A New Type of Professional Farmers' Digital Literacy Ability Evaluation Algorithm Based on Two-Step Clustering

Feng Wu(✉), Wenjing Liu, Zhixiong Liu, and Jingwen Zeng

Guangdong Polytechnic of Science and Trade, Guangzhou 510430, China
46036530@qq.com

Abstract. In order to solve the problem of evaluating the digital literacy ability of new professional farmers, the definition and classification of urban and rural new professional farmer groups are integrated with existing research, the evaluation index of digital literacy ability of new professional farmers is constructed, and the secondary clustering method is applied to subdivide and evaluate the research population. Using the survey data and case analysis, SPSS is used to calculate and automatically determine the optimal number of clusters. The digital literacy of new professional farmers is divided into three levels: high ability, general ability and low ability. The most important indicator for improving digital literacy is digital governance, followed by digital application. Then using the chi-square analysis, it is found that there is no significant difference in gender in the digital literacy of the three types of new professional farmers; The digital literacy of the low age group is significantly higher than that of the high age group; The digital literacy of people with high education is significantly higher than that of people with low education. In the future, we need to enhance the awareness of personal digital rule of law; Strengthen the training of digital application, implement the post of agricultural digital technician, etc.

Keywords: Professional farmers · Digital literacy · Two step clustering

1 Introduction

With the accelerating digitalization of the world economy, digital literacy is the core literacy that every social person should have, and it is also an essential basic skill for personal survival in the digital age. It profoundly affects people's daily life and communication. Therefore, the level of digital literacy ability directly affects the quality of life in the future. In addition, digital literacy is a key indicator of international competitiveness and soft power, which has also become the focus of discussion by many scholars around the world.

Author Information: Wu Feng (1976-), Male, Master, associate professor, Research Direction: Rural Vocational Education

N. Akhtar et al. (Eds.): PMIS 2023, AHIS 8, pp. 12–23, 2023.
https://doi.org/10.2991/978-94-6463-200-2_3

In the vast rural areas, due to the partial geographical location avoidance, information occlusion, farmers' own low cultural quality, weak information awareness and other objective factors, the digital literacy of farmers in rural areas is generally not high, which can not adapt to meet the needs of modern agricultural information development. In order to give full play to the role of new professional farmers in rural revitalization and agricultural modernization development, we should constantly improve the digital literacy ability of professional farmers and cultivate them into new professional farmers. This paper plans to construct the digital literacy ability model of new professional farmers, evaluate and study the literacy ability situation of professional farmers, find out the problems and deficiencies through the two-step clustering method, and put forward training suggestions from the age and education, so as to promote the development of rural digitalization.

2 Related Literature Review

2.1 New Type of Professional Farmers

According to the opinions of relevant experts at home and abroad, new professional farmers refer to professional groups engaged in agricultural production and operation activities. In 2007, the State Council put forward a new type of farmers with "culture, technology and business skills" in several opinions on promoting the construction of new rural areas in society. Referring to the definition of many domestic experts, the so-called new professional farmers refer to the new generation of farmers who take agriculture as their occupation, have certain culture, higher ideology, know science and technology, and are good at management, and have strong production and management ability, market information acquisition, analysis ability and innovation ability.

2.2 Digital Literacy

In 1997, the American Gilster first proposed the concept of "digital literacy", which defined it as a comprehensive ability to acquire, understand, organize, and criticize digital information [1]. Eshet-Alkalai further expanded the definition of digital literacy to include the survival skills needed for residents to live, study, and work in the digital age, including image-image literacy, reproduction literacy, branch literacy, information literacy, social and emotional literacy, and real-time thinking skills [2]. The European Commission emphasizes the ability to critically and creatively use information tools in its work and life, including five literacy domains: information domain, river basin area, content creation domain, security domain and problem solving domain [3]. On this basis, the UNESCO project team Law in the global digital literacy framework defined digital literacy as "participate in the economic and social life through digital technology and networking devices, security, properly define, acquisition, management, integration, dissemination, evaluation, and create information ability", including operation domain, information domain, watershed, content creation domain, security ethics, problem solving domain and specific occupation related domain seven aspects of literacy domain. Greene proposes that digital literacy consists of two dimensions: "ability" and "knowledge", namely, the ability to obtain effective information and the knowledge to review and integrate the information obtained, which are applied to libraries and in the field of digital literacy education.

Although domestic scholars have not yet reached a consensus on the core elements of digital literacy, their overall understanding is constantly improving. For example, Zhang Lingqian explores the current situation and causes of digital literacy gap, proposed to pay attention to the cultivation of information awareness, improve the level of information construction, eliminate computer anxiety, build digital citizen education, and promote the balanced development of education. Zeng Hong studied the digital divide of trade circulation in rural areas, and put forward to improve the infrastructure and promote the construction of a new trade circulation system based on the "Internet +" thinking [4]. Yang Shuang et al. has constructed a digital literacy evaluation index system for university teachers, which was analyzed through five dimensions: digital technology use, digital information management, digital content creation, digital community construction and digital security capability [5]. Su Lanlan and other explorers have built a farmer digital literacy evaluation index system in four aspects: digital general literacy, digital social literacy, digital creative literacy and digital security literacy, which can effectively activate the coordinated development of rural "digital infrastructure, digital industry, digital life, digital ecology, digital governance and digital governance" [6]. Jiang Minjuan and other scholars build the "five forces" model of civic digital literacy, including perception power, integration power, absorption power, practical power and development power, and put forward four major projects of building foundation, enabling, expansion and strengthening governance, which play a role in promoting the digital transformation of society.

Based on the above research and analysis of digital literacy, the research objects focus on college students, teachers, farmers, library staff, civil servants; the research direction includes current situation investigation, model construction, implementation path, improvement strategy, but less research on the ability of digital literacy of new professional farmers. This paper holds that professional farmers' digital literacy should have a collection of digital acquisition, production, use, evaluation, interaction, sharing, innovation, security, ethics and a series of qualities and abilities. Improving their digital literacy focuses on model building through four aspects: digital device operation, digital application, digital security specifications, and digital innovation.

3 Subdivision Algorithm for Digital Literacy Ability Evaluation of New Professional Farmers

3.1 First-Level Index Design

Through literature analysis method, this study compiled four primary indicators of new professional farmers: digital equipment operation (B1) refers to the basic ability to use and operate digital equipment, mainly reflected in practical ability; digital application (B2) refers to use digital software, such as APP commonly used in mobile terminals; digital governance (B3) refers to the ability to comply with the code of conduct and digital virtual space; digital innovation (B4) refers to digital learning ability.

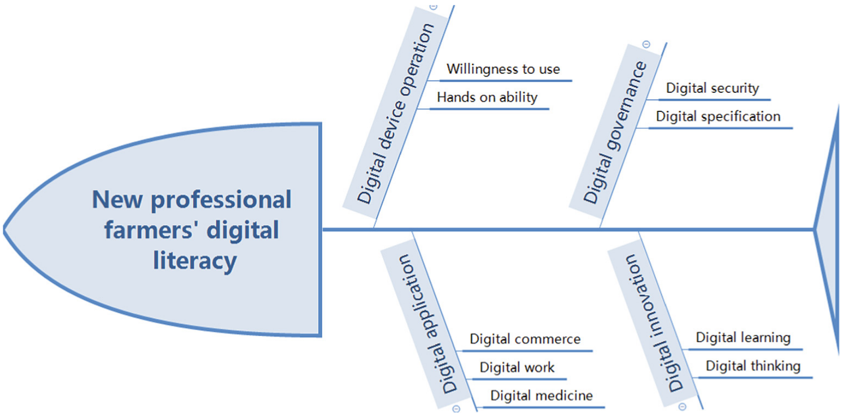

Fig. 1. Evaluation index system of digital literacy ability of new professional farmers

3.2 Design of Secondary Indicators

The four dimensions of the framework description of new professional farmers. To make the model more operational, the four dimensions, according to digital work, digital commerce, digital application capability, digital security and digital specification, and digital innovation capability, including digital learning and digital thinking. The evaluation index system of digital literacy ability of new professional farmers is shown in Fig. 1.

3.3 Two-Step Cluster Algorithm

Two-step clustering is a kind of intelligent clustering analysis method developed in recent years, suitable for large number and complex structure analysis, the most important is can handle classification data and quantitative data, by the SPSS software automatically find the optimal number of cluster categories, also can output more indicators to assist analysis, including cluster quality judgment and predict the importance of clustering, etc. Two-step Cluster is divided into two stages. The first stage is called pre-clustering, which clusters the source data into multiple sub-categories; The second stage uses the clustering results of the first stage to cluster again, and finally clusters these small clusters into the expected number of clusters. The specific calculation steps are as follows:

a) The cluster feature tree is established, and the first record in the data set is placed on a leaf node initiated by the tree root, which contains all the variable information in this record. The distance measurement is used as the similarity criterion. According to its similarity with the current node, the existing node is merged to generate a new node, and the clustering feature tree is established by recursion. The distance measurement model adopts logarithmic similarity, and the calculation formula is:

$$d(i, j) = \delta_i + \delta_j + \delta_{<i,j>} \tag{1}$$

Where: d(i,j)is the distance between two clusters i and j; δ_i and δ_j are the likelihood function values of cluster i and j respectively; $< i,j >$ is a new cluster generated by the combination of clusters i and j.

$$\delta_v = -N_V\left(\sum_{k=1}^{K^A}\frac{1}{2}\log\left(\overset{\Lambda}{\delta k}^2 + \overset{\Lambda}{\delta vk}^2\right) + \sum_{k=1}^{K^B}\overset{\Lambda}{E}_{vk}\right) \quad (2)$$

Where: N_V is the number of data records of cluster v; K^A is the total number of all continuous variables; K^B is the total number of all classification variables; $\overset{\Lambda}{\delta_k}^2$ is the estimated variance of the k-th continuous variable in the whole data set; $\overset{\Lambda}{\delta_{vk}}^2$ is the estimated variance of the k-th continuous variable in cluster v; $\overset{\Lambda}{E}_{vk}$ is the estimated mean value of the k-th continuous variable in cluster v.

$$\overset{\Lambda}{E}_{vk} = -\sum_{l=1}^{L_k}\frac{N_{vkl}}{N_v}\log\frac{N_{vkl}}{N_v} \quad (3)$$

Where: L_k represents the number of categories of the k-th classification variable; N_{vkl} is the number of data records of the k th classification variable in cluster v in the l classification.

b) By combining the leaf nodes with the merge clustering algorithm, a group of clustering schemes with different number of clusters can be generated. According to the BIC criterion, various clustering schemes are compared, and the number of clusters is automatically selected to optimize the clustering scheme. For clusters, the calculation formula of BIC is:

$$BIC(j) = -2\sum\nolimits_{j=1}^{j}\delta_j + m_J\log\ N \quad (4)$$

$$m_j = J\left\{2K^A + \sum_{k=1}^{K^A}(L_k - 1)\right\} \quad (5)$$

Where: N is the number of records in the dataset; $\sum_{j=1}^{j}\delta_j$ is the maximum value of the likelihood function; m_J is the number of model parameters.

4 Example Solution

The questionnaire is designed with the sample background information, mainly including gender, age and education background, and 18 core quantity index items. After the designed questionnaire was adjusted for the preliminary test, 150 questionnaires were issued again, and 136 questionnaires were recovered, with a recovery rate of 90.6%.

4.1 Reliability and Validity Test

To ensure the reliability of the questionnaire, the reliability and validity. First, the reliability coefficient of the consistency coefficient of the core item is tested. The value is 0.905, greater than 0.9, and the coefficient of each dimension is greater than 0.7, indicating the high credibility of the research data.next, To construct the structural validity analysis of the questionnaire, The first exploratory factor analysis of the rotation matrix results found the core item of item 4, Lower factor loading coefficient for questions 6,7, and 14, And has the "double load" phenomenon, Therefore, the four items were deleted, After a second, exploratory factor analysis, Good results, The resulting validity of 4 factors was eventually formed, The value of the KMO is 0.871, The Bartlett spherical test value is 1014.168 (sig = 0.000), Therefore, the core items are very suitable for factor analysis, It indicates that the questionnaire has high structural validity.

4.2 Two-Step Cluster Analysis

This study was conducted by using a two-step clustering to analyze the sample data, and the analysis results are shown in Table 1.

Table 1 shows the results of the sample cluster distribution generated using the two-step cluster analysis, and we can see that a total of 3 cluster categories were generated, with each sample size (N) of 55,61 and 20, respectively. The sample size of category 3 was too small, representing only 14.7%, and it was renamed according to the results of the later chi-square analysis.

Figure 2 can see that the model summary results input 4 variables, the cluster is 3 categories, and the quality of the cluster is good. On the whole, the model fits is ideal and is acceptable.

Figure 3 shows that among the four factors of the predictive variables, digital governance is the most important to the model, followed by digital application and digital device operation. Relatively speaking, the digital innovation factor has the least effect on cluster modeling.

Table 1. Two-step Cluster Distribution

		N	% of Combined	% of Total
Cluster	1	55	40.4%	40.4%
	2	61	44.9%	44.9%
	3	20	14.7%	14.7%
	Combined	136	100.0%	100.0%
Total		136		100.0%

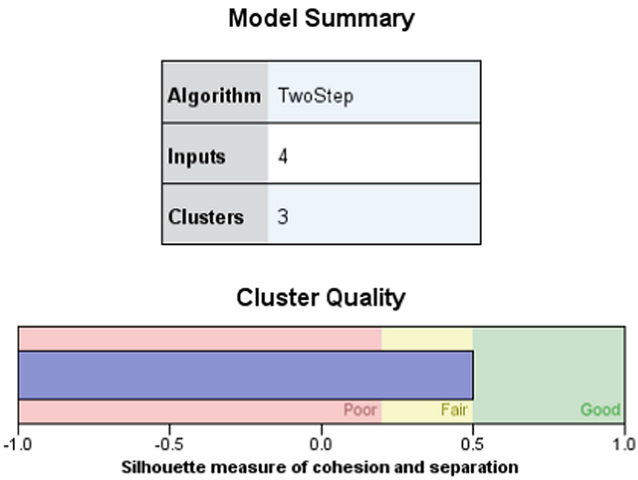

Fig. 2. Two-step cluster summary and cluster quality result diagram

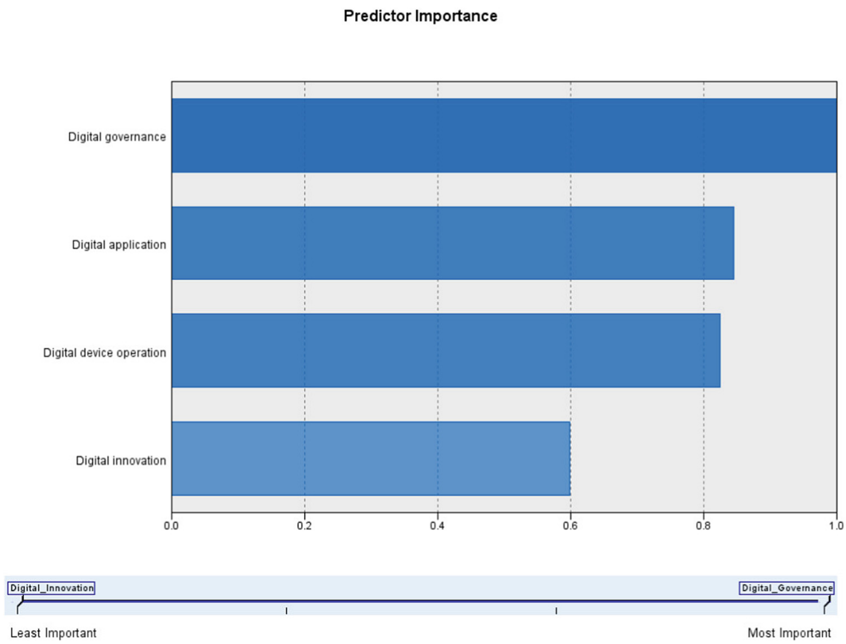

Fig. 3. Results plot of predictive variables

Table 2. Two-step cluster samples with factor ANOVA results

	class 1(N = 55)	class 2(N = 61)	class 3(N = 20)	*F*	*P*
Digital Device Operation	1 + 0.0	1.04 + 0.13	1.72 + 0.31	193.95	0.00^{**}
Digital application	1 + 0.0	1.34 + 0.31	1.75 + 0.28	82.98	0.00^{**}
Digital Governance	1 + 0.0	1.21 + 0.24	1.74 + 0.29	105.16	0.00^{**}
Digital Innovation	1 + 0.0	1.22 + 0.35	1.78 + 0.38	58.63	0.00^{**}

* $P < 0.05$; ** $P < 0.01$

5 Analysis of Digital Literacy Characteristics of New Professional Farmers

5.1 Variance Analysis

Rename the 3 categories of samples based on the results of the two-step clustering. In this paper, we analyze the differences of four variables (i. e., four factors of the digital literacy of new professional farmers), distinguish them by ANOVA, and then rename the cluster samples. The summary table of ANOVA is shown in Table 2.

The questions designed in this questionnaire are all answered with "yes" or "no", If you choose Yes, you will get 1 point, Select No, score 2 points, From the results in Table 2, we can see that the digital literacy of the new professional farmers is significantly different in the digital equipment operation, digital application, digital governance and digital innovation, Looking at the characteristics of the three categories, it can be found that: Class 1 performs very well in digital literacy in all four aspects, Named it as a "high-digital literacy farmer"; Class 3 was very weak in all aspects of digital literacy ability, Especially weaker in digital innovation, digital applications, Named it as a "low-digital literacy farmer"; Class 2 Except for the good operation performance of digital equipment, Other aspects are mediocre, Named it as a "General-digital literacy farmers". Considering the above analysis, these three types of professional farmers can be clearly distinguished. Combined with the quality evaluation of the two-step clustering analysis, the final clustering effect is good.

5.2 Analysis of Age and Educational Background of Different Types of Professional Farmers

It further analyzes whether the performance of the three types of new professional farmers is different in age and educational background, and whether it is more convenient for training institutions to carry out stratified and classified digital literacy ability improvement training. For example, Table 3 conducts chi-square analysis in terms of gender, age and educational background.

Table 3. Results of Chi-square analysis on gender, age and education of New Professional

	option	Clustering categories			Total	X^2	P
		high-digital literacy farmer	General-digital literacy farmers	low-digital literacy farmer			
sex	**man**	26 (47.3%)	16 (26.2%)	7 (35%)	49 (36%)	5.568	0.062
	woman	29 (52.7%)	45 (73.8%)	13 (65%)	87 (64%)		
age	**Under 25**	24 (43.6%)	16 (26.2%)	3 (15%)	43 (31.6%)	10.12*	0.038
	26–45 Years old	29 (52.7%)	37 (60.7%)	16 (80%)	82 (60.3%)		
	Over 45 years old	2 (3.6%)	8 (13.1%)	1 (5%)	11 (8.1%)		
record of formal schooling	**High school and below**	12 (21.8%)	13 (21.3%)	13 (65%)	38 (27.9%)	15.998**	0.000
	University and above	43 (78.2%)	48 (78.7%)	7 (35%)	98 (72.1%)		

* $P < 0.05$; ** $P < 0.01$

The three cluster samples included in Table 3 (high digital literacy, general digital literacy, and low digital literacy) showed no significant difference in gender, indicating that there is no relationship between the level of digital literacy of new professional farmers and gender. There was a significant difference in age, with a P-value of 0.038, less than 0.05, at the 0.05 significance level, indicating that professional farmers in the three age groups differ in digital literacy ability. In recent analysis, we can see that professional farmers under 25 years old have the most high-quality literacy, professional farmers aged 26–45 years old have the most digital literacy, and the samples over 45 years old are relatively small, so the research has little significance. In terms of education, highly educated professional farmers and low educated professional farmers have obvious differences in three categories, under the 0.01 significance level, the P value of 0.000, far less than 0.01, shows that college and above new professional farmers (that is, high degree) generally have higher digital literacy ability, high school and below new professional farmers (i. e., low degree) digital literacy ability in general or low.

6 Conclusions and Suggestions

6.1 Study Conclusion

First, the focus of cultivating the digital literacy of the new professional farmers is to strengthen the digital governance capacity. Improving digital governance ability is mainly manifested in identifying illegal behaviors such as online fraud or online pyramid scheme, can realize the seriousness of online game addiction and short video addiction, can identify vulgar and indecent language on online platform, can report or complain about digital information that violates cultural taboos, reflects social ugliness and unhealthy ideas; conforms to network behavior norms and application boundaries; can set and protect personal digital identity and online personal privacy.

Second, the new professional farmers have good performance in the operation of digital equipment, but the weakest is the digital innovation ability, for the general digital literacy people, especially the APP applications based on mobile terminals, such as online shopping and mobile payment; online ordering of hotels, tickets or travel tickets; online conference or online office; online registration or online consultation. The improvement of digital innovation ability should start from digital learning, cultivate the digital thinking of new farmers, use digital tools to evaluate and adjust and optimize the optimization of agricultural production technical solutions; be able to creatively use digital technology to develop new products or new processes; and creatively solve technical problems in agricultural production in digital environment.

Third, In terms of age, young and new professional farmers generally have higher digital literacy ability than that of the elderly. People aged 25 and under are much more capable than those over the age of 26 in digital device operation, digital application, digital governance and digital innovation. Especially in terms of digital applications, the digital application ability of young people is far higher than that of the elderly, and the difference is very obvious.

Fourth, Improving the educational background of the new professional farmers is an effective means to improve the digital literacy ability. According to the study, it is found that people with high educational level have strong digital literacy ability, and those with low educational level have relatively weak digital literacy ability, and the two are positively correlated. Through education, to enhance and enhance the understanding of learning, to improve the educational level of professional farmers, to stimulate their inner desire to learn, and to promote the improvement of digital literacy.

6.2 Suggested Measures

First, improve the digital governance system and enhance the awareness of the rule of law in rural areas. In view of the most prominent privacy protection, urban and rural digital divide, digital copyright, data security and other issues of management, we should improve the relevant laws and regulations, optimize the rural mechanism, and establish a scientific and effective decision-making thinking. At the government level, we will promote government data sharing and openness, and cooperate with digital technology enterprises to empower the construction of smart villages. At the enterprise level, the legal ownership of platform data is determined, the responsibilities are defined, and the obligations of digital platform self-discipline and social responsibility are formed. For the new professional farmers, we should clarify and effectively protect the legal rights and interests of individuals such as personal information and digital assets, and abide by the language norms of the digital age.

Second, Strengthen digital application and digital innovation training, and quickly promote and implement the implementation of agricultural digital technician positions. Digitalization is gradually integrated into the daily life of rural areas, in urgent need of personnel engaged in the application, promotion and service activities of agricultural production and rural life. To carry out digital agriculture and rural field talents to the countryside activities, Targeted training on digital application and digital innovation level of "agriculture, rural areas and farmers" and new professional farmers, Select outstanding personnel to serve as "agricultural digital technicians", The main task is to collect data in the field, Based on these data, using digital technology, Analyze the targeted solutions; Also need to explain to the farmers around, demonstration of digital agricultural production machines, facilities, To guide and help them in scientific breeding, demonstration planting, pest prevention and control, large-scale promotion, They are both the leaders of digital technology into the countryside, It is also the escort to help the smart countryside.

Third, School, enterprise and government multi-master linkage to improve the educational level of new professional farmers. Referring to the successful experience of developed countries, the improvement of digital literacy ability cannot be separated from the coordination of the government, enterprises, universities and other departments, giving full play to the forces of various parties to formulate personalized education improvement plans. Chinese government department has promoted the for migrant workers, veterans, laid-off workers, new professional farmers 4 higher vocational personnel "millions" the plan of action, on the training mode through "study", "engineering alternation" teaching "door", according to the enterprise demand tailored talent training plan, which expanded the group of high education in our country, covers the national of different ages, different professional groups, realize the universality and comprehensiveness of vocational education. Colleges and universities of higher learning should play a main role in the implementation of the curriculum plan and training content of digital literacy for urban and rural residents. For example, higher vocational colleges should set up a digital literacy training center, which is specially responsible for the decision-making and overall design of the digital literacy education strategy, and urge and coordinate the relevant departments to organize and implement the teaching objectives, course content design, teaching schedule, students' learning requirements and assessment methods of the digital literacy courses. Through the joint efforts of various parties, to ensure the quality and effectiveness of training programs for urban and rural residents, and promote the improvement of national digital capacity.

Acknowledgments. the "14th five year plan" for the development of philosophy and Social Sciences in Guangzhou, "Research on the cultivation of digital literacy of urban and rural residents in Guangzhou under the goal of common prosperity" (2022gzgj190); The key project of "climbing plan" of Guangdong Province's scientific and technological innovation strategy in 2022 (pdjh2022a0872) "investigation and analysis of the current situation of digital literacy of new professional farmers under Rural Revitalization".

References

1. GILSTER .P. Digital literacy [M].New York: Wiley Publish-ing,1997:25–48.
2. ESHET-ALKALAI Y. Digital Literacy: A Conceptual Framework for Survival Skills in the Digital era [J]. Journal of Educational Multimedia and Hypermedia, 2004, 13(3): 93-106.
3. European Commission. DigComp 2.1: The digital competence framework for citizens with eight proficiency levels and examples of use[EB/OL]. [2021–11–23]. https://publications.jrc.ec.europa.eu/repository/handle/JRC106281.
4. Zeng Hong. The digital gap and improvement path of commercial circulation in rural areas [J]. Business Economics Research, 2021 (10): 124-127.
5. Yang Shuang, Zhou Zhiqiang. Research on the construction of digital literacy evaluation indicators for university teachers [J]. Modern Information, 2019, 39 (03): 59–68+100.
6. Su Lanlan, Peng Yanling. Farmers' digital literacy, rural elite identity and rural digital governance participation [J]. Agricultural Technology and Economy, 2022 (01): 34-50.