represents class label, and m different values. The number of samples in class C_i is S_i . If the samples need to be classified, the required expected information value expression is shown in Eq. (1):

$$I(S_1, \dots, S_m) = - \sum_{i=1}^m p_i \log_2 p_i \quad (1)$$

In Eq. (1), p_i is used to describe the probability value of sample C_i , which can be estimated by S_i/s . S_1, \dots, S_v represents v subsets of attribute A , and S_j are some samples with value a_i on A in s . These subsets correspond to node branches of set S . If A selects these subsets as test attributes, the quantity of samples of class C_i in subset S_j is S_{ij} . The entropy divided into subsets by A is shown in Eq. (2):

$$E(A) = - \sum_{i=1}^v \frac{S_{ij} + \dots + S_{mj}}{S} I(S_{ij}, \dots, S_{mj}) \quad (2)$$

In Eq. (2), the weight of the j subset can be expressed as $\frac{S_{ij} + \dots + S_{mj}}{S}$, which is equivalent to the quantity of samples in the subset (that is, the A value is a_i) divided by the overall sample size. The purity of the subgroup increases as entropy decreases. For a given subset S_j , the Eq. (3) can be obtained:

$$I(S_{1j}, \dots, S_{mj}) = - \sum_{i=1}^m p_{ij} \log_2 p_{ij} \quad (3)$$

The likelihood that the samples in S_j belong to class C_i is $p_{ij} = S_{ij}/S_j$. The information gain ratio expression obtained on attribute A is shown in Eq. (4):

$$SplitInfo(S, A) = - \sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (4)$$

When a node chooses an attribute branch, the C4.5 algorithm evaluates each attribute's information gain ratio and chooses the feature that the greatest information gain ratio as the test attribute. Create a node, set different values of this attribute as different branches, use this attribute to identify this node, and classify samples. Random forest can combine bagging and random attribute selection. It mainly uses bagging to form a training set for each tree, and then each self-service sample is integrated into a single classification tree. Then, a simple majority voting method is used to obtain the results of random forest.

Compared with the decision tree, the random forest model's capacity to generalize has been improved, and the classification precision has also been significantly enhanced.

One of the most common machine learning methods is the artificial neural network algorithm. The back-propagation neural network (BPNN) algorithm is a typical artificial neural network method. In the course of outputting neurons, the total value and threshold of the accumulated neurons are compared, and the final output of neurons is generated through activation function.

3.3 Model Construction and Performance Comparison

Establish the model of distribution mode selection, select 12 indicators, such as population, population density, daily average express delivery volume, and the number of nearby express convenience stores, as the measurement attributes to judge the terminal distribution point to select the terminal distribution mode, and discretize some attributes. Through field investigation, it is found that the current terminal distribution mode in rural areas mainly includes third-party collection points, door-to-door delivery, self-delivery at the store and other forms. After determining the attributes that affect the distribution mode, the survey data will be normalized, the abnormal data will be eliminated, and the neural network and random forest models will be created. Fig. 1 depicts the construction of the two models.

The algorithm is implemented using R language, and 70% of the input data is randomly calculated as the practice set, and 30% of the remainder the input data is used as the verification set. Cross-validation is used to verify the reliability of the data, and two models are used to select and classify the terminal distribution methods for the validation set data. Accuracy, recall, and F-value are selected as evaluation indicators to obtain the capability of the two models under the validation data set. Fig. 2 depicts the findings.

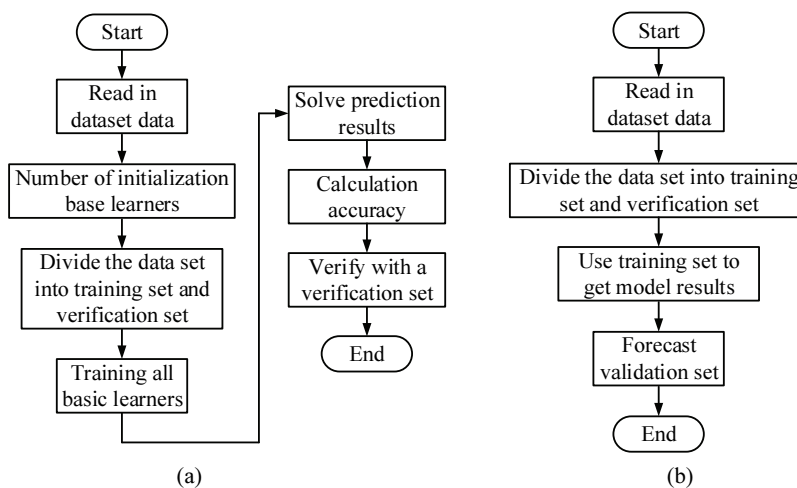


Fig. 1. The process of building two models: (a) stochastic forest model and (b) neural network model.

As seen in Fig. 2, the accuracy, recall and F-value of the random forest model are higher than those of the artificial neural network model, that is, the random forest algorithm should be selected as the model algorithm for the selection of logistics terminal distribution mode in this study. In a rural environment, the random forest model is utilized to examine the terminal distribution mode. Data such as the population number, population density, daily average express delivery volume, and the number of nearby express

convenience stores can be put into the model are collected and input into the model. The model of "self-delivery + door-to-door delivery" is suitable for the terminal logistics distribution in this rural area.

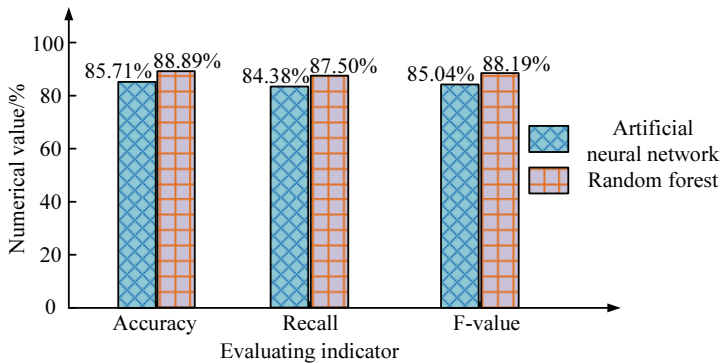


Fig. 2. Model performance comparison.

4. Development Countermeasures of Rural E-commerce Logistics under the Big Data and IoT

4.1 Integrate Courier Company Resources to Coordinate Distribution

In rural areas of China, the volume of logistics and distribution business is very small, which leads to the waste of labor costs and material resources, resulting in high logistics costs for enterprises. Therefore, logistics enterprises can also cooperate with each other, integrate advantageous resources, form a common and efficient logistics distribution alliance, and achieve efficient interaction and sharing of real-time logistics information between mobile logistics companies and local postal outlets [9].

4.2 Big Data and IoT to Establish a Rural Logistics Information Platform

To ensure that the operators and consumers of traditional agricultural products can query the relevant agricultural product logistics information in actual time, dynamically and accurately during the whole process of entering the transaction order and inquiring the commodity distribution and transportation information, it is vital to boost rural contemporary logistics Internet infrastructure planning and development, as well as build a more effective modern logistics signal base station with network coverage, and achieve wider information network coverage. Meanwhile, we should take note of the use of information resources [10,11].

4.3 Strengthen Joint Distribution at Logistics Terminals

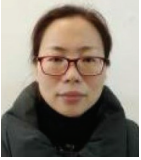
In view of the high cost, long distance, inconvenient transportation, and other reasons of rural logistics terminal distribution, it is necessary to select the logistics terminal distribution mode based on data mining, analyze the terminal distribution in rural areas by using random forest algorithm, and select the logistics terminal distribution mode suitable for rural regions to address the issue of rural logistics terminal distribution.

5. Conclusion

The expansion of rural online shopping has been aided by the introduction of big data and IoT. However, as time goes by, backward modern logistics also constrains the expansion of rural online shopping. Thus, it is necessary to improve the current situation of logistics, optimize the logistics service environment, develop a new O2O model, speed up logistics informatization, integrate logistics resources, and improve the rural logistics terminal joint distribution model based on data mining. Driving the modernization of rural logistics and the expansion of the rural economy are crucial. In this study, rural logistics terminal data are obtained by means of research, and the amount of data is limited, which has some impact on the training and accuracy of the model. To enhance the accuracy of the model, it is required to combine a large number of rural logistics terminal data training models. In the future, research will also be carried out on rural logistics express collection to further improve rural logistics efficiency.

References

- [1] Y. Q. Li and J. C. Wang, "Research on the development countermeasures of rural e-commerce under the background of Internet finance: taking Jiaying as an example," *Agriculture Network Information*, vol. 2016, no. 8, pp. 28-33, 2016.
- [2] S. Q. Li, "Research on the status quo, problems and countermeasures of China's tourism e-commerce development under the background of Internet," in *Proceedings of 2018 4th International Conference on Innovative Development of E-commerce and Logistics (ICIDEL)*, Zhengzhou, China, 2018, pp. 164-171.
- [3] R. Zhao, G. Gu, and Z. Yang, "logistics supply chain management under the background of big data," in *2021 International Conference on Applications and Techniques in Cyber Intelligence: Applications and Techniques in Cyber Intelligence (ATCI 2021)*. Cham, Switzerland: Springer International Publishing, 2021, pp. 796-802.
- [4] P. Sun and L. Gu, "Optimization of cross-border e-commerce logistics supervision system based on Internet of Things technology," *Complexity*, vol. 2021, article no. 4582838, 2021. <https://doi.org/10.1155/2021/4582838>
- [5] B. Huang, "Research on optimization of e-commerce supply chain management process based on Internet of Things technology," *Journal of Physics: Conference Series*, vol. 2074, no. 1, article no. 012070, 2021. <https://doi.org/10.1088/1742-6596/2074/1/012070>
- [6] X. Peng and X. Zhang, "Research on the current situation of rural e-commerce in the age of "Internet+" and its countermeasures: taking Jiangxi County as an example," in *Proceedings of 2018 International Conference on Economics, Politics and Business Management (ICEPBM)*, Nanjing, China, 2018, pp. 158-165.
- [7] Y. L. Zhao, "Analysis of the development strategy of rural e-commerce logistics from the perspective of economics," *Modern Economic Information*, vol. 2018, no. 8, pp. 334-334, 2018.
- [8] H. C. Zhu, Y. Y. Li, X. Y. Du, and H. Guo, "Development status, problems and promotion countermeasures of rural E-commerce logistics in China," *Logistics Engineering and Management*, vol. 40, no. 1, pp. 1-2, 2018.
- [9] X. M. Zhou, and L. F. Lu, "Research on intelligent logistics based on internet of things and big data analysis," *China Logistics and Procurement*, vol. 2020, no. 1, pp. 59-60, 2020.
- [10] L. Xiao, "Analysis of the development model of intelligent logistics based on the Internet of Things and cloud computing in the era of big data," *Science and Informatization*, vol. 2018, no. 7, pp. 35-39, 2018.
- [11] S. Liu and X. D. Sun, "Prospect analysis of logistics real estate under the background of big data," *China Real Estate*, vol. 4, no. 11, pp. 53-59, 2021.



Xianfeng Zhu <https://orcid.org/0000-0003-2562-4939>

She was born in Dengfeng, Henan Province, China, in 1979. She received an undergraduate degree from Henan Agricultural University, and a graduate degree from Henan Polytechnic University. Now, she works in Jiaozuo University. Her research interests include logistics management and agricultural economy.