Received November 26, 2018, accepted November 30, 2018, date of publication December 5, 2018, date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2885069

Evaluation of Regional Rural Information Environment Based on Fuzzy Method in the Era of the Internet of Things

YANFENG JIN[1,2], GANG LI[1], AND HUIFENG ZHANG[2]
[1]School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China

Corresponding author: Gang Li (114560731@qq.com)

This work was supported by the Hebei Province Higher Education Science and Technology Research Project under Grant QN2018304.

ABSTRACT Internet of Things technology has brought new opportunities for rural informatization. The quality of rural information environment is directly related to the level of rural informatization. So accurately understanding the status and level of information environment in rural areas is a prerequisite for rural informatization. This paper assesses differences in rural information environment by developing a framework based on a fuzzy comprehensive evaluation method to measure information environment changes. The assessment focuses on three aspects: one is the evaluation of the overall rural information environment based on macro-factors in time dimension, the other is in 31 different provinces in space dimension, and the third is the evaluation of regional rural information environment based on macro-and subjective factors. The Probit model reveals the influence of various factors on the rural information environment, which constitutes the primary and secondary indicators of the evaluation system. The fuzzy comprehensive evaluation method is adopted to evaluate the rural information environment, and the cluster analysis is applied to the results. The results show that the differences in rural information environment exist objectively, which are the result of a variety of subjective and objective factors. In addition, the differences in rural information environment promote the diversified development path of rural information construction. These findings have clear policy implication for the rural ministry, since understanding the changes and differences in the rural information environment is important for their successful development. Results of the experimental applications of this evaluation system are given to illustrate the proposed method.

INDEX TERMS Cluster analysis, Fuzzy comprehensive evaluation, Internet of Things, Probit model, rural information environment.

I. INTRODUCTION

During the past decade, the Internet of Things has developed in a variety of directions It has fully applied the new generation of IT technology to various industries, connected the physical network with the Internet, and realized the interconnection of all things in human society. The Internet of Things itself is a ''limited network'' for specific management objects. It aims at realization, control and management. It connects management objects through sensors, recognizers and networks to realize intelligent management and control of information perception, recognition, intelligence processing, situation judgment and decision-making execution. The application of the Internet of Things will bring a huge amount of data to the communication network. And the Internet and information processing technology will bring a magnitude order demand growth. The development of the Internet of things has provided unprecedented opportunities for the realization of rural informatization. Rural informatization is a process in which communication technology and computer technology are widely applied and popularized in rural production, life and social management. As a part of social informationization, It is first of all a social and economic form and a conceptual description of the rural economic development to a specific process. From the perspective of technology, rural informatization not only agricultural information technology, but also microelectronics technology, communication technology, optoelectronic technology in rural production, life, management and other aspects of universal and

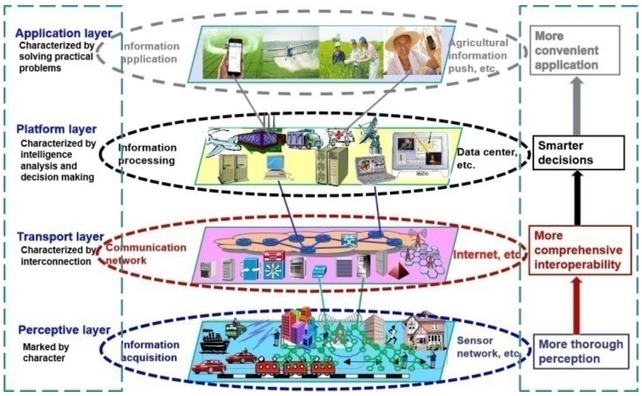

FIGURE 1. Internet of things - from transmission to perception and application for analysis and processing.

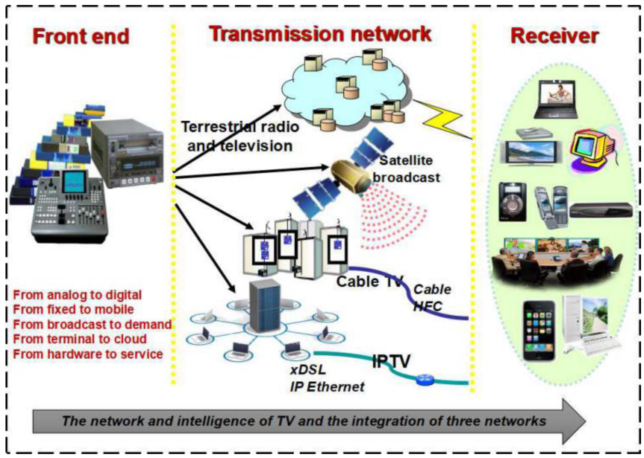

FIGURE 2. Integration of three networks to promote the transformation of information environment.

systematic application process. On the other hand, from the perspective of process, rural informatization includes the process from traditional agriculture to modern agriculture and then to information agriculture, and also includes the process from primitive society to capital society and then to information society [1]–[3].

However, there are few studies on the rural information environment at home and abroad, and most of them are around the status of the information environment, the related elements and the optimization strategy. Feng(2006) thought that the advantages and disadvantages of the rural information environment were directly related to the degree of informatization, analyzed the role of the rural information environment and the related elements, and put forward the strategy of optimizing the rural information environment. Huang(2006) carried out an empirical analysis based on the questionnaire survey of rural information environment in Sichuan. He studied the status, causes and problems of rural information environment, and put forward the countermeasures and suggestions to improve the rural information environment. Liu (2009) analyzed the status of rural information in Guangzhou and found out the existing problems. It was believed that the optimization of information environment in rural areas of Guangzhou needed to be led by the government. And constructed a scientific and rational organization, training system and mechanism to increase the integration of agricultural information resources and to promote the innovation of the rural information application system. This study mainly used qualitative analysis method to discuss, and there was a few empirical research based on questionnaire data. But no scholars have conducted empirical research on the macro rural information environment based on panel data, especially the evaluation and research on the advantages and disadvantages of rural information environment [4].

It is always the most important task to solve three dimensional rural issues in China. We should adhere to the priority development of agricultural and rural areas. Establish and improve the system mechanism and policy system of urban and rural integration development in accordance with the general requirements of industrial prosperity, ecological livability, rural wind civilization, effective governance and rich life, so as to accelerate the modernization of agricultural and rural areas [9]–[13]. From the information development process, the advent of the ''Internet accelerated speed'' era has provided an unprecedented good environment for the development of rural informatization. The rapid development of information technology characterized by intelligence, networking and digitalization provides strong technical support for the rural informatization. The country's deployment of the innovative action plan, and the internet accelerated speed, and the big data strategy, and many other things provide policy guarantee for the development of rural informatization. Farmers' information consciousness and information ability have been greatly improved. Rural information infrastructure is also constantly improving. All of these provide space for the development of rural informatization in the two aspects of rural information demand and supply. Of course, the rural informatization in our country is still at the initial stage. The information environment of rural areas is uneven. And the development of regional rural informatization is uneven too. Therefore, in the exploration of the development path of rural informatization, we cannot generalize. Many factors affect the rural informatization environment, such as different regions, different levels of economic development, different social cultures and so on [5]–[8].

At the same time, in the era of ''Internet accelerated speed,'' the rural environment as well as farmers' production and life style have undergone profound changes. The biggest change is the way farmers get information. Traditional information channels such as rural broadcasting, neighborhood communication and rural libraries are gradually replaced by the Internet and mobile Internet, computers and smart phones. In this context, the level of rural informatization is closely related to the popularity of the Internet. And the difference between information and environment have also become the main cause of the uneven level of information in the rural areas. Therefore, in the context of ''Internet accelerated speed,'' the regional rural information environment was studied in order to provide reference for the differentiation of rural information development path.

Due to the unclear understanding of the rural information environment, the formulation of rural informatization policies is not targeted. Therefore, it is necessary to evaluate the rural information environment and formulate different rural informationization policies with the evaluation results. In the next section, after a statement of the basic problem, various situations involving possibility knowledge are investigated: first, a discrete Probit model is proposed; then the evaluation index system of rural information environment is established; the comprehensive evaluation method based on fuzzy theory is studied; lastly, the information environment of rural areas is evaluated.

II. ANALYSIS OF RURAL INFORMATION ENVIRONMENT AND ITS INFLUENCING FACTORS

A. DEFINITION OF RURAL OF INFORMATION ENVIRONMENT

The information environment is an environment in which a society or a group is exposed to possible information and the overall composition of its communication activities. Although the concept of information environment is common and very important, the definition is not clear enough. Feng (2006) gave the definition of information environment from a macro point of view: the information environment referred to the sum of all natural, social and psychological factors related to human information activities. He believed that the information environment was a system composed of people, information resources and information technology. Only when the three achieved a high degree of coordination and unification could we achieve a good cycle of information environment. Du (2009) proposed a more complete definition from the microscopic point of view: information environment corresponding to a specific subject. And it was the sum total of all information content factors and information media factors that revolved around and may affect the subject at a time,. She believed that the main factors that affected the information environment included information content factors, information user factors and information spread time and space factors. For the rural information environment, there is no doubt that the most important subject is the peasant. While the rural areas are limited in the geographic location of the information environment, and the improvement of the rural infrastructure and information technology determines the effectiveness of the information carrier. Therefore, rural information environment can be defined as the sum of subjective information consciousness and behavior of rural information subjects and their objective influencing factors [14]–[18].

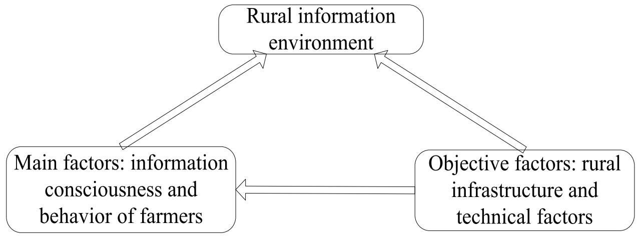

FIGURE 3. Definition of rural information environment.

B. THE INFLUENCING FACTORS ANALYSIS OF RURAL INFORMATION ENVIRONMENT

From the macro level, the rural information environment is affected by many factors, such as the economic environment, the political environment, the natural environment, the geographical position and the historical conditions. At the micro level, the popularity of the Internet, the number of computer access, the number of mobile phones, the popularity of television and radio have become important factors affecting the rural information environment. The advantages and disadvantages of the rural information environment reflect the allocation of rural information resources to a certain extent. According to the definition of the rural information environment, the influence factors of rural information environment are analyzed from two aspects of the internal and external.

(1) The macro factors that affect the rural information environment. ''Internet accelerated speed'' is a new form of Internet development under the innovation 2.0. It is the new form of the evolution of the Internet form and the development of the economic and social development driven by the knowledge society innovation 2.0. ''Internet + agriculture'' has greatly promoted the Internet of agricultural information. New farmers can use the Internet to obtain the latest information, so as to grasp the price trend of agricultural products in time. They can arrange agricultural production activities reasonably, and enter the ecological chain of rural e-commerce through the Internet platform, and increase their own income. Therefore, ''Internet accelerated speed'' is very obvious for the improvement of the rural information environment. According to the main research results of the information evaluation system at home and abroad, the following aspects can be analyzed.

First of all, rural information infrastructure and technology are the basic resources of rural information environment, and it is also an important prerequisite for the optimization. And its construction level is an important part of the rural information environment. The main medium of communication between the countryside and the outside world is the mobile phone and the fixed telephone. The telephone line infrastructure is the main way of Internet access. The proportion of the established administrative villages reflects the construction degree of the telephone infrastructure in the countryside. In addition, the coverage rate of the TV comprehensive population can reflect the basis of information resources to a certain extent. The development of modern optical networks reflects the popularity of modern Internet. Secondly, the optimization of rural information environment needs high-quality professional talents to achieve the goal, so as to continuously promote the improvement of rural informatization level. Rural science and technology personnel is an important force in the construction of rural information environment. It is an important guarantee for the continuous promotion of rural information. Rural college students have high professional quality and business ability. They are also an important force in the construction and optimization of the information environment in the countryside in the future. Thirdly, the rural information network coverage reflects the application status of the rural information infrastructure. And the last is the input and output of rural informatization.

(2) The main factors affecting the rural information environment. The most important subject of the rural information environment is the peasants. And the information consciousness and ability of the farmers are the most important external factors of the rural information environment. These factors include the peasants' age, sex, education level, income level, family status, geographical position, the cognition of information, and the awareness of information, and the ability to accept, and many other aspects.

C. THE MODEL OF INFLUENCE FACTORS TESTING OF RURAL INFORMATION ENVIRONMENT

Many objective factors affecting the rural information environment have been approved by experts and scholars in the field. And it has been analyzed and verified in many other documents. This section mainly aims at the subjective factors affecting the rural information environment. Through the questionnaire survey, the Probit model is used to find out the key factors and abandon the unrelated or lower factors [19].

1) DISCRETE PROBIT MODEL

The Probit model is a generalized linear model which obeys normal distribution. The simplest Probit model means that the interpreted variable Y is a 0, 1 variable, and the probability of the event depends on the explanatory variable, that is, P (Y=1) =f (x), the probability of Y=1 is a function of X, in which F () obeys the normal distribution. From the perspective of support decision, the evaluation of influencing factors can be regarded as a discrete variable. The size of the factors affecting rural information environment is designed according to the Li Kate scale. It is divided into five grades: 1 is completely without influence, 2 indicates almost no impact, 3 indicates a certain influence, 4 indicates a greater impact,5 indicates a very big impact. Therefore, the Probit model of rural information environment can be constructed [20]–[23]:

$$Y = \alpha X_i + \beta \tag{1}$$

Y is the decision-making variable of the rural information environment. The larger the Y, the stronger the impact on the rural information environment. X is an independent variable, which represents the age, sex, education level, income level and family status of farmers. The vector matrix is made up of the objective factors that affect the rural information environment. Beta represents a random error variable, which obeys normal distribution, indicating the total effect of the model ignored. α represents the estimated coefficient vector. Given the variables such as age, sex, education level, income level and family status, the probability of access to alpha coefficient can be expressed as:

$$P(Y_i = 1) = P(Y_i < \mu_1) = \phi(\mu_2 - X_i'\alpha) \tag{2}$$
$$P(Y_i = 2) = \phi\left(\mu_2 - X_i'\alpha\right) - \phi(\mu_1 - X_i'\alpha) \tag{3}$$
$$P(Y_i = 3) = \phi\left(\mu_3 - X_i'\alpha\right) - \phi(\mu_2 - X_i'\alpha) \tag{4}$$
$$P(Y_i = 4) = \phi\left(\mu_4 - X_i'\alpha\right) - \phi(\mu_3 - X_i'\alpha) \tag{5}$$
$$P(Y_i = 5) = 1 - \phi\left(\mu_4 - X_i'\alpha\right) \tag{6}$$

The likelihood function is as follows:

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m} [P(Y_i = j)]^{d_{ij}} \tag{7}$$

Log on both sides:

$$\ln L = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} \ln [P(Y_i = j)] \tag{8}$$

According to the formula (2) - (8), the maximum likelihood solution of α can be obtained, and the estimated values of the parameters are calculated by the least square method. And then the Probit probability distribution model can be constructed, and the model about the influence of the rural information environment can be obtained.

2) VARIABLE DESIGN AND PROPOSITIONAL HYPOTHESIS

a: VARIABLE DESIGN

Based on the Probit model of the rural information environment, the corresponding variables and indicators are designed and defined, as shown in Table 1.

b: PROPOSITIONAL HYPOTHESIS

Based on the variables needed in the model, analyze the relevant factors affecting the rural information environment and make the following propositional assumptions::

Hypothesis 1: the quality of rural information environment is positively related to the age factor of farmers.

Hypothesis 2: the advantages and disadvantages of rural information environment are related to rural sex factors.

Hypothesis 3: the quality of rural information environment is positively related to the education level of farmers.

Hypothesis 4: there is a positive correlation between the advantages and disadvantages of rural information environment and the number of peasant family members.

Hypothesis 5: there is a positive correlation between the advantages and disadvantages of rural information environment and farmers' family income.

Hypothesis 6: the advantages and disadvantages of rural information environment are related to farmers' marital status.

Hypothesis 7: the advantages and disadvantages of rural information environment are related to the source of farmers' income.

Hypothesis 8: there is a positive correlation between the advantages and disadvantages of rural information environment and the development of rural residents' transportation.

TABLE 1. Design and definition of variables.

variable	Variable name	Variable value	The meaning of variable value
Y	Influence degree of information environment	{1,2,3,4,5 }	1= had no effect at all, 2= had little effect, 3= had some effect, 4= had a great influence, 5= had a great influence.
X1	age	{1,2,3,4,5 }	1=18 years old and below, 2=19 to 28 years old, 3=29-38 years old, 4=39 to 48 years old, 5=49 years old and above
X2	sex	{1,2}	1=male，2=female
X3	Educational level	{1,2,3,4,5 }	1= primary school and below, 2= junior high school, 3= high school or technical secondary school, 4= specialist, 5= undergraduate and above.
X4	Family income	{1,2,3,4,5 }	1=1000 and below, 2=1000-3000, 3=3000-5000, 4=5000-7000, 5=7000 above
X5	Number of family members	{1,2,3,4,5 }	1=1, 2=2-3, 3=4-5, 4=6-7, 5=8 and above
X6	Marital status	{1,2,3,4}	1= unmarried, 2= married, 3= divorced, 4= widow
X7	Source of income	{1,2}	1= agricultural income, 2= non agricultural income
X8	traffic condition	{1,2,3,4,5 }	1= is very inconvenient, 2= is inconvenient, 3= is general, 4= is convenient, 5= is very convenient.
X9	Willingness to accept information	{1,2,3,4,5 }	1= conflict, 2= is unwilling, 3= is general, 4= is willing, 5= is very willing.
X10	Degree of information cognition	{1,2,3,4,5 }	1= is very confused, 2= doesn't understand, 3= is general, 4= understands, 5= knows very well.
X11	Information acceptance ability	{1,2,3,4,5 }	1= is very bad, 2= is bad, 3= is general, 4= is strong, 5= is very strong.

Hypothesis 9: the advantages and disadvantages of rural information environment are positively related to farmers' willingness to accept information.

Hypothesis 10: the advantages and disadvantages of rural information environment are positively related to farmers' cognition of information.

Hypothesis 11: there is a positive correlation between the advantages and disadvantages of the rural information environment and the information acceptance ability of the farmers themselves

c: DATA SOURCES

The data needed in this section comes from a questionnaire survey of farmers in 30 administrative villages of 15 townships in Lingshou County, Hebei. The choice of Lingshou county mainly considers that the county's 15 towns cover three landforms in plain, hilly and mountainous areas, and the data are strongly representative. According to the above assumptions, 11 related questions were designed in the questionnaire. The answer options for each question were set according to the Likert scale. 1 means no effect at all, 2 means there is some influence, 3 means there is influence, and 4 means there is comparison. Great impact, 5 means very big impact. The questionnaire is mainly accomplished by combining network questionnaire and field visit. In the process of data collection, a preliminary survey is made using the network questionnaire. According to the feedback situation, the survey of the administrative villages with relatively centralized data coverage is carried out in the field, and the data collection is finished through the assistance of the village committee. Through the arrangement of the questionnaire, A serious questionnaire with missing information was discarded, and 955 valid questionnaires were collected, and the specific distribution of the questionnaires is shown in Table 2.

TABLE 2. Sample distribution and quantity in the study.

Town name	Administrative village	Sample size
Lingshou	Xituo	43
	New village	46
Sanshengyuan	Tongxia	37
	Sanshengyuan	40
Beiwa	Dangjiazhuang	41
	Zhushi	27
Niucheng	Niucheng	31
	Hudong	29
Goutai	Beigoutai	35
	Nanzhule	21
Tashang	Tashang	39
	Wanli	42
Yanchuan	Nanyanchuan	33
	Guanzhuang	23
Qingtong	Nanqingtong	29
	Hujiatuan	25
Ciyu	Ciyu	48
	Dongbaishan	29
Chatou	Xichatou	38
	Taitou	19
Tanzhuang	Beitanzhuang	56
	Shanmenkou	48
Nanzhai	Naizhai	21
	Majiazhuang	24
Chenzhuang	Zhong village	41
	Xiazhuang	17
Zhaitou	Zhaitou	25
	Zhangjiazhuang	16
Nanying	Nanying	21
	Manshan	11
	Total	955

The distribution of samples is related to the development of regional economy. Generally speaking, the villages with developed economy and superior geographical position have relatively dense population and relatively large number of samples. On the contrary, the villages with poor economy and inconvenient transportation have less population and less sample size. For example, villages in plain areas have larger sample sizes, while villages in less developed mountain areas have smaller samples.

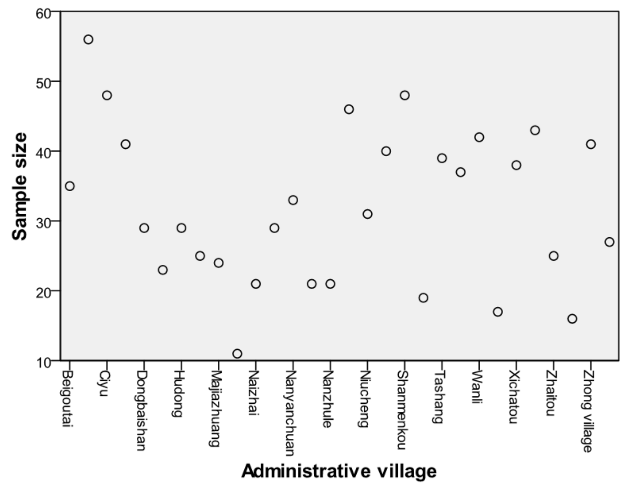

FIGURE 4. The sample size and distribution.

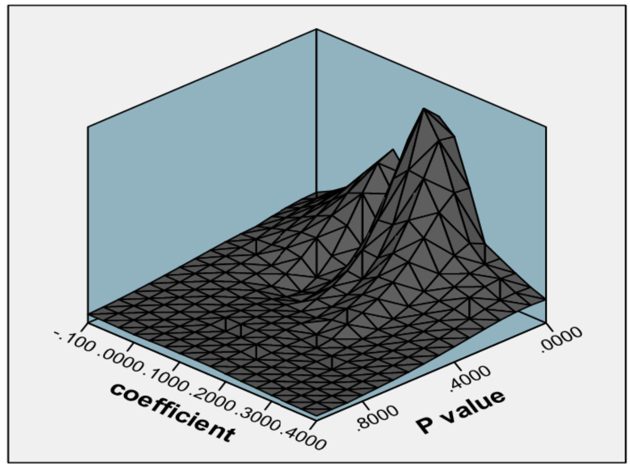

FIGURE 5. Three dimensional density relation between coefficient and P value.

TABLE 3. Test results of various influencing factors to information environment.

variable	coefficient	P value	Correlation
X1	0.0561**	0.0017	Positive correlation
X2	-0.0224	0.8568	Unrelated
X3	0.02815***	0.0004	Positive correlation
X4	0.1073**	0.0041	Positive correlation
X5	0.0205*	0.0263	Positive correlation
X6	0.1848	0.0875	Unrelated
X7	0.3734*	0.0181	Positive correlation
X8	0.1791*	0.0203	Positive correlation
X9	0.1733**	0.0956	Positive correlation
X10	0.233**	0.0148	Positive correlation
X11	0.169**	0.0177	Positive correlation

(Note: *, * * and * * * respectively indicate significant at 10%, 5%, and 1% levels)

3) MODEL ANALYSIS

Eleven factors such as age, sex and education level are taken as independent variables. The rural information environment is used as the dependent variable, and the effective Probit model is tested based on the questionnaire data. The following results are obtained:

From the above results, we can see that there is no correlation between the two variables of X2 and X6 and the dependent variables. It indicates that sex and marital status have no obvious influence on the rural information environment. There is no positive or negative correlation between them. Both the sex and marital status can be removed from the influence factors. Therefore, it is easy to prove that hypothetical proposition 2 and hypothetical proposition 6 are false.

At the same time, all the independent variables exceptX2 and X6 have a positive correlation with the dependent variables. Moreover, it indicates that 9 factors all have a significant positive impact on the rural information environment such as age, education, family income, family members, income sources, traffic conditions, information acceptance, information recognition and information acceptance. Therefore, hypothetical proposition 1, hypothetical proposition 3-5, hypothetical proposition 7-11 have been all established.

III. INDEX SYSTEM CONSTRUCTION AND EVALUATION MODEL SELECTION OF RURAL INFORMATION ENVIRONMENT

A. CONSTRUCTION OF EVALUATION INDEX SYSTEM

According to the analysis of the factors, evaluate the rural information environment from rural objective infrastructure, technical factors and farmers' subjective behavior and consciousness factors. Design 6 first grade indicators and 23 two level indicators to evaluate the rural information environment with reference to the research results at home and abroad [24].

B. WEIGHT DETERMINATION AND MODEL SELECTION

1) WEIGHT CALCULATION BASED ON INFORMATION ENTROPY

The application of information entropy first appeared in thermodynamics, which was used to represent the physical state of chaos in molecular state. Then the concept of entropy was introduced into the field of information. Information entropy was used to solve the problem of information measurement. At the same time, information entropy can also express the uncertainty of things happening. For multi-attribute problems, the weight of an attribute can be represented by information entropy [25]. The specific calculation process of information entropy is as follows:

The first is to construct a decision matrix $X = (x_{ij})_{n\times m}$, Using the formula (9) to normalize the matrix, a new matrix is obtained $Y = (y_{ij})_{n\times m}$.

$$y_{ij} = \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{9}$$

TABLE 4. Evaluation two-level indicator system of rural information environment.

First level index	Two level index
Rural information infrastructure and technology X1	The proportion of fixed line covering administrative villages X11
	The number of cable TV per thousand X12
	Fiber length per 100 square kilometers X12
	Television comprehensive population coverage X14
Rural information talents X2	The number of information personnel per 10000 people X21
	Number of college students per thousand X22
Coverage of rural information network X3	Every hundred households in the countryside have the number of computers X31
	The number of TV sets per hundred households in the countryside X32
	The number of mobile phones per hundred households in the countryside X33
	The number of Internet users per 10000 people X34
The input and output of rural informatization X4	Per capita postal volume in rural areas X41
	Fixed investment in the telecommunication industry accounts for the proportion of investment in the whole society X42
	Fixed investment in the information industry accounts for the proportion of fixed assets investment in the whole society.X43
Individual and family status of farmers X5	Age X51
	Educational level X52
	Family income X53
	Number of family members X54
	Source of income X55
	Convenience of residential traffic X56
Consciousness and ability of farmers X6	Willingness to accept information X61
	Degree of information cognition X62
	Information acceptance ability X63

According to the definition, the information entropy of each index is obtained by formula (2)$E_j (j = 1, 2, \cdots, m)$.

$$E_j = -\frac{1}{\ln n}\sum\nolimits_1^n \frac{y_{ij}}{\sum_1^n y_{ij}} \ln\left(\frac{y_{ij}}{\sum_1^n y_{ij}}\right) \tag{10}$$

When $\frac{y_{ij}}{\sum_1^n y_{ij}} = 0$, $\lim\left(\frac{y_{ij}}{\sum_1^n y_{ij}} - \ln\left(\frac{y_{ij}}{\sum_1^n y_{ij}}\right)\right) = 0$.

Calculating the weight value of the attribute according to the formula (3):$\omega = (\omega_1, \omega_2, \cdots, \omega_m)$.

$$\omega_j = \frac{1 - E_j}{\sum_1^m (1-E_k)} \tag{11}$$

According to the above steps, we can get the weight of all indicators in each level, and set up the sets of index weight based on information entropy [26]. The weight of the first layer factor is expressed as a calculation: $\omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6)$. The weight of the underlying factor is expressed as a calculation: $\omega_1 = (\omega_{11}, \omega_{12}, \omega_{13}, \omega_{14})$, $\omega_2 = (\omega_{21}, \omega_{22})$, $\omega_3 = (\omega_{31}, \omega_{32}, \omega_{33}, \omega_{34})$, $\omega_4 = (\omega_{41}, \omega_{42})$, $\omega_5 = (\omega_{51}, \omega_{52}, \omega_{53}, \omega_{54}, \omega_{55}, \omega_{56})$, $\omega_6 = (\omega_{61}, \omega_{62}, \omega_{63}, \omega_4)$. This method is an objective method to determine weights. Compared with analytic hierarchy process (AHP), it reduces the influence of subjective factors.

2) THE ESTABLISHMENT OF MULTILEVEL FUZZY COMPREHENSIVE EVALUATION MODEL

Fuzzy comprehensive evaluation method is a comprehensive evaluation method based on fuzzy mathematics. This method converts qualitative evaluation into quantitative evaluation according to the membership degree theory of fuzzy mathematics. When the domain A × B is a finite set, the fuzzy relation can be expressed in matrix form. The matrix element r_{ij} is only in closed interval [0, 1], which is $0 \leq r_{ij} \leq 1$. This matrix is called fuzzy judgment matrix. When it is used to express fuzzy relations, r_{ij} represents the degree of association between the i element in the set A and the j element in the set B.

$$R = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{pmatrix}, \quad 0 \leq r_{ij} \leq 1 \tag{12}$$

For the multilevel fuzzy comprehensive evaluation problem, it is necessary to solve the judgment matrix of the influencing factors by layer. The model involves two levels, so the first level is carried out, that is, the elements in the $U_i = \{U_{ij}\}$ are subordinate to the evaluation of the set domain. It is expressed as:

$$R_i = \begin{pmatrix} r_{i1v_1} & \cdots & r_{i1v_s} \\ \vdots & \ddots & \vdots \\ r_{imv_1} & \cdots & r_{imv_s} \end{pmatrix} \tag{13}$$

The e presents the number of first layer factors, and the m represents the number of the underlying factors.

The multi-level fuzzy evaluation is carried out from the bottom to the high level. Therefore, according to ω_i and R_i, the second level fuzzy evaluation set is obtained: $A_i = \omega_i \cdot R_i$, which represents the membership degree or evaluation grade of the first layer factor set U. The normalization of A_i is carried out to get the final fuzzy evaluation matrix $R = (A_1, A_2, A_3)$.

IV. MODEL VERIFCATION AND ANALYSIS

A. DATA SOURCE

According to the established index system, 2007-2016 of the study interval is selected in the evaluation study of time dimension. The rural areas of the 31 provinces are selected as the research object in the evaluation study of the spatial dimension. And choose Lingshou County of Hebei Province as the research object of regional comprehensive evaluation research. The data used in this paper are mainly based on the simulated test data of the Chinese Statistical Yearbook, China's Rural Statistical Yearbook, China Information Yearbook and other simulation test data. Due to the different sizes of the above index units, there is a large difference between the data, so the original data should be standardized first.

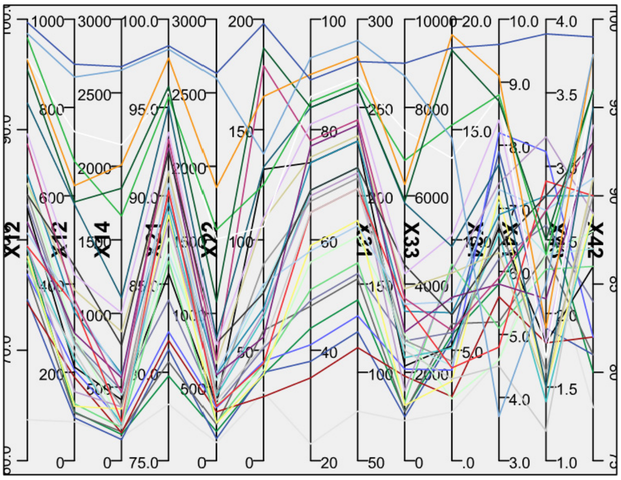

FIGURE 6. Parallel relations between indicators.

B. EVALUATION AND COMPARATIVE ANALYSIS OF SPATIAL DIMENSIONS BASED ON OBJECTIVE FACTORS

In the process of evaluating the country's rural information environment, the impact of major subjective factors is negligible. Therefore, in the selection of the evaluation index, only the influence of macro factors is considered. Then select the first 4 first-level indicators and the corresponding 13 two-level indicators. Due to the availability of more data sheets during the 2007-2016 time interval, the test data will be fully calculated in 2016. In other years, only the calculation results are given, and the calculation process is similar, and will not be described again. The choice of data is basically in line with the law of regional economic development. There are subtle differences between the different indicators, but the overall trend is consistent. It can be seen from the comparison of the two indicators in Figure 5. For example, there are microscopic differences in the specific index values of indicators X21 and X34, but the overall trend is similar.

(1) Data standardization. Because the unit dimension of each index data is different and there is a great difference in the quantity level, so we need to use the formula (9) to normalize the original data in Table 7 and convert it into the standard data.

(2) According to the multilevel fuzzy comprehensive evaluation model, the decision matrix R_1, R_2, R_3 are constructed by using table 4 data.

The comprehensive evaluation resultis92.47, so the comprehensive assessment of the national rural information environment in 2016 is 92.47. The results of the comprehensive evaluation of the rural information environment in the past 2007-2015 years were calculated similarly. The specific results are shown in Table 5.

The evaluation trend of rural information environment is shown in Figure 1. From the chart, we can see that the rural information environment has been optimized and improved. From the two level evaluation index data of rural information environment, the input growth of rural information infrastructure and technology is remarkable: the coverage rate of the fixed telephone, the number of cable TV, the length of the fiber, the comprehensive coverage of TV, and so on. The number of rural college students and the network coverage of rural information has a qualitative leap in breadth and depth: the number of computers in each hundred households in rural areas, the number of TV sets per hundred households in rural areas, the number of mobile phones per 100 households in rural areas, the number of Internet users per 10000 people, and the input and output of rural informatization are significantly raised. High: per capita postal volume in rural areas, fixed investment in telecom industry accounts for the proportion of total social investment, and fixed investment in information industry accounts for the proportion of fixed assets investment in the whole society.

TABLE 5. Results of comprehensive evaluation of rural information environment in 2007-2016 years of the whole country in China.

Year	2007	2008	2009	2010	2011
Evaluation results	73.65	77.85	81.79	82.33	85.49
Year	2012	2013	2014	2015	2016
Evaluation results	86.32	88.46	89.91	91.68	92.47

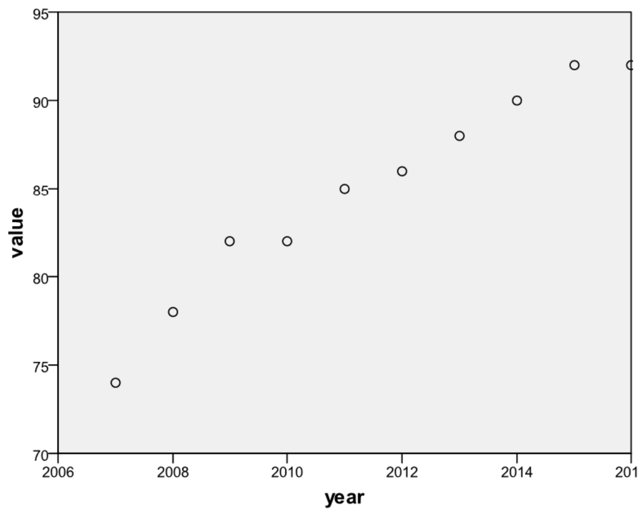

FIGURE 7. The trend of information environment change in 2007-2016 rural areas in China.

C. COMPREHENSIVE EVALUATION AND ANALYSIS OF SPACE BASED ON MACRO FACTORS

For the comprehensive evaluation of rural information environment in space, the rural information environment of 31 provinces and cities is evaluated by fuzzy comprehensive evaluation based on the difference of macro factors. And the cluster analysis is based on the evaluation results. As the calculation methods for the evaluation of rural information environment in the 31 provinces and cities are same, the detailed calculation process is given only in Hebei Province.

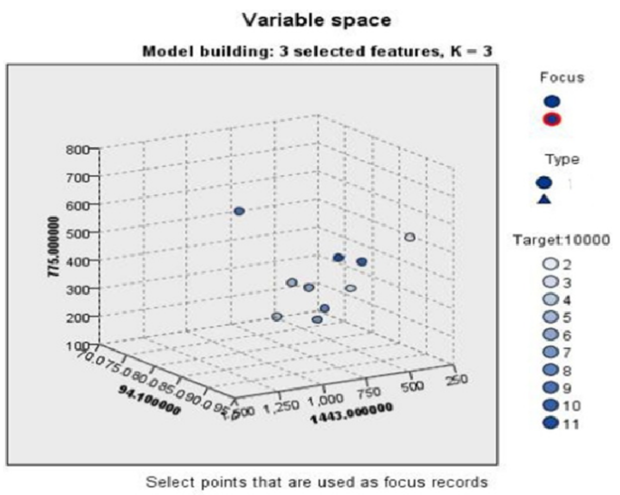

FIGURE 8. Lower dimensional projection of 13 variables.

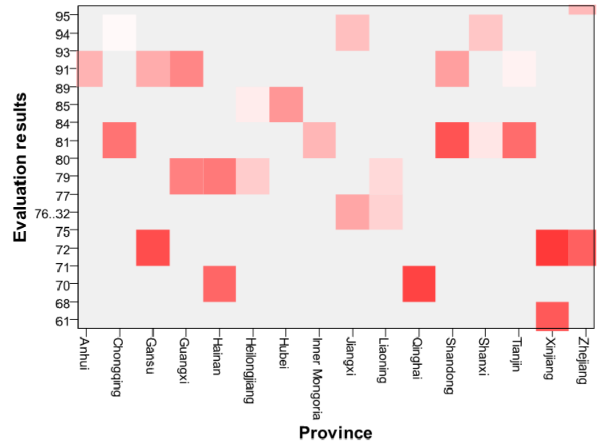

FIGURE 9. Thermal maps of evaluation results of provinces.

The calculation results of other provinces and cities are given directly.

(1) Original data collection. Firstly, collect the relevant index data from different cities in Hebei province according to needs. The data of 13 two grade indexes of 11 cities in Hebei, Shijiazhuang, Baoding, Tangshan, Handan, Zhangjiakou, Chengde, Langfang, Qinhuangdao, Cangzhou, Xingtai and Hengshui are collected.

(2) The comprehensive score of rural information environment in other provinces and cities in 2016 is calculated similarly. The specific results are shown in Table 6.

TABLE 6. The results of rural information and environment score in each province.

Province	Evaluation results	Province	Evaluation results
Beijing	93.72	Hubei	85.22
Tianjin	90.56	Hunan	81.34
Hebei	84.25	Guangdong	90.56
Shanxi	81.33	Guangxi	79.35
Inner Mongoria	80.23	Hainan	78.66
Liaoning	78.96	Chongqing	81.38
Jilin	75.39	Sichuan	79.66
Heilongjiang	76.66	Guizhou	70.23
Shanghai	94.28	Yunnan	71.28
Jiangsu	92.51	Tibet	60.85
Zhejiang	94.63	Shaanxi	80.49
Anhui	89.45	Gansu	71.58
Fujian	88.78	Qinghai	68.24
Jiangxi	76..32	Ningxia	70.27
Shandong	88.69	Xinjiang	72.01
Henan	84.12	***	***

(3) Cluster analysis based on K-Means algorithm

The K-Means algorithm has many possible convergence conditions. For example, the search can be terminated in a classification error and does not change during the redistribution. It indicates that this division may be local optimal. Another termination condition may be a pre defined number of iterations. The K-Means algorithm can be viewed as a gradient decent process. It starts with an initial clustering center set. And iteratively update the cluster containing the sample points to reduce errors. The sample point of the K-Means algorithm is N, the properties of each sample point is m-dimensional, the number of cluster type is K, the time complexity of T iterations is O(T $*$ K:$*$ N $*$ M). K-Means clustering algorithm is based on a clustering method. The specific approach is to divide the N object data set into K division in Euclidean multidimensional space ($k \ll n$), where each partition separately represents a cluster. First, the algorithm requires the user to specify the number of clusters as K, and select the K data object as initial cluster center according to initial clustering center selection strategy. According to the Euclidean distance of the respective cluster centers, the non-initial cluster center data objects are placed in the nearest cluster. Then, recalculate average value of all data objects in each cluster and form a new clustering center. This clustering iterative process is repeated until a predetermined termination condition is met. The steps of K-Means clustering algorithm are described as below [17]–[20]:

Grouping: assign samples to their nearest central vector and reduce the value of the objective function.

$$1 = \sum_{i=1}^{n} \min \left\| x_i - p_j \right\|^2 \tag{14}$$

in which $j = 1, 2, \cdots, k$.

Determining center: The objective is to minimize the value of the target function.

$$\sum_{i=1}^{m} \left\| x_i - p_j \right\|^2 \leq \sum_{i=1}^{m} (\|y_i - \bar{y}\|^2 + \|\bar{y} - w\|^2) \tag{15}$$

The necessary and sufficient condition for equality(2) is that (3) was established.

$$w = \bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i \tag{16}$$

Description of the improved K-Means algorithm:

Input: DS: data sets contain N data objects, K is the number of clusters. Output: the set of K clusters.

Algorithm:

① Using outlier detection strategies to remove outliers;

② Arbitrarily select K data objects from the DS as the initial clustering center;

③ Repeat the above steps;

④ Assign each object to the most similar cluster according to the similarity between the data object and the K cluster;

⑤ Calculate the mean of all objects in each cluster to form a new cluster center;

⑥ Until a predetermined termination condition is satisfied.

The steps described above in ⑥ returns to step ④ loop execution, and when the predetermined termination condition is satisfied, the algorithm ends.

K-means clustering method is used to deal with the evaluation results of provinces. The calculated results are shown in table 13. From the clustering results, we can see that there are certain differences in rural information environment in spatial. It can be roughly divided into four levels. The first is a good area of rural information environment, which includes 9 provinces, such as Beijing, Tianjin, Shanghai, Guangdong, Zhejiang, Jiangsu, Fujian, Anhui, Shandong, etc, and the two is the better area of rural information environment, including Hebei, Henan, Inner Mongolia, Shanxi, Jiangsu and other 8 provinces, and the three is the medium of rural information environment. Region, including Sichuan, Liaoning, Jilin, Heilongjiang, Jiangxi, Hainan, Guangxi and other 7 provinces; the four is a relatively backward rural areas of information and environment, including Gansu, Qinghai, Ningxia, Xinjiang, Guizhou, Yunnan, Hainan and other provinces. According to the above steps 1-6 of the K-means clustering algorithm, specific clustering results are obtained, as shown in Table 7.

TABLE 7. Cluster analysis of evaluation results by provinces.

Cluster 1		Cluster 2	
Beijing	93.72	Hebei	84.25
Tianjin	90.56	Shanxi	81.33
Shanghai	94.28	Inner Mongoria	80.23
Jiangsu	92.51	Hubei	85.22
Zhejiang	94.63	Hunan	81.34
Guangdong	90.56	Chongqing	81.38
Anhui	89.45	Shaanxi	80.49
Fujian	88.78	Henan	84.12
Shandong	88.69	***	***
Cluster 3		**Cluster 4**	
Liaoning	78.96	Guizhou	70.23
Jilin	75.39	Yunnan	71.28
Heilongjiang	76.66	Tibet	60.85
Guangxi	79.35	Gansu	71.58
Hainan	78.66	Qinghai	68.24
Sichuan	79.66	Ningxia	70.27
Jiangxi	76.32	Xinjiang	72.01

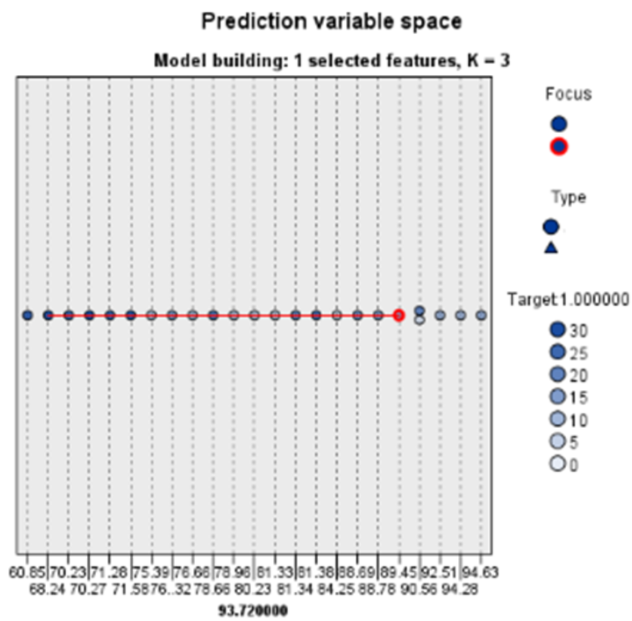

FIGURE 10. Correlation analysis of clustering results.

D. REGIONAL COMPREHENSIVE EVALUATION BASED ON OBJECTIVE AND SUBJECTIVE FACTORS

To assess the entire rural information environment in the time and spatial dimensions of the rural information environment, it is more important to consider the impact of macro factors. For a smaller range of county space, it is mainly based on macro factors and subjective factors to carry out comprehensive evaluation. Taking Lingshou County of Hebei Province as an example to carry out the regional comprehensive evaluation.

(1) Original data. According to the evaluation indicators, the corresponding macro factor data is collected from the statistical yearbook data, and the data related to subjective factors are obtained through surveys and interviews. In this case, data were collected from 15 townships in Lingshou County, Hebei Province.

(2) Data standardization. Since the unit dimension of each indicator data is different, the quantity level is very different, so we need to use equation (9) to normalize the original data and convert it into standard data.

TABLE 8. Weight calculation results based on information entropy for indicators.

First level index	First level index weight	Two level index	Two level index weight
Rural information infrastructure and technology X1	0.25	The proportion of fixed line covering administrative villages X11	0.25
		The number of cable TV per thousand X12	0.35
		Fiber length per 100 square kilometers X12	0.20
		Television comprehensive population coverage X14	0.20
Rural information talents X2	0.16	The number of information personnel per 10000 people X21	0.60
		Number of college students per thousand X22	0.40
Coverage of rural information network X3	0.09	Every hundred households in the countryside have the number of computers X31	0.37
		The number of TV sets per hundred households in the countryside X32	0.26
		The number of mobile phones per hundred households in the countryside X33	0.24
		The number of Internet users per 10000 people X34	0.13
The input and output of rural informatization X4	0.10	Per capita postal volume in rural areas X41	0.39
		Fixed investment in the telecommunication industry accounts for the proportion of investment in the whole society X42	0.38
		Fixed investment in the information industry accounts for the proportion of fixed assets investment in the whole society.X43	0.23
Individual and family status of farmers X5	0.18	Age X51	0.14
		Educational level X52	0.17
		Family income X53	0.21
		Number of family members X54	0.26
		Source of income X55	0.09
		Convenience of residential traffic X56	0.13
Consciousness and ability of farmers X6	0.22	Willingness to accept information X61	0.53
		Degree of information cognition X62	0.21
		Information acceptance ability X63	0.26

(3) Calculate the weight of the indicators. The amount of information contained in the entire assessment is different, so the contribution to the goal is different. The weight of each indicator can be calculated by the methods of information entropy (10) and the formula (11), and the settlement result is calculated in table 8.

(4) According to the multilevel fuzzy comprehensive evaluation model, use table 14 data to construct decision matrix R_1, R_2, R_3, the decision matrix R_1, R_2, R_3 are normalized to obtain the matrix R'_1, R'_2, R'_3 which satisfies the normalization requirements. The comprehensive evaluation result is 79.54, so the comprehensive evaluation of the rural information environment in Lingshou County of Hebei is 79.54.

V. CONCLUSION

The aim of this paper is to provide methods to evaluate the rural information environment. This study has identified that the rural information environment is influenced by macro and subjective factors. The research has also shown that there are big differences in different areas of rural information environment. The findings of this research provide insights for rural informatization path, which have significant implications for the understanding of how to speed up the construction of regional rural informatization. These findings from the study make several contributions to the current literature. First of all, the information environment in rural areas has been greatly improved from time to time. Secondly, the spatial difference of regional rural information environment is objective. Third, the rural information environment is affected by both macro and subjective factors. Despite its exploratory nature, this study offers some insight into evaluation system of rural information environment. The study is limited by the lack of information on Macro data affecting rural information environment. An additional uncontrolled factor is the possibility that the difference of rural information subject. It is suggested that the association of these factors is investigated in future studies. And further investigation and experimentation into rural information environment is strongly recommended.

REFERENCES

[1] R. Khajouei, S. G. Hajesmaeel, M. Mirzaee, "Comparison of two heuristic evaluation methods for evaluating the usability of health information systems," *J. Biomed. Inf.*, vol. 80, pp. 37–42, Apr. 2018.
[2] A. E. Elshaikh, X. Jiao, and S.-H. Yang, "Performance evaluation of irrigation projects: Theories, methods, and techniques," *Agricult. Water Manage.*, vol. 203, pp. 87–96, Apr. 2018.
[3] J.-K. Park, S.-K. Lee, and J.-H. Kim, "Development of an evaluation method for nuclear fuel debris–filtering performance," *Nucl. Eng. Technol.*, vol. 50, no. 5, pp. 738–744, Jun. 2018.
[4] A. M. Valente, H. Binantel, D. Villanua, and P. Acevedo, "Evaluation of methods to monitor wild mammals on Mediterranean farmland," *Mammalian Biol.*, vol. 91, pp. 23–29, Jul. 2018.
[5] Y. Jin and G. Li, "Application of improved K-means algorithm in evaluation of network resource allocation," *Boletin Tecnico*, vol. 55, no. 5, pp. 284–292, 2017.
[6] G. Zhang, C. Zhang, and H. Zhang, "Improved K-means algorithm based on density canopy," *Knowl.-Based Syst.*, vol. 145, pp. 289–297, Apr. 2018.
[7] Y. Han, J. Li, X.-L. Yang, W.-X. Liu, and Y.-Z. Zhang, "Dynamic prediction research of silicon content in hot metal driven by big data in blast furnace smelting process under Hadoop cloud platform," *Complexity*, vol. 2018, pp. 1–16, Oct. 2018.
[8] H. S. Loh, Q. Zhou, V. V. Thai, Y. W. Wong, and K. F. Yuen, "Fuzzy comprehensive evaluation of port-centric supply chain disruption threats," *Ocean Coastal Manage.*, vol. 148, pp. 53–62, Nov. 2017.
[9] W. Yang, K. Xu, J. Lian, L. Bin, and C. Ma, "Multiple flood vulnerability assessment approach based on fuzzy comprehensive evaluation method and coordinated development degree model," *J. Environ. Manage.*, vol. 213, pp. 440–450, May 2018.

[10] A.-M. Yang, X.-L. Yang, J.-C. Chang, B. Bai, F.-B. Kong, and Q.-B. Ran, "Research on a fusion scheme of cellular network and wireless sensor for cyber physical social systems," *IEEE Access*, vol. 6, no. 99, pp. 18786–18794, Apr. 2018.

[11] W. Li, W. Liang, L. Zhang, and Q. Tang, "Performance assessment system of health, safety and environment based on experts' weights and fuzzy comprehensive evaluation," *J. Loss Prevention Process Ind.*, vol. 35, pp. 95–103, May 2015.

[12] X. Yu, W. Meng, and L. Xiang, "Comprehensive evaluation chronic pelvic pain based on fuzzy matrix calculation," *Neurocomputing*, vol. 173, pp. 2097–2101, Jan. 2016.

[13] W. Cai, L. Dou, M. Zhang, W. Cao, J.-Q. Shi, and L. Feng, "A fuzzy comprehensive evaluation methodology for rock burst forecasting using microseismic monitoring," *Tunnelling Underground Space Technol.*, vol. 80, pp. 232–245, Oct. 2018.

[14] T.-T. Gao and S.-M. Wang, "Fuzzy integrated evaluation based on HAZOP," *Procedia Eng.*, vol. 211, pp. 176–182, 2018.

[15] R. Fattahi and M. Khalilzadeh, "Risk evaluation using a novel hybrid method based on FMEA, extended MULTIMOORA, and AHP methods under fuzzy environment," *Saf. Sci.*, vol. 102, pp. 290–300, Feb. 2018.

[16] A. Antunes, D. Bonfim, N. Monteiro, and P. M. M. Rodrigues, "Forecasting banking crises with dynamic panel probit models," *Int. J. Forecasting*, vol. 34, no. 2, pp. 249–275, Apr. /Jun. 2018.

[17] S. Han J. Edward, and J. Vytlacil, "Identification in a generalization of bivariate probit models with dummy endogenous regressors," *J. Econometrics*, vol. 199, no. 1, pp. 63–73, Jul. 2017.

[18] H. Fu, Z. Li, Z. Liu, and Z. Wang, "Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization," *Sustainability*, vol. 10, no. 7, p. 2488, 2018.

[19] S. Jiang, M. Lian, C. Lu, Q. Gu, S. Ruan, and X. Xie, "Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope," *Complexity*, vol. 2018, Aug. 2018, Art. no. 1048756, doi: 10.1155/2018/1048756.

[20] P. Mozharovskyi and J. Vogler, "Composite marginal likelihood estimation of spatial autoregressive probit models feasible in very large samples," *Economics Lett.*, vol. 148, pp. 87–90, Nov. 2016.

[21] T. Yen Steven and M. E. Zampelli, "Religiosity, political conservatism, and support for legalized abortion: A bivariate ordered probit model with endogenous regressors," *Social Sci. J.*, vol. 54, no. 1, pp. 39–50, Mar. 2017.

[22] A. Antunes, D. Bonfim, N. Monteiro, and P. M. M. Rodrigues, "Forecasting banking crises with dynamic panel probit models," *Int. J. Forecasting*, vol. 34, no. 2, pp. 249–275, Apr. /Jun. 2018.

[23] L. Kang, H. L. Du, H. Zhang, and W. L. Ma, "Systematic research on the application of steel slag resources under the background of big data," *Complexity*, vol. 2018, Oct. 2018, Art. no. 6703908.

[24] F. Liu, Y. Jin, D. Jin, X. Jia, and T. Wang, "Research on workshop-based positioning technology based on Internet of Things in big data background," *Complexity*, vol. 2018, pp. 1–11, Oct. 2018.

[25] J. Wu, L. Zhang, S. Yin, H. Wang, G. Wang, and J. Yuan, "Differential diagnosis model of hypocellular myelodysplastic syndrome and aplastic anemia based on the medical big data platform," *Complexity*, vol. 2018, pp. 1–12, Nov. 2018.

[26] J. Yanfeng, L. Gang, and Z. Chang, "Study of data mining in social network analysis," *J. Inf. Comput. Sci.*, vol. 12, no. 16, pp. 5947–5955, Nov. 2015.

YANFENG JIN was born in Hebei, China, in 1981. He received the bachelor's degree in information and computing science, in 2005, and the master's degree in computer application technology, in 2007. He is currently pursuing the Ph.D. degree in management science and engineering with the Beijing University of Posts and Telecommunications.

Since 2007, he has been a Teaching Assistant in data mining with the China Post and Business Correspondence Research Center. In 2009, he was hired as a Lecturer, engaged in teaching and research work in the major of big data. In 2015, he was promoted to an Associate Professor with the Beijing University of Posts and Telecommunications. His research interests include big data analysis, rural informatization, and the Internet of Things. He has published more than 30 papers in these fields.

GANG LI is the first doctor of philosophy and social sciences to teach at the Beijing University of Posts and Telecommunications. He is a well-known young and middle-aged scholar in the field of network culture management and social economic philosophy in China. He is currently a Professor and Doctoral Supervisor with the Beijing University of Posts and Telecommunications. He is the representative of the One Hundred Personnel Project of Excellent Social Science Theory Talents in Beijing in the New Century. In 2001, he was a recipient of the title of Teacher with Outstanding Contribution to Teaching by the Beijing University of Posts and Telecommunications, and in 2006, he was a recipient of the Famous Teacher Award by the Beijing University of Posts and Telecommunications. In 2007, he was chosen by the Ministry of Education's New Century Talents Support Program.

HUIFENG ZHANG received the bachelor's degree in bioengineering from the Hebei University of Science and Technology, in 1999, and the master's degree in business administration from the Beijing Institute of Technology, in 2005. Since 2005, he has been a Lecturer with the Post Express Research Center, where he was mainly involved in teaching and research on informationization of postal enterprises. He was promoted to Associate Professor in 2015. His main research interests include postal enterprise information management, the Internet of Things, and logistics supply chain. He has published 17 papers in these fields. He was awarded the honorary titles of "Young Technological Talents in Transportation" and "Excellent Technological Workers of China Post Corporation."