Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

Exploration into Students' Information Literacy Training Model Based on Graph Theory-regression Algorithm

Minmin Xiao
College of Intelligent Manufacturing and Information Engineering, Shaanxi Energy Institute, Xianyang, Shaanxi, China

Abstract:

Research of the problems related to the cultivation of students' information literacy is mainly theoretical research on the concept of information literacy, cultivation modes and methods, cultivation strategies and approaches, which is highly subjective and lacks specific and explicit data models to provide data support for the implementation analysis of the research, process tracking, and result evaluation. To this end, this paper collects students' real data by means of questionnaires, and based on the data, conducts research on the training model of students' information literacy. First, the reliability and validity of the collected data are analyzed to prove the consistency, reliability and energy efficiency of the data. Second, students' information literacy at this stage is analyzed, the linear regression algorithm is used to analyze the information literacy data of students, thus forming a preliminary regression training model. Finally, in order to reduce the problem of data paranoia due to the singleness of the data collection group and the lack of large data volume, this paper analyzes the in-out degree between the factors influencing students' information literacy based on the graph theory algorithm, using it as the relative influence weight of each factor. Combining the regression training model, a student information literacy training model of graph theory-regression type that integrates graph theory algorithm is proposed. The model is based on student data, uses linear regression method to study the basic composition of the training model, and uses graph theory algorithm for secondary correction of each influencing factor. The analysis results show that the model can not only provide data support for the cultivation of students' information literacy, but also objectively reflect the real information literacy of students at this stage, thus providing data basis for the subsequent training and evaluation of students' information literacy.

Keywords*: Linear regression, Graph theory algorithm, Information literacy, Model building.*

I. INTRODUCTION

In order to meet the needs of lifelong education in modern society, an important goal of modern higher education is to cultivate information literacy of college students. "Information literacy" is a basic ability and an ability to adapt to the information society. Students' information literacy level is an essential skill for students' scientific research, innovation and lifelong learning, and knowledge structure updating.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

The cultivation of information literacy has become a major theoretical and practical topic concerned by the educational circles of all countries in the world and even all sectors of the society. Research institutions worldwide and scholars at home and abroad have carried out extensive exploration and in-depth research on how to improve information literacy, and put forward a series of new insights into the definition, connotation and evaluation criteria of information literacy. In 1989, the American Library Association (ALA) proposed that information literacy involves three levels: cultural literacy, information awareness and information skills. People with information literacy can know when information is needed, and can effectively retrieve, process and use information[1]. The most influential and famous foreign experimental research on information literacy is the "Big6" skill, which was founded in 1988 by Dr. Mike Eisenberg and Dr. Bob Berkowitz in the United States[2]. Professor Sang Xinmin believes that the internal structure and goal system of information literacy cultivation can be established from three levels[3]. Information literacy research is carried out in various forms in China, among which the most common method is to use various guides to conduct online information literacy education[4]. At present, the scanty research is mostly limited to theoretical discussion and various investigations and explorations with the teaching of a certain course as the starting point. Based on the questionnaire survey, Wang Rong et al. analyzed the characteristics of students' information behavior, and put forward the countermeasures for improving the ability of information literacy education in the library of higher vocational colleges[5]. Taking higher vocational students as the main research objects, Li Kai et al. analyzed the current status of information literacy education in higher vocational colleges and the status quo of information literacy of higher vocational students through investigation and research, deeply analyzed the problems existing in current higher vocational education, and put forward suggestions for improving the information literacy of higher vocational students[6]. Liu Xiaowen sorted out and analyzed the core concepts closely related to information literacy in order to define the difference between them and find out the connection between them, thereby promoting further research in this field[7]. Mu Xiangwang analyzed the rich connotation of information literacy from different levels of knowledge, technology, consciousness and ethics, and according to the connotation requirements of information literacy, completed the research on the methods and ways of strengthening students' information literacy training in colleges and universities. He then proposed countermeasures and suggestions from three aspects: strengthening hardware facilities and information resource construction, innovating teaching ideas and teaching content, and strengthening information ethics education[8]. Li Jianli studied the development and evolution of information literacy and the evaluation criteria. At the same time, taking graduate students in some colleges and universities in Anhui Province as an example, he analyzed the current situation of information literacy, and put forward effective improvement strategies for existing problems, thus providing useful reference for further implementation of information literacy education[9].

This paper takes higher vocational students as the research object, and uses questionnaire scale to collect the students’ basic data. The scale consists of six aspects: information awareness, information knowledge, information skills (6 components), information ethics, database awareness, and improvement awareness. This paper uses reliability and validity analysis to reflect the consistency of this questionnaire, adopts linear regression method to explore the preliminary model for objective cultivation of students'

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

information literacy based on data, and integrates the graph theory algorithm to introduce students' essential characteristics to form the final student information literacy training model.

II. FACTORS INFLUENCING INFORMATION LITERACY

ALA[1] proposes cultural literacy, information awareness and information skills as the three focuses of information literacy. Li Kedong proposed three basic points of information literacy: the information technology application skills, the ability to criticize and understand information content, the ability to use information and the attitude and ability to integrate into the information society. Sang Xinmin put forward the three levels of attention to information literacy: the ability to control information, the ability to use information technology to learn and communicate efficiently, the need to cultivate and improve the ethics, emotion, legal awareness and social responsibility of citizens in the information age[3]. Zhang Yibing et al. summarized the connotation of information literacy from the perspectives of technology, psychology, sociology, and cultural studies: information processing, information problem solving, information exchange, and the multiple construction ability in information culture[10].

On the basis of the above research, this paper takes into account the current situation of higher vocational students, and focuses on comprehensive analysis and research regarding the impact on information literacy from six aspects. The influencing factors and components are shown in Fig 1.

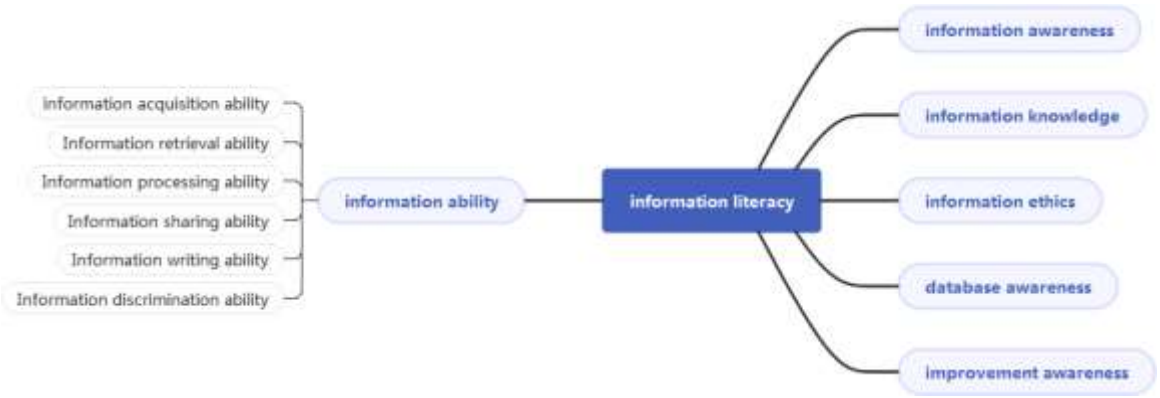

Fig 1: Factors influencing information literacy

III. CONSTRUCTION OF GRAPH THEORY-REGRESSION TRAINING MODEL

3.1 Graph Theory Algorithm

Graph Theory is a branch of discrete mathematics, which is widely penetrated in the field of computer science and provides simple and systematic solutions for many difficult problems[11-12]. Graph theory algorithms can transform complex discrete problems into graph theory problems and solve them using graph theory algorithms.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

In graph theory algorithms, the associated lines with vertex as the center are called degrees. The connection from this point to another point is called out-degree; conversely, the connection from other points to this point is called in-degree[13-14].

First, use the derivation path to calculate the in-out degree value of each factor, and then normalize the in-out degree value to calculate the corresponding weight. Suppose the influencing factors are A, B, and C, then the derivation path is shown in Fig 2. Then its corresponding in-out degree adjacency matrix is G, and the normalized relative weight matrix is ω_G . Where, the calculation of the in-out degree adjacency matrix G is shown in formula (1), and the calculation of the weight matrix ω_G is shown in formula (2).

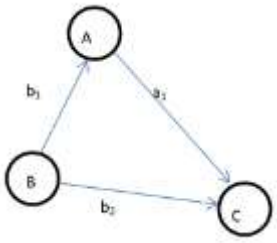

Fig 2: The derivation path between influencing factors

$$G = \begin{bmatrix} 0 & 0 & a_1 \\ b_1 & 0 & b_2 \\ 0 & 0 & 0 \end{bmatrix} \tag{1}$$

$$\omega_G = \begin{bmatrix} a_1 \\ b_1 + b_2 \\ 0 \end{bmatrix} / (a_1 + b_1 + b_2) \tag{2}$$

Where, a_1 is the derivation path between influencing factor A-> influencing factor C, b_1 is the derivation path between influencing factor B-> influencing factor A, and b_2 is the derivation path between influencing factor B-> influencing factor C. If the path exists, the degree is 1. Otherwise, it is 0. Here, a_1, b_1, b_2 are used to represent the in and out degree values of the path: 1. That is, A has an in-degree of 1 and an out-degree of 1; B has an in-degree of 0 and an out-degree of 2; C has an in-degree of 2 and an out-degree of 0.

3.2 Linear Regression Algorithm

Linear regression is a statistical analysis method that uses mathematical statistics and regression analysis to determine the interdependent quantitative relationship between two or more variables[15-16]. In regression analysis, according to the real needs of the analysis, there will be one or more independent variables, and there is a close linear relationship between the dependent variable and the independent

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

variable[17-18].

In general, there are often more than one factor that affects the dependent variable y. Suppose there are n factors such as x_1, x_2, ..., x_n, then the general linear learning function $h_\beta(x)$ is shown in formula (3), and the influence degree matrix β is shown in formula (4).

$$h_\beta(x) = \sum_{i=0}^{n} \beta_i x_i = \beta^T x \tag{3}$$

$$\beta = \begin{bmatrix} \beta_0 & \beta_1 & \dots & \beta_n \end{bmatrix}^T \tag{4}$$

Where, x_0 takes a value of 1. When there is sufficient training data, the predicted value learned by the learning function $h_\beta(x)$ is infinitely close to the actual value[19-20].

For the single sample error $j(\beta)$, as shown in formula (5), the error of all its samples: loss function, can be calculated by formula (6).

$$j_i(\beta) = \alpha(h_\beta(x(i)) - y(i)) \tag{5}$$

$$J(\beta) = \alpha \sum_{i=1}^{n} j_i(\beta) \tag{6}$$

Where, α is the parameter. In the subsequent calculation, the search algorithm is used to select α to minimize the sum of errors and avoid falling into the local optimum.

3.3 Construction of Graph Theory-regression Training Model

Firstly, the correlation between the factors is determined through the graph theory algorithm: the relative influence weight matrix ω_G , and then the learning function $h_\beta(x)$ is obtained using the linear regression algorithm. Where, β is the influence degree matrix of each influencing factor regarding information literacy. Finally, multiply the relative influence weight matrix ω_G and the absolute influence degree matrix β to obtain the final influence matrix ω , as shown in formula (7).

$$\omega = \chi \omega_G{}^T \beta \tag{7}$$

Where, β does not take a value of β_0, and is χ he coefficient. Since the weight and parameters are both less than 1, χ takes a value of 100. That is, the final learning function can be calculated by formula (8):

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

$$h_{\omega}(x) = \sum_{i=1}^{n} \omega_i x_i = \omega^T x \tag{8}$$

IV. INFLUENCING FACTORS AND MODEL ANALYSIS OF STUDENT LITERACY

4.1 Data Reliability and Validity Analysis

All questionnaires in this survey are Likert scale. Reliability and validity are two key indexes to measure whether the questionnaire is usable. Reliability measures the consistency of all items within the questionnaire, while validity analyzes the energy efficiency of each item. The value of Cronbach's Alpha coefficient is usually used to test reliability, and exploratory factor analysis is used to describe validity. If the Cronbach's Alpha value is 0.7-0.9, it means the questionnaire is usable with good consistency. If the KMO value exceeds 0.6 and the significance is< 0.05, it means the questionnaire is suitable for exploratory factor analysis. Whether the specific validity is good or not needs to be determined according to the cumulative variance value. The reliability and validity test values of the questionnaire are shown in Table I and Table II.

TABLE I. Reliability and validity test

Coefficient	Value
Cronbach’s Alpha coefficient	.881
Kaiser-Meyer-Olkin measures sampling adequacy	.758
Significance	.000

TABLE II. Illustrated variance statistics

	Initial eigenvalue			Extraction Sums of Squared Loading			Circulation Sums of Squared Loadings		
Element	Statistics	% of variance	Cumulative %	Statistics	% of variance	Cumulative %	Statistics	% of variance	Cumulative %
1	2.756	45.939	45.939	2.756	45.939	45.939	2.099	34.988	34.988
2	1.054	17.569	63.508	1.054	17.569	63.508	1.711	28.520	63.508
3	.818	13.633	77.141						
4	.601	10.013	87.154						
5	.448	7.471	94.624						
6	.323	5.376	100.000						

From Table I, it can be found that the Cronbach's Alpha coefficient is 0.881, indicating that the questionnaire has good internal consistency and can be used. The KMO value is 0.758 and the significance is 0.000, indicating that the validity analysis of the questionnaire can be done using exploratory factor analysis. From Table 2, it can be found that the cumulative variance is >60%, which further indicates that

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

each item of the questionnaire is reliable, the dimensions are divided appropriately, each item is essential, and the energy efficiency is high.

4.2 Status Quo Analysis of Students' Information Literacy

Statistics on the current situation of students' information literacy is carried out through descriptive statistical methods. The statistical results are shown in Figure 3. In the questionnaire scale, 1: excellent, 2: good, 3: moderate, 4: poor, 5: very poor. According to the chart data and trend, it can be found that the students' information awareness, information retrieval ability, and information ethics average are all in a good stage currently, database awareness and improvement awareness are in a moderately good stage, and the average information knowledge and information discrimination ability are in a moderately poor stage, and there is big standard deviation in information sharing ability and improvement awareness among the students.

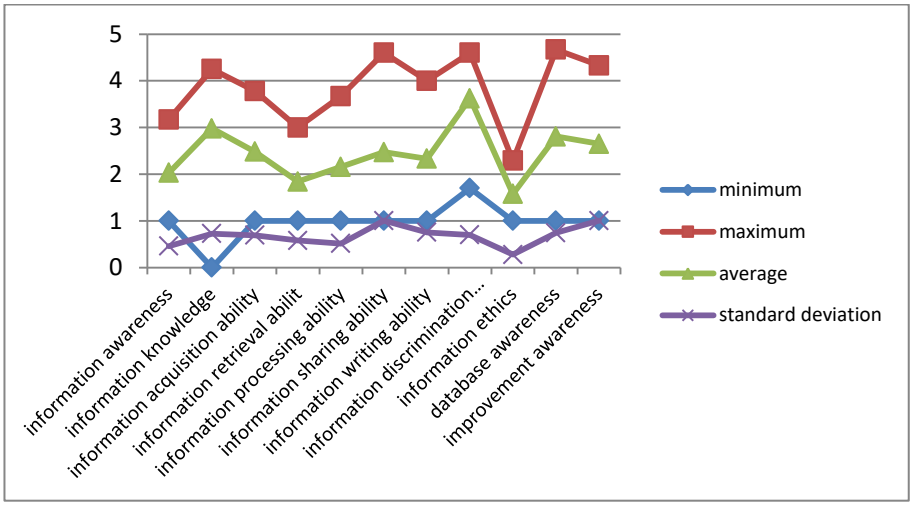

Fig 3: Statistics on the status quo of students' information literacy

4.2 Application of Graph Theory-Regression Model in Student Information Literacy Training

(1) Establish the information literacy learning function. Based on the valid data obtained from the questionnaire, SPSS software was used to calculate the correlation coefficient of the learning function according to the linear regression algorithm[21], as shown in Table III.

TABLE III. Regression coefficient

Model	Non-standard coefficient		Standard coefficient	T	Significance	VIF
	B	Standard error	Bate			
(constant)	-.746	-.746	-.746	-.746	.000	

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

Information Awareness (A)	.155	.155	.155	.155	.000	1.261
Information Knowledge (B)	.007	.031	.010	.229	.819	1.416
Information acquisition ability (C1)	.162	.035	.226	4.645	.000	2.021
Information processing ability(C2)	.049	.040	.051	1.213	.228	1.500
Information discrimination ability (C3)	.107	.029	.151	3.751	.000	1.384
Information retrieval ability (C4)	.140	.038	.163	3.689	.000	1.669
Information sharing ability (C5)	.112	.021	.227	5.226	.000	1.610
Information writing ability (C6)	.134	.027	.204	4.999	.000	1.511
Information Ethics (D)	.080	.068	.045	1.185	.239	1.224
Database Awareness (E)	.007	.028	.011	.264	.793	1.542
Improvement awareness (F)	.015	.022	.031	.685	.495	1.756

$R^2 = 0.877$

F = 68.969

$P = .000^b < 0.05$

Dependent variable: Information literacy level (INFOABL)

According to Table III, it can be found that the linear fitting degree is good this time, R2 = 0.877 > 0.6, indicating that the learning function can reflect the influence of various factors on information literacy level. The VIF corresponding to each factor <5 in the table indicates that there is no collinearity among the related variables. The learning function is significant, F = 68.969, P = .000b < 0.05, which means that among multiple factors, there must be one that can significantly affect the information literacy level. Where, according to Table 3, it can be inferred that the learning function calculation formula for the students' information literacy level INFOABL is formula (9).

$$\begin{aligned} h(INFOABL) = &-0.746 + 0.155 * A + 0.007 * B + 0.162 * C1 + 0.049 * C2 \\ &+ 0.107 * C3 + 0.140 * C4 + 0.112 * C5 + 0.134 * C6 \\ &+ 0.080 * D + 0.007 * E + 0.015 * F \end{aligned} \tag{9}$$

The significance of parameters such as information writing ability, information ethics, database awareness, and improvement awareness greatly exceeds 0.05, indicating that there is poor correlation between data and dependent variables, which is different from the previous analysis. Therefore, this paper will perform further weighted correction based on graph theory weights.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

(2) Calculate the relative influence weight matrix ω_G. According to the graph theory algorithm, analyze the derivation relationship between the factors influencing information literacy to obtain the list of in-out degrees as shown in Table IV.

TABLE IV. In-out degree values of each factor influencing information literacy

Influence factor	Out degree	In degree	Influence factor	Out degree	In degree
A	10	6	B	9	8
C1	4	7	C2	2	7
C4	5	6	C5	5	6
C6	0	4	C3	8	6
E	6	6	F	9	5
D	6	3			

The in-out degree values in Table IV are normalized to obtain the weight value ω_G of in-out degree for each factor, as shown in Table V.

TABLE V. In-out degree weight of each factor

A	B	C1	C2	C3	C4
0.131	0.139	0.090	0.074	0.090	0.115
C5	C6	D	E	F	
0.090	0.033	0.066	0.107	0.090	

(3) Model application. Through formula (7), parameter value of the final influence matrix ω can be determined, as shown in Table VI.

TABLE VI. Influence weight of each factor

A	B	C1	C2	C3	C4
2.0305	0.0973	1.458	0.3626	1.26	1.008
C5	C6	D	E	F	
0.4422	1.2305	0.528	0.063	0.1605	

From Table VI, the information literacy training model expression can be determined as formula (10).

$$\begin{aligned} h(INFOABL) = &-0.746 + 2.031 * A + 0.097 * B + 1.458 * C1 + 0.363 * C2 \\ &+ 1.26 * C3 + 1.008 * C4 + 0.442 * C5 + 1.2305 * C6 \\ &+ 0.528 * D + 0.063 * E + 0.161 * F + \varepsilon \end{aligned} \tag{10}$$

Where, ε is the adjustment parameter.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

V. CONCLUSION

(1) Through this research, it can be found that most higher vocational students have relatively good information awareness, information retrieval ability and information ethics at present, but the students' information knowledge and information discrimination ability are generally moderately poor, indicating that the students' information knowledge and discrimination ability still need to be strengthened at present. It is necessary to strengthen the learning of professional courses, improve the level of information knowledge, and better discriminate the information. According to the data, it is found that there is a great deviation in the students' information sharing ability and improvement awareness, indicating that learning of general courses is needed.

(2) This research analyzes students' information literacy based on six aspects, and studies the derivation relationship between the influencing factors according to the graph theory algorithm to determine the relative relationship between each factor. Regression algorithm is used to explore the relationship between information literacy level and various influencing factors to form a preliminary functional relationship.

(3) Due to the strong subjectivity of some data in the data collection, the limited data scale and the problem of dirty data, in data fitting analysis, the model coefficients show that information knowledge and students' information literacy level are disconnected, which is obviously inconsistent with the actual analysis. Therefore, this study subsequently uses graph theory algorithm to make secondary corrections to the coefficients, and will continue to conduct correction research for cases involving large data volume. So far, it has been in line with the real situation of students and teaching, so real problems are solved.

(4) This research provides feedback on students' information literacy based on data. Based on data, the theoretical discussion method is changed, and a training model is constructed in combination with algorithm analysis, which maximizes the training efficiency of students' information literacy, and also provides certain reference basis for evaluation of students' information literacy.

ACKNOWLEDGEMENTS

This research was supported by Key project of Shaanxi Vocational and Technical Education Society: Construction of Information Quality Training System for Higher Vocational Students (Grant No. 2021SZXZD05).

REFERENCES

[1] President committee on information literacy: final report. [2011-10-9]. http://www.ala.org/ala/acrl/acrlpubs/whitepapers/presidential.html
[2] Mike E. A Big6 skills overview. http://www.big6.com/showarticle.php?id=16
[3] Zhu C, Li Y (2014) Resorting the concept of information literacy. Journal of Schooling Studies (Shanghai) 011(003): 72-77

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 143-153
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

[4] Hu YY (2010) Investigation and Analysis of Information Literacy of Soochow University Graduate Students. Soochow University
[5] Wang R, Gu XQ, Zhang YY, et al. (2017) Investigation on information behavior of higher vocational students and discussion on information literacy education. Journal of Medical Informatics 38(9): 5
[6] Li K, Li XB, Wen DF, et al. (2013) Investigation and Reflection on Information Literacy of Higher Vocational Students in the Information Age. Journal of Jiamusi Vocational Institute 2013(7): 253
[7] Liu XW (2008) Academic Review of Information Literacy and Related Concepts. Library Journal 27(4): 12-13
[8] Mu XW (2006) On the connotation and training path of information literacy. Improving the scientific literacy of the whole people and building an innovative country – 2006 Proceedings of the China Association for Science and Technology 1997-2000
[9] Li JL (2012) Research on Information Literacy of Postgraduates in the Network Environment. Anhui University
[10] Zhang YB, Li Y, et al. (2003) New Definition of "Information Literacy". Educational Research 024(003):78-81
[11] Karin RS (2021) Graph theory: an introduction to proofs, algorithms, and applications. United States: CRC Press
[12] Diestel R (2000) Graph Theory. Mathematical Gazette 173(502): 67-128
[13] Li YL, Li G (2014) Graph Theory and Its Application in Image Processing, Tsinghua University Press
[14] Bondy JA, Murty U (1977) Graph theory with applications. Journal of the Operational Research Society 28(419): 237-238
[15] Cohen J, Cohen P, West SG, Aiken LS (2003) Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates
[16] Wang SG (1999) Linear Statistical Model: Linear Regression and Analysis of Variance. Higher Education Press
[17] Wang HW, Meng J (2007) Predictive modeling method of multiple linear regressio. Journal of Beijing University of Aeronautics and Astronautics 33(4): 5
[18] Yang ZJ, Wang WB, Huang P, et al. (2021) Linear Regression Classifier and Face Recognition Based on Local Weighted Representation. Computer Science 48(S02): 9
[19] Gu BT, Zhu JM, Gong L (2020) Empirical Analysis of Factors Influencing Grain Yield in China Based on Multiple Linear Regression. Natural Science Journal of Harbin Normal University 36(3): 6
[20] Liao WH, Huang YQ, He ZF, et al. (2021) Robustness comparison and application of parameter estimation in linear regression models. Journal of Applied Statistics and Management 40(5): 11
[21] Du JJ, Chen ZW (2010) A method for path analysis using SPSS linear regression. Bulletin of Biology 2010(2): 3