Research Article

Using Spectral Clustering Association Algorithm upon Teaching Big Data for Precise Education

Yongfu Zhou, 1 Zhi Zeng, 2 and Huabin Wang 2

1 Institute of Electronic and Information Engineering, Heyuan Polytechnic, Heyuan 517000, China
2 School of Computer Science and Engineering, Huizhou University, Huizhou 516007, China

Correspondence should be addressed to Zhi Zeng; zengzhi@hzu.edu.cn

Received 22 June 2022; Accepted 20 August 2022; Published 26 September 2022

Academic Editor: jianguo duan

Copyright © 2022 Yongfu Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous deepening of the application of educational OA, massive educational data has been produced. Hence, the application of teaching big data (TBD) has a certain theoretical basis, practical methods, and research methods in the field of education. How to fully play the leading role of education on TBD in professional education, guide and recommend students to carry out personalized learning, change the teaching mode, enrich the teaching evaluation, then further improve the quality of talent training is a current issue, which has yet to be solved. Based on the analysis and mining of big data, this paper uses the spectral clustering algorithm to construct the curriculum association classification model and realizes the clustering of core courses. Then, through the analysis of the academic achievements of previous professional core courses, we can master the current situation of students’ learning, construct the model as the portrait of students’ individual, curriculum, and professional characteristics through deep learning, so as to realize the precise referral of personalized learning courses, provide students with targeted academic guidance, and further dynamically adjust the teaching syllabus, including the teaching methods and teaching means. Vice versa, we can improve the core courses clustering to further feedback on the curriculum association classification model by analyzing the job position technical requirements. Experiments show that the proposed model using a spectrum clustering algorithm could be better provided strong technical support for the decision-making of precision education in colleges and universities.

1. Introduction

Nowadays, with the wide spread and application of the big data concept, its connotation is constantly evolving and extending. People are gradually realizing that big data is not only a kind of technology but also a kind of ability, such as the to find meaningful associations from massive and complex data, mine the changing rules of things, and accurately predicting the development trend of things. In addition, big data is also a way of thinking; it can let data speak, and let data become the basic starting point of human thinking and decision-making behavior. However, TBD refers to the data set generated in the whole process of educational activities and collected according to the needs of education, which can be used for the development of education and create great potential value [1]. TBD has a clear goal orientation, that is, to the development of education. It can play an effective role in improving the quality of education, promoting educational equity, realizing personalized learning, optimizing the allocation of educational resources, and assisting scientific decision-making of education [2, 3]. The ultimate value of big data should be reflected in the deep integration with the mainstream business of education and the continuous promotion of the intelligent reform of the education system [4, 5]. At present, there are some innovative application cases of education big data at home and abroad, covering teaching, management, evaluation, service, and other fields [6–8]. With the support of TBD, upon the mining and analysis of students’ teaching data, we can more accurately predict learners’ action trajectory, and then, effectively support students’ personalized adaptive learning, which is the key to realize precision education [9]. As early as
the 1990s, some scholars studied the referral service. With the continuous development and improvement of the referral algorithm, Resnick, and others proposed the concept of personalized referral, and personalized referral service is respected [10]. It was first used in the commercial field, mainly represented by American business giants such as Amazon and Walmart. It collects relevant data from users, calculates the similarity of various goods, and connects users with goods through the referral algorithm, which greatly improves the personalized and commercial level of the referral service. With the continuous application of computer technology, there are many information referral systems in different fields, such as web watcher, lira, Letizia, and other personalized referral systems [11]. The application of personalized referral services in the field of education has gradually developed. For example, the e-learning system supported by online course resources [12], or the learning resources provided according to the dynamic progress of students [13], and the e-book Package based on the information of learners’ personal, preference, academic, performance, and so on [14]. TBD covers the knowledge of educational data mining and learning analytics [15, 16]. However, in precision teaching, in addition to promotion services, the learning effects are also a part that we must pay attention to, so the early warning of academic performance based on teaching big data is also the focus of this research. When it comes to academic achievement early warning, is naturally related to data mining and analysis. TBD [17], whether it involves students or teaching managers, should focus on finding hidden valuable information [18], so as to provide effective guidance for future teaching activities, which is of great significance. Academic early warning is an important means for colleges and universities to strengthen students’ learning management and improve the level of education and teaching management. Similarly, in the 1990s, many foreign universities used students’ academic achievement as the data source for academic early warning [19]. In China, it started relatively late to predict and explain the learning performance with academic risk through the first semester scores and college entrance examination scores, so as to help implement academic early warning [20]. However, the mechanism and method of early warning are relatively simple, and once problems are found, we are unable to solve them timely with an effective response [21].

Therefore, based on the score data of software majors in the first semester, this paper focuses on the three aspects of teaching quality and establishes the school’s running orientation and quality objectives in line with the actual situation of the school in order to meet the needs of economic and social development and students’ personal development [22].

Specifically, implementing precise teaching referral, for students, mainly includes precise referral of students’ chosen courses and precise warning of great academic fluctuations. The core of precise teaching referral is a precise prediction of student-course achievement to get professional referral courses. Through the multidimensional teaching quality evaluation model, we can adjust the professional courses. Considering that the student courses’ score is a comprehensive problem involving all aspects, in which the more complex curriculum factors include: syllabus, teaching focus, and assessment difficulties, etc., and some other accidental factors, such as whether the examination is good or not. Therefore, it is a challenging issue to precisely predict student-course achievement and professional course referral. In the process of solving this problem, this paper focuses on solving the following key technical problems, as shown in Figure 1: refining the professional portrait, core curriculum portrait, and students’ personal portrait; constructing a curriculum association classification model to complete the correlation measurement between various professional courses and job positions; modeling the similarity of student achievement distribution; and multidimensional constructing and improving precise courses referral. On the whole, the work of this paper includes:

1. To construct a frequent pattern spectral clustering classification model of professional core courses.
2. To study the potential relationship between professional courses, the characteristic model of students’ individual, curriculum, and a professional portrait, and to use the deep learning method to train the same professional achievements in the past, so as to realize the prediction of the related curriculum achievements between the lower grade and the higher grade. We should give early warning of the courses that may be difficult and guide teachers to provide targeted guidance to students.
3. To build a multidimensional precise referral of teaching course, dynamic adjustment to guide teachers in the adjustment of the syllabus, teaching methods, and teaching means.

2. Theoretical Methods

2.1. Spectral Clustering Algorithm. Spectral clustering is a dimensionality reduction clustering method based on graph theory, which has a smaller calculation amount, is easy to implement, and is good at processing high-dimensional data [23]. Generally speaking, the difficulty of extracting professional portraits from the syllabus and training plan with complex structure lies in the description of the complex structure of the professional training plan, especially from the professional plan to the curriculum. We can use frequent pattern spectral clustering to classify courses and determine the weight proportion of professional courses. According to the feedback effect of the teaching quality evaluation, the professional courses should be adjusted appropriately. Therefore, we must further obtain the job skills required by the Internet through crawlers and determine the curriculum system to support such skills. The spectral clustering algorithm is used to construct the relationship between job position skills and professional core courses.

Spectral clustering (SC) is a clustering method based on graph theory, which divides the weighted undirected graphs into two or more optimal subgraphs to make the subgraphs as similar as possible and the distance between subgraphs as far as possible, so as to achieve the purpose of common
clustering. Among them, the best means that the optimal objectives function is different. It can be the smallest cut and the best cut of the cut edge, such as the normalized cut in Figure 2.

In this way, spectral clustering can identify any shape of sample space and converge to the global optimal solution. The basic idea is to use the feature vectors obtained from the feature decomposition of the sample data similarity matrix (Laplace matrix) to cluster. For the space vector and item-user matrix, see Table 1.

Now, according to Figure 3(a), we can calculate the similarity between items, and then we can get a similarity matrix with only items. Furthermore, we regard items as vertex \((V)\) in the graph \((G)\), and the similarity between items as an edge \((E)\) in \(g\). In this way, we can get our common concept of the graph. Then we get the adjacency matrix \(e\) as shown in Figure 3(b), where \(e_{ij}\) represents the weights of VI and the edges of VI, \(e\) is a symmetric matrix, and the elements on the diagonal are 0. Thus, the Laplacian matrix as shown in Figure 3(c) is obtained: \(l = D - e\), where \(di\) (the sum of row or column elements).

First, consider an optimal image segmentation method. Take bisection as an example, divide the graph into \(S\) and \(T\) parts, which is equivalent to the following loss function cut \((s, t)\), as shown in formula (1), that is, the minimum (weighted sum of cut edges).

\[
\text{cut}(S, T) = \sum_{i \in S, j \in T} e_{ij}, \quad (1)
\]

\[
q = \left[ q_1, q_2, \ldots, q_n \right]^T, \quad (2)
\]

\[
q = \begin{cases} 
C_1, & i \in S, \\
C_2, & i \in T.
\end{cases} \quad (3)
\]

Suppose that the two categories are divided, \(S\) and \(T\), and \(Q\) as shown in formula (2) is used to represent the classification, and \(Q\) satisfies the relationship of formula (3) used for class identification.

\[
\text{cut}(S, T) = \sum_{i \in S, j \in T} e_{ij} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}(q_i - q_j)^2}{2(c_1 - c_2)^2}, \quad (4)
\]

where,

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}(q_i - q_j)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}(q_i^2 - 2q_iq_j + q_j^2)
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} -2q_iq_je_{ij} + \sum_{i=1}^{n} q_i^2 \left( \sum_{j=1}^{n} e_{ij} \right)
\]

\[
= 2q^T (D - E)q = 2q^T Lq,
\]
Where $D$ is the sum of row or column elements in the diagonal matrix, and $L$ is the Laplacian matrix. From formula (5):

$$q^T L q = \frac{1}{2} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij} (q_i - q_j)^2 \right).$$

(6)

$$q^T q = n.$$

We have,

(a) $L$ is a symmetric positive semidefinite matrix, which guarantees that all eigenvalues are greater than or equal to 0;

(b) $L$-matrix has a unique 0 eigenvalue and its corresponding eigenvector is 1.

2.2. Similarity Calculation of Professional Core Courses. According to PCC (Pearson’s correlation coefficient), within the given threshold range, the calculation of PCC between previous students’ academic achievement and current students’ academic achievement is to realize the curriculum referral and early warning of academic achievement, as well as the job referral. PCC is widely used to measure the degree of correlation between two variables, and its value ranges from -1 to 1. It evolved from a similar but slightly different idea, put forward by Francis Galton in the 1880s by Carl Pearson. This correlation coefficient is also called the Pearson product-moment correlation coefficient.

PCC between two variables is defined as the quotient of covariance and standard deviation between two variables. We have,

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}. \quad (7)$$

The above formula defines the overall correlation coefficient, which is usually represented by the Greek small letter $\rho$. PCC can be obtained by estimating the covariance and standard deviation of the sample, which is usually expressed by the letter $r$.

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}. \quad (8)$$

Similarly, it can be estimated from the mean value of the standard fraction of $(x, y)$ sample points to obtain the equivalent expression of the above formula.

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{X_i - \bar{X}}{\sigma_X} \right) \left( \frac{Y_i - \bar{Y}}{\sigma_Y} \right). \quad (9)$$

Where $(X_i - \bar{X}/\sigma_X)$, $\bar{X}$ and $\sigma_X$ are the standard fraction of sample points $X_i$, the mean value and standard deviation, respectively.

3. Proposed Models

Considering that the spectral clustering algorithm is the method of dimensional reduction, it is more suitable for high-dimensional data processing and does not need to consider the shape of the sample space. The algorithm only needs to calculate the similarity matrix between the data sets to achieve clustering. Compared with the traditional clustering method, it is more efficient in dealing with the high-dimensional and sparse matrix of university curriculum scores.

In order to realize the spectral clustering algorithm of course grades, first, the undirected weight graph based on similarity metrics is constructed. Then the graph is divided into different subgraphs according to the trimming rules, and finally to realize clustering. The clustering algorithm is described as follows:

4. Testing and Experiments

In Python, we developed a big data visual platform for the overall overview of the college of computer science, and our experiments are conducted on a server with an Intel Core5i@1600HZ and 16 GB of memory. The platform is as shown in Figures 4(a)–4(c), which includes student portraits, teacher portraits, professional portraits, and so on. Also we took the academic achievement from TBD as an example to analyze the personalized learning, to implement precise teaching referral and precise warning of great academic fluctuations.

Therefore, we studied the relationship between student attributes and job positions, and mined the association rules of student groups, namely feature vectors [24]. Herein, let the feature vectors are the basis of student clustering. According to the attribute value of historic achievement, assume that the E.Score or Re.score divided into 5 levels, more than 90 is ‘A’, range from 80 to 90 is ‘B’, 70 to 80 is ‘C’, 60 to 70 is ‘D’, and less than 60 is ‘E’. We generalize the values of each attribute for historic achievement, as shown in Table 2, technical capability on job position got from web mining in Table 3, and a sample of employment historic data in Table 4.
Figure 4: Continued.
There are two kinds of courses: compulsory and optional. More specifically, compulsory courses include some professional basic courses and professional core courses. Therefore, as far as computer-related professional majors are concerned, job skills needs should pay more attention to the professional core courses. Now, assuming that the core courses of job requirements have been mined from the external recruitment website, we can calculate the correlation degree between the job and these courses by PCC according to the professional core courses. In this way, we can regard each course as a vertex and the correlation between courses as weighted edges. Spectral clustering uses the graph segmentation method to divide the weighted undirected graph of all courses into several optimal subgraphs so that the weight of edges connecting different courses is as low as possible, and the weight of edges connecting the same courses is as high as possible. Table 5 is an example of a core curriculum set for technical capability in a job position.

Let the min-support threshold and credibility be 0.1, respectively, and the undirected graph of spectral clustering is built as Figure 5.

From Figure 4, we can see the clustering results of professional core courses are constructed by the spectral clustering algorithm. Now, taking the above Table 2 as an example, there are four types of job positions, which need to possess the related professional courses. To calculate the distribution between job positions and the professional courses by formula (1), we adjusted the core courses of the major specified in the training program, such as increasing or reducing the existing courses.

\[
H(P) = - \sum_{i=0}^{m} p_i \log p_i .
\]  

(10)

Among them, \( p_i \) represents the distribution of professional courses. Figure 6 describes the distribution of job positions’ intervals, and courses are divided into first intervals, indicating that such courses mainly belong to the first type of job positions, and so on.

For the purpose of validation of the professional core courses, we can calculate the correlation by PCCs for different job positions, choose the maximum value for the core courses according to the final result, and subsequently adjust the professional courses for the training program. Figure 7 shows the referral process for professional core courses for learners.
Table 2: Sample of historic achievement.

<table>
<thead>
<tr>
<th>StuNo</th>
<th>Name</th>
<th>Sem.-year</th>
<th>Sem</th>
<th>M.ID</th>
<th>Major</th>
<th>C.No</th>
<th>E.Score</th>
<th>Re.Score</th>
<th>C.Char</th>
<th>S.Num/P</th>
</tr>
</thead>
<tbody>
<tr>
<td>019120801</td>
<td>Q.H W</td>
<td>2014–2015</td>
<td>1</td>
<td>1109</td>
<td>ST.</td>
<td>C1</td>
<td>D</td>
<td></td>
<td>Comp</td>
<td>0.5/1.5</td>
</tr>
<tr>
<td>2019130801</td>
<td>Y.F Zh</td>
<td>2018–2019</td>
<td>2</td>
<td>1108</td>
<td>NT.</td>
<td>C3</td>
<td>D</td>
<td></td>
<td>Comp</td>
<td>0.5/1.5</td>
</tr>
<tr>
<td>2019100802</td>
<td>Q.H W</td>
<td>2018–2019</td>
<td>1</td>
<td>1107</td>
<td>CS.</td>
<td>C2</td>
<td>A</td>
<td></td>
<td>Comp</td>
<td>0.5/1.5</td>
</tr>
</tbody>
</table>

Table 3: Sample of technical capability on job position.

<table>
<thead>
<tr>
<th>J-position</th>
<th>Course1</th>
<th>Course2</th>
<th>Course3</th>
<th>Course4</th>
<th>Course5</th>
<th>Course6</th>
<th>Course7</th>
<th>Memo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web.Dev</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C6</td>
<td>C7</td>
<td>C5</td>
<td>C10</td>
<td></td>
</tr>
<tr>
<td>Arm.Dev</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C5</td>
<td>C8</td>
<td>C11</td>
<td>C15</td>
<td></td>
</tr>
<tr>
<td>Net.Eng</td>
<td>C1</td>
<td>C2</td>
<td>C4</td>
<td>C6</td>
<td>C9</td>
<td>C10</td>
<td>C13</td>
<td></td>
</tr>
<tr>
<td>Net.Secu</td>
<td>C1</td>
<td>C2</td>
<td>C4</td>
<td>C7</td>
<td>C9</td>
<td>C10</td>
<td>C16</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Sample of employment historic data.

<table>
<thead>
<tr>
<th>StuNo</th>
<th>Name</th>
<th>Grad.Y</th>
<th>Emp/Stu.</th>
<th>Major</th>
<th>Corp./Sch</th>
<th>City</th>
<th>Salary</th>
<th>Memo</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014120812</td>
<td>Q.H W</td>
<td>2016</td>
<td>Emp</td>
<td>NT.</td>
<td>N.Grid.C</td>
<td>HY</td>
<td>4000.0</td>
<td></td>
</tr>
<tr>
<td>2014120802</td>
<td>Z.Z</td>
<td>2018</td>
<td>Stu</td>
<td>CS.</td>
<td>PKU.</td>
<td>PK</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2014120826</td>
<td>Y.F Zh</td>
<td>2020</td>
<td>Emp</td>
<td>ST.</td>
<td>IBM</td>
<td>SZ</td>
<td>12000.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Sample of core curriculum for differ major.

<table>
<thead>
<tr>
<th>StuNo</th>
<th>Major</th>
<th>Curriculums</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019120801</td>
<td>ST.</td>
<td>[C1, C2, C3, C6, C7, C10, C15]</td>
</tr>
<tr>
<td>2019120802</td>
<td>ST.</td>
<td>[C1, C2, C3, C6, C8, C10, C12]</td>
</tr>
<tr>
<td>2019130801</td>
<td>NT.</td>
<td>[C1, C2, C3, C5, C8, C11, C14]</td>
</tr>
<tr>
<td>2019120803</td>
<td>ST.</td>
<td>[C1, C2, C3, C7, C10, C15]</td>
</tr>
<tr>
<td>2019130801</td>
<td>NS.</td>
<td>[C1, C2, C3, C5, C8, C11, C15]</td>
</tr>
<tr>
<td>2019100801</td>
<td>CS.</td>
<td>[C1, C2, C3, C7, C11, C13, C16]</td>
</tr>
<tr>
<td>2019100802</td>
<td>CS.</td>
<td>[C1, C2, C3, C7, C9, C13, C16]</td>
</tr>
</tbody>
</table>

Figure 5: Minicut of professional core courses.

Figure 6: The distribution of job positions’ intervals.
Generally speaking, we can predict the course performance of the new students and recommend the course to realize personalized learning or calculate the pass rate of the course according to the grade performance of the previous students, so as to realize the adjustment of teachers or syllabus.

In order to qualitatively analyze the similarity between students, due to the uncertainty of the two group data dimensions, PCC can be calculated according to the historical scores of a student in the last term and the current students’ scores to measure the similarity of their learning characteristics.

In general, to use the Pearson correlation coefficient, it is necessary to determine whether two variables are linearly correlated. In other words, the Pearson correlation coefficient can explain the degree of linear correlation between variables. If it is not clear whether it is a linear relationship, PCC is meaningless. If PCC equals 0, it is unable to show that the two variables are not related, and it may meet a more complex relationship. Therefore, before using the Pearson correlation coefficient, it is suggested to draw a scatter diagram to roughly judge whether the overall linear relationship between the two variables is reflected.

Assume that the scores of six major core courses are extracted, the two groups of data are calculated first, and their linear correlation is then obtained. Then, the PCC is calculated. The value range is $[-1,1]$, the greater the absolute value of the correlation coefficient, the more significant the correlation is, and the positive and negative correlation are respectively expressed by positive and negative values.

From Table 6, we can see that most data do not satisfy a linear relationship, but we still should calculate the correlation coefficient. All individuals are used to calculate the overall Pearson correlation coefficient directly. We can use Matlab to analyze and summarize the statistics into a matrix. Figure 8 shows the distribution trend of scores and employment between a previous and newer student. The Std.D with Avg scores could satisfy the function relationship of a normal distribution with courses therefore it is in the line with the expectations of the teaching law, see Figure 8(a).

Careful observation shows that there is a certain linear relationship between the average score and the correlation rate of professional employment, see Figure 8(b).

The following matrix reflects the PCC for students’ scores between ranks.

$$
\begin{pmatrix}
1.0000 & 0.0556 & -0.3217 & 0.0546 & -0.2192 & 0.0945 \\
0.0556 & 0.0864 & 0.0316 & -0.0462 & 0.0675 & 1.0000 \\
-0.3217 & 0.0864 & -0.0719 & -0.3217 & 0.0945 \\
0.0546 & 0.0316 & 0.0316 & -0.0214 & 0.0348 & 0.0546 \\
-0.2192 & -0.0462 & -0.0214 & 0.0316 & -0.0462 & 0.0348 \\
0.0945 & 1.0000 & 0.0675 & 0.0348 & 0.0316 & 1.0000 \\
\end{pmatrix}
$$

(11)

Through PPC calculation, we can predict professional scores with previous students’ scores as individual students,
Table 6: PPC of the historic score in each course between previous and fresher students.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>72</td>
<td>65</td>
<td>78</td>
<td>66</td>
<td>74</td>
<td>76</td>
</tr>
<tr>
<td>S.E.</td>
<td>3.5</td>
<td>2.0</td>
<td>6.0</td>
<td>2.5</td>
<td>5.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Med.N</td>
<td>70</td>
<td>67</td>
<td>79</td>
<td>68</td>
<td>76</td>
<td>78</td>
</tr>
<tr>
<td>Mode</td>
<td>72.5</td>
<td>68</td>
<td>80</td>
<td>70</td>
<td>78</td>
<td>80</td>
</tr>
<tr>
<td>Std.D</td>
<td>3.087</td>
<td>5.026</td>
<td>3.615</td>
<td>5.263</td>
<td>5.033</td>
<td>5.031</td>
</tr>
<tr>
<td>Vari.</td>
<td>9.5</td>
<td>8.4</td>
<td>7.2</td>
<td>8.5</td>
<td>8.47</td>
<td>8.42</td>
</tr>
<tr>
<td>Max.V</td>
<td>92</td>
<td>86</td>
<td>95</td>
<td>89</td>
<td>92</td>
<td>90</td>
</tr>
<tr>
<td>Min.V</td>
<td>44</td>
<td>37</td>
<td>47</td>
<td>32</td>
<td>23</td>
<td>34</td>
</tr>
</tbody>
</table>

Figure 8: The distribution trend between previous and fresher students. (a) Std.D with Avg scores (b) Professional relevance with Avg scores.

Figure 9: Precision, score and professional consistency for position of prediction. (a) Prediction of scores for individual (b) Precision of prediction on individual-class. (c) Precision of prediction on professional consistency for position.
course referrals, and job positions for every portrait model. Figure 8 shows the precision and score of prediction on individual-class. According to the previous student's scores, the current score of every course is the same as prediction, see Figure 9(a). Also, we can see the precision of the individual and average score in a class for one course keep the high level, see Figure 9(b). According to the scores, the prediction of professional consistency for the position was seen in Figure 9(c). The main reason is that professional technology is decided by many courses, especially since the technical content is relatively high, such as the course on network security.

5. Conclusions
To establish the internal quality management platform in colleges or universities, implementing teaching quality evaluation and rectification systems helps meet the practical needs of the separation of management, operation and evaluation, and the transformation of functions of educational administrative departments. It also helps meets the need for vocational education to actively adapt to the new economic development, independently guarantee quality, and enhance core competitiveness. Therefore, the analysis of the data of graduating and employed graduates can directly reflect whether the teaching, decision-making, and management can meet the skill requirements of the new era.

In general, the evaluation and improvement of the teaching quality assurance system must adhere to the concept of total quality management, focus on playing the main role of quality, implement all staff, all process, and all factor evaluation and improvement, integrate the preplanning, in-process monitoring, and postimprovement, and establish a normalized information feedback analysis and improvement mechanism. Finally, the quality assurance system and working mechanism of three-year complete education programed are established.

In future work, we will devote time to improve the predict accuracy and efficiency, optimize the curriculum association classification model. Moreover, there are still several unsolved problems. The key point of the predicted accuracy mentioned here is that it is related to the accuracy of the portraits of students’ individuals, curriculum and professional portraits, and the similarity of technical requirements of positions mined from the Internet. Of course, besides that, the accuracy of the proposed model is also related to the continuous correction of the feedback model. Hence, the universal consistency between the individual and the whole for precision teaching service is a direction of our efforts in the future.

Data Availability
The data used to support the findings of this study have not been made available because of student and teachers’ private datasets including their ID and achievements.

Disclosure
The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Acknowledgments
This research was partly funded by the project of Guangdong Provincial Science and Technology department (No. KTP20200292), the Science and Technology Project of Huizhou City of Guangdong province (No.2016X0423038), the Teaching Reform Project of Huizhou University (No. SZJG2020062), and the Research Center of Big Data Application Engineering Technology of Heyuan City (2017, No.74).

References


[18] H. Barwick: The four Vs of Big Data Implementing Information Infrastructure Symposium [EB/OL], [2012-10-02].


