ISSN 2522-6398 Vol. 8, Issue 10: 219-223, DOI: 10.25236/FER.2025.081032

Big Data Analysis and Cluster Research on Vocational Professional Setting and Artificial Intelligence Career Based on Gaussian Mixture Model

Jiangyi Lv[1], Cherry Jiang[2], Nan Wang[1]

[1]*Auto. College, Beijing Polytechnic University, Beijing, China*
[2]*Teachers College, Columbia University, New York, USA*

Abstract: *This Paper is mainly the analysis of data on the establishment of Vocational Education Bachelor's Degree Programs and research on employment orientation. Vocational education in undergraduate colleges and universities is an important part of the national modern education system. A new thing in the field of modern vocational education, and is currently in a critical of getting started. 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. To achieve self-adaptive transformation, we must adhere to the core principles of vocational education, balancing value rationality with instrumental rationality. Entropy Power method is a method that uses information entropy theory to calculate. This Gaussian Mixture Model Big Data Analysis and cluster Research on Vocational Education Professional Setting and Artificial Intelligence Career Orientation just Based on Entropy Power Method.*

Keywords: *Vocational Education, Cluster Research, Big Data Analysis, Artificial Intelligence, Gaussian Mixture Model*

1. Introduction

The collaboration in educational forms is crucial. 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. To achieve self-adaptive transformation, we must adhere to the core principles of vocational education, balancing value rationality with instrumental rationality [1]. 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. Establishing a multi-faceted talent training target is essential, with a focus on cultivating high-level skilled talents equipped with "digital literacy and craftsmanship." This approach ensures adaptation to the advancements of the artificial era [2]. Constructing a modern vocational education system represents a significant development strategy for global vocational education, influencing the future direction of vocational education development. The undergraduate professional data analysis will be discussed in the following sections [3].

2. Data Description

2.1 Data Filtering

This study uses professional research data from 32 higher vocational colleges in 20 provinces and cities. The data 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. Among them, the target industries of school majors belong to, including the first, second and third industries such as service industry, manufacturing, software and information technology, finance and construction [4][5].

ISSN 2522-6398 Vol. 8, Issue 10: 219-223, DOI: 10.25236/FER.2025.081032

2.2 Research Method

The Gaussian Mixture Model (GMM) uses multiple Gaussian distribution mixing as a parameter model and solves the algorithm through expected maximization, so that the data is characterized by multi-Gaussian mixing distribution. GMM can be regarded as a generalization of K-means clustering algorithm. Compared with K-means, GMM can not only describe the position relationship of data clusters, but also better describe the distribution shape of data clusters [6].

2.3 Algorithm Design

The Gaussian Mixture Model (GMM) uses multiple Gaussian distribution mixing as a parameter model and solves the algorithm through expected maximization, so that the data is characterized by multi-Gaussian mixing distribution. GMM can be regarded as a generalization of K-means clustering algorithm. Compared with K-means, GMM can not only describe the position relationship of data clusters, but also better describe the distribution shape of data clusters [7].

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

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

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

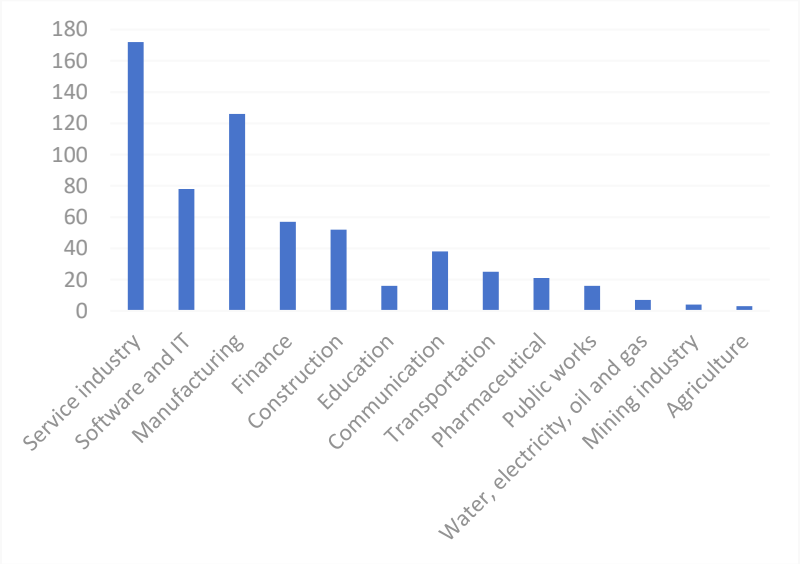

Fig.1 Distribution of the number of majors in different industries.

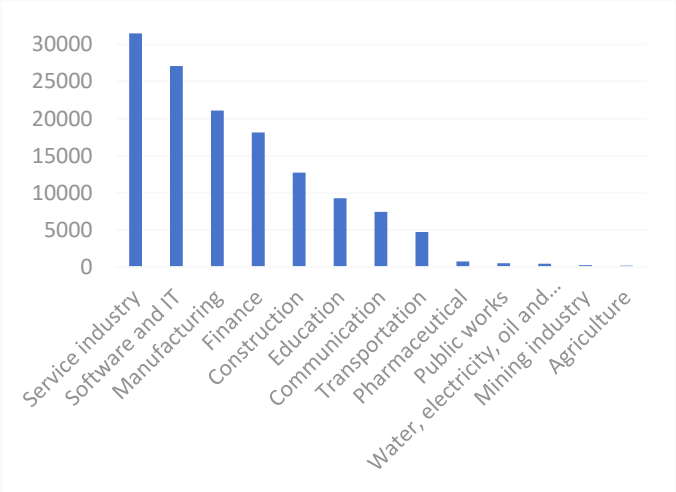

Fig.2 Number of students in different industries.

ISSN 2522-6398 Vol. 8, Issue 10: 219-223, DOI: 10.25236/FER.2025.081032

2.4 Entropy Law

Entropy weight method is a method that uses information entropy theory to calculate multidimensional data and comprehensively evaluate data weights. In information theory, entropy is a measure of the disorder of the system. The smaller the information entropy of a set of data, the greater the dispersion of the data, the greater the information content. When comprehensively weighting evaluation with other indicators, the corresponding weight will be greater. 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 [8].

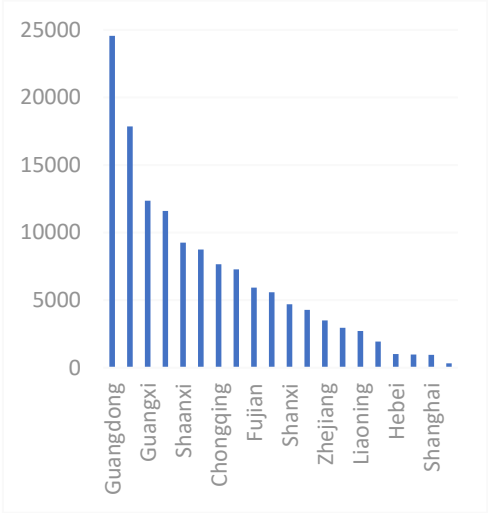

Fig.3 Number of students in different provinces

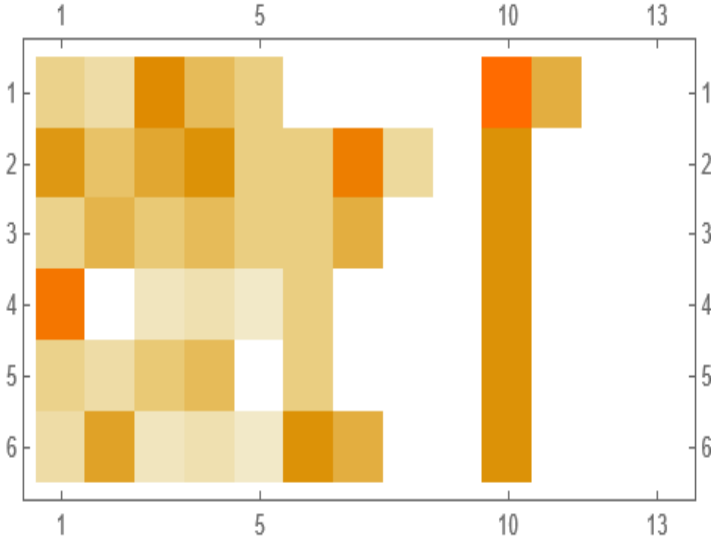

Fig.4 Category 1: {Fujian, Hainan, Henan, Sichuan, Xinjiang, Chongqing}.

2.5 Number of industry layouts

Count the number of all majors in each industry, and the specific number is sorted as Figure 1 as shown. It can be seen that the top five professional settings of higher vocational colleges are service industry, manufacturing, software and information technology, finance and construction.

Count the number of all students in each industry, and the specific number is as Figure 2 as shown. It can be seen that the top five higher vocational colleges are the service industry, software and information technology, manufacturing, finance and construction, which are basically the same as the number of professional settings. This shows that the majors of service industry, software and information technology, manufacturing, finance and construction occupy an important position in higher vocational colleges, and these industries are also in greater demand for high-quality technical skills [9] [10].

ISSN 2522-6398 Vol. 8, Issue 10: 219-223, DOI: 10.25236/FER.2025.081032

Count the number of all students in each province, and the specific number is as Figure 3. It can be seen that the top four provinces of higher vocational colleges are Guangdong, Shandong, Guangxi and Jiangxi.

3. The layout of professional types in different provinces

Summarize the professional industry layout of different provinces and get data:

$$Xi, j, 1 \leq i \leq 20, 1 \leq j \leq 13 \tag{2}$$

Among them, the data is summarized by various provinces and listed as 13 industry specialties, namely, service industry, manufacturing, software and information technology, financial construction, communications, transportation, pharmaceutical manufacturing, public works, education, hydropower, oil and gas, mining industry, and agriculture. The following is a clustering algorithm to analyze the structural differences between provinces [11].

First of all, standardize the data and make

$$x_{\{i,j\}} = \frac{x_{i,j}}{max\ x_{\{i,j\}1 \leq i \leq 20}} \tag{3}$$

The Gaussian Mixture Model (GMM) algorithm is used to cluster the data. The distance function is Bray-Curtis distance to obtain 5 clusters. The results are as follows Figure 4.

Class 1 and Class 3 are similar in professional layout. They are professionally distributed in service industry, manufacturing, software and information technology, financial construction, communications, transportation, pharmaceutical manufacturing and education. The number of majors in cluster 1 is relatively small, while the number of majors in cluster 3 is relatively large. Compared with other clusters, cluster 2 has a differentiated layout in the fields of public works and water, electricity, oil and gas, cluster 4 has a distinctive layout in pharmaceutical manufacturing, and cluster 5 provinces mainly focuses on the top 7 industries. Compared with the distribution of students in school, cluster 3 is the province with the top 4 in the number of students, while cluster 1 is mainly the province with 5 to 10 students. Among the top ten provinces with the number of students, only Shanxi does not have a professional layout in the education industry [12].

4. Conclusion

Achieving coordinated adjustments, optimization, and meeting regional industrial development needs are imperative for building a modern vocational education system. This is also a vital method to promote talent training that aligns with economic development demands. The article, grounded in vocational education professional settings and talent training, employs data analysis and modeling technology to conduct mathematical clustering and statistical analysis. It explores the realization path of coordinated professional settings in vocational education and regional industrial development, further advancing the high-level integration of industry and education within vocational education. The undergraduate professional data analysis has been discussed in the aboving sections. From the analyse indicate 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.

Acknowledgment

Project approved by the Municipal Education Commission - Sino German Vocational Education International Industry Education Integration Promotion Action (Project Number 2024H068-KQ).

References

[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[C]. 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.

ISSN 2522-6398 Vol. 8, Issue 10: 219-223, DOI: 10.25236/FER.2025.081032

[3] Song Huizuan, Wang Jiangnan. Adjustment and development of vocational college majors based on industrial development needs [J]. Knowledge Window (Teacher Edition), 2024, (10): 5-58.
[4] Jiangyi Lv, Cherry Jiang , Xiaolong Ren; Construction and Analysis of Higher Vocational College Teaching Evaluation System Resources Based on Data Mining Function[C]. 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.
[5] Wang Lina. on the adaptability of major settings in vocational colleges in the Guangdong-Hong Kong-Macao Greater Bay Area [J]. Journal of Hebei Energy Vocational Technical College, 2024, 24(03): 5-9 .
[6] Yuan Xiaohua, Zhang Miao. Upgrading and Digital Transformation of Vocational Education Majors: Elements and Optimized Strategies[J]. China Vocational and Technical Education, 2024, (20): 80-87.
[7] 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.
[8] Han Yu, Xu Han, Zhao Chu. Digitalization of vocational education professional ability standards: ability map and application scenario development[J]. Educational Research and Experiment, 2024, (05): 99-107.
[9] 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.
[10] Deng Zhenhua. Setting of Undergraduate-Levelational Education Majors: Current Situation, Reflection, and Suggestions - Analysis Based on Data from 2019 to 2023[J]. Journal Tianjin Zhongde University of Applied Technology, 2024, (03): 21-30. DOI: 1016350/j.cnki.cn12-1442/g4.2024.03.004.
[11] 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.
[12] Zhou Na. Study on the Integration of Major Settings in Higher Vocational with Regional Economic Development - A Case Study of Guangdong Province[J]. Journal of Taiyuan City Vocational and Technical College, 2024,03): 5-9. DOI: 10.16227/j.cnki.tycs.2024.015