Cluster Analysis on Polytechnic Professional Settings and AI Work Based on Entropy Power Method

Jiangyi Lv[1], Cherry Jiang[2] and Zhiwang Gan[3,*]

[1]Beijing Polytechnic University, Beijing, China
[2] Columbia University, New York, US;
[3] Tianjin Peisheng Education Technology Co., Ltd, Tianjin, China
Email:ljybuaa@163.com; Corresponding author email:mhi2008@yeah.net

Abstract. Vocational education in undergraduate colleges and universities is an important part of the national modern education system. We aim for comprehensive literacy in training goals, reinforcing the integration of industry and education within professional contexts, and fostering organizational innovation in governance capabilities. Entropy Power method is a method that uses information entropy theory to calculate. This Entropy Power Model BigData Analysis and cluster Research on Vocational Education Professional Setting and Artificial Intelligence Career Orientation just Based on Entropy Power Method.

Keywords: Cluster Analysis, Polytechnic Professional Setting, AI (Artificial Intelligence), Entropy Power Method

1 Instruction

We aim for comprehensive literacy in training goals, reinforcing the integration of industry and education within professional contexts, and fostering organizational innovation in governance capabilities. We should leverage artificial intelligence technology to facilitate a quantum leap in vocational education formats and enhance the compatibility between the supply side of professional settings and the demand side of industry. This approach ensures adaptation to the advancements of the artificial era. Constructing a modern vocational education system represents a significant development strategy for global vocational education, influencing the future direction of vocational education development.

1.1 Data description

The data table header includes more than 30 indicators such as school name, province, professional name, professional code, corresponding industry, number of students, number of students in the first grade, number of students in the second grade, etc., with a total of 616 data records.

R. Chitrakar et al. (eds.), *Proceedings of the 5th International Conference on New Computational Social Science (ICNCSS 2025)*, Advances in Social Science, Education and Humanities Research 954,
https://doi.org/10.2991/978-2-38476-456-3_29

Serial number	First-level indicator	Secondary indicators
	Basic information	School name
2		Province
3		Professional name
4		Professional code
5		Corresponding industry
6	Student information	Number of full-time higher vocational students
7		Number of students in the first grade
8		Number of students in the second grade
9		Number of students in the third grade
Ten		2023 enrollment plan
11		The actual number of admissions in 2023
12		The number of new students in 2023
13		2023 freshman registration ratio (%)
14		The number of students from the province in 2023 (person/major)
15		The proportion of students from the province in 2023 (%)
16	Employment information	Number of graduates in 2023 (person/major)
17		Initial employment rate of 2023 graduates (%)
18		Employment ratio of 2023 graduates in this province and city (%)
19		The employment rate of 2023 graduates (%)
20	Teacher information	Number of full-time teachers in the school (person/major) (school data)
21		The quality ratio of full-time teachers (%) (school-wide data)
22		Total number of part-time teachers in the near 3 years academic year (person/professional)
23		The proportion of part-time teachers' teaching hours in the near 3 years academic year is (%)
24	Practical training resources	Number of on-campus training bases (unit/professional)
25		Frequency of use of on-campus training bases in the near 3 years academic year (human time)
26		Number of off-campus internship training bases (one/professional)

27		The number of internship students accepted by the off-campus internship training base for half a year in the near 3 years academic year (person/major)
28		Total number of cooperative enterprises (unit/professional)
29		Total number of cooperative enterprise order training (person/professional)
30		The total number of courses jointly developed by co-operative enterprises (men/professional)
31		The total number of part-time teachers supported by cooperative enterprises (person/professional)
32		The total number of students receiving internships in cooperative enterprises (person/major)
33		Total value of equipment donated by cooperative en-terprises (10,000 yuan/professional)
34		Total value of equipment donated by cooperative en-terprises (10,000 yuan/professional)
35		The total number of employment of 2023 graduates accepted by cooperative enterprises (person/major)
36		Total number of employees trained for enterprises (person/professional)

Table 1 Data indicators

2 Research method

2.1 Gaussian hybrid model

The Gaussian Mixture Model (GMM) uses multiple Gaussian distribution mixing as a parameter model and solves the algorithm through expected maximization. GMM can be regarded as a generalization of K-means clustering algorithm.

2.2 Bray-Curtis distance

Bray-Curtis distance is a common distance measurement method in ecology. The Bray-Curtis distance in n-dimensional space is defined as:

$$d(x,y)=\frac{\sum_{i=1}^{n}|x_i-y_i|}{\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}y_i}$$

Its value is between [0,1], which can be used to calculate the similarity between samples.

Entropy Law

. Entropy weight method is a method that uses information entropy theory to calculate multidimensional data and comprehensively evaluate data weights. The smaller the information entropy of a set of data, the greater the dispersion of the data, the greater the information content. The entropy weight method is to calculate the weight of data based on the entropy of data information and conduct a comprehensive evaluation of multiple indicators with weight.

1. Analysis of professional layout of schools and provinces

Data analysis of students and the results obtained are as Figure 6 to 4 as shown.

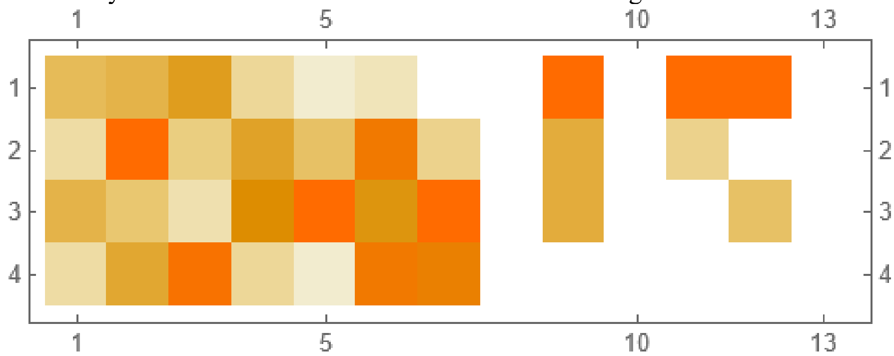

Figure1. C Class : {Gansu, Hebei, Shanxi, Shaanxi}

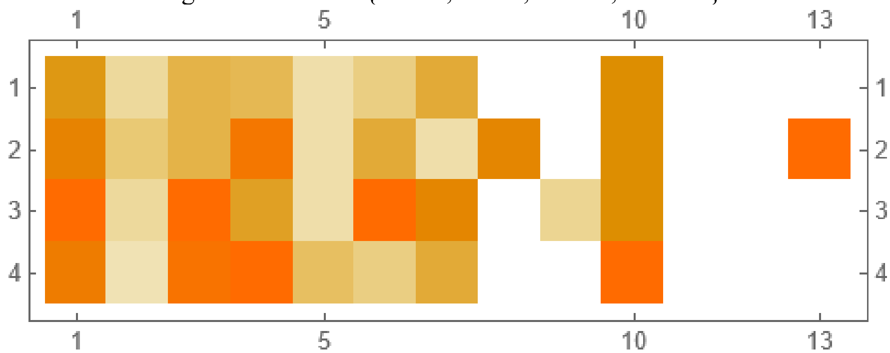

Figure 2 Class : {Guangdong, Guangxi, Jiangxi, Shandong}

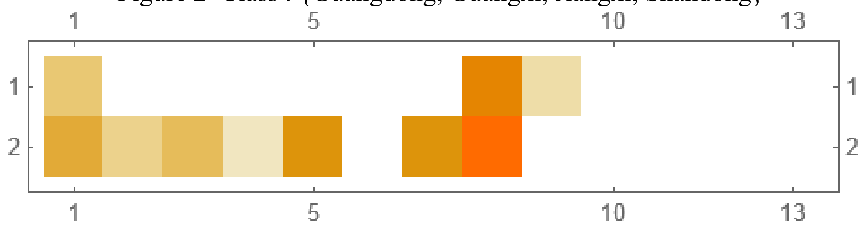

Figure3 Class : {Guizhou, Zhejiang}

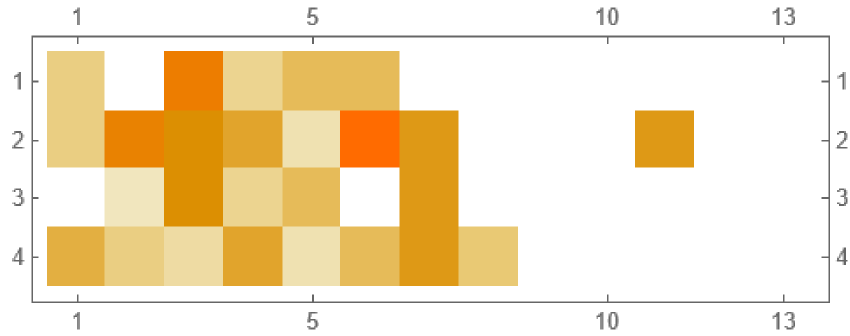

Figure 4 Class : {Hunan, Jiangsu, Liaoning, Shanghai}

Professional development analysis

（1）. Analysis of regional training internship resources

According to the province where the school is located, summarize and count the data of training and internship resources in different regions, including the number of on-campus training bases, the frequency of on-campus training bases in the near 3 years academic year, the number of off-campus internship training bases, etc., and $S_{i,j}, 1 \leq i \leq 20, 1 \leq j \leq 14$ obtain regional training resource data.

The following uses the entropy power method to evaluate the training resources. First of all, standardize the data and make

$$s_{i,j} = \frac{S_{i,j}}{MAX\{S_{i,j}\}_{1 \leq i \leq 20}}$$

Bring $s_{i,j}$ Scored as various second-level practical training indicators. Right $s_{i,j}$ Carry out normalization,

$$\bar{s}_{i,j} = \frac{s_{i,j}}{\sum_i s_{i,j}}$$

Calculate information entropy

$$K_{i,j} = \begin{cases} 0, \bar{s}_{i,j} = 0 \\ \bar{s}_{i,j} \cdot \ln \bar{s}_{i,j} \end{cases}$$

$$e_j = -\frac{\sum_i K_{i,j}}{\ln 20}$$

Calculate the utility value of information entropy

$d_j = 1 - e_j$

Determine the weight of sub-indexes through information entropy utility value

$$w_j = \frac{d_j}{\sum d_j}$$

Specific weights such as
Table2 As shown.

W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13
0.03	0.02	0.06	0.07	0.05	0.11	0.09	0.04	0.04	0.15	0.14	0.06	0.15

Table2 The weight of the second-level indicators of practical training resources

The maximum weight is w_{10} ·The corresponding index is the total value of equipment donated by cooperative enterprises, indicating that in terms of the difference in training resources. The total value of equipment donated by cooperative enterprises in schools in various provinces is the largest.Use the weight to weight the scoring of practical training resource indicators in each province, and the results obtained are as Figure as shown.

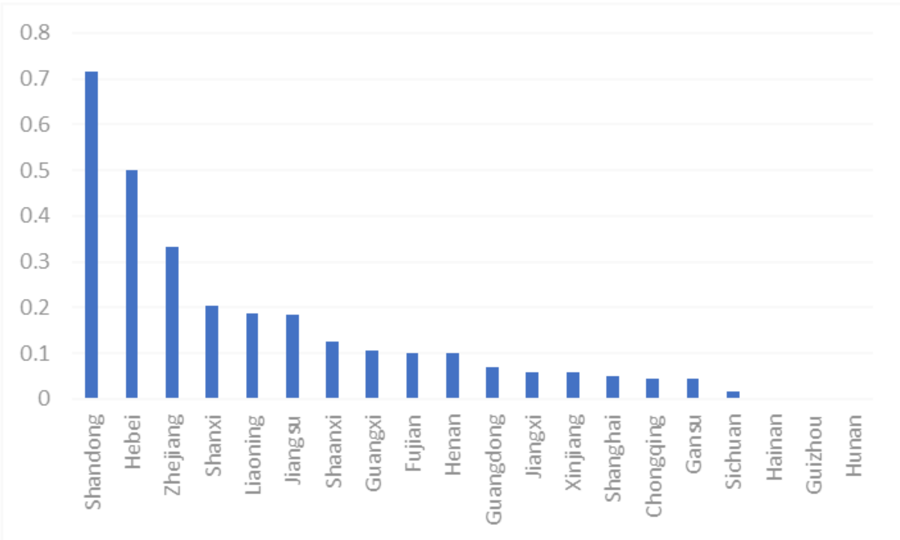

Figure 5 Evaluation results of regional training resources

Analysis of industry training internship resources

. According to the corresponding industry of the major, summarize and count the training and internship resource data of different industries, including the number of on-campus training bases, the frequency of use of on-campus training bases in the near 3 years academic year, the number of off-campus internship training bases, etc., and obtain regional training resource data.

$$F_{i,j}, 1 \leq i \leq 13, 1 \leq j \leq 14$$

The following uses the entropy power method to evaluate the training resources. The specific calculation process is similar to the previous section, which can be obtained. The weight of the sub-index is:

W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13
0.05	0.04	0.06	0.06	0.04	0.09	0.05	0.08	0.04	0.18	0.14	0.06	0.11

Table3 Weight of industry training resource indicators

The maximum weight is still w_{10}, consistent with the weight of regional training and internship resources, in terms of practical training resources, the total value of equipment donated by cooperative enterprises varies the most.Use weights to weight and sum the scores of practical training resource indicators in various industries, and the results obtained are as Figure 6 as shown.

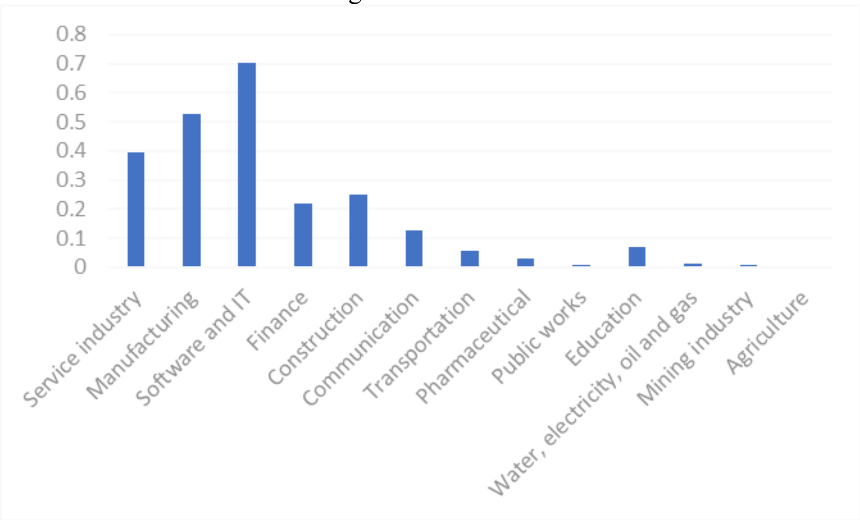

Figure 6 Evaluation results of industry training resources

3 Conclusion

The top 3 training resources are software and information technology, manufacturing and service industry, which is significantly different from the number of professional layout and the distribution of students in school.It is indicating that enterprises and schools are willing to invest more practical training resources in software and information technology majors, and prefer software and information in the job market of technical professional talents.

Reference

1. Yi Dexin, Shan Xuerong. The survival challenges, self-adaptation, and realization paths of education in the era of artificial intelligence [J]. Vocational Education Forum, 2024, 40(11): 108-114;
2. Jiangyi Lv, Cherry Jiang, Zhiwang Gan. Analysis on teacher teaching evaluation system based on clustering constant modulus， ITEME 2023 SHS Web of Conferences, 2023(3): 274-279. The 2023 International Conference on Information Technology in Education and Management Engineering (ITEME 2023), Volume 140, March 2th-30th, 2023. Shanghai, China, EDP Sciences;
3. Jiangyi Lv, Cherry Jiang, Xiaolong Ren; Construction and Analysis of Higher Vocational College Teaching Evaluation System Resources Based on Data Mining Function;Proceedings of the Global Conference on Robotics,Artificial Intelligence and Information Technology (GCRAIT 2023) ， 2023 : 529-531. .July 30-31,2023 Chicago,USA ,536-539,
4. Han Yu, Xu Han, Zhao Chu. Digitalization of vocational education professional ability standards: ability map and application scenario development . Educational Research and Experiment, 2024, (05): 99-107;
5. Yuan Xiaohua, Zhang Miao. Upgrading and Digital Transformation of Vocational Education Majors: Elements,ities, and Optimized Strategies[J]. China Vocational and Technical Education, 2024, (20): 80-87.
6. Wang Yong, Yang Hengyue, Qin Zhen. Study on the Match between Major Settings of Higher Vocational Colleges and Regional Industrial Development under Background of Industry-Education Integration[J]. Technology Wind, 2024, (19): 63-65. DOI: 1.19392/j.cnki.1671-7341.202419021.
7. Wei Heng, Deng Jing. Analysis of the Impact of Industrial Structure Adjustment on Vocational Education Majors - A Case Study of Shenzhen[J]. Science,, and Education, 2024, (12): 155-158. DOI: 10.1681/j.cnki.kjwh.2024.12.036.
8. Zhang Jie, Chen Zhixin. The Logic and Path of Matching Major Settings in Vocational Education with Local Industries[J]. Industrial Innovation Research, 204, (08): 166-168.