International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-14

Article

A Practical Study on the Integration of Learner Profiling and Teachers' Digital Literacy in Foreign Language Classrooms Driven by Big Data

Lixia Jing[1,*]

[1] School of Foreign Languages, Shanghai Zhongqiao Vocational and Technical University, Shanghai, 201514, China

* Correspondence author: Nicexiaxia678@163.com

Abstract: The development of big data technology provides new opportunities for the development of education and gives new connotations and missions to precision teaching, and the integration of learner portraits and teachers' digital literacy practices in foreign language classrooms is now studied using college English classroom teaching as an example. Based on the learning data in the foreign language classroom, the study constructs individual portraits of learners in the foreign language classroom in terms of learner styles, learner behaviors and learner outcomes, and generates group portraits by using the K-means algorithm based on Canopy and maximum-minimum distance to output the learner portraits in the foreign language classroom. The results of the learning portraits are then embedded into the teaching process, realizing the accurate teaching practice that is organically integrated with teachers' digital literacy. The foreign language classroom presents four types of learners: "potential learners", "excellent learners", "borderline learners" and "striving learners", with their respective proportions being 46.00%, 19.80%, 13.40% and 20.80%. Teaching practice shows that the method not only improves students' learning ability and interest in learning, but also improves students' innovation and thinking ability.

Keywords: canopy algorithm; k-means algorithm; learner profiling; teacher quality literacy

1. Introduction

With the acceleration of globalization and frequent cross-cultural exchanges, the importance of foreign language education is becoming more and more prominent. Under the development and application of big data technology, foreign language education has also gradually explored and attempted to apply big data technology in learner profiling and personalized teaching, which has injected new vitality into the reform and innovation of foreign language classroom [1-2]. Learner portrait is driven by big data, through the description and analysis of learners' learning behaviors and learning process, and based on this, it analyzes and predicts learners' learning characteristics and learning behaviors, so as to help foreign language teachers optimize their teaching strategies, improve their teaching quality, and realize personalized teaching [3-6].

Digital literacy refers to the ability of teachers to master and utilize digital technologies, including general software, subject software, digital educational resources, intelligent education platform, intelligent analysis and evaluation tools, intelligent classrooms, etc [7-8]. Teachers' digital literacy plays a key role in the implementation of personalized teaching under learner profiling in foreign language classrooms. In the application of learner portraits, teachers' digital literacy can help them better understand students' learning and learning needs, and provide students with personalized learning content and learning styles through digital tools and resources [9-11]. For example, teachers can provide students with learning resources of different levels of difficulty and types according to their learning progress and interests, so that each student can make progress in a learning environment that suits him or her. This requires foreign language teachers to continuously learn and improve their digital technology skills and master new digital tools and resources [12-13]. At the same time, they also need to continuously reflect on and improve their teaching practices to integrate classroom learner profiling with digital literacy to create more effective and meaningful teaching activities [14-15].

In this study, after collecting and preprocessing the data related to 500 students' learning in English classroom, the Canopy and Maximum Minimum Distance based K-means algorithm was used to output the foreign language classroom learner portrait from three dimensions: learner style, learner behavior and learner outcome. The traditional K-means algorithm is first improved in terms of the determination of the number of clusters and the selection of the initial clustering centers, and then the K-means algorithm based on Canopy and Maximum Minimum Distance is used to cluster and analyze the data of learner behaviors and learner outcomes, and the clustering effect is evaluated by the profile coefficients. In this paper, we construct and analyze learner portraits to explore the hidden learning laws, accurately identify the learning characteristics and learning needs of each learner, and design precise teaching intervention strategies that integrate learner portraits and teachers' digital literacy in the foreign language classroom, with a view to providing suggestions and references for frontline teachers.

2. Constructing Learner Profiles in Foreign Language Classrooms Driven by Big Data

Based on visual learning analytics technology, this study adopts a quantitative user portrait construction method based on qualitative research, and proposes a visual learning analytics-based research learning student portrait process framework, and the foreign language classroom learner portrait construction process is shown in Fig. 1, in order to specifically guide the analysis, portrayal, and presentation of the learning portrait.

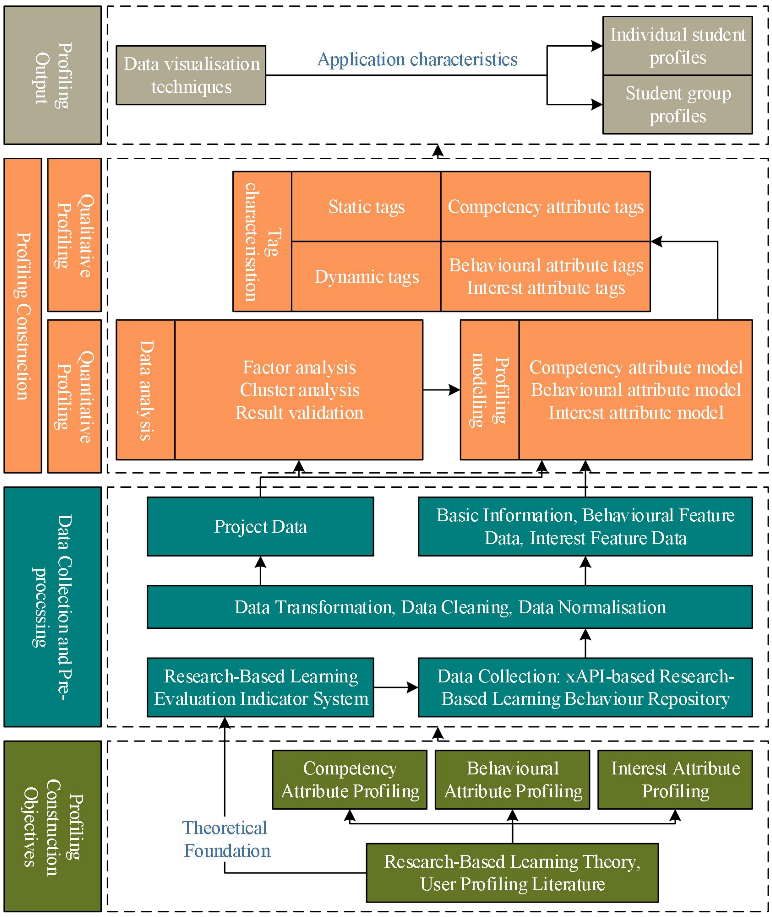

Figure 1. The foreign language class learner portrait construction process

2.1. Data Acquisition and Preprocessing

2.1.1. Data acquisition

Data collection is the foundation of learning analysis and portrait construction. In order to ensure the quality of proficiency portrait construction, this study preconstructed a research study evaluation index system to guide the specific collection of learning data, which was collected from the xAPI-based research study behavioral record library.

In this study, data related to 500 students' learning in English classroom were collected and preprocessed. The preprocessed data were classified into four categories, i.e., students' basic information, learning styles, learning behavior data, and learning outcome data. Among them, subject data are students' process data based on research scaffolding; learning interaction behavior data include routine operations (logging in, logging out, clicking on access, browsing and downloading, saving and deleting, etc.), collaborative exchanges (teacher-student and student-student), viewing, sharing, and commenting on resources (topics and learning resources, etc.), and note-taking; and interest profile data include searching, liking, and bookmarking resources, and results of Multiple Intelligences tests.

2.1.2. Data pre-processing

(1) Data cleaning

(a) Dealing with missing values

In this study, students' course grades are first considered as missing values and then interpolated using Lagrange interpolation.

First, solve the $n-1$ th polynomial of the course grade data of the known over n students, which is calculated as shown in Equation (1):

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} \tag{1}$$

Substituting the course grade coordinates $(x_1, y_1), (x_2, y_2), (x_3, y_3) \cdots (x_n, y_n)$ for n students into the polynomial function yields Equation (2):

$$\begin{aligned} y_1 &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{n-1} x_1^{n-1} \\ y_2 &= a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{n-1} x_2^{n-1} \\ &\cdots \\ y_n &= a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_{n-1} x_n^{n-1} \end{aligned} \tag{2}$$

Next, the Lagrange interpolation polynomial is solved as shown in Eq. (3):

$$\begin{aligned} L(x) &= y_1 \frac{(x-x_2)(x-x_3)\cdots(x-x_n)}{(x_1-x_2)(x_1-x_3)\cdots(x_1-x_n)} \\ &+ y_2 \frac{(x-x_1)(x-x_3)\cdots(x-x_n)}{(x_2-x_1)(x_2-x_3)\cdots(x_2-x_n)} \\ &+ y_n \frac{(x-x_1)(x-x_2)\cdots(x-x_{n-1})}{(x_n-x_1)(x_n-x_2)\cdots(x_n-x_{n-1})} \\ &= \sum_{i=0}^{n} y_i \prod_{j=0, j\neq i}^{n} \frac{x-x_j}{x_i-x_j} \end{aligned} \tag{3}$$

Finally, the Lagrange interpolation method is applied to interpolate the missing values existing in the course grades, firstly, five unmissing grade data before and after the missing values are selected, the independent variable $x_i (i = 1, 2, \cdots, 10)$ and the dependent variable $y_i (i = 1, 2, \cdots, 10)$ of the interpolating polynomials are determined, and then the course grades containing the missing values of the course grade defined as x is substituted into the interpolating polynomial to obtain an approximation of the missing grade $L(x)$.

(b) Handling outliers

In the process of data cleaning, the outliers that exist in the basic student data are converted or not processed, and the existing student data are used to fill in the data containing outliers, and if it is found that the outliers containing outliers are valid data, the data containing outliers are analyzed directly.

(c) Data consistency analysis

In order to maintain the consistency of the data, we integrated the basic attribute data and the dynamic behavior data by using the method of statistical analysis. In the experimental process, each student is regarded as a data object, and each data object contains different basic attributes and dynamic behavioral categories, so we obtain more complete behavioral data of various types.

(2) Data conversion

Data conversion is mainly to normalize the behavioral data of the data object, and normalizing the behavioral data of the data object is a fundamental work in the process of data mining.

(1) Deviation normalization

The deviation normalization refers to the minimum-maximum normalization, which refers to the behavioral data of the data object of the linear transformation processing, mapping its behavioral data in the $[0,1]$ interval, the conversion is shown in the formula (4):

$$x^* = \frac{x - \min}{\max - \min} \tag{4}$$

where x^* denotes the value of the behavioral data of the behavioral data object after min-max normalization; x denotes the initial value of the behavioral data of the behavioral data object; $\max$ denotes the maximum value of the behavioral data of the behavioral data object; $\min$ denotes the minimum value of the behavioral data of the behavioral data object; $\max - \min$ denotes the extreme difference of the behavioral data of the data object in the behavioral data set.

(b) Standard deviation normalization

Standard deviation normalization refers to zero-mean normalization, while the mean value of behavioral data after standard deviation normalization processing is 0, and the standard deviation is 1. The conversion method is shown in Equation (5):

$$x^* = \frac{x - \overline{x}}{\sigma} \tag{5}$$

where x^* denotes the behavioral data value of the behavioral data object after standard deviation normalization; x denotes the initial value of the behavioral data; $\overline{x}$ denotes the mean value of the behavioral data; and σ denotes the standard deviation of the behavioral data.

Currently, the classification algorithm of data mining requires the input data to be in the form of classified attributes, and it is necessary to convert the attribute values of continuous data into classified attribute value intervals, i.e., discretize the continuous behavioral data.

(c) Data Statute

Data statute refers to improve the efficiency of data analysis and mining by constructing a new data set with smaller data volume and can maintain the integrity of the original data. The commonly used methods of data statute are divided into two categories: attribute statute and numerical statute.

2.2. Correlation analysis

Correlation analysis research refers to the method of using statistical analysis of indicators to represent the correlation between random data objects, such as the correlation between students' course grade data and cafeteria consumption data, students' library loan data and course grade data and students' cafeteria consumption data and library loan data are all under the correlation analysis research issues. Commonly used correlation analysis methods include drawing scatter plots, scatter plot matrices and calculating correlation coefficients. In order to more accurately judge and describe the degree of correlation between data objects, we analyze by calculating the correlation coefficient between data objects.

Pearson correlation coefficient is generally used to analyze the correlation between the behavioral data of data objects, and the calculation method is shown in Equation (6).

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{6}$$

where the correlation coefficient r has a value range of $-1 \le r \le 1$, and different values of r indicate the degree of correlation between the behavioral data of the data objects.

$$\begin{cases} r > 0 \text{ Indicates a positive correlation} \\ r < 0 \text{ Indicates a negative correlation} \\ |r| = 0 \text{ Indicates no linear relationship} \\ |r| = 1 \text{ Indicates a perfect linear relationship} \end{cases} \tag{7}$$

Among them, the correlation coefficient $|r|$ takes values in the range of $0 < |r| < 1$, and different values of r indicate different degrees of linear correlation between the behavioral data of the data objects:

$$\begin{cases} r > 0 \text{ Indicates a positive correlation} \\ r < 0 \text{ Indicates a negative correlation} \\ |r| = 0 \text{ Indicates no linear relationship} \\ |r| = 1 \text{ Indicates a perfect linear relationship} \end{cases} \tag{8}$$

2.3. Learner clustering based on improved K-means algorithm

2.3.1. Determination of the number of clusters for clustering

Canopy algorithm is a fast clustering technique that gives the optimal number of clusters although the accuracy is low and does not give the exact cluster results.Canopy algorithm needs to set two variables T1 and T2, randomly select the initial clustering centers, calculate the Euclidean distance between the samples and the initial centers, and divide the initial dataset into a number of imperfectly overlapping subsets according to the thresholds T1 and T2. The basic principle of the algorithm is shown in Fig. 2.

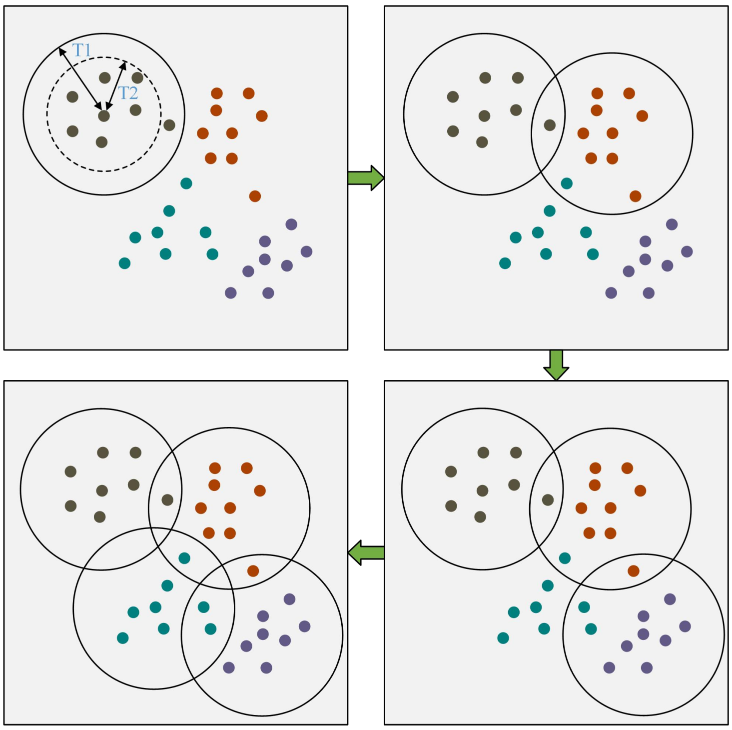

Figure 2. The principle of the canopy algorithm

Canopy algorithm compared to K-means algorithm, although the accuracy is lower, but has a great advantage in speed and does not need to specify the number of clusters in advance. Some scholars have proposed that the Canopy algorithm can be used for “coarse” clustering of data, and then the number of clusters and cluster centers obtained by the Canopy algorithm can be used as the input parameters of the K-mean algorithm to complete the “fine” clustering of data. However, using the cluster center of Canopy as the initial clustering center of K-means algorithm may cause K-means algorithm to fall into local optimum. Therefore, only the Canopy algorithm is chosen to determine the number of clusters K, the specific process is as follows:

The specific algorithm is as follows:

Input: dataset $X = \{x_1, x_2, \cdots, x_n\}$

Step 1: Determine two distance thresholds T_1 and T_2 by cross-validating the tuning parameter or prior knowledge, where T_1 is greater than T_2.

Step 2: Randomly select a sample point P from dataset X as the first Canopy clustering center and add P to set C. Remove P from X.

Step 3: Randomly select a sample Q from the set X and compute the Euclidean distances from Q to each of the Canopy clustering centers in the set C, choosing the smallest of these distances.

Step 4: Compare T_1 with the distance d. If d is greater than T_1, the sample point Q is added to the Canopy with the smallest distance from it and labeled with a weak marker. If d is less than T_1, set Q as the new Canopy clustering center, add the sample point Q to the set C, and remove Q from the dataset X.

Step 5: Comparison will be made with the distance d. If d is less than T_2, attach a strong marker to it, update the center position of all strongly marked samples to the cluster center of this Canopy, and remove the sample point Q from the dataset X. If d is greater than T_2, add Q to the current Canopy set.

Step 6: Repeat steps 3 through 5 until the dataset X is empty.

Output: The number of samples in the cluster center set C is K.

2.3.2. Selection of initial clustering centers

In the K-means algorithm, the selection of initial clustering centers is random. Random selection of initial clustering centers may cause the algorithm to fall into a local optimum, and if the initial clustering centers are clustered together, it will increase the number of iterations of the algorithm and take more time, so the selection of initial clustering centers is an important part of the improvement of the K-means algorithm.

The maximum minimum distance algorithm is a trial-based clustering algorithm that can use the Euclidean distance to select the furthest possible K clustering centers. Using the maximum minimum distance algorithm for K-means algorithm to select as far as possible sample points as the initial clustering center, can avoid K-means algorithm falls into the local optimum, and can reduce the number of iterations of the algorithm. The flowchart for determining the clustering center using the maximum minimum distance algorithm is shown in Figure 3, where $K \geq 3$.

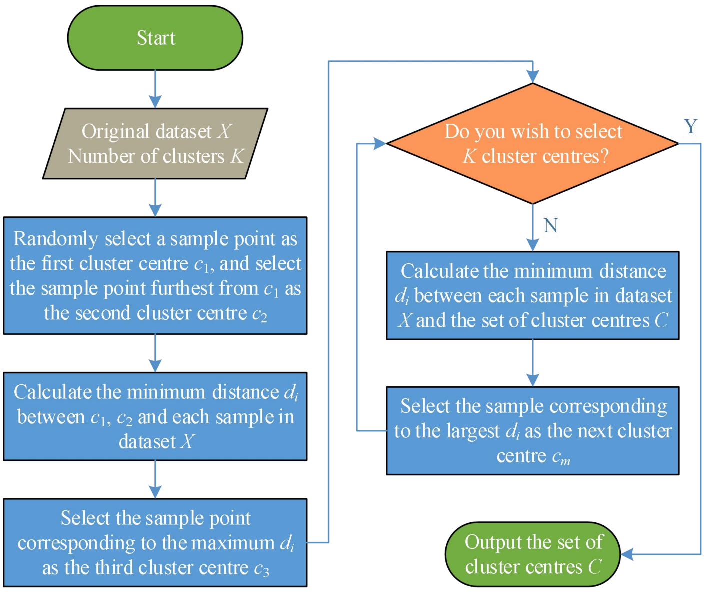

Figure 3. The maximum minimum distance algorithm determines the cluster center

The specific algorithm is as follows:

Input: dataset $X = \{x_1, x_2, \cdots, x_n\}$, number of clusters $K(K \geq 3)$.

Step 1: Take a random sample from the dataset X as the first cluster center c_1.

Step 2: Using Euclidean distance, select the sample furthest from c_1 as the second clustering center c_2.

Step 3: Calculate the Euclidean distance d_{ij} between each clustering center and the samples in the dataset X (where $i = 1, 2, \cdots, n$; $j = 1, 2$), and find the smallest value of them. Select the sample point corresponding to the maximum value d_l in $\{d_1, d_2, \cdots, d_n\}$ and set it as the third clustering center c_3. Let $m = 4$.

$$\begin{cases} d_{ij} = \left\| x_i - c_j \right\| \\ d_i = \min\left[d_{i1}, d_{i2} \right] \\ d_l = \max\left\{ d_1, d_2, \cdots, d_n \right\} \end{cases} \tag{9}$$

Step 4: Select the m th clustering center. For $j = 1, 2, \cdots, m-1$, calculate the distance d_{ij} from the samples in the dataset X to all the clustering centers in the set $\{c_1, c_2, \cdots, c_{m-1}\}$, respectively, and compute d_l by using Eq. (9). The sample corresponding to d_l is chosen as the m th clustering center c_m.

$$d_l = \max_i \left[\min\left[d_{i1}, d_{i2}, \cdots, d_{ij} \right] \right] \tag{10}$$

Step 5: If m is equal to K, output the clustering center set C; otherwise $m = m + 1$, repeat step 4.

Output: initial clustering center set $C = \{c_1, c_2, \cdots, c_k\}$.

3. Foreign language classroom learner profile output

The presentation of student portraits as a form of visual representation relies on data visualization techniques. Iconic graphics can better support the presentation of text-based data and portrait labels. Geometry is the research direction in the field of structured data visualization, and bar charts, line graphs, pie charts, etc. are commonly used forms of data visualization in the field of education for presenting multi-dimensional data attributes. Therefore, in the output part of the portrait, specifically combining the characteristics of data types, appropriate visualization techniques are selected to design and present the individual portrait for students and the group portrait for teachers. The goal of portrait construction is to design a learner portrait model that includes three dimensions: learning style, learning behavior, and learning outcome; learning style includes four dimensions: information input, information perception, information processing, and information comprehension; learning behavior includes three dimensions: participation, interaction, and completion; and learning outcome is classified into three dimensions: learning comprehension, application practice, and transfer innovation.

3.1. Analysis of Indicators for Foreign Language Classroom Learners

3.1.1. Data normalization

The cluster portrait mainly presents the level and characteristics of different learner groups, in which the style characteristics are static indicators, so this type of data is used as a separate criterion for clustering. The data indicators of behavioral characteristics and ability level are selected for clustering, in order to ensure the consistency of data standards, the six indicators involved in clustering are firstly normalized by Min-Max, and the mean and standard deviation of learners' data after normalization are shown in Table 1, which obtains the mean and standard deviation of each data of students, among which the normalized value of concentration is the highest 0.701.

Table 1. The average and standard deviation of the learner's data

Dimension	Mean	SD
Participation	0.241	0.255
Interaction degree	0.288	0.155
Concentration	0.701	0.222
Learning comprehension	0.622	0.189
Practical practice	0.523	0.235
Migration innovation	0.501	0.211

3.1.2. Data mapping

According to the above data processing results, using principal component analysis for dimensionality reduction, PC1 and PC2 explain the cumulative percentage of 64.24%, which is representative and can be calculated by subgrouping. A two-dimensional space is constructed from these two PCA factors as the X/Y axis, and the flat spatial distribution of learning data points mapped to learning data within this space is shown in Figure 4, where the data distribution is relatively uniform.

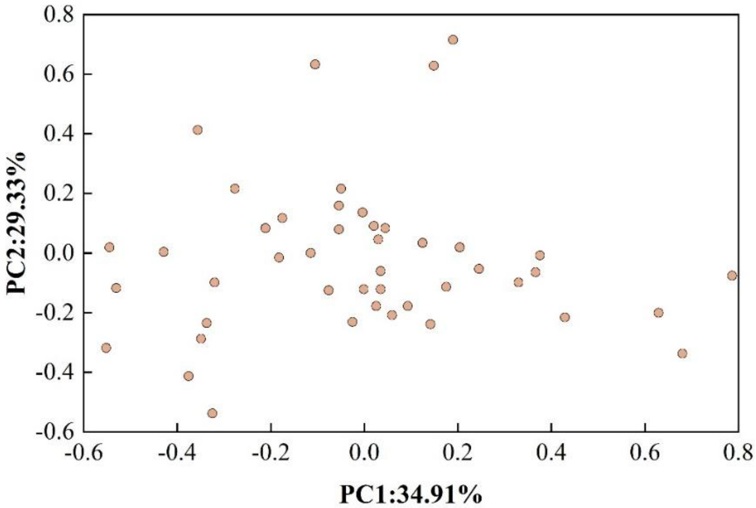

Figure 4. The plane space distribution of the learning data

3.1.3. Correlation analysis of indicators

In order to explore the correlation between the variables, this study used Pearson correlation analysis, the analysis was done through python, and the correlation matrix is shown in Figure 5. The ten dimensions of information input, information perception, information processing and information comprehension, participation, interaction and completion, learning comprehension, application practice and transfer of innovation are noted as D1~D10.

Statistically, there is a significant correlation between some of the variables, and there is a significant positive correlation between students' information processing and information perception, so that students with high information perception have a greater tendency to utilize it for learning. There is a significant positive correlation between students' information comprehension and information input. Academic performance has a positive correlation with all behavioral data indicators, and the correlation coefficient of the two indicators, especially the completion of the learning behavior dimension and the learning understanding of the learning outcome, came to 0.946, which shows a strong correlation between the two. Therefore, students' learning behaviors can affect academic performance, and academic performance can be improved by intervening in students' behaviors, especially to improve the degree of classroom interaction and learning concentration.

From the human perspective, students' learning effectiveness is not only related to students' own efforts, learning attitudes, and ability levels, but also has an important relationship with teachers' teaching methods, which is presented in this study as an important effect of encouraging teaching methods on students; from the classroom perspective, the degree of students' practice in class and the degree of participation in class have an important effect on learning effectiveness; from the perspective of learning styles, in this study whether or not modern computing was used for learning had no significant effect on student learning outcomes.

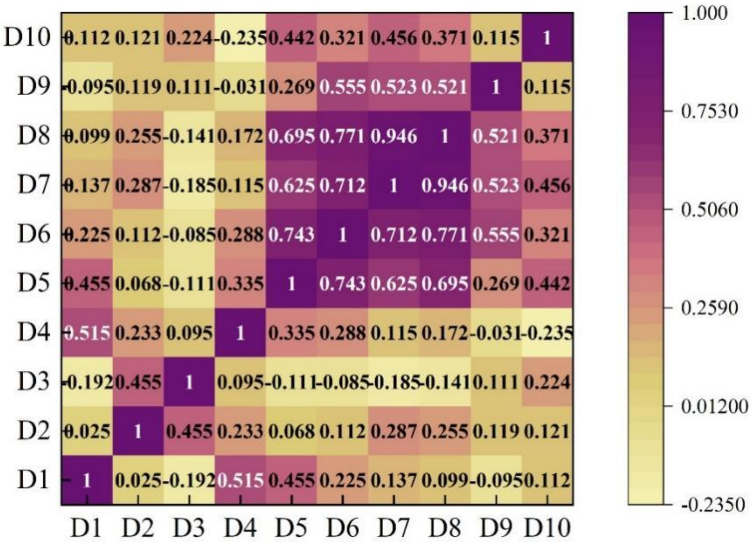

Figure 5. Correlation matrix

3.2. Improved K-means clustering of foreign language learners

Clustering is done using K-means clustering, before that the value of K, i.e. the number of classification values, needs to be determined. The SSE value (sum of squared residuals) is calculated using Python and the K value is set to iterate from 1 to 10, and the optimal K value is determined by the elbow rule. The elbow rule is based on the degree of distortion of the cluster to determine the number of clusters, for a certain degree of differentiation of the data, the degree of distortion will be substantially improved when a certain threshold is reached, and the downward trend gradually becomes slower. The elbow rule determines the optimal K value as shown in Table 2, when K is 4, there is an obvious trend change, at this time the SSE value is 6.3346, which determines the optimal K value of 4, that is, dividing 4 groups of learners, using the Chiplot tool to map the results of the sub-cluster to the previous two-dimensional space to get the clustering scatterplot shown in Figure 6.

Table 2. The elbow rule determines the optimal k value

K	1	2	3	4	5
SSE	12.2523	10.4456	8.6625	6.3346	5.3215
K	6	7	8	9	10
SSE	5.0742	4.6562	4.2235	3.7014	3.3325

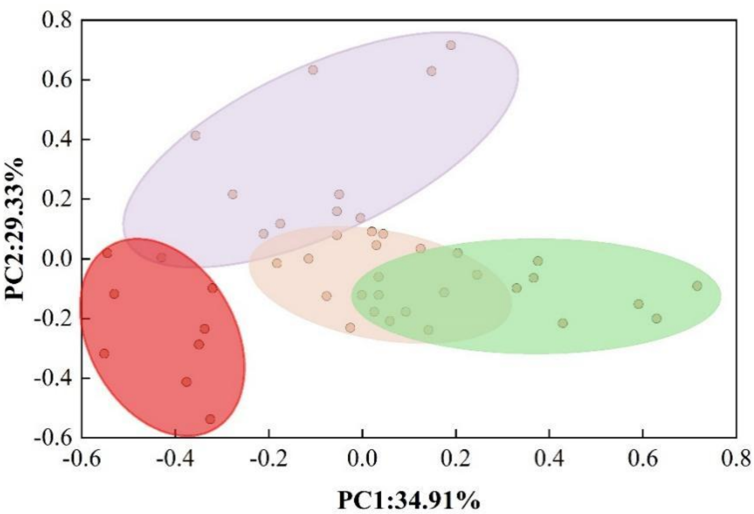

Figure 6. The learner's portrait group results

3.3. Construction of Learning Portrait in Foreign Language Classroom

After the above K-means algorithm clustering based on Canopy and Maximum Minimum Distance, this study divides the students in the class into four category groups, the number of people in the four category groups is counted, and the results of the clustering results of the class portrait are analyzed as shown in Table 3. There are 230 people in category group 1, the learning style takes the big value, and it is defined as active + sensing + visual type, the behavior is mostly medium-involved learners, and the ability level is basically at level 3. There are 99 people in category group 2, the learning style is defined as active + sensing + holistic type, the high-involved learners are mostly, the ability is at level 1 or level 2, and it is the best performance in terms of the behavior and the ability. Cluster 3 has 67 people with learning styles defined as active + intuitive + visual, mostly low-involvement learners in terms of behavior, and competence level grades located at levels 3, 4, and 5, with the worst performance. Cluster 4 has 104 people with learning styles defined as active + visual + sequential, behaviors mostly as medium to high engaged learners, and competency level grades are mostly located at level 4.

Table 3. The group portrait clustering results analyzed

Group, number and proportion		Group 1 (N=230,46.00%)	Group 2 (N=99,19.8%)	Group (N=67,13.4%)	Group (N=104,20.8)
Learning style	Active/reflective	3/0	5/2	3/1	2/0
	Feeling type/intuitive type	4/1	3/1	0/1	2/2
	Visual/verbal type	5/2	3/3	4/0	2/1
	Sequence type/overall type	4/4	0/2	1/1	1/0
Learning behavior	High input	0	67(67.7%)	0	52(50%)
	Medium input	220(95.7%)	32(33.3%)	10(14.9%)	52(50%)
	Low input	10(4.3%)	0	57(85.1%)	0
Learning results	Level 1	0	32(33.3%)	0	0
	Level 2	0	67(67.7%)	0	0
	Level 3	176(76.5%)	0	22(32.8%)	10(9.6%)
	Level 4	54(23.5%)	0	22(32.8%)	84(80.8%)
	Level 5	0	0	23(34.4%)	10(9.6%)

Calculate the average of each cluster in each dimension, get the average of the four clusters in these six indicators and the overall average, the average of the cluster portrait in each indicator is shown in Table 4, combined with the clustering statistics and the four cluster portraits in Table 3, the four categories of learners are assigned cluster labels. The ability level of cluster 1 is slightly higher but the behavioral commitment is lower, this kind of learner has strong comprehension and expression but poor discipline on their own behavior, which leads to the final learning performance can not show the best results, it is defined as “potential learners”, and the performance of cluster 2 is the best, but the degree of participation and interaction is somewhat lacking, it is defined as Cluster 2 is the best performer, but lacks engagement and interaction, so it is defined as a “good learner”, Cluster 3 is the worst performer, and even though “learning understanding” is more outstanding, it is still insufficient compared to the other clusters, so it is defined as a “marginal learner”, and Cluster 4 does not show the best results in terms of performance. “Cluster 4 actively participated in the behavior, but scored low in terms of correct answers to the questions in the resource, and the quality of the exercises and assignments, which may be ineffective participation or weak learning ability, and was therefore classified as a ”struggling learner".

Table 4. The average of a group of portraits on each index

		Group1	Group2	Group3	Group4	Mean
Learning behavior	Participation	0.131	0.565	0.084	0.168	0.237
	Interaction degree	0.302	0.31	0.227	0.305	0.286
	Completion degree	0.814	0.77	0.344	0.865	0.698
Learning results	Learning comprehension	0.648	0.827	0.524	0.487	0.621
	Applied practice	0.404	0.799	0.519	0.358	0.52
	Migration innovation	0.551	0.773	0.252	0.414	0.498

4. Practice of Integrating Learner Portrait and Teacher Digital Literacy in Foreign Language Classroom

4.1. Instructional Pathways that Integrate Learner Profiles and Teacher Digital

Literacy

(1) Define the goal of building accurate portraits of teachers and students

In the rapid development of the information age and the rapidly changing environment of education, society has put forward higher and higher requirements for all aspects of education and teaching. The overall goal of this paper can be summarized as to improve the quality of teaching, promote the development of classroom education, and promote the common development of teachers' digital literacy. Specifically, the construction of accurate portraits of teachers and students in classroom teaching is to help teachers examine their own problems, improve the effectiveness of teachers' classroom teaching, and help students reflect on the learning process, so as to effectively improve students' learning.

(2) Implementation of hierarchical teaching

The results of the group portrait can lay the foundation for the implementation of "accurate stratification" in the classroom. Teachers can carry out stratified teaching design from teaching objectives to teaching evaluation according to the output of the portrait. Taking the above categorization results as an example, in the process of designing teaching objectives, in order to avoid the phenomenon of "one-size-fits-all", teachers can integrate the characteristics of the results of the student group portrait, design three levels of quantifiable teaching objectives according to the level of learning results of the student group orientation, and make clear the proportion of different types of students before and after each objective is set. In the arrangement of teaching activities, for the progressive and inefficient students, teachers need to build a dynamic support system of "learning scaffolding and scaffolding removal" in terms of learning strategies and resources; for the highly efficient and progressive students, they need to realize the supply of expanding resources in order to build personalized enhancement channels. In the arrangement of teaching interventions, teachers need to arrange collaborative learning between rational copers and highly interactive learners.

(3) Teachers' Digital Literacy Enhancement

In the English classroom, teachers still play the roles of instructional activity designers and content lecturers, but based on the results of group profiling, teachers can change their roles. For efficient and aggressive learners, teachers can turn into personalized learning guides with the help of teaching resource libraries, guiding students to plan their learning paths; for rational and inefficient learners, teachers can turn into learning motivators and supporters, providing support and motivation in terms of resources, tools, and strategies, for example, by using the flipped classroom, cooperative learning, and point-scoring mechanisms to enhance students' English cognition and intrinsic motivation.

4.2. Empirical analysis

To address the effectiveness of the application of the precise teaching model integrating learner profiling and teacher digital literacy in college English teaching, the author conducted a quasi-experimental study based on the utilization of the model in a college English course. The participants in the study were non-English majors in the class of 2020. The effective number of students in both classes is 50, and they are assigned with the same lecturer. Before the experiment, the author first tested the English proficiency of the students in the two classes taught, randomly set the two classes as an experimental class and a control class, and conducted a one-year teaching experiment. During the experiment, the experimental class received the precise teaching mode, while the control class received the regular teaching mode. At the end of the experiment, the author again tested the English proficiency of the two classes by sub-tests and drew conclusions.

4.2.1. Comparative analysis of learning effectiveness

The improvement of learners' English proficiency by comparing the precise model of college English was conducted. Achievement tests were conducted on the experimental and control classes before and after the experiment. The English proficiency of the students in the two classes taught was tested before the experiment, and Figure 7 shows the comparison of the total scores of the experimental and control classes on the pre-test and post-test sides, the experimental and control classes were comparable before the experiment, and the mean value of the students in the control class was a little higher than the mean value of the experimental class, but there was no significant difference between the two ($p>0.05$). After the experiment, there is a significant difference between the average scores of the experimental class and the average scores of the control class (sig.=0.000), and the scores of the experimental class are significantly better than the scores of the control class, with the difference between the average scores of the two classes amounting to 10. And the standard deviation of the post-test scores of the experimental class is narrowed, which indicates that the class's English scores are more average, and that the bifurcation of the scores has been improved to a certain extent.

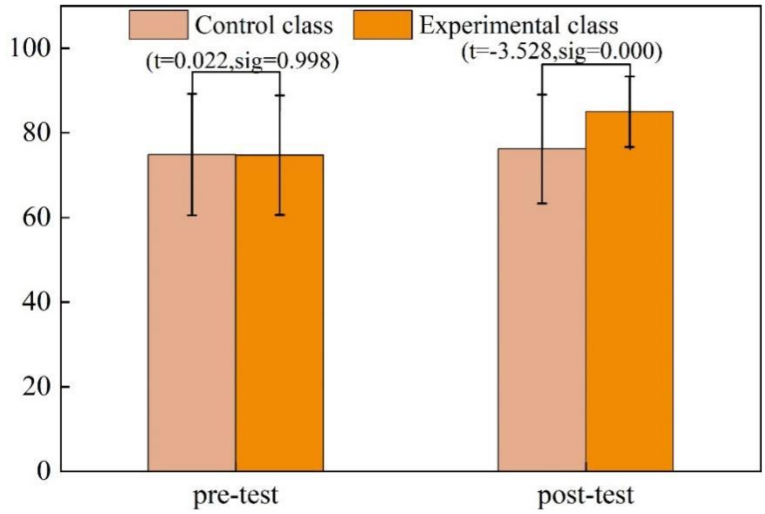

Figure 7. Comparison of results in the previous test class and the comparison class

In order to better determine the differences in students' performance in the English subcategories, the author divided the post-test questions into five parts: listening, reading, writing and translating, and Table 5 shows the subcomparison of the experimental class and the control class at the end of the semester. Under the integration of learner profiling and teachers' digital literacy teaching model, the experimental class showed significant improvement in listening, reading, writing and translating compared with the control class. Among them, there are significant differences in listening, reading, writing and translation ($p<0.05$), indicating that the precision teaching has improved students' listening, reading, writing and translation. The reason for this is that under the precise teaching mode, the results of the English reading proficiency test are more accurate, and the students adjust their learning state in time according to the test data. The difference between the experimental class and the control class in speaking is not significant and does not show a significant difference ($p>0.05$), indicating that the precise teaching mode is not obvious enough to improve listening and speaking. The main reason is that it is more difficult to test the speaking ability under the precise mode, and the data are not precise enough. In addition, the English language environment is lacking, and students have less independent training.

Table 5. The grades of grades were compared in the end of the two classes

	Control Class	Experimental Class	Independent sample t test	
	Mean±S.D.	Mean±S.D.	t	Sig.
Listening	14.995±12.331	17.221±6.335	-3.221	0.000
Speaking	15.928±0.533	16.887±8.265	-2.786	0.771
Reading	14.485±13.222	17.111±9.123	-4.495	0.005
Writing	14.414±12.155	16.705±7.235	-3.377	0.002
Translation	15.655±10.332	17.122±6.666	-3.012	0.004

4.2.2. Comparative Analysis of Comprehensive Competency Enhancement and Learning Experience

After the completion of the experiment, questionnaires were administered to the experimental and control classes for the dimensions of learning interest, learning goals, independent learning ability, learning methods and teamwork ability to detect the improvement of students' comprehensive ability.

There is a significant difference ($p<0.05$) between the control class and the experimental class in the four aspects of learning interest, learning goals, learning methods and independent learning ability, which clarifies the learning goals for students and lays the foundation for students' selective learning; the precision test further promotes independent learning, and the learning data stored in the big data system can be used as the basis for students to understand the learning progress and knowledge mastery, and to utilize the big data to make a comprehensive and objective assessment of the learning effect over a period of time. The learning data stored in the big data system can be used as a basis for students to understand their learning progress and knowledge mastery, and the use of big data to make a comprehensive and objective understanding of the learning effect for a period of time, which is convenient for students to adjust their learning methods in a timely manner, improve their learning efficiency, and thus increase their learning interest. However, there is no significant difference in the index of “teamwork ability”, indicating that precision teaching has limited improvement in teamwork ability. The reason for this is that the precise teaching mode emphasizes more on students' independent learning and precise testing, and the special characteristics of the English language subject do not require much teamwork, and there are fewer team tasks set up, so students' teamwork ability is not significantly improved.

Table 6. Students' overall ability to improve

	Control Class	**Experimental Class**	**Independent sample t test**	
	Mean±S.D.	Mean±S.D.	t	Sig.
Interest in learning	2.95±1.23	4.08±1.23	-3.341	0.000
Clear learning goals	2.55±1.31	3.62±1.11	-3.445	0.000
Autonomous learning	2.44±1.12	3.71±1.17	-4.421	0.000
Learning method improvement	2.78±1.21	3.92±1.33	-3.582	0.000
Team ability improvement	2.51±1.38	2.95±1.47	-1.445	0.158

5. Conclusion

Big data-driven education paradigm shift from “evidence-based” to “data-based”. Precision teaching is a personalized teaching paradigm that implements targeted teaching based on a full understanding of each student's characteristics and learning needs. This study constructed a learner profile model from a goal-oriented perspective, collected data to create the profiles, and used the profiling results to diagnose teaching issues. The learner profiles in foreign language classrooms classified learners into "potential learners", "excellent learners", "borderline learners", and "striving learners". By deconstructing the English learning characteristic portraits of different clusters, the differentiated learning characteristics and needs of different learner groups are presented, and based on that, the precise practical suggestions for integrating the learner portraits of foreign language classrooms and teachers' digital literacy are proposed. Practice shows that there is a significant difference between the control class and the experimental class in the improvement of the five aspects of academic performance, learning interest, learning goals, learning methods and independent learning ability, which suggests that it has a significant role in promoting the learning effect of English learners, and demonstrates the feasibility as well as practicability of precision teaching. However, due to the limited time and conditions, the study fails to fully reflect the dynamics of teaching and learning, and it is necessary to carry out continuous research in the future to further verify the reasonableness of the intervention design framework and the learner portrait model, as well as the effectiveness of the portrait-based precision teaching intervention strategy, so as to better meet the personalized learning needs of the learners and the needs of future teaching.

Funding

Shanghai Zhongqiao Vocational and Technical University 2025 Qinglan Project — Social Sciences General Project: A Study on the Integration of Learner Profiles and Teacher Digital Literacy in Big Data-Driven Foreign Language Classrooms (Project No.: 2025ZQSK63).

References

1. Nie, Y. (2023). Application of multimodal multimedia information and big data technology in teaching chinese as a foreign language course. International Journal of Digital Multimedia Broadcasting, 2023(1), 2257863.
2. Ji-Hua, F. (2018). On computer and foreign language teaching and learning in big data era. International Journal of Emerging Technologies in Learning (Online), 13(5), 236.
3. Hamim, T., Benabbou, F., & Sael, N. (2019, October). Student profile modeling: an overview model. In Proceedings of the 4th international conference on smart city applications (pp. 1-9).
4. Sael, N., Hamim, T., & Benabbou, F. (2019). Implementation of the Analytic Hierarchy Process for student profile analysis. International Journal of Emerging Technologies in Learning (Online), 14(15), 78.

5. Hamim, T., Benabbou, F., & Sael, N. (2019, October). Toward a generic student profile model. In The Proceedings of the Third International Conference on Smart City Applications (pp. 200-214). Cham: Springer International Publishing.
6. Dasmana, A., Wasliman, I., & Yoseptry, R. (2022). Implementation of integrated quality management strengthening character education in realizing Pancasila student profiles. IJGIE (International Journal of Graduate of Islamic Education), 3(2), 361-377.
7. Nguyen, L. A. T., & Habók, A. (2024). Tools for assessing teacher digital literacy: a review. Journal of Computers in Education, 11(1), 305-346.
8. Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. Phi Delta Kappan, 99(6), 27-32.
9. Çam, E., & Kiyici, M. (2017). Perceptions of Prospective Teachers on Digital Literacy. Malaysian Online Journal of Educational Technology, (4), 29-44.
10. Rahmawati, S., Abdullah, A. G., & Widiaty, I. (2024). Teachers' digital literacy overview in secondary school. International Journal of Evaluation and Research in Education, 13(1), 597-606.
11. Jung, J., Choi, S., & Fanguy, M. (2024). Exploring teachers' digital literacy experiences. International Review of Research in Open and Distributed Learning, 25(2), 41-59.
12. Aluko, F. R., & Ooko, M. (2022). Enhancing the Digital Literacy Experience of Teachers to Bolster Learning in the 21st Century. Journal of Learning for Development, 9(3), 420-435.
13. Zayas, J. D. M., & Rofi'ah, N. (2022). The effect of digital literacy skills on improving teacher creativity. INSECTA: Integrative Science Education and Teaching Activity Journal, 3(2), 168-174.
14. Záhorec, J., Hašková, A., & Munk, M. (2019). Teachers' Professional Digital Literacy Skills and Their Upgrade. European Journal of Contemporary Education, 8(2), 378-393.
15. Falloon, G. (2020). From digital literacy to digital competence: the teacher digital competency (TDC) framework. Educational technology research and development, 68(5), 2449-2472.