Evaluation of Cultivated Land Productivity Based on the Perspective of Big Data

Mingliang Duan(✉) and Min Zhu

School of Humanities, Shanghai Institute of Technology, Shanghai 201418, China
gnailnaud@163.com

Abstract. In order to objectively evaluate the indicators affecting cultivated land productivity, under the background of big data, taking Jiaozuo City as an example, data mining is used to analyze the cultivated land productivity indicators, and finally the cultivated land in Jiaozuo City is divided into three types, and four indicators that have an impact on the cultivated land types in Jiaozuo City are screened: average annual rainfall, effective soil thickness, total nitrogen and latitude.

Keywords: Jiaozuo City · big data · big data mining · evaluation of cultivated land productivity

1 Introduction

Big data has been widely applied in agriculture, such as agricultural production, management, and services, especially in the operation system and management system of modern agriculture [1]. It is an ultra-large-scale data collection, which dramatically exceeds traditional database software tools' capacity range in the storage, management, and analysis. However, the application in the evaluation of cultivated land productivity is not deep enough. Traditional evaluation of cultivated land productivity rest on the basis of small data, which cannot carry out comprehensive evaluation, while big data, by integrating various types of structured, and unstructured data, can carry out Spatial-temporal long- term tracking, identify various characteristics of farmland. With the advent of the information age, the application of big data technology can provide data support for evaluation of cultivated land productivity, the author excavated the big data of cultivated land in Jiaozuo City, aiming to screen out the land fertility indicators that have an impact on the type of cultivated land in Jiaozuo City.

2 Literature Review

Big data has the following four application advantages in evaluation of land productivity [2]. First, big data can solve the problem of improper articulation and inconsistent standards in traditional evaluation. The application of big data technology realizes information sharing among departments, makes more reasonable decisions, establishes practical agricultural spatial economic land, and establishes a good system operation

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 533–539, 2023.
https://doi.org/10.2991/978-94-6463-198-2_55

platform. Second, big data can improve the spatial evaluation system of farmland. The relevant departments have limitations in grasping the functions and roles of farmland spatial evaluation, and related evaluation functions are not perfect. The lack of a perfect spatial evaluation system for farmland before affects the coordination and wholeness of spatial evaluation. Therefore, applying big data technology can better grasp the current utilization status and development trend of farmland, build up a holistic spatial evaluation system of farmland, and build up a top-down spatial coordination planning approach. Through the information processing of big data, the most reasonable evaluation plan can be selected, and the rational evaluation of cultivated land productivity can be completed through regulation means.

Third, Big data can ensure scientific planning [3]. The big data technology of geographic information and the big data system of geographic information integrate all the information related to farmland, construction, etc., and provide the natural geographic information in the farmland area, including soil conditions, light conditions, temperature, and humidity conditions, which is the basis of evaluation of cultivated land productivity [4]. Moreover, the big data technology presents the past construction information of farmland, including the crops planted, water conservancy, and irrigation projects. Furthermore, big data technology can provide market information, especially the sales information of agricultural products, and feed market information to decision-makers in time to provide a reference for farmland planning and design. With the help of big data technology, the relevant units can use the data platform to break the limitation of time and space in the revision or evaluation of cultivated land productivity that has not been fully considered, providing a more convenient office environment for staff at all levels, improving office efficiency, enhancing the scientific nature of evaluation, and improving the management level. Fourth, big data can enhance the supervision of farmland resource utilization [5].

3 Empirical Study

3.1 Research Objects and Methods

10 county towns in Jiaozuo City were used as research areas, and the arable land evaluation database established by the research was analyzed by using the clustering model in the statistical model, and the main factors affecting cultivated land productivity were screened. Using SPSS Modeler tools for data mining, the Two Step-AS clustering model was used to classify cultivated land in Jiaozuo City. Data mining standard process CRISP-DM model [6], consisting of six stages (Fig. 1).

3.2 Data Wrangling

In this study, the potential classification indicators for cultivated land productivity evaluation are divided into: landform type, aspect, and texture. The number of categories of each index is: the landform type is divided into 60 categories, the slope is 70 categories, and the texture is 60 categories. After adjustment by SPSS modeler data mining: landform type indicators are divided into three categories: mountains, hills and plains;

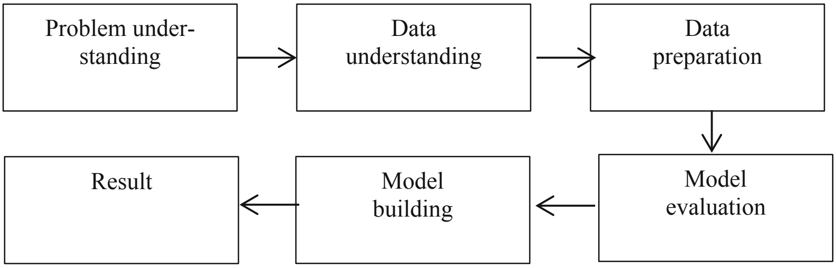

Fig. 1. CRISP-DM model process

The aspect type indicators are divided into 8 categories, which are east, south, west, north, southeast, southwest, northwest, northeast; The texture category indicators were divided into six categories: aquifer soil, brown soil, clay soil, medium brown loam, sandy soil, and amphibious, of which the missing value of landform type was 11.55%, 75% of texture indicators are other categories, so their classification is meaningless. For the missing value data, the author uses the method of replacing the missing value to deal with it, the comparison of indicators before and after classification is shown in Table 1.

Due to the large number of continuous ground force indicators, it is more complicated to process the data, and this study mainly preprocesses the data of outliers, extremes and missing values. In this study, 4 continuous indexes of soil nutrient were extracted,

Table 1. Number of various indexes of data before and after adjustment

Items	Landform types	Slope orientation	Texture
Number of indexes before adjustment	60	70	60
Number of indexes after adjustment	3	9	6

Table 2. Soil nutrient index grouping

Group	Organic matter (g·kg-1)	Available phosphorus (g·kg-1)	Total nitrogen (g·kg-1)	Available potassium (g·kg-1)	pH
1	>25	>30	>1.05	>180	<3.50
2	20–25	10–30	0.85–1.05	130–180	3.50–4.5
3	15–20	8–10	0.65–0.85	80–130	4.50–5.50
4	10–15	4–6	0.35–0.65	30–80	5.50–6.50
5	5–10	2–4	0.1–0.35	10–30	6.50–7.50
6	<5	<2	<0.1	<110	>7.50

Table 3. Grouping of site indexes

Group	Average annual rainfall	Longitude	Latitude	Effective soil layer thickness	Altitude
1	<1400	104.52–105.08	16.39–16.84	35–50	0–50
2	1400–1500	105.08–105.65	16.84–17.29	50–65	50–95
3	1500–1600	105.65–106.22	17.29–17.75	65–80	95–135
4	1600–1700	106.22–106.78	17.75–18.20	80–95	135–1
5	1700–1800			>95	185–230
6	>1800				>230

including organic matter, available phosphorus, total nitrogen, and pH, see Table 2; 6 of site indexes were extracted, including average annual rainfall, longitude, latitude, effective soil layer thickness, and altitude, see Table 3. Among them, the soil nutrient indicators were grouped with reference to the "Methods for Survey, Monitoring and Evaluation of Cultivated Land Quality". The altitude and effective soil thickness are divided according to the fixed width, and the average annual rainfall and longitude and latitude reference values are divided by a fixed width [7].

4 Results and Analysis

Preliminary calculation, from the perspective of the importance of indicators, the top 9 indicators are shown in Fig. 2. After data mining, taking 0.2 as the critical value, the ground force index greater than the feature importance of >0.2 is used as the eigenvalue of the model cluster, the important indicators affecting the soil fertility index are the annual average rainfall, total nitrogen, effective soil thickness, and latitude, the annual average rainfall importance is 0.95, the importance of latitude is 0.88, the importance of total nitrogen is 0.79, and the importance of effective soil thickness is 0.74, Since the organic matter content is related to total nitrogen, the organic matter content and total nitrogen are hidden when the critical value is 0.2.

According to the soil fertility indicators (annual average rainfall, latitude, total nitrogen, effective soil thickness), the Two Step-AS clustering model was established [8]. The conventional number of clusters in the final model is three, that is, the cultivated land types in Jiao zuo City can be divided into three types: mountainous, hilly and plain, and the characteristics of each cultivated land type are shown in Table 4.

As important indicators to measure the effect of cluster analysis, the overall model goodness obtained by using the TwoStep-AS clustering algorithm in this study is 0.28 [9], and the goodness and importance of each cluster are shown in Table 5. According to the analysis in Table 5, the goodness of each type is medium, but the importance is good, and the cluster model used in the study is generally cohesive and has good separation.

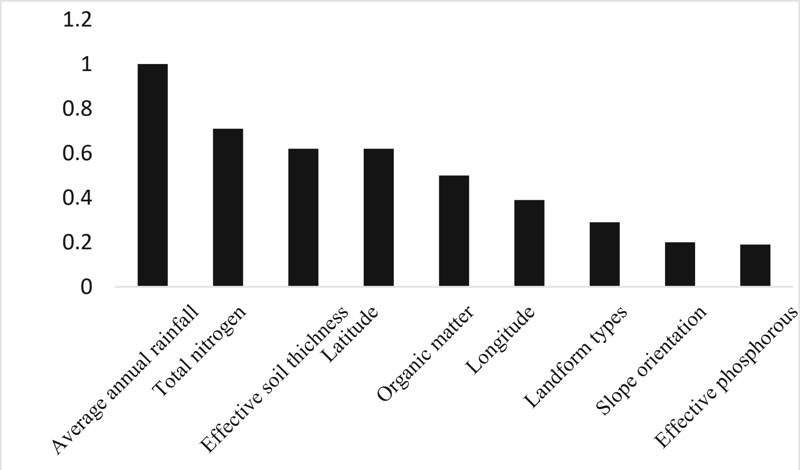

Fig. 2. The initial characteristic importance of land productivity

Table 4. The clustering center for three types of farmlands

Land productivity indexes	Cluster-1	Cluster-2	Cluster-3
Average annual rainfall/mm	1500–1600	1600–1700	>1800
Total nitrogen/(g·kg)$^{-1}$	0.35–0.65	0.65–0.85	0.1–0.35
Latitude/(°)	17.75–18.20	17.29–17.75	16.39–16.84
Effective soil thickness/cm	65–95	65–95	50–65

Table 5. Model quality

Clustering types	Number of sample plots	Goodness of fit	Importance
Cluster-1	1534	0.174	0.82
Cluster-2	1445	0.479	0.573
Cluster-3	2432	0.498	0.92

5 Conclusion

In this study, the factors affecting the productivity of cultivated land in Jiaozuo City were evaluated by big data mining, which greatly improved the work efficiency, and the cultivated land in ten county-level cities of Jiaozuo City was divided into three types of cultivated land: Cluster-1, Cluster-2 and Cluster-3, and the most significant feature of Cluster-1 is that the average annual rainfall is between 1500–1600 mm. The total nitrogen is between 0.35–0.65 and the effective soil thickness is between 65–95; the most significant feature of Cluster-2 is that the average annual rainfall is between 1

600–1 700 mm, the total nitrogen is between 0.65–0.85, and the effective soil thickness is 65–95; the most significant feature of Cluster-3 is that the annual average rainfall is greater than 1800, the total nitrogen is between 0.1–0.35, and the effective soil thickness is between 50–65. The most important soil fertility indicators affecting the quality of cultivated land in Jiaozuo City are: annual average rainfall, total nitrogen, latitude, and effective soil thickness. Today, Chinese research on big farmland data is mainly based on theoretical descriptions, propaganda of laws and regulations or application prospects, and a lack of practical applications [10]. To solve these difficulties, we have some policy suggestions:

First, strengthen policy guidance support. The application of big data technology is the focus of the development of modern information technology. Therefore, the government's support for big data technology is the main factor in using big data technology to evaluation of cultivated land productivity. Second, establish a big data system for evaluation of cultivated land productivity. The government needs to establish an evaluation of cultivated land productivity with reference to big data technology so as to deepen the integration of big data technology in various departments, it is necessary to deeply explore the intrinsic relationship between land productivity evaluation and big data technology, optimize the traditional land productivity evaluation with the help of big data technology, and construct relevant prediction models. In short, borrowing big data technology to evaluation of cultivated land productivity is an inevitable requirement for the development of modern agricultural technology.

References

1. A. Cravero and S. Sepúlveda, "Use and adaptations of machine learning in big data—Applications in real cases in agriculture," Electronics, vol. 10, no. 5, p. 552, 2021.
2. A. Weersink, E. Fraser, D. Pannell, E. Duncan, and S. Rotz, "Opportunities and challenges for big data in agricultural and environmental analysis," Annual Review of Resource Economics, vol. 10, no. 1, pp. 19–37, 2018.
3. D. Kamrowska-Załuska, "Impact of ai-based tools and urban big data analytics on the design and planning of cities," Land, vol. 10, no. 11, p. 1209, 2021.
4. Zhang and N. Cao, "Application and research progress of Geographic Information System (GIS) in agriculture," in 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 2019, pp. 1–5.
5. J. Astill, R. A. Dara, E. D. G. Fraser, B. Roberts, and S. Sharif, "Smart poultry management: Smart sensors, big data, and the internet of things," Computers and Electronics in Agriculture, vol. 170, p. 105291, 2020.
6. C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," Procedia Computer Science, vol. 181, pp. 526–534, 2021.
7. J. Peng, C. Tang, and C. Jiang, "Evaluation of Land Productivity in Rubber Plantations Based on Data Mining," Journal of Tropical Biology, pp. 380–386, 2019.
8. A. A. Said, L. A. Abd-Elmegid, S. Kholeif, and A. A. Gaber, "Classification based on clustering model for predicting main outcomes of breast cancer using hyper-parameters optimization," International Journal of Advanced Computer Science and Applications, vol. 9, no. 12, 2018.

9. T. Wendler and S. Gröttrup, Data mining with SPSS modeler: theory, exercises and solutions. Springer, 2016.
10. C. Liu, Z. Zhang, and S. Zhang, "Smart Initiatives for Land Resource Management: Perspectives and Practices from China," Journal of Urban Technology, pp. 1–19, 2020.