Applying K-Means Algorithm to Students' Online Learning Data Mining: From Perspective of "1+X" BIM Vocational Skill Level System Reform

Lili Zhang(✉), Xiaorong Chen, Dunhua Huang, Mengyuan Lin, and Tingting Liu

School of Mechanical and Electrical Engineering, Beijing Polytechnic, Beijing, China
zll0823@126.com

Abstract. Currently, vocational skills training is getting more and more attention. China is in full swing with the "1+X" BIM vocational skills level system reform. However, there is little research on student learning assessment in the context of this reform. To address these issues and gain practical insights into student learning analysis in the context of the "1+X" BIM vocational skills leveling system reform, this paper first collects an online learning dataset of vocational skills levels from a specific university in China. Second, a clustering algorithm called K-Means was used to evaluate the student learning analysis in the context of the "1+X" BIM vocational skill level system reform. Based on the obtained online learning data, we found that the optimal number of clusters was 3 (silhouette coefficient of 0.54). This led to 3 levels of student online learning evaluation: excellent (23.33%), good (53.67%) and fair (23%). Based on these results, we further propose corresponding policy recommendations. The results of this paper provide meaningful practical insights into the national policy and government strategy for reforming the "1+X" BIM vocational skill level system.

Keywords: data mining · clustering · "1+X" reform · BIM · online learning

1 Introduction

In early 2019, China's State Council released the National Implementation Plan for Vocational Education Reform. The plan calls for the launch of the "1+X" certificate system in vocational colleges and applied universities starting in 2019 [1]. The "1+X" certificate system is a powerful tool and instrument for vocational education reform [2]. It encourages students to actively obtain multiple vocational skills certificates along with academic certificates [3]. The system can effectively promote the reform of technical skills training and evaluation mode, and effectively improve the quality of talent training [4].

Since the implementation of "1+X" BIM certificate examination on November 21, 2019, in the process of organization and implementation, how to better mobilize students to actively participate in the examination, improve the passing rate of students and promote the improvement of students' vocational skills, there are the following problems:

V. Balakrishnan et al. (Eds.): ICMEIM 2022, AHSSEH 6, pp. 441–449, 2022.
https://doi.org/10.2991/978-94-6463-044-2_56

The teaching content is disconnected from the actual knowledge and skills, the enthusiasm and initiative of students to participate in BIM certificate forensics examination is not high, the teaching resources of teachers in the process of guiding students in the forensics examination are insufficient, and the skills of teachers need to be improved, leading to the poor BIM certificate passing rate [5]. In view of these problems, the reasons are as follows:

- **Students' awareness and participation of "1+X" BIM certificate need to be improved:** Based on the questionnaire survey of students from relevant majors in our school about their thoughts on taking the "1+X" BIM certificate examination (as shown in Fig. 1), it can be seen that students are not very interested and active in taking the examination, and their participation needs to be improved. The reasons are as follows: on the one hand, the "1+X" certificate system is still in the pilot stage, the participation of enterprises is not enough, the supervision, management and service mechanism of vocational skill level certificate and training evaluation organization still needs to be improved, the system and framework of certificate recognition needs to be further defined, and the social recognition and gold content of "1+X" BIM certificate needs to be improved; On the other hand, as the "1+X" certificate system is a new measure of national vocational education reform, colleges and universities are in the exploratory stage of personnel training policy implementation under the "1+X" system, resulting in students' insufficient understanding of "1+X" BIM certificate and low awareness. Lack of effective system restriction, in the practice process, even if take measures such as after-class guidance, check the missing, still only a small number of students actively participate in.
- **Students' knowledge transfer ability needs to be strengthened**: According to the analysis of learning situation and the characteristics of vocational college students, most of the students' learning and thinking consciousness, initiative and enthusiasm need to be improved, self-learning ability needs to be strengthened, and their knowledge and skills are more from teachers' classroom indoctrination, because the classroom teaching time is limited, and students can't pressure themselves in their spare time. Therefore, some students can only complete the corresponding tasks according to the task list given by the teacher, which cannot be flexibly applied to the actual project, and the transfer ability of knowledge into skills needs to be strengthened.
- **The professional ability of teachers needs to be improved urgently**: Most of the teachers in vocational colleges are fresh graduates, so there is a certain gap between the teaching content and the actual job needs of the enterprise. Besides, the daily teaching tasks of teachers are heavy, and the time for in-depth study and research in the enterprise is limited, so it is difficult to effectively connect the teaching content with the enterprise standard. And BIM certificate is one of the first five pilot areas under the "1+X" certificate system, and its implementation is a process from scratch, so its curriculum setting, curriculum standards, curriculum resources, teaching content

and so on need to be improved. In addition, BIM technology runs through the whole life cycle of the actual project, covering a wide range of professional areas, requiring the coordination of multiple majors, which has high requirements for the professional diversification of BIM technology teachers, and the professional ability of teachers urgently needs to be improved.

Moreover, in recent years, COVID-19 has posed a more serious challenge to education, and online education has become popular worldwide. In this context, students' online learning performance is even more important for achieving the "1+X" educational goals. Therefore, this paper uses a K-Means clustering method to analyze students' online learning performance under "1+X" vocational skills training based on the collected students' online learning dataset, and proposes policy recommendations based on the results obtained. The results of this paper have important practical implications for higher education institutions and "1+X" vocational skills training programs [6].

The reset of this paper is organized as follows: Sect. 2 provides a detailed discussion of the methodology of this paper. The obtained results are discussed in Sect. 3. Finally, Sect. 4 summarizes the full paper and suggests future research directions.

2 Methods and Materials

2.1 Students' Online Learning Data

This paper collects the learning dataset of 300 students in a university's "1+X" vocational skills training e-course. The dataset consists of three parts: course audio and video completion, chapter quiz results, and final exam results, as shown in Table 1.

Further, we show the statistical description of the obtained data in Table 2.

It is clear that the data obtained in this paper differ significantly due to the variables having different weights. However, these differences may cause large performance differences for our method. To obtain more reasonable results, later, we will first preprocess the data in order to normalize them to the [0,1] interval, as shown in Eq. (1).

$$x_{ij}^{'} = \frac{\max\{x_{ij}, ..., x_{nj}\}}{\max\{x_{ij}, ..., x_{nj}\} - \min\{x_{ij}, ..., x_{nj}\}}, \tag{1}$$

where x_{ij} represents the value of the j-th indictors of the i-th sample.

Table 1. Data collected in this paper

Meaning	Symbols
course audio and video completion	x_1
chapter quiz results	x_2
final exam results	x_3

Table 2. Statistical description of the obtained data

	x_1	x_2	x_3
Min	0.00	0.00	0.00
Mean	57.19	2.62	4.48
Median	75.00	2.85	4.80
Max	75.00	5.00	20.00
Std	30.76	1.82	4.07

2.2 K-Means Clustering Method

K-Means algorithm is a simple but powerful clustering algorithm [7]. It using the distance as the metric and given K classes in the dataset, calculate the distance mean, giving the initial centroid, with each class described by the centroid. In this paper, we specify the distance measure as the Euclidean distance, as shown in Eq. (2):

$$dis = \sum_{k=1}^{k}\sum_{i=1}^{n}||(x_i - u_k)||^2, \tag{2}$$

where k represents K cluster centers, u_k represents the k-th center, and x_i represents the i-th point in the data set. Usually, the solution to u_k in Eq. (2) can be expressed as follows:

$$\begin{aligned}\frac{\partial}{\partial u_k} &= \frac{\partial}{\partial u_k}\sum_{k=1}^{k}\sum_{i=1}^{n}(x_i - u_k)^2 \\ &= \sum_{k=1}^{k}\sum_{i=1}^{n}\frac{\partial}{\partial u_k}(x_i - u_k)^2 \\ &= \sum_{i=1}^{n}2(x_i - u_k).\end{aligned} \tag{3}$$

If set Eq. (3) equals to 0, it's easy to obtain $u_k = \frac{1}{n}\sum_{i=1}^{n} x_i$.

As known to all, the K in K-Means algorithm is a pre-determined parameter. And it's greatly affecting the performance of clustering results. Thus, before we get start to clustering, we first find the optimal K value via silhouette coefficient via Algorithm 1.

In Algorithm 1, $a(i)$ is the distance between sample i and other samples in the cluster, and $b(i)$ is the average distance between sample i and the samples in the nearest cluster.

Algorithm 1: Silhouette Coefficient

Input: dataset, X = dataset.shape
Output: S(i), k
1. for i in X, C, D:
2. $a(i) = \frac{\sum_n C_n - i}{n}$; $b(i) = \frac{\sum_n D_n - i}{n}$;
3. for $a(i) \to \min, i \in C; b(i) \to \max, i \notin D, do$
4. $s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$;
5. $if\ a(i) < b(i), s(i) = 1 - \frac{a(i)}{b(i)}$;
6. $if\ a(i) = b(i), s(i) = 0$;
7. $if\ a(i) > b(i), s(i) = \frac{b(i)}{a(i) - 1}$;
8. for k=2,3,4,5,6,7,8,9,10 do
9. labels=KMeans(n_clusters=k).fit(x).labels_;
10. return S(i), k;

3 Results and Discussions

3.1 Pre-processing

We first normalize the obtained data to the [0,1] interval to eliminate the order-of-magnitude inconsistency problem between the data and to facilitate further clustering work. The statistical description of the normalized data is shown in Table 3.

3.2 Silhouette Coefficient Results

Then, we input the above normalized data set into the K-Means algorithm. In this step, we used a step as shown in Algorithm 1, specifying positive integers with K values between [2, 10].

We obtain the values of the silhouette coefficients once during each iteration of K. Finally, we plot all the changes of the silhouette coefficients when K takes a positive integer between [2, 10], as shown in Fig. 1.

It is clear that the optimal K value (number of clusters) here is 3. Therefore, later, we specify K = 3 for the K-Means algorithm.

3.3 Clustering Results

After the above processing steps, we can further obtain 3 groups of student groups as shown in Table 4. It is clear from Table 4 that students were significantly classified into three categories of academic performance: Excellent, Good, and Fair.

Table 3. Statistical description of the pre-processed data

	x_1	x_2	x_3
Min	0.00	0.00	0.00
Median	1.00	0.57	0.24
Mean	0.76	0.52	0.22
Max	1.00	1.00	1.00
Std	0.41	0.36	0.20

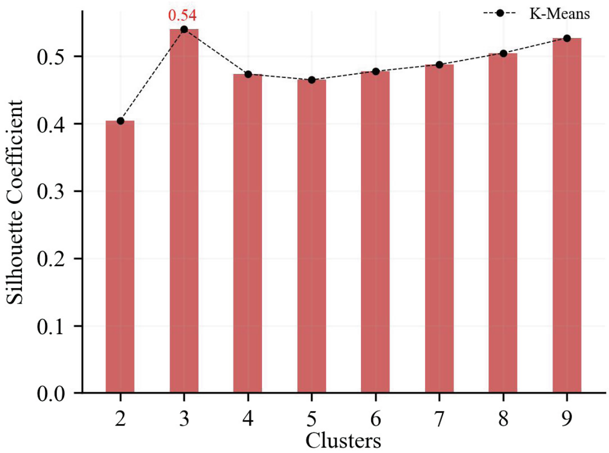

Fig. 1. The silhouette coefficient when k is between [2, 10]

Table 4. CLUSTERING RESULTS

labels	**Cluster 1**	**Cluster 2**	**Cluster 3**
x_1	0.03	0.97	1.00
x_2	0.03	0.68	0.65
x_3	0.01	0.23	0.42
Count	69	161	70
Performance	Excellent	Good	Fair

It can be found that the number of students evaluated as excellent is 70 (23.33%); the number of students evaluated as good is 161 (53.67%); and the number of students evaluated as average is 69 (23%). At the same time, all three indicators of students who rated as average were close to 0, indicating that their participation in the course was very low. Based on these observations, in the next subsection, we further propose corresponding policy suggestions.

3.4 Policy Suggestions

Based on the above research results, we further propose the following policy suggestions:

1. Improve student awareness and social value of "1+X" BIM certificate. The recognition of enterprises is one of the important factors determining the gold content of "1+X" BIM certificate. The government can actively create conditions to guide powerful enterprises to actively become the makers and influentials of "1+X" BIM certificate standards, and encourage enterprise technical personnel to participate in the development of "1+X" BIM certificate assessment standards and talent training programs. Make the vocational skills of students match the talent needs of enterprises, shorten the time of induction training, and save time and cost for enterprises. The school and enterprises are closely connected to build employment and entrepreneurship platform for the students who have obtained "1+X" BIM certificate, promote the in-depth participation of enterprises in collaborative education, and improve the gold content and social recognition of "1+X" BIM certificate.
2. Reconstruct the training system of curriculum and certificate integration talents. School level overall planning shall be formulated by the secondary school of "1" and "X" certificate (BIM) fusion depth of talent cultivation system, play a "1+X" certificate of BIM to guide teaching, skill level certificate request into the talent training scheme, through collaboration between colleges, curriculum setting and the corresponding standards for professional fusion penetration, implementation class accommodation.
3. Strengthen the training of teachers with dual qualifications. Based on BIM certificate examination standard and expert consulting enterprises, combining with the network resource, integration of teaching content, enterprise typical task module is blended in among them, establishment situation to combine theory with practice, the formation of BIM primary modeling, BIM intermediate technology application and BIM module courses, training and other related theory and practice to develop curriculum standards. Modify and improve the talent training program, integrate the EXEMPTION mechanism of BIM certificate-related courses, formulate the selection mechanism of excellent talents, absorb them to join the scientific research team, and form a three-level progressive talent training mode of primary, middle and senior skills. In-depth enterprise research, combined with "1+X" building Information Model (BIM) vocational skill level certificate evaluation outline, the ASSESSMENT standards of BIM level certificate are summarized and condensed into teaching knowledge points, forming BIM primary modeling, BIM intermediate technology application and BIM practical training and other relevant theoretical and practical module courses, and developing the course standards. Develop a series of integrated teaching materials and problem sets corresponding to the knowledge points that are compatible with BIM certificate standards. Using the network online teaching platform, create BIM technology online courses, develop the question bank and automatic scoring system for simulating BIM certificate examination, self-test BIM certificate pass rate, and formulate corresponding improvement measures according to the self-test results. Formulate a teacher growth assessment mechanism, encourage teachers to improve their technical skills through part-time training, training and further study, project

practice on the scientific research platform, and obtaining BIM trainer certificate, etc., and build a team of teachers with "double teacher quality". Adopt online and offline mixed learning quantity teaching method, encourage students to study independently and learn from each other through incentive measures, and improve the pass rate of obtaining evidence. Finally, teachers are evaluated by the degree of integration of teaching content and curriculum objectives, the updating and maintenance of curriculum network resources and the timeliness of evaluation feedback, and the cutting-edge of teaching content and the intersection of knowledge.

4 Conclusions

Currently, as vocational skills training getting more and more attention. China is in full swing with the "1+X" BIM vocational skills level system reform.

However, there is little research on student learning assessment in the context of this reform. To address these issues and gain practical insights into student learning analysis in the context of the "1+X" BIM vocational skills leveling system reform, this paper first collects an online learning dataset of vocational skills levels from a specific university in China. Second, a clustering algorithm called K-Means was used to evaluate the student learning analysis in the context of the "1+X" BIM vocational skill level system reform.

Based on the obtained online learning data, we found that the optimal number of clusters was 3 (silhouette coefficient of 0.54). This led to 3 levels of student online learning evaluation: excellent (23.33%), good (53.67%) and fair (23%). Based on these results, we further propose corresponding policy recommendations. The results of this paper provide meaningful practical insights into the national policy and government strategy for reforming the "1+X" BIM vocational skill level system.

Acknowledgment. Beijing Education Science Planning Project (priority attention): Research on skills evaluation method in the construction of "1+X certificate" system (Project No. 2020Z190-SST); Research on reconstruction of module curriculum System of Higher Vocational Colleges based on BIM Certificate Grade Standard (Project No. 2020Z183-SXZ).

References

1. S. Sun, Some Understanding of 1+X Certificate System, China Vocational and Technical Education, vol. 7, pp. 72–76, 2019.
2. H. Dai, The Significance of Implementing 1+X Certificate System in Higher Vocational Colleges and Analysis of Problems Needing Attention, Education Modernization, vol 81, pp. 101–201, 2019.
3. K. Wu, Research and practice of BIM technical personnel Training Mode under the background of 1+X certificate System Pilot, vol 27, pp. 13–16, 2019.
4. J. Du, Construction and Practice of Professional Skills Evaluation System for Local Normal University students, Journal of Higher Education, vol 19, pp. 153–155.
5. Y. Zhao, The Influence of "1+X" Certificate System on Higher Vocational Education, Vocational Education Communication, p. 235, 2019.

6. L. Jin, Construction of Double-qualified Teachers in Higher Vocational Colleges Based on Competency Theory, Education and Career, vol 24, pp. 53–56, 2019.
7. X. Ran, X. Zhou, M. Lei, W. Tepsan, W. Deng, A novel k-means clustering algorithm with a noise algorithm for capturing urban hotspots. Applied Sciences, vol. 11(23), pp. 11202., 2021.