International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 243-257

https://doi.org/10.70517/ijhsa46317

Cluster Analysis-Based Student Ability Portrait and Accurate Cultivation Path in Internet International Trade Competition in Higher Vocational Colleges and Universities

Jianing Du[1,*] and Meizhen Zhang[1]
[1] School of Digital Economy & Trade, Wenzhou Polytechnic, Wenzhou, Zhejiang, 325035, China
Corresponding authors: (e-mail: du_jianing@163.com).

Abstract The development of vocational education cannot be separated from skill competitions, and skill competitions in higher vocational colleges and universities are an important platform for testing the improvement of students' skills. In this paper, we take the students of S higher vocational colleges in Guangdong Province, China, who are in the Internet international trade competition for higher vocational colleges as the research sample, and choose descriptive statistical analysis and correlation analysis as the data statistical methods of this research. From the perspective of learner portrait, we constructed a student ability portrait model based on the objective function of clustering algorithm FCM, calculated fuzzy division coefficients through the number of classifications and attribution probability matrix, determined the appropriate number of student clusters, and obtained the classification results of the portrait. In the analysis of student ability portrait, the optimal number of clusters is determined to be 4, and four groups of learners, namely, active catch-up, potential constructor, active collaborator and passive receiver, are classified, and at the same time, based on the group characteristics presented by different learning ability portraits, the precise cultivation paths are proposed for different groups of learners in preparation for the process of Internet International Trade Competition.

Index Terms student ability portrait, FCM algorithm, correlation analysis, international trade competition

I. Introduction

Higher vocational colleges and universities specializing in international trade mainly cultivate applied talents who are adapted to the needs of the development of the market economy and engaged in international trade. With the rapid development of the digital economy, digital trade and cross-border e-commerce business is rapidly expanding, which puts forward higher requirements for international trade talents, mainly in the form of theoretical + practical + scientific and technological composite talents, and emphasizes the practice and theory in parallel [1], [2]. In higher vocational colleges and universities, international trade to literacy, knowledge, skills as the guiding ideology of talent training, “post, certificate, class” trinity of talent training model, focusing on knowledge, skills, literacy as the focus of professional training, emphasizing the entire teaching process to cultivate students' excellent professionalism, solid professional foundation, excellent Modern business service skills. Traditional international trade personnel training is often carried out in a theoretical or practical manner, the former “heavy theory light practice”, the latter “heavy practice light theory”, which is a low match with the concept of international trade personnel training in higher vocational colleges and universities and social demand [3], [4].

Nowadays, all kinds of skills competitions have become an important part of the test of students' practical skills and comprehensive ability, such as the Internet + international trade competition, which enhances the students' understanding of the operation of various positions in international trade, promotes the students' ability to establish cooperative relationships, and improves the students' job skills and vocational literacy [5]-[7]. In addition, today's educational requirements pay attention to the personalized and precise cultivation of students, and through the union of various data analysis techniques, the establishment of students' ability portrait provides an effective path for cultivating precise talents [8]. Among them, cluster analysis, as a powerful data analysis technique, plays a key role in discovering hidden patterns in data sets and supporting decision making, and has been used in education for student group classification, student behavior analysis, teaching evaluation, and precise resource recommendation [9]-[12]. Based on this, clustering analysis is applied to the processing of student competition data to construct a student ability portrait, which provides a decision-making basis for the precision of international trade talent training and better meets the needs of enterprises in the foreign trade industry.

Taking students of S higher vocational colleges and universities in Guangdong Province, China, as a research sample, this study aims to explore the division of students' ability portraits and the corresponding precise cultivation paths in Internet international trade competitions. Data preprocessing is realized through data cleaning and data conversion, and three data statistical methods, namely descriptive statistical analysis, correlation analysis and cluster analysis, are proposed respectively to provide a methodological basis for the analysis of student ability portrait model and precise cultivation path analysis. Proposed. The objective function-based clustering algorithm FCM is used to construct a student ability portrait model, compare the size of the fuzzy division coefficients of different numbers of classifications, and determine whether the number of clusters is appropriate. The fuzzy division coefficients are normalized to determine that there is a unique maximum value in the interval of the number of clusters, and the number of clusters corresponding to the maximum value is obtained to output the classification results of the student ability portrait. Using the student ability portrait model proposed in this paper, the student ability portraits of S higher vocational colleges and universities under the Internet international trade competition are analyzed, and student clusters are classified and named based on the portraits. According to the group characteristics presented in the student ability portrait, the precise cultivation paths for different student groups in the process of Internet international trade competition are proposed.

II. Research data collection

II. A. Data collection

Data is not only the basis for profile analysis, but also the prerequisite for drawing learner profiles. In daily learning activities, students' learning behaviors generate a large amount of data, which are accurately recorded in the background database. Based on the research needs, three main types of key data were collected and integrated, as shown in Table 1. The first type of data is the students' basic information, the second type of data is the students' learning behaviors, and the third type of data is the students' total academic performance.

Table 1: Data categories and data indicators

Data category	Data indicators
Basic information	Name
	The region
	Professional
Online learning behavior	Number of participation activities
	Watch the duration of the course
	Number of courses participated
	Number of submissions of learning reports
	Number of submissions of academic summaries
Learning results	Academic Achievements

II. B. Data pre-processing

The research sample of this study was students in S higher vocational colleges in Guangdong Province, China, and a total of 128 valid data were collected. These data not only summarized the students' personal information and learning behavior records, but also included the total academic performance. In order to ensure the accuracy and ease of use of the data, all the collected data were organized and stored in an Excel sheet to facilitate the subsequent data analysis and research work.

At this stage, the first and foremost aspect was the pre-processing of the collected data to address issues such as data clutter, inconsistent formatting, and omissions in order to ensure data accuracy and usability. Common data preprocessing methods include data integration, data cleansing, data normalization, and data transformation, which help to improve the overall quality and usability of data.

1) Data Cleaning

Due to systematic reasons, certain fields may be missing when collecting or transmitting data, or data with certain parameters may be missing. In order to ensure the integrity of the data, data cleansing is performed on the basic information table and the training teacher learning behavior data table.

2) Data Conversion

In order to avoid errors triggered by different data units and to standardize the format of behavioral data, standardization of different data types is executed here. This involves a linear transformation of the original data to map the results to the range of $[0,1]$, i.e., the online learning behavior-related data becomes values between $[0,1]$, and the linear transformation strategy adopted here is to make use of the maximum-minimum normalization method, and the following are the relevant detailed formulas:

$$x'_{(normalized)} = \frac{x - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where x denotes the learning data to be processed, and $X_{\min}$ describes the minimum value of the sample data and $X_{\max}$ describes the maximum value of the sample data, which can be retrieved and accessed in an orderly manner. This method is applicable given that the online learning data table has been successfully acquired and standardized, and no new data will be added to the sample, nor will it result in a change of $X_{\max}$ and $X_{\min}$ in the sample.

Finally, after data preprocessing, 126 valid data were obtained, which can be used for subsequent studies.

II. C.Statistical methods

1) Descriptive statistical analysis

Descriptive statistical analysis was conducted on the basic information and online learning behavior data of the study participants. In the analysis, the basic information variables involved included the teachers' regions, subjects taught, and school divisions. In order to analyze these data in depth, SPSS statistical analysis software was used, and the function of "Analysis - Descriptive Statistics - Frequency" was used to describe the percentile indicators of these variables. Through this analysis, a more comprehensive understanding of the characteristics and behaviors of the research subjects can be achieved, providing effective data support for subsequent research. Through the collection of online learning behavior data, including participation in meetings, participation in activities, watching the length of the course, teaching design and project summary and other information, and in the SPSS software "analysis - descriptive statistics - descriptive" operation to analyze the data, to get the maximum value, minimum value, the average value, the standard deviation, the variance and other statistical results, in order to comprehensive understanding of the distribution and characteristics of learning behavior data.

2) Correlation Analysis

Correlation coefficient is an important indicator to measure the degree of correlation between attributes, which helps to filter the variables that need to be explored in depth, and at the same time, exclude the attributes that are not correlated or have multiple covariance, so as to ensure that only the valid attributes are retained for analysis. For continuous measurement data, Pearson coefficient is more applicable [13]. Therefore, in this study, the Pearson coefficient was chosen to be used for correlation test and the formula is as follows:

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \tag{2}$$

In the above correlation coefficient r takes the value of $[-1,1]$ (1 is a perfect positive correlation, -1 is a perfect negative correlation), when r is greater than 0 it indicates a positive correlation, and when it is less than 0 it indicates a negative correlation. If the absolute value of the correlation coefficient $|r|$ is between 0.3 and 0.5, it indicates a low correlation between the variables; if it is between 0.5 and 0.8, it indicates a medium correlation; if it is between 0.8 and 1, it indicates a high correlation between the variables.

III. A model of student ability portrait based on cluster analysis

As an important branch of analysis, learner portrait is committed to describing the whole picture of learner characteristics from static, dynamic and result dimensions, which can play an important role in learning assessment and teaching decision-making [14]. In this paper, we will apply the principle of learner profile to the training of students in the Internet international trade competition in higher vocational colleges and universities, and construct a student ability profile.

III. A. Learner Profiling Process

Based on learning analytics related theories and existing research, this study proposes a learner portrait construction process for students in Internet international trade competitions in higher vocational colleges and universities, as shown in Figure 1. The learner portrait process includes four steps: portrait construction goal, data collection and pre-processing, portrait construction, portrait visualization, and the application service of the portrait is precise teaching intervention.

The goal of image construction can guide the division of elements and the input and output of the image, guide the whole process of image construction, and ensure the significance of the image. The goal of image construction in this study is to achieve the first goal of the research part, that is, to realize the objective and comprehensive evaluation of the learner through the image, and then provide the basis for realizing the precision and personalization of teaching. After determining the goal, we collect the data generated in the learning process according to the portrait model, and then construct the portrait after data screening, cleaning and normalization. The construction process involves data analysis, learner portrait modeling and portrait labeling, extracting meaningful labels, outputting the portrait results corresponding to the goal, and presenting them in a visualized way. The overall portrait can reflect the overall level of learners, which helps teachers to grasp the overall learning; the group portrait constructed by category subgroups classifies learners differently, which ensures that the subsequent precise teaching interventions are more targeted and effective; and the individual portrait is the most precise, which can help teachers to analyze and predict the learning trend of individual learners, and support the self-regulation of individual learners. The visualization results of the portraits are finally applied to teaching and guiding the implementation of interventions. Due to the dynamic nature of the data, the portraits need to be continuously updated in the process of application.

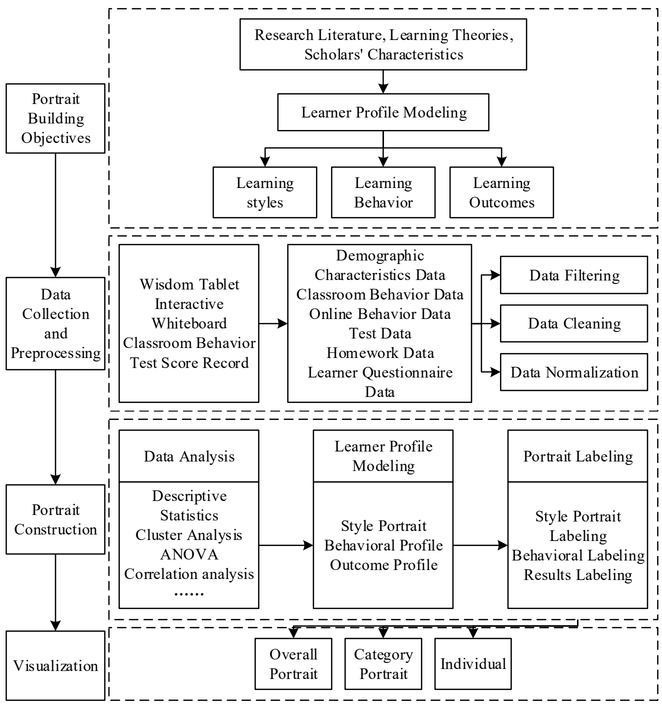


Figure 1: Learner portrait process

III. B. Student Ability Portrait Based on FCM Algorithm

III. B. 1) FCM clustering algorithm

Fuzzy C mean (FCM) is a clustering algorithm based on the objective function, the FCM algorithm is improved from the ordinary C mean algorithm, FCM is usually used to analyze the data for clustering [15]. The hard division of the ordinary C mean algorithm will strictly divide the original data into certain class clusters, and use 0 and 1 to indicate whether the data belongs to the class cluster, instead of using the value between 0 and 1 to indicate the degree of affiliation of the data belonging to the class cluster, so the results obtained by the ordinary C mean analysis will often be in error with the real situation. However, the FCM algorithm obtains a fuzzy delineation result, which is more in line with the relationship between data objects and classes in reality.

If the data sample set to be analyzed is X, the data sample needs to be divided into C categories, $C_j \in X, 1 \le j \le C, c_j$, $1 \le j \le C$, c_j is the center of C_j, and the i data sample x_i belongs to C_j membership is u_{ij}, The goal of the FCM clustering algorithm is to divide all data samples into C categories, so that each sample has the highest membership degree to the classification in which it is located. The membership degree of each data sample is multiplied by the Euclidean distance from the sample to the center point of the cluster, and then the sum is performed to obtain the objective function, where m is the parameter that controls the flexibility of the algorithm, that is, the mildness of the sample, which is generally set to 2:

$$J=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}\left\|x_j-c_i\right\|^2 \tag{3}$$

$$\sum_{i=1}^{c}u_{ij}=1, j=1,2,\cdots,n \tag{4}$$

FCM algorithm through the constraints then can ensure that the samples belong to different classes of affiliation and 1, so for the initial affiliation matrix, need to be normalized to meet the requirements. Algorithm iteration process is to find the attribution probability u_{ij} and the center of mass c_j, iteration termination is based on the need to set a termination value of e, so that the equation (4) holds, that is, the first k round and the first round of iteration of the $\mathrm{k}+1$ rounds of u_{ij} difference in the maximum value of the difference between e is less than e, its The geometric implications will be such that J_m minimizes or lies in the saddle shape:

$$\max\left\{\left|u_{ij}^{k+1}-u_{ij}^{k}\right|\right\}<e, 1\le i\le n, 1\le j\le C \tag{5}$$

The specific steps of the traditional FCM algorithm are

1) Determine the number of categories n for clustering, the flexibility parameter m of the algorithm, and the termination condition of the algorithm.

2) Perform the initialization of the affiliation matrix, and the affiliation matrix needs to satisfy the constraints.

3) According to the affiliation matrix U, carry out the calculation of the clustering centroid C, the calculation formula is shown in Equation (6), i.e., for a certain category j, calculate the ratio of the sum of the product of all the data nodes x_i and their probability of belonging to the category u_{ij} (weighted) to the sum of the probability of belonging to the category (weighted) again:

$$c_j=\frac{\sum_{i=1}^{n}u_{ij}^{m}\times x_i}{\sum_{i=1}^{n}u_{ij}^{m}} \tag{6}$$

4) Update the affiliation matrix by C, the update formula is shown in Eq. (7), which is the sum of the ratios of the similarity (weighted) of the data point x_i to the center of mass c_j and the similarity (weighted) of x_i to the centers of mass of other categories, then:

$$u_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{\left\|x_i-c_j\right\|}{\left\|x_i-c_k\right\|}\right)^{\frac{2}{m-1}}} \tag{7}$$

5) Carry out the judgment of iteration termination condition, if the condition is satisfied then terminate the calculation, if not then continue the iteration, continue from step (3).

6) Obtain the clustering results.

III. B. 2) FCM Cluster Number Determination Methods

One of the most difficult tasks in cluster analysis is to determine the appropriate number of clusters. In fuzzy clustering, the fuzzy partition coefficient can be calculated by using the number of classifications N and the attribution probability matrix, comparing the size of the fuzzy partition coefficient for different numbers of classifications, and then determining whether the number of clusters is appropriate. The formula for calculating the fuzzy division coefficient without normalization is shown in equation (8):

$$F(U)=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{2} \tag{8}$$

The normalization of $F(U)$ can be calculated by equation (8), after normalization, the value range of F_c is from 0 to 1. When F_c is 0, it represents that the data samples are completely fuzzy, and the classification effect according to the cluster center is poor under the current number of clusters; when F_c is 1, it represents that the classification effect is optimal according to the cluster center under the current number of clusters. $F_c(U)$ has a unique maximum value in a certain interval of determined number of clusters, and the number of clusters corresponding to the maximum value is the optimal number of clusters:

$$F_c(U)=\frac{F(U)-\left(\frac{1}{N}\right)}{1-\left(\frac{1}{N}\right)} \tag{9}$$

III. B. 3) FCM-based group portrait construction methods

The process of group portrait construction is to determine the optimal number of clusters for clustering based on the group portrait source data and $F_c(U)$ metrics, and then carry out the clustering division, and then finally get the classification result and the clustering center which is the result of the group portrait. The flowchart of group portrait construction based on FCM is shown in Fig. 2, which mainly includes the following seven steps.

1) Set the value range of the number of clusters according to the sample size of the device group portrait source data.

2) Calculate the value of the indicator in the case of different number of clusters according to the formula of $F_c(U)$.

3) Select the number of clusters with the largest value of $F_c(U)$ as the optimal number of clusters of the FCM algorithm in the previous step.

4) Determine the size of the affiliation matrix based on the optimal number of clusters and the number of samples, and generate the affiliation matrix by randomization.

5) Calculate the clustering center point c_j according to the affiliation matrix.

6) Update the affiliation matrix according to c_j, and determine whether the iteration termination condition has been satisfied, if not, return to step 5), otherwise output the clustering results and the clustering center location.

7) Obtain the device group portrait classification results.

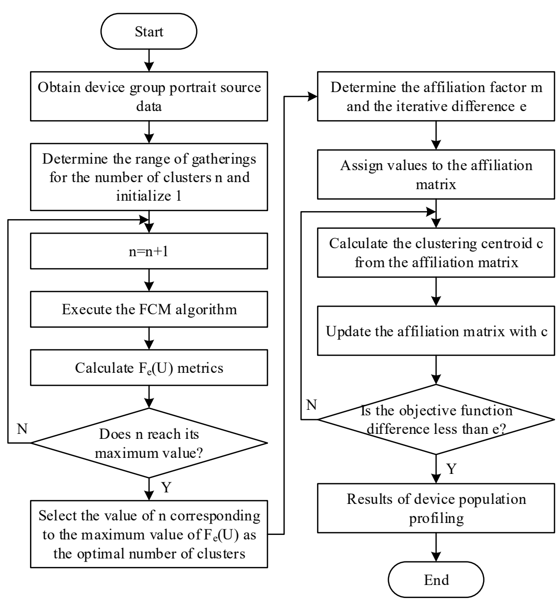


Figure 2: The construction flow chart of group portraits

IV. Analysis of the ability profile of students in international trade competitions on the Internet

In the context of the Internet International Trade Competition held in the Internet, this chapter will take the collected and preprocessed student data as the basis for analyzing the student ability portrait of S higher education institutions in conjunction with the student ability portrait model based on the FCM algorithm proposed in this paper.

IV. A. Portrait Analysis of Students' Ability Attributes

The performance of each attribute of learner competence was analyzed by descriptive statistics as shown in Table 2. The degrees of the four dimensions of interaction attribute, interest attribute, competence attribute and knowledge attribute have the same range of values, between 0 and 1. The mean values of learners' scores on ability attribute (M=0.37, SD=0.22), interaction attribute (M=0.48, SD=0.25), interest attribute (M=0.58, SD=0.21), and knowledge attribute (M=0.81, SD=0.19) showed a trend of increasing, which shows that the learners have a strong interest in learning the course, but their performance in online communication is not active enough. However, the performance in online communication is not active enough, and there is still much room for improvement in the ability attribute. In addition, the dispersion of learners' scores on the four attributes (SD<0.3) is not large, indicating that their performance on each attribute is stable.

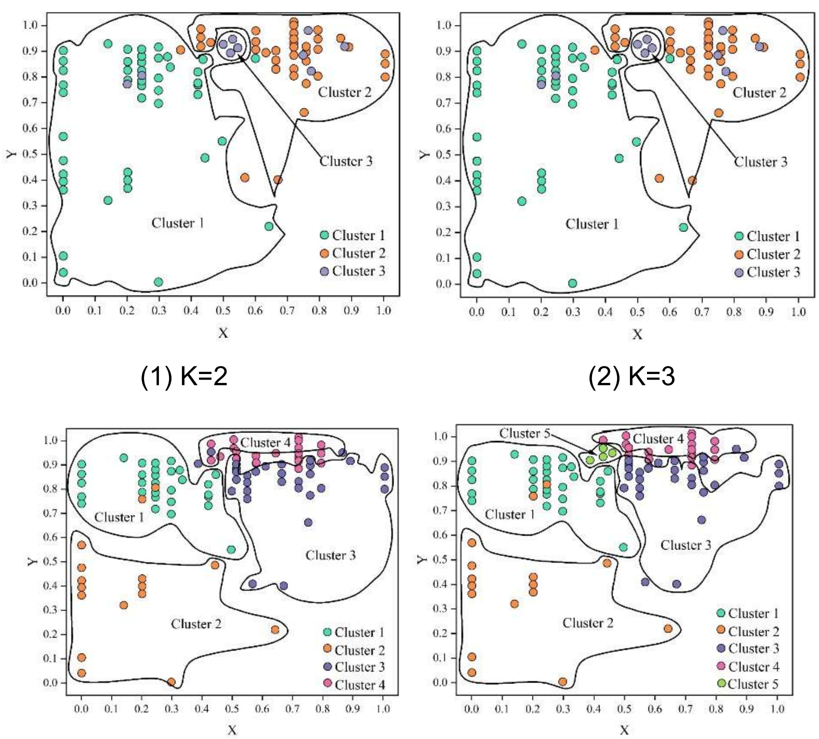
Table 2: Descriptive statistics of learners' attributes

Dimensions	Mean	SD
Interactive attributes	0.48	0.25
Interest attribute	0.58	0.21
Ability attribute	0.37	0.22
Knowledge property	0.81	0.19

IV. B. Clustering Portrait Analysis

IV. B. 1) Optimal selection of clustering centers

In this paper, four attributes of learners will be clustered, with the help of python to complete the update and acquisition of clustering points, respectively, take K = 2, 3, 4, 5 when the clustering scatter plot,specifically shown in Figure 3. Figures (a)~(d) correspond to the cases of K=2, 3, 4, 5 respectively. From the figure, it can be preliminarily judged that the clustering results are clearer when the K value is 2 or 4.

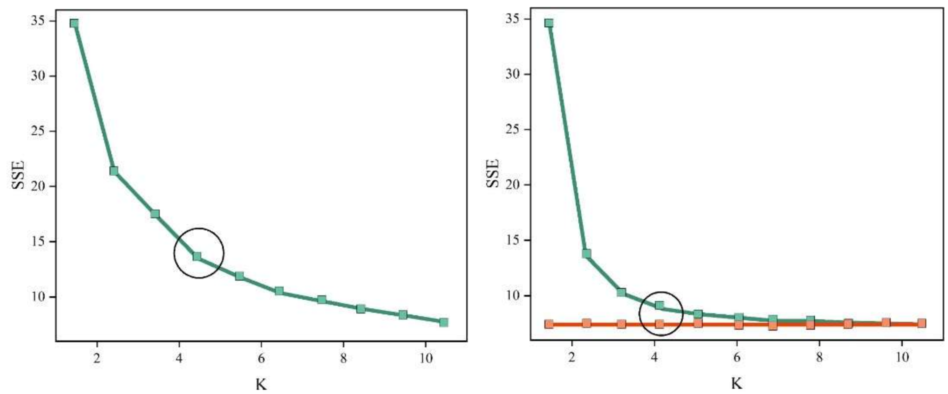


(1) K=2 (2) K=3

(3) K=4 (4) K=5

Figure 3: clustering scatter plot

The K-value and corresponding SSE value are calculated, and in order to consider the advantages and disadvantages of the K-value of the number of clusters more comprehensively, the intra-class distance and the inter-class distance are taken into account, as shown in Fig. 4, and the curve close to the X-axis on the right side indicates the trend of the inter-class distance with the rise of the number of clusters. Figures (a) and (b) show the K value and the corresponding SSE value and their calculation results after taking into account the intraclass distance and interclass distance, respectively. Obviously, the clustering effect is best when the K value is 4, so the optimal number of clusters should be 4 for this data set.



(a) Results of K value and the corresponding SSE value

(b) Results after taking into account intra-class distance and inter-class distance

Figure 4: K value and SSE value

IV. B. 2) Cluster characterization

K-means clustering (K=4) was performed on S higher education students and finally iterated 4 times until convergence. The test results of clustering results are shown in Table 3. The F-test found that the clustering results of the four attributes showed significance, which verified the validity of the results of this clustering, and the order of the importance of each attribute to the clustering results was as follows: knowledge attribute (F=168.778)>ability attribute (F=127.716)>interaction attribute (F=72.874)>interest attribute (F=93.802).

Table 3: Clustering results test results

ANOVA	Clustering		Error		F value / significance
	Mean square	Degree of freedom	Mean square	Degree of freedom	
Interactive attributes	2.157	3	0.024	181	93.802***
Interest attribute	1.544		0.022		72.874***
Ability attribute	2.153		0.018		127.716***
Knowledge property	1.508		0.008		168.778***
*** at the 0.001 level (double tail), the difference is significant.					

The headcount statistics for the four clusters are shown in Table 4. Cluster 2 and Cluster 3 accounted for the largest proportion of the number of people, 47.6% and 30.2%, respectively, while Cluster 1 (18, 14.3%) and Cluster 4 (10, 7.9%) had a relatively small number of learners, and an in-depth analysis of the characteristics of the various clusters will be carried out below.

Table 4: Statistics of the number of people in various clusters

Groups	Number of students	proportion
Group 1	18	14.3%
Group 2	60	47.6%
Group 3	38	30.2%
Group 4	10	7.9%

The statistics of the mean values of the attributes of each cluster are shown in Figure 5. Taxon 1 has higher scores in interest attribute and ability attribute, Taxon 2 has more stable performance in interest attribute, interaction attribute, ability attribute and knowledge attribute, Taxon 3 has relatively better performance in interaction attribute and knowledge attribute, while Taxon 4 has lower scores in four attributes of interaction, interest, ability, and knowledge, and it is a more special type of group. Therefore, taxon 1 can represent the group of learners with strong interest and outstanding ability; taxon 2 is used to represent the group of learners with learning potential, obvious interest in learning, and good mastery of knowledge and skills; taxon 3 represents the group of learners who are good at communicating and active in their efforts; and taxon 4 represents the group of learners who are not good at interaction, interest, ability, and knowledge. Based on the attribute characteristics of the above four clusters, this paper identifies and names the four types of learners, and classifies Cluster 1 as active catch-up, Cluster 2 as potential constructor, Cluster 3 as active collaborator, and Cluster 4 as passive receiver.

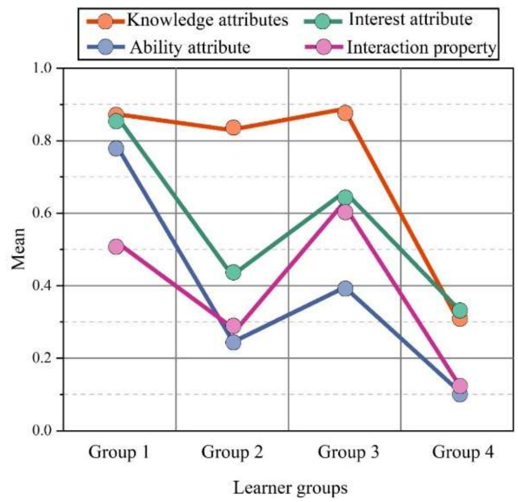


Figure 5: Statistics of attribute mean score of each group

IV. B. 3) Cluster depth analysis

In order to explore in depth the key elements in each category that affect the overall level of learners, this paper uses linear regression analysis to analyze each of the four clusters in depth. In SPSS, the command "Analysis-Regression-Linear" is used to establish a linear regression equation model with interaction attributes, interest attributes, ability attributes, and knowledge attributes as independent variables, and the overall level as the dependent variables, R is used to represent the degree of linear correlation between the two variables, R^2 and the adjusted R^2 can represent the degree of interpretation of the dependent variable by the independent variable, and are used to judge the advantages and disadvantages of the linear regression model. 4. The data modeling of active catch-ups, potential constructors, active collaborators, and passive recipients respectively is dedicated. The model construction of regression equations for each type of group is specifically shown in Table 5. After F-test, the construction effect of model 1, 2 and 3 all show significance (p<0.001), model 3 has the best construction effect, which is able to explain about 71.8% of the variance variance of the overall level through four attributes, model 4 and model 1 have a general construction effect, and model 2 is only able to explain about 28.8% of the variance of the overall level, and the specifics of each type of group will be analyzed in depth in the following.

Table 5: Construction of regression equation model of various groups

Model	R	R^2	Adjusted R^2	F value / significance	Independent variable	Dependent variable
1	0.677	0.462	0.433	16.518***	Interaction attribute,interest attribute,ability attribute,knowledge attribute	Overall level
2	0.585	0.345	0.288	6.372***		
3	0.872	0.723	0.718	54.475***		
4	0.683	0.451	0.265	2.153/0.138		
*** at the 0.001 level (double tail), the difference is significant.						

1) Depth Analysis of Active Catch-Up Classes

The correlation of each attribute of the active catchers is specifically shown in Figure 6. A1~A5 in the figure represent interaction attribute, interest attribute, knowledge attribute, ability attribute, and overall level, respectively. From the correlation between each attribute, there is a strong linear correlation between the interest attribute and the ability attribute, which indicates that when active catch-upers improve their interest in learning, their learning ability is also likely to be improved, which in turn improves the overall level of performance, which shows that the improvement of the ability is very important for active catch-upers to maintain high-quality learning.

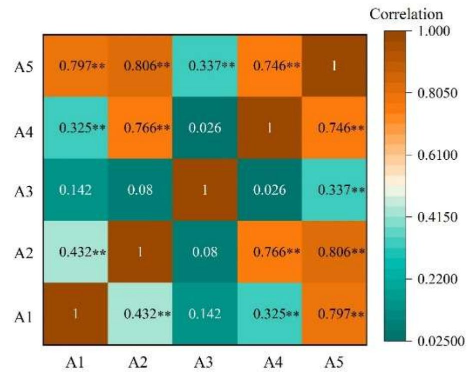


Figure 6: Active catcher

2) Depth analysis of potential constructor taxa

The correlation matrix between the attributes of potential constructors was plotted, as shown in Figure 7. Only four pairs of learning attributes showed significant correlations, with the strongest linear correlation between the ability attribute and the interest attribute, and the weakest correlation between the interaction attribute and the overall level, suggesting that the performance of potential constructors in terms of engaging in the basic tasks of the classroom and in engaging in creative activities are influenced by each other. It was further found that the interaction attribute showed different degrees of negative correlation with each of the three attributes, namely the knowledge attribute, the ability attribute, and the interest attribute, suggesting that when potential constructors generate too much online interaction behavior, it may also be detrimental to their attainment of a better overall level, which deserves to be brought to the attention of the relevant pedagogical stakeholders.

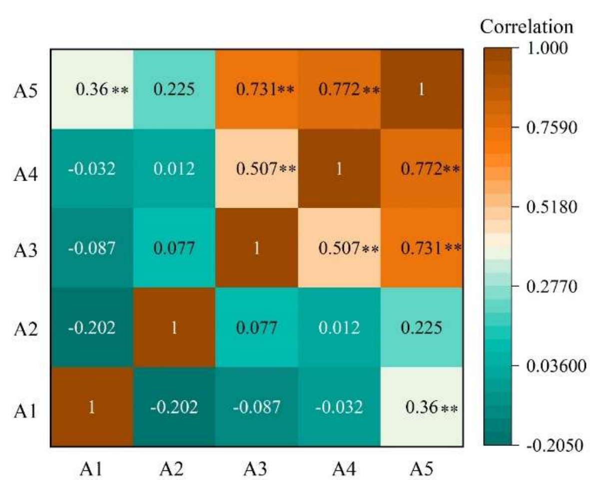


Figure 7: Potential constructors

3) Depth analysis of active collaborator taxa

The correlation performance of each attribute of the active collaborator taxon is specifically shown in Figure 8. It can be seen that the three pairs of attributes, namely, knowledge attribute and ability attribute, knowledge attribute and interest attribute, ability attribute and interest attribute, show significant correlation, and the correlation between knowledge attribute and interest attribute is the highest, which can be hypothesized that the more active active the active collaborators are in terms of learning interest, the higher their scores on knowledge attribute are likely to be, so that the active collaborators should gradually improve their online learning interest, Therefore, active collaborators should gradually increase their interest in online learning, make full use of the learning resources, and better master the course knowledge and achieve the learning objectives.

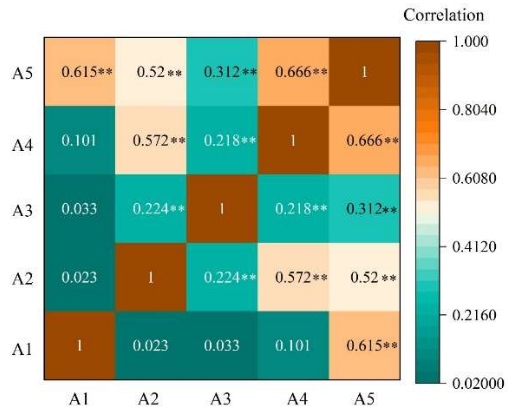


Figure 8: Active collaborators

4) Depth analysis of passive recipient taxa

The correlation of each attribute of passive receivers is specifically shown in Figure 9. In terms of attribute correlation, there is a strong correlation between the knowledge attribute and the ability attribute of the passive recipients, but they show a low linear negative correlation between the interaction attribute and the interest attribute, which indicates that they produce a lot of unconscious and meaningless interaction behaviors under the plague of passive learning, and their interest in learning may not be high.

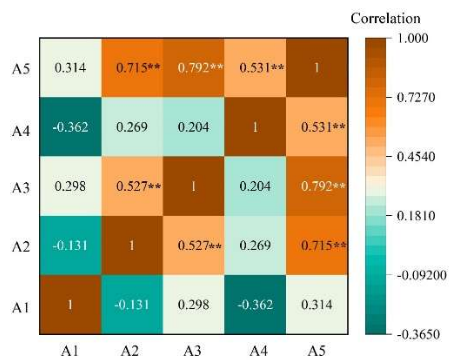


Figure 9: Passive recipients

V. Analysis of precise training paths based on students' ability portraits

This chapter will analyze the reasons for the group's learning results based on the group portraits of active catch-up, potential constructors, active collaborators, and passive receivers output from the previous paper, and give the precise cultivation paths in the process of preparing for the Internet International Trade Competition based on the group characteristics presented in the portraits.

V. A. Active Catch-Ups

The group of active catch-up learners belongs to the group with the best performance among the four groups of learners. In order to analyze the main influencing factors of the performance of the group of active catch-up learners, the correlation between the group's total performance and 11 indicators such as online participation and computer use hours was analyzed, and the correlation of their performance influence is shown in Table 6. It can be seen that the group of active catching up does not show obvious correlation between online participation, computer use time and other online learning data indicators, and the online learning behavior is not active. To address this situation, the corresponding precise cultivation path is proposed:

1) Teachers need to focus on the online international trade course-related learning activities of this group of students, prompting students to combine online learning and offline learning methods to further improve their learning results.

2) Teachers can pay attention to the offline learning mode of this group of students, play a positive role in leading the learning, prompting this group of students to promote their own offline learning mode and tap their potential for online learning, which will lead to the progress of other types of learner groups.

Table 6: Main influencing factors of the performance of active catch-up groups

Influencing factors	Pearson correlation coefficient
Online participation	-0.301
Computer use duration	-0.43
Reading duration	0.173
Number of classroom interactions	-0.327
Number of test questions completed	-0.095
Correct rate	0.039
The proportion of reading	0.446
Total test questions	-0.099
Total objective questions	-0.107
Teacher evaluation	-0.216
R^2 ranking	0.146

V. B. Potential constructors

The Potential Constructors group belongs to the General Adaptation group among the four groups of learners, which has better learning behaviors but worse performance. The correlation between the performance of the potential constructor group is shown in Table 7, which shows that the number of completed test questions and the total number of test questions of the potential constructor group are positively correlated with their performance. Based on this situation, the corresponding proposed precise cultivation path is:

1) Increase the amount of practice of the problems. But not a mechanical sea of questions, teachers can guide students to understand the knowledge points of international trade competition, sharpening the knife is not wrong.

2) Guide students to establish the international trade competition wrong question bank, regular review of the wrong question bank can be more solid grasp of knowledge. By the end of the semester, teachers can list all the knowledge points of the course, so that students can assess whether they have mastered all the knowledge points. This will help students to identify their blind spots and improve the effectiveness of teaching.

Table 7: Main influencing factors of the performance of Potential constructor group

Influencing factors	Pearson correlation coefficient
Online participation	-0.316
Computer use duration	-0.088
Reading duration	0.045
Number of classroom interactions	-0.412
Number of test questions completed	0.763*
Correct rate	-0.256
The proportion of reading	-0.075
Total test questions	0.386*
Total objective questions	-0.156
Teacher evaluation	-0.006
R2 ranking	-0.143

V. C. Active collaborators

The group of active collaborators is one of the better performers in terms of academic performance, which is not outstanding, but has a positive and active attitude towards learning. The correlation of the factors affecting performance is shown in Table 8. There is a significant negative correlation between the number of classroom interactions of students in the active collaborator group and their grades, and students' performance is related to the number of interactions they have in class. Based on this correlation, the proposed precise cultivation paths are:

(1) Teachers need to pay attention to the actual situation of online participation of this group of students, and observe the students' concentration in the process of participation, in order to prevent ineffective time piling up learning.

(2) Teachers need to pay attention to the quality of practice and practice time in and out of class in international trade-related courses to determine the percentage of effective learning time, and improve the utilization of students' online time through the improvement of teaching methods.

(3) Help students to find more appropriate learning methods, improve the current inefficient learning mode, promote the improvement of learning results, and better prepare for the Internet international trade competition.

Table 8: Main influencing factors of the performance of active collaborators group

Influencing factors	Pearson correlation coefficient
Online participation	0.084
Computer use duration	0.178
Reading duration	0.092
Number of classroom interactions	-0.748*
Number of test questions completed	-0.674
Correct rate	-0.112
The proportion of reading	-0.005
Total test questions	-0.06
Total objective questions	-0.663
Teacher evaluation	-0.674
R^2 ranking	-0.528

V. D. Passive recipients

Students in the passive recipient group are among the more disadvantaged of the four learner groups and have great potential for improvement in learning outcomes. The correlation of the impact on the performance of the passive recipient group is shown in Table 9. From the data, there is a significant positive correlation between the correct rate of the online practice questions and the performance of the “potential” learners, and the performance of the students is related to the correct rate of the online practice questions. The proposed precise cultivation paths for the passive receiver group are as follows:

1) Teachers need to guide students to actively participate in classroom activities by setting up the teaching environment and changing the teaching methods, actively cooperate with teachers to participate in teaching interactions, and enhance students' enthusiasm to participate in the Internet international trade competition.

2) Teachers need to set up and check the students' practice in class and after class in a timely manner, and can set up a certain proportion and number of subjective questions for students to urge them to consolidate what they have learned in class through the topic of practice.

3) Pay attention to the quality of students' practice, improve the correct rate of students' test questions for the purpose of reasonable arrangement of the test content, and urge students to revise and review the wrong questions in a timely manner.

Table 9: Main influencing factors of the performance of the passive recipient group

Influencing factors	Pearson correlation coefficient
Online participation	-0.318
Computer use duration	-0.014
Reading duration	-0.482
Number of classroom interactions	-0.183
Number of test questions completed	-0.452
Correct rate	0.763*
The proportion of reading	-0.415
Total test questions	-0.453
Total objective questions	-0.412
Teacher evaluation	-0.01
R^2 ranking	-0.176

VI. Conclusion

In this paper, we take the students of higher vocational colleges and universities in Guangdong Province, China, as the research samples, take the students of Internet international trade competitions as the main subjects of this research, establish a student ability portrait model based on cluster analysis, classify the student groups, and based on the characteristics of the clusters, put forward the precise cultivation paths of the students oriented to the Internet international trade competitions.

In the analysis of students' ability portraits, the optimal number of clusters is determined to be 4 by calculating the K value and the corresponding SSE value, and taking into account the intra-class distance and inter-class distance, and the attributes of students' ability portraits are ranked in order of importance, and the results are as follows: Knowledge attribute (F=168.778)>Ability attribute (F=127.716)>Interaction attribute (F=72.874)>Interest attribute (F=93.802)>Attribute of interest (F=93.802). 93.802). Based on the attribute characteristics of different groups, four groups were divided and named as active catch-up, potential constructor, active collaborator, and passive receiver respectively.

On the basis of the portraits of the four groups, the correlation method was used to explore the factors affecting the performance of the different groups, and the corresponding precise cultivation paths were proposed. The group of active catch-up learners is not highly active in online learning behaviors, and the online learning data indicators such as online participation and computer usage hours do not show obvious correlations. We should focus on their online learning activities, combine online and offline learning, and deeply explore students' online learning potential. The potential constructor group shows a positive correlation between the number of completed test questions and the total number of test questions and the performance, so it should increase the amount of practice of exercises and establish a database of wrong questions in international trade competitions to consolidate students' knowledge. The active collaborator group shows a significant negative correlation between the number of classroom interactions and performance, and needs to pay attention to the actual situation of their online participation, reduce the ineffective learning time, and improve the learning efficiency of students through the improvement of teaching methods and learning methods. As for the passive recipient group, there is a significant positive correlation between the correct rate of online exercises and performance, so we should guide students to actively integrate into the classroom through changes in the teaching environment and teaching methods, set up and check the students' exercises in and out of the classroom in a timely manner, and pay attention to the quality of the exercises, so as to enhance the students' motivation to learn.

Funding

Higher Vocational Education Research Project of Wenzhou Polytechnic: Reform and Practice of International Trade Skills Training in Higher Vocational Colleges under the "Promoting Teaching through Competition" Model—A Case Study of the "Internet + International Trade Comprehensive Skills Competition" (WZYGJzd202103).

About the Authors

Jianing Du was born in Luoyang, Henan. China, in 1991. She obtained a master's degree from Kyoto University in Japan. She is currently serving as a lecturer at the School of Digital Economy & Trade, Wenzhou Polytechnic. Her main research direction is international trade and investment.

Meizhen Zhang was born in Xian, Shaanxi. China, in 1995. She obtained a master's degree from Northwestern Polytechnical University in China. She is currently serving as a lecturer at the School of Digital Economy & Trade, Wenzhou Polytechnic. Her main research direction is management science.

References

[1] Zhang, P. (2021). Research on practical teaching reform of international economy and trade major in higher vocational colleges under the background of cross border e-commerce. Int. J. Soc. Sci. Univ, 4(1).

[2] Pan, Y. (2023). Research on the Versatile Talent Training Mode of International Trade in Vocational University. The Educational Review, USA, 7(7).

[3] Qian, P. (2020). New thinking on the teaching mode of international economy and trade under the demand of innovative international talents. Int. J. Comput. Eng, 5(3).

[4] Zhang, R., Liu, H., & Zhang, J. (2017). A brief study on the problems and development countermeasures of China's international trade talents training. Modern Economy, 8(12), 1575-1579.

[5] Jun, Y., Xiao-Hui, S., & Yan, G. (2021). The "Competition" for Training Cross Border E-commerce Talents in China on the Background of "Internet Plus". Journal of Higher Education Theory & Practice, 21(6).

[6] Hu, Q. (2020, July). The Vocational Skills Competition Based on Big Data Analysis Promotes the Research of Students' Vocational Ability. In 2020 International Conference on Communications, Information System and Computer Engineering (CISCE) (pp. 279-282). IEEE.

[7] Mi, J. (2022). Analysis on the Design of Higher Vocational International Trade Specialty Course System Based on "Post, Class, Contest, Certificate" Integration. The Educational Review, USA, 6(6).

[8] Bo, Y., Wang, C., Li, L., & Lu, X. (2022, January). Student ability portrait construction research based on Big Data. In 2022 International Conference on Big Data, Information and Computer Network (BDICN) (pp. 156-160). IEEE.

[9] Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering analysis for classifying student academic performance in higher education. Applied Sciences, 12(19), 9467.

[10] Battaglia, O. R., Di Paola, B., & Fazio, C. (2016). A new approach to investigate students' behavior by using cluster analysis as an unsupervised methodology in the field of education. Applied Mathematics, 7(15), 1649.

[11] Gu, P., & Zheng, Y. (2018, August). Cluster analysis on the teaching evaluation data from college students. In 2018 13th International Conference on Computer Science & Education (ICCSE) (pp. 1-6). IEEE.

[12] Zhao, Y., Zhao, C., Wang, Z., & Min, Z. (2024). Cluster analysis and its application in teaching resources of university curriculum: A personalized method. Computer Applications in Engineering Education, 32(2), e22696.

[13] Ji Ting Mo,Fang Liu & Xin Xing Deng. (2025). A group decision making model with non-reciprocal fuzzy preference relations based on Pearson correlation coefficient. Expert Systems With Applications,276,127095-127095.

[14] Zhang Zongbiao. (2022). Construction and Empirical Study of Learner Portrait in Online General Education Course. Discrete Dynamics in Nature and Society,2022.

[15] Mu Liang,Xuechao Du,Mengya Lu,Shen Tian,Yingying Qiao,Yang Gao... & Chongxin Shan. (2025). Ppb-level photoacoustic sensor system for SF[formula omitted] decomposition component HF using fuzzy c-means algorithm based on baseline correction. Sensors and Actuators: B. Chemical,430,137319-137319.