ISSN: 2688-8645

Research on the Evaluation of College Students' Algorithmic Literacy based on Factor Analysis and Cluster Analysis

-- Taking Anhui Province as an Example

Ruichao Nie

School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, 233030, China

Abstract

This paper takes 16 prefectures and cities under Anhui Province as the research object, and constructs an index group that affects the level of algorithm literacy. First, based on the qualitative research on the quality level of algorithm literacy in Anhui Province, a combination model of the influencing factors of algorithm literacy level was established by factor analysis, and the comprehensive score of algorithm literacy was used as its representative index; secondly, the algorithm literacy level was established by cluster analysis method. To explore the development model of algorithm literacy in different regions. The final research results show that: in the combined model, the three factors of regional education comprehensive strength, per capita education level and material resource input have a greater impact on the algorithm literacy level of regional college students; in the classification model, Hefei City is the core of Anhui Province. In the city, the algorithm literacy level of its college students ranks first. Finally, research on the commonality of the two models is conducted to reasonably evaluate the algorithm literacy of college students in various cities in Anhui Province, and provide feasible suggestions for improving the algorithm literacy of college students in Anhui Province.

Keywords

Algorithm Literacy; Factor Analysis; Cluster Analysis; Comprehensive Educational Strength.

1. Introduction

With the rapid update of algorithm technology, the rapid rise of new algorithm platforms and the emergence of "new media" have greatly changed the way of information transmission, the amount of transmission and its concealment. In order to cope with the "algorithmic" challenges of the new era, college students must have matching algorithmic literacy. However, at present, my country's research system on "algorithmic literacy" is still not perfect, and the related research on "algorithmic literacy" cannot be transferred to the field of education. There are problems such as lack of "algorithmic literacy" ability among college students in my country. Therefore, it is particularly important to find the shortcomings of contemporary college students in the field of "algorithmic literacy" and propose improvement plans.

Algorithmic literacy has been mentioned many times as a countermeasure to solve algorithmic problems. Fang Jiuxian (2020) proposed to discuss the cultivation of user media literacy from the perspective of recommendation algorithms on the research of algorithm literacy, the purpose is to provide a reference for effectively solving the drawbacks brought by the algorithm recommendation technology in the era of intelligent media, and to play the role of media literacy in correcting technical deviations. In 2017, a report titled "Code Dependence: The Pros and Cons of the Algorithmic Age" pointed out the "growing demand for algorithmic literacy,

ISSN: 2688-8645

transparency, and oversight" in the algorithmic age, which proposed that algorithm designers should be ethical. Training on the other hand can help directly improve the algorithmic literacy efficiency of the general population. Xiao Tian (2019) used media literacy and Lasswell's "5W" model as the theoretical basis to construct secondary indicators to describe the current situation of algorithm literacy of contemporary young netizens.

2. Theory and Method

2.1. Factor Analysis

2.1.1. The Principle of the Method

As a data reduction technique, factor analysis can integrate multiple variables into a few factors to analyze the original variables. That is, using a few independent non-observed variables to reflect most of the important information given in the original data.

2.1.2. Theoretical Model

Let P have original variables Z_1, Z_2, ..., Z_m, and the mean of each variable is 0, and the standard deviation is n. Now each original variable is represented by a linear combination of p common factors F_1, F_2, ..., Fp, the basic model of factor analysis is as follows:

$$\begin{cases} Z_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1p}F_p + c_1U_1 \\ Z_2 = a_{12}F_1 + a_{22}F_2 + \cdots + a_{2p}F_p + c_2U_2 \\ \cdots \\ Z_m = a_{m1}F_1 + a_{m2}F_2 + \cdots + a_{mp}F_p + c_mU_m \end{cases}$$

Represented in the form of a matrix as: $Z = AF + CU$

2.2. Cluster Analysis

2.2.1. The Principle of the Method

According to different research methods, clustering methods can be divided into fast clustering, hierarchical clustering and two-stage clustering and other clustering methods. This paper intends to use the fast-clustering method, that is, the K-means clustering method to cluster the samples.

2.2.2. Theoretical Mode

The Phi square distance is used to measure the distance, and the model formula is as follows:

$$d_{xy} = \sqrt{\frac{\sum_{i=1}^{k}\frac{[x_i - E(x_i)]^2}{E(x_i)} + \sum_{i=1}^{k}\frac{[y_i - E(y_i)]^2}{E(y_i)}}{n}}$$

3. Construction of the Indicator System

3.1. Construction Principles

In order to accurately quantify the algorithm literacy level of college students in Anhui Province, this paper follows the five principles of scientificity, relevance, integrity, flexibility and practicality when selecting influencing factors. Scientificity refers to the authenticity and reliability of data sources, such as the official government statistical website of the data source, to ensure that the final evaluation results are true and reliable; relevance refers to retaining relevant indicators and excluding relevant indicators under the condition that various indicators have been selected completely. Indicators with a very low degree of sexuality ensure

ISSN: 2688-8645

that the model has a certain pertinence after it is established; completeness means that the selected indicators are sufficient to cover the core indicators for measuring the level of algorithm literacy of college students; flexibility means that according to the local characteristics of different regions Selected representative indicators with local characteristics, Anhui Province, as a city in central my country, has a certain lag in the development of its education level, so the selection of indicators must conform to the characteristics of local education levels; Practicality refers to the final mathematical model established It is of practical significance to reasonably evaluate the algorithm literacy level of college students in Anhui Province, and provide feasible suggestions for the development of higher education in Anhui Province. The above five principles are the basis for the construction of the indicator system in this paper, and provide a guarantee for the construction of the model.

3.2. Evaluation Object

The goal of this study is to comprehensively evaluate the algorithm literacy level of college students in Anhui Province. Therefore, a total of 16 cities in Anhui Province are selected as the evaluation objects, including Hefei City, Huaibei City, Bozhou City, Suzhou City, Bengbu City, Fuyang City, Huainan City, Chuzhou City, Lu'an City, Ma'anshan City, Wuhu City, Xuancheng City, Tongling City, Chizhou City, Anqing City, Huangshan City.

3.3. Definition of Evaluation Index System

Table 1. Evaluation system of algorithm literacy level of college students

Level indicators	The secondary indicators	Index sign
Overall size	Years of education per capita	X_1
	Total number of regular universities	X_2
	Number of college graduates	X_3
	Higher education per 100000 people	X_4
faculty	Number of college teacher	X_5
	The proportion of the teacher and students	X_6
Technological achievements	Number of scientific papers published	X_7
	Full-time researcher	X_8
	The number of patents granted	X_9
Economic strength and investment	Research and experiment funds for colleges and universitis	X_{10}
	GDP per capita	X_{11}
	Educational funds expenditure	X_{12}
	Investment in fixed assets in education increased	X_{13}

The selection of indicators requires that such indicators cover representative indicators sufficient to measure the algorithm literacy level of college students in Anhui Province. By synthesizing the previous research and the development status of education in Anhui Province, and taking into account the selection requirements of aggregate and average indicators. Therefore, this paper divides the selected indicators into four categories: overall scale, faculty strength, scientific and technological achievements, economic strength and investment. Among the overall scale, the per capita education years (X_1), the number of ordinary colleges and universities (X_2), The number of graduates of higher education (X_3) and the population of higher education per 100,000 people (X_4) are four indicators. Among the faculty, the number of full-

ISSN: 2688-8645

time teachers in higher education (X_5) and the proportion of teachers and students (X_6) are selected. From the scientific and technological achievements, three indicators are selected: the number of published scientific papers (X_7), the full-time researchers (X_8), and the number of patents granted (X_9),from the economic strength and financial investment, four indicators are selected: the investment in research and experimental development of colleges and universities (X_{10}), the per capita GDP (X_{11}), the education expenditure (X_{12}), and the growth of fixed self-examination of your investment in education (X_{13}), a total of 13 indicators were selected as the indicators of this research.

4. A Comprehensive Evaluation Model for the Algorithm Literacy Level of College Students

4.1. Evaluation of the Algorithmic Literacy Level of College Students based on Factor Analysis

4.1.1. Model Suitability Check

In this paper, there are many variables and data selected to study the algorithm literacy level of college students in Anhui Province. If the correlation between the variables is insufficient, factor analysis cannot be carried out. Therefore, in order to verify the applicability of factor analysis to the index group, KMO and Bartlett tests were performed on each index in advance, and the test results are shown in Table 1: the moderate test value of KMO is 0.654>0.5, and the sphericity test value of Bartlett is Sig. is 0<0.01, rejecting the null hypothesis that the correlation coefficient matrix is an identity matrix. It shows that there is a correlation between the variables, and factor analysis can be carried out.

Table 2. KMO and Bartlett's test

KMO		0.654
Sphercity test of Bartlett	The approximate chi-aquare	372.159
	df	78
	Sig.	0.000

4.1.2. Extract Common Factors

According to the principle that the cumulative variance contribution is greater than 85%, three common factors can be extracted. The results show that the eigenvalues of the first three factors are greater than 1, the variance contribution rate of the first factor is 68.278%, the variance contribution rate of the second factor is 12.566%, and the variance contribution rate of the third factor is 8.795%. The cumulative variance contribution rate of the three factors is 89.639%>85%. Therefore, these three factors can represent 89.639% of all index information, and it also shows that these three factors are sufficient to explain the algorithm literacy level of the college students in Anhui Province.

Table 4 shows the factor loading values after rotation. It can be seen from Table 4 that the principal component 1 is mainly composed of variables X_2, X_3, X_5, X_7, X_8, X_9, X_{12}. These variables all reflect the students in colleges and universities in Anhui Province. The comprehensive strength of algorithm literacy is set as the common factor F1; the principal component 2 is mainly composed of variables X_1, X_4, X_{11}, which mainly reflect the per capita education level, and is set as the common factor F2; the principal component 3 is mainly composed of variables X_6, X_{10}, X_{13}, which mainly reflects the input of material resources in higher education, and is set as the common factor F3

ISSN: 2688-8645

Table 3. Total variance explained

Initial eigenvalue			Extract the sum of aquares and load			Rotate squares and load		
sum	variance%	accumulate%	sum	variance%	accumulate%	sum	variance%	accumulate%
8.876	68.278	68.278	8.876	68.278	68.278	5.853	45.021	45.021
1.634	12.566	80.844	1.634	12.566	80.844	3.714	28.567	73.588
1.143	8.795	89.639	1.143	8.795	89.639	2.087	16.051	89.639
0.775	5.958	95.597						
0.274	2.109	97.706						
0.169	1.296	99.003						
0.067	0.519	99.522						
0.027	0.204	99.726						
0.021	0.159	99.885						
0.012	0.092	99.977						
0.002	0.014	99.991						
0.001	0.009	99.999						
6.729E-005	0.001	100.000						

Table 4. Rotation component matrix

variable		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
ingredient	1	0.55	0.862	0.838	0.44	0.841	-0.186	0.675	0.797	0.866	0.593	0.143	0.964	-0.166
	2	0.619	0.472	0.519	0.877	0.511	-0.079	0.519	0.579	0.477	0.273	0.944	-0.061	0.027
	3	0.257	0.099	0.125	0.111	0.142	-0.644	0.47	0.066	0.09	0.668	0.087	0.067	0.926

4.1.3. Model Building

The calculation formulas of F1, F2 and F3 can be obtained from Table 5:

$$F_1 = -0.017X_1 + 0.169X_2 + 0.142X_3 - 0.138X_4 + 0.144X_5 + 0.005X_6 + 0.050X_7 + 0.109X_8 + 0.170X_9 + 0.087X_{10} - 0.276X_{11} + 0.396X_{12} - 0.149X_{13}$$

$$F_2 = 0.170X_1 - 0.013X_2 + 0.021X_3 + 0.382X_4 + 0.013X_5 + 0.079X_6 + 0.043X_7 + 0.079X_8 - 0.011X_9 + 0.101X_{10} + 0.528X_{11} - 0.377X_{12} - 0.017X_{13}$$

$$F_3 = 0.045X_1 - 0.058X_2 - 0.045X_3 - 0.057 - 0.035X_5 - 0.354X_6 + 0.169X_7 - 0.082X_8 - 0.064X_9 + 0.316X_{10} - 0.054X_{11} - 0.032X_{12} + 0.551X_{13}$$

In order to facilitate subsequent research, this paper defines the comprehensive score of factor analysis, that is, the algorithm literacy level of college students in Anhui Province as F, and combines the three common factors F1, F2, and F3 obtained by factor analysis to establish a linear function of F and F1, F2, and F3. According to the relationship, the comprehensive score of algorithm literacy of college students in various cities in Anhui Province is obtained:

ISSN: 2688-8645

Table 5. Component score coefficient matrix

variable	ingredient		
	1	2	3
X_1	-0.017	0.170	0.045
X_2	0.169	-0.013	-0.058
X_3	0.142	0.021	-0.045
X_4	-0.138	0.382	-0.057
X_5	0.144	0.013	-0.035
X_6	0.005	0.079	-0.354
X_7	0.050	0.043	0.169
X_8	0.109	0.079	-0.082
X_9	0.170	-0.011	-0.064
X_{10}	0.087	-0.101	0.316
X_{11}	-0.276	0.528	-0.054
X_{12}	0.396	-0.377	-0.032
X_{13}	-0.149	-0.017	0.551

$$F = (0.45021F_1 + 0.28567F_2 + 0.16051F_3)/0.89639$$

Then, according to the calculation formulas of F, F1, F2, and F3, the comprehensive scores of the algorithm literacy of college students in various cities are obtained as shown in Table 6 below:

Table 6. Comprehensive score and ranking of algorithm literacy of college students

city	F_1	F_2	F_3	F	ranking
He fei	3.22355	1.63766	0.23217	2.18	1
Wu hu	-0.13368	1.2732	0.24104	0.38	2
Huai bei	-0.80754	-0.13164	3.13001	0.11	3
Ma anshan	-0.81025	1.1972	0.42205	0.05	4
Chu zhou	-0.22876	0.30963	0.29414	0.04	5
Fu yang	0.89158	-1.97276	0.54163	-0.08	6
Ben gbu	-0.1735	-0.08209	0.09737	-0.1	7
An qin	0.06119	-0.28619	-0.28361	-0.11	8
Huai nan	-0.08541	-0.39321	-0.14491	-0.19	9
Tong lin	-0.85609	0.77755	-0.13336	-0.21	10
Lu an	0.30844	-1.02904	-0.32784	-0.23	11
Su zhou	0.25821	-0.90803	-0.49943	-0.25	12
Bo zhou	0.33113	-1.45855	-0.30774	-0.35	13
Huang shan	-0.74333	0.40942	-0.74775	-0.38	14
Xuan cheng	-0.42584	0.23684	-1.59897	-0.42	15
Chi zhou	-0.80969	0.42002	-0.91481	-0.44	16

ISSN: 2688-8645

4.2. Quality Classification of Algorithm Literacy Level of College Students based on Cluster Analysis

The final cluster analysis results are shown in Table 7 below. Through the cluster analysis method, the algorithm literacy levels of college students in 16 prefectures and cities in the province can be divided into three categories:

The first category: Hefei City.

The second category: Huaibei City, Bengbu City, Huainan City, Ma'anshan City, Xuancheng City, Tongling City, Chizhou City, Huangshan City.

The third category: Bozhou City, Suzhou City, Fuyang City, Chuzhou City, Lu'an City, Anqing City, Wuhu City.

As the city with the highest level of higher education construction in the province, Hefei's development model and various human and material resources are different from other cities in the province and have certain uniqueness. The other 15 prefectures and cities are roughly divided into two categories, and each has its own advantages and disadvantages in the level of algorithm literacy of college students.

Table 7. Clustering results

Serial number	city	cluster	distance
1	He fei	1	0.000
2	Huai bei	2	21853.798
3	Bo zhou	3	100306.367
4	Su zhou	3	22525.733
5	Beng bu	2	220373.957
6	Fu yang	3	331828.087
7	Huai nan	2	114332.975
8	Chu zhou	3	120691.279
9	Lu an	3	21346.014
10	Ma Anshan	2	37660.026
11	Wu hu	3	103922.361
12	Xuan chen	2	57784.419
13	Tong ling	2	118445.046
14	Chi zhou	2	119809.579
15	An qing	3	28475.612
16	Huang shan	2	145899.356

5. Result Analysis

5.1. Analysis of the Results of the Evaluation of the Algorithm Literacy Level of College Students based on Factor Analysis

From the results of factor analysis, it can be seen that there are obvious differences and non-equilibrium in the level of algorithm literacy among college students in 16 prefectures and cities in Anhui Province. Among them, the regional comprehensive strength factor F1 has the greatest influence, which plays a major role in promoting the algorithm literacy level of regional college students in Anhui Province; followed by the per capita education level F2, and finally the

ISSN: 2688-8645

material resource investment in regional education F3, so the algorithm for college students in Anhui Province. The cultivation of literacy should focus on improving the comprehensive strength of the region rather than just pursuing a single direction of development.

It can also be seen intuitively in the final comprehensive ranking that, as the capital city of Anhui, the algorithm literacy level of its college students is much higher than that of other cities. Hefei City not only has many well-known domestic universities, but also has a regional educational resource investment and a regional human resource level that is far higher than other cities, which also leads it to the highest level in the final evaluation of the algorithm literacy level of college students; Secondly, Wuhu City, Huaibei City, Maanshan City and Chuzhou City, the comprehensive scores of these four regions are positive, indicating that these four regions are still in a trend of rapid growth; The algorithm literacy level of the remaining 11 prefectures and cities is in the middle and lower reaches of the level, which cannot reach the average level of Anhui Province, and there is still huge room for improvement in the future.

5.2. Result Analysis of the Classification of Algorithm Literacy Level of College Students based on Cluster Analysis

The classification results of the algorithm literacy level of college students show that there are aggregations and differences between the algorithm literacy of college students in various cities in Anhui Province. Among the first echelon, Hefei takes the lead. This is the same as the result of factor analysis. Hefei's comprehensive strength in higher education is indeed far more than other cities, and it is in the forefront in all aspects; in the second batch, Huaibei City, Bengbu City, Huainan City, Ma'anshan City, Xuancheng City, Tongling City, Chizhou City, and Huangshan City have similar levels of development in the level of algorithmic literacy of college students, which is mainly reflected in the fact that the investment in education is large, but the number of high-quality and high-quality institutions of higher learning is relatively small; In the third batch, Bozhou City, Suzhou City, Fuyang City, Chuzhou City, Lu'an City, Anqing City, and Wuhu City belong to the same category. The overall performance is that although the level of algorithm literacy is slightly higher, the overall investment is insufficient.

According to the above analysis, among these three categories, except for the first category of Hefei, there are still many deficiencies in the overall development of the other two categories of cities.

5.3. Comprehensive Analysis Results based on Factor Analysis and Cluster Analysis

Based on the analysis results of the two models, the algorithm literacy level of college students in various cities in Anhui can be divided into five levels. The algorithm literacy level of college students in 16 cities is divided into five levels: high level, medium-high level, medium level, medium-low level and low level.

Table 8. Urban Hierarchy

grade	city	Education quality level
The first level	He fei	High level
The second level	Huai bei	High and middle level
	Ma Anshan	High and middle level
	Wu hu	High and middle level
The third level	Chu zhou	High and middle level
	Fu yang	High and middle level
	Beng bu	High and middle level

ISSN: 2688-8645

	An qing	High and middle level
	Huai nan	High and middle level
The fourth level	Tong ling	Low and middle level
	Lu an	Low and middle level
	Huang shan	Low and middle level
The fifth level	Xuan cheng	Low level
	Chi zhou	Low level
	Su zhou	Low level
	Bo zhou	Low level

6. Relevant Countermeasures and Suggestions

6.1. Coordinated Development of Regional Educational Resources

As an important city in the central region of my country, Anhui Province plays an important role in the development of the algorithm literacy level of its college students for the regional construction of Anhui Province. However, according to the above analysis, the distribution of higher education resources in Anhui Province is uneven, and there are sufficient educational resources in some areas, but the level of higher education in most areas is very backward. As the capital city of Anhui Province, Hefei's higher education development ranks among the highest in Anhui Province. It not only has the University of Science and Technology of China, Hefei University of Technology, Anhui University and other well-known domestic institutions of higher learning, but also has more than 50 higher education institutions. The number of colleges and universities exceeds that of the other 15 cities combined. At the same time, Hefei's investment in educational resources is also very abundant. Not only does it have sufficient self-investment, but it also ranks first in Anhui Province in terms of teaching staff and infrastructure construction. The development of the algorithm literacy level of the overall college students in Anhui Province is inseparable from the coordinated development of various regions. In the future, Anhui Province should pay attention to the distribution of educational resources, which can not only ensure the full use of human resources in different regions, but also promote the level of regional construction and improve in an all-round way. The comprehensive strength of college students in Anhui Province.

6.2. Realize Balanced Investment and Improve the Comprehensive Strength of Higher Education in Anhui Province

The improvement of regional education level requires comprehensive development in all aspects. The reason why Hefei City can become the apex of higher education in Anhui Province is that the construction of higher education in Hefei City is carried out simultaneously from multiple aspects. Therefore, in the construction of higher education in different regions of Anhui Province, we should focus on improving the comprehensive strength of the regional education level. In terms of capital investment, we should not only increase the construction of campus fixed assets, but also increase the investment in related scientific research and education; not only in the field of talents. To strengthen teachers and improve the ratio of teachers and students, it is also necessary to formulate relevant strategies for talent training and talent introduction; in the construction of higher education institutions, on the one hand, it is necessary to move into well-known domestic institutions, but at the same time, it is also necessary to strengthen the higher education institutions in the region. School construction, to create a unique well-known institutions of higher learning in Anhui Province.

ISSN: 2688-8645

Therefore, in the process of the construction of higher education in Anhui Province, we should not only pursue unilateral total growth, but should develop in an all-round way, from the aspects of talents, teachers, and campus construction. Only in this way can Anhui Province be improved as a whole. level of higher education quality.

6.3. Driven by Government Policies, Providing Strong Guarantees

Government support is also an important force to promote the construction of algorithmic literacy level of college students in Anhui Province. For private and different colleges and universities, it is difficult to rely on their own strength to complete the construction of funds, human resources and campuses. This requires the strong support of the government. At the same time, the government also has a strong administrative method, which plays the most important role in the distribution of regional educational resources, and has a good catalyst role in easing the tension of higher education resources in different regions. At the same time, the government is also leading the construction of an important bridge between regions, strengthening strategic cooperation with regions rich in educational resources, and formulating corresponding talent introduction strategies, which are also important measures to improve the algorithmic literacy level of college students in Anhui Province.

Acknowledgments

Fund Project: 2022 Undergraduate Scientific Research and Innovation Fund Project of Anhui University of Finance and Economics "Evaluation and Research of Algorithmic Literacy of University Students in the Era of Artificial Intelligence" (XSKY22026ZD).

References

[1] Fang Jiuxian. Cultivation of user media literacy from the perspective of recommendation algorithm [J]. Science and Technology Communication, 2021, 13(20): 120-122. DOI: 10.16607/j.cnki.1674-6708. 2021.20.037.

[2] Hong Bingcun. Teaching of high school information technology "algorithm and program design" project based on core literacy [J]. China Information Technology Education, 2021(19):54-57.

[3] Kang Fuzhu, Tian Menglong. Improve user media literacy to avoid algorithm bias [J]. News and Communication, 2021(01): 4-5+8.

[4] Yu Xuan, Xiao Tao. New challenges and new requirements of public media literacy under algorithmic recommendation mechanism [J]. Audiovisual, 2020(04):185-187. DOI: 10.19395/ j. cnki. 1674-246x.2020.04 .098.

[5] Luo Yanfei. Media information reception literacy: A new dimension of media literacy in the context of algorithm recommendation [J]. Research on Communication Power, 2020, 4(02): 4-5+8.

[6] Xiao Tian. Research on algorithm literacy evaluation of young netizens in the era of artificial intelligence [D]. Jinan University, 2019. DOI: 10.27167/d.cnki.gjinu.2019.001241.