Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 14 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i14.41921

PAPER

Process Evaluation for Diversified Academic Assessment Mechanism in Higher Education Institutions by Use of Data Mining

Heng Liang(✉)

School of Information Engineering, Xuchang University, Xuchang, China

liangheng@xcu.edu.cn

ABSTRACT

A diversified academic assessment mechanism can effectively improve students' learning motivation, make up for the possible blind spots of a single assessment method, and better guide students' learning and teachers' teaching. Using data mining methods to process evaluation data for diversified academic assessment mechanisms in colleges and universities can discover patterns in students' learning, find key factors affecting academic performance, and provide a basis for teaching reform. Most of the current process evaluation data mining methods focus on hard skills, such as academic performance and classroom participation, but it is difficult to evaluate soft skills such as critical thinking and teamwork. To this end, this paper studies the process evaluation data mining methods for a diversified academic assessment mechanism in colleges and universities. It constructs an indicator system for process evaluation of diversified academic assessment mechanism in colleges and universities, gives a quantitative method for indicators, and performs fuzzy comprehensive evaluation based on AHP-entropy weight method. For the evaluation of text-based indicators, a consistency training method is introduced to train the process evaluation correlation mining model using a large amount of unlabeled process evaluation examples, which effectively solves the problems of lack of labeled data, high labeling cost, and changes in data distribution, and improves the performance and availability of the model. The experimental results verify the effectiveness of the proposed method.

KEYWORDS

college assessment mechanism, diversified academic assessment, process evaluation, data mining

1 INTRODUCTION

With the modernization of education and the continuous development of information technology, people have begun to question traditional educational evaluation methods, especially in higher education environments. A single summative assessment method is found to be inadequate to comprehensively evaluate students'

Liang, H. (2023). Process Evaluation for Diversified Academic Assessment Mechanism in Higher Education Institutions by Use of Data Mining. *International Journal of Emerging Technologies in Learning (iJET)*, 18(14), pp. 200–214. https://doi.org/10.3991/ijet.v18i14.41921

Article submitted 2023-04-03. Resubmitted 2023-06-02. Final acceptance 2023-06-03. Final version published as submitted by the authors.

© 2023 by the authors of this article. Published under CC-BY.

learning process and academic development [1–6]. More and more research shows that diversified academic assessment mechanisms and process evaluation models can better reflect students' comprehensive development and learning depth [7–11]. At the same time, the development of big data and data mining technology makes it possible to track and analyze students' learning process in detail over the long term [12–17]. Diversified academic assessment mechanisms can effectively improve students' learning motivation, make up for the possible blind spots of a single assessment method, and better guide students' learning and teachers' teaching [18, 19]. Process evaluation can record the student's learning process and construct the student's long-term learning trajectory to provide a more comprehensive perspective on the student's learning growth. Using data mining methods to process evaluation data for diversified academic assessment mechanisms in colleges and universities can discover patterns in students' learning, find key factors affecting academic performance, and provide a basis for teaching reform.

Multiple choice tests with binary scoring are one of the most commonly used assessment methods in undergraduate education. Determining students' views on different types of choice item test formats is important for effective assessment. Williams et al. [20] compared two choice item test formats used in a required second-year chemistry course: (i) Immediate Feedback Assessment Technique (IFAT®) and (ii) Personal Partial Credit (PPA). Both test methods allow partial credit but only IFAT® provides students with immediate feedback on their responses. MacDonald and Rozaklis [21] evaluated an institution's efforts to overcome this experience gap by providing students with opportunities to participate in three authentic user experience client projects. Surveys of students and clients over four academic years provided a set of lessons learned and recommended practices for incorporating project-based learning opportunities into UX curricula. Given the current theoretical and empirical support for student-generated question learning methods and the advantageous characteristics of web technologies, several online student-generated question learning systems with peer assessment components have been developed. Nevertheless, all existing systems are limited in the types of communication modes they allow for peer assessment. The online discourse experience as well as the quantity and quality of interactions may vary depending on the particular interaction modes students are exposed to. Due to this, as well as the possibility and potential ideal in diverse learning spaces at various stages of learning and instruction, academic evaluation data often comes from multiple sources, including but not limited to teacher scores, online learning platforms, self-evaluation, etc. The quality, completeness and consistency of these data may be problematic, affecting the results of data mining. How to translate the results of data mining into operable teaching strategies and methods to truly serve teaching practice is an important challenge being faced at present. Most of the current process evaluation data mining methods focus on hard skills, such as academic performance and classroom participation, but it is difficult to evaluate soft skills such as critical thinking and teamwork. To this end, this paper studies the process evaluation data mining methods for diversified academic assessment mechanism in colleges and universities. Firstly, in section 2, an indicator system for process evaluation of diversified academic assessment mechanism in colleges and universities is constructed, a quantitative method for indicators is given, and fuzzy comprehensive evaluation is performed based on AHP-entropy weight method. In section 3, for the evaluation of text-based indicators, a consistency training method is introduced to train the process evaluation correlation mining model using a large amount of unlabeled process evaluation examples, which effectively solves the problems of lack of labeled data, high labeling cost, and changes in data distribution, etc., and improves the performance and availability of the model. The experimental results verify the effectiveness of the proposed method.

2 QUANTITATIVE METHODS

The process evaluation of the diversified academic assessment mechanism in colleges and universities can be divided into quantifiable indicators and non-quantifiable indicators. This paper constructs the following quantifiable process evaluation indicator system for diversified academic assessment mechanism in colleges and universities.

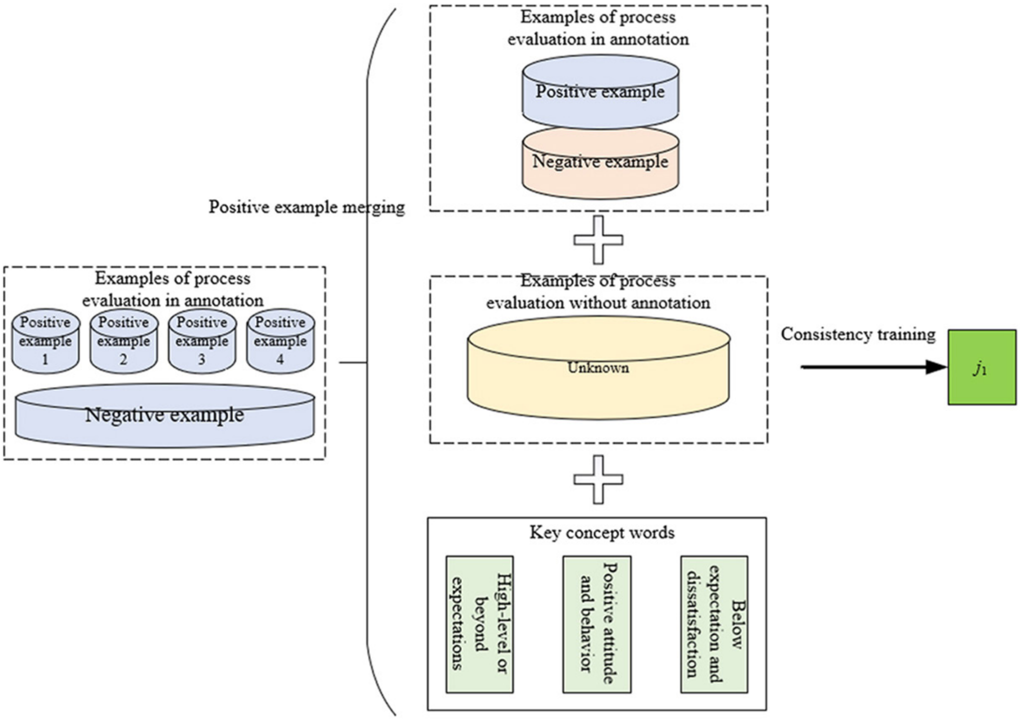

Fig. 1. Flowchart of positive/negative example recognition model

The first-level indicators include: 1. Academic performance; 2. Learning skills; 3. Participation; 4. Innovation ability; 5. Teamwork. Each first-level indicator can be divided into several second-level indicators. Specific second-level indicators under academic performance: mid-term and final exam scores; course paper and report scores; experimental and practical scores. Second-level indicators under learning skills: self-learning ability; problem-solving ability; critical thinking. Second-level indicators under participation: classroom participation; participation in club activities; participation in school activities. Second-level indicators under innovation ability: innovative thinking ability; innovative projects or scientific research achievements. Second-level indicators under teamwork: results of team cooperation projects; peer evaluation. Figure 1 shows the constructed evaluation indicator system.

For the quantification method of each indicator, the following methods can be referred to: 1) Academic performance: use scores, percentages or grading systems to quantify. 2) Learning skills: can be quantified through specific assessment tools and tests. 3) Participation: can be quantified by the number or duration of activity participation. 4) Innovation ability: can be quantified by the number of innovative projects, the impact factor of scientific research achievements or the number of citations, etc. 5) Teamwork: can be quantified through peer evaluation or teacher evaluation, or reflected through the outcomes of team projects.

The fuzzy comprehensive evaluation method based on *AHP*-entropy weight method combines analytic hierarchy process (*AHP*) and entropy weight method.

It aims to comprehensively and scientifically evaluate diversified assessment mechanisms through quantification and fuzzification. This method can comprehensively consider multiple evaluation indicators and consider the relationship between these indicators. At the same time, it allows the evaluation results to have a certain fuzziness, which is more in line with the actual situation.

The specific steps of this evaluation method are as follows:

Step 1: Determine the first-level indicators and second-level indicators according to the actual needs and construct the evaluation indicator system. For example, in the diversified academic assessment mechanism of colleges and universities, the first-level indicators include academic performance, learning skills, participation, etc., and each first-level indicator has corresponding second-level indicators. i evaluation indicators are represented by $G = (g_1, g_2, ..., g_i)$;

Step 2: Define the judgment levels and corresponding fuzzy subsets for each indicator. For example, the judgment levels for academic performance can be set as excellent, good, average and poor, and each level corresponds to a fuzzy subset. The judgment set L of process evaluation is represented by $L = (l_1, l_2, ..., l_j)$, and each level in the set corresponds to a fuzzy subset;

Step 3: Establish a fuzzy relation matrix according to the judgment levels of each indicator and actual data. Each element in the matrix represents the degree of membership of the corresponding indicator in the corresponding judgment level. The membership degree matrix is given by the following formula:

$$F = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1j} \\ f_{21} & f_{22} & \cdots & f_{2j} \\ \cdots & \cdots & \cdots & \cdots \\ f_{i1} & f_{i2} & \cdots & f_{ij} \end{bmatrix} \tag{1}$$

In the fuzzy relation matrix F, the element f_{nm} in the nth row and mth column represents the degree of membership of the evaluated student to the fuzzy subset l_m from the perspective of the process evaluation indicator g_n.

Step 4: Determine the weight vector of the process evaluation indicators. Suppose the final combined weight is represented by $Y^T = (Y_1, Y_2, ..., Y_n)^T$, the combined weight calculation formula of *AHP*-entropy value method is given by the following formula:

$$Y_m = \frac{l_m z_m}{\sum_{m=1}^{j} l_m z_m} \tag{2}$$

Step 5: According to the fuzzy relation matrix and weight vector, obtain the fuzzy comprehensive evaluation result through fuzzy operation, and then determine the final evaluation result based on the principle of maximum membership degree of the fuzzy comprehensive evaluation result. Let the combined weight matrix of the four second-level indicators under the first-level indicator Z_1 be represented by $Z_1 = [z_1, z_2, z_3, z_4]$, then the fuzzy comprehensive evaluation result of Z_1 can be obtained by the following formula:

$$P_1 = Z_1 F_1 \tag{3}$$

Similarly, the fuzzy comprehensive evaluation results of all first-level indicators can be obtained. Based on all the results, a first-level comprehensive evaluation matrix $P = [P_1 P_2 P_3]^T$ can be constructed. Multiplying P by the weight allocation matrix Q of the first-level indicators can obtain the final evaluation result.

$$R = QP \tag{4}$$

3 PROCESS EVALUATION CORRELATION MINING BASED ON TEXT PROCESSING

In the process evaluation of diversified academic assessment mechanisms, some important indicators may be difficult to quantify directly, but they still have a significant impact on the evaluation results. These indicators mainly involve students' learning behaviors, attitudes and soft skills, which usually need to be analyzed and evaluated through text data. Specifically, including learning attitude, course feedback, personal development goals, communication and expression ability, communication and interpersonal ability. The evaluation of these text indicators relies on text analysis techniques. Through text processing technology, valuable information and patterns can be extracted from these text data to understand and evaluate students' learning more comprehensively.

At the same time, there may be complex correlations between different evaluation indicators. For example, students' learning attitudes may affect their learning skills and academic performance. Students' communication and expression abilities may affect their teamwork and innovation abilities. Through text processing technology, these correlations can be discovered to improve the accuracy and effectiveness of the evaluation.

In actual situations, obtaining labeled process evaluation examples often requires a lot of manpower and time. Therefore, available labeled data may be very limited. In contrast, unlabeled process evaluation examples are usually easier to obtain and larger in number. Consistency training can utilize these unlabeled data to increase the amount of training data and improve model performance. This paper introduces a consistency training method to train the process evaluation correlation mining model using a large number of unlabeled process evaluation examples, which effectively solves the problems of lack of labeled data, high labeling cost, changes in data distribution, etc., and improves the performance and availability of the model.

A batch of labeled process evaluation examples is represented by $C = \{(c_u, t_u): u \in (1, ..., U)\}$, where U is the number of process evaluation examples in the batch. c_u is the uth labeled process evaluation example in the batch, and t_u is the corresponding label of c_u, representing one of five possible correlation prediction results: positive correlation, negative correlation, nonlinear correlation, indirect correlation, and complex correlation, represented by *ME*, *AD*, *IN*, *EF* and *NE* respectively. A batch of unlabeled process evaluation examples is represented by $A = \{a_u: u \in (1, ..., \omega U)\}$, where ωU is the number of process evaluation examples in the batch. ω is a hyperparameter used to represent that the number of unlabeled process evaluation examples A in each batch is ω times that of labeled process evaluation examples C. $t_j(g|c)$ represents the prediction of the correlation category distribution of the input process evaluation example c by the process evaluation correlation mining model j. The values in the vector represent the probabilities of c being predicted as a correlation category by model j. $Q(\cdot)$ represents advanced data enhancement operations to obtain the corresponding process evaluation examples of c after data enhancement.

The supervised loss SS can be calculated using the prediction result $t_j(g|c_u)$ of model j on the labeled c_u and its corresponding label t_u. Assuming the cross-entropy loss function is represented by GH, the calculation formula is as follows:

$$SS = \frac{1}{U}\sum_{u=i}^{U} GH(t_u, t_j(g|c_u)) \tag{5}$$

The unsupervised loss V_A can be calculated using the prediction results $t_j(g|c_u)$ and $t_j(g|Q(a_u))$ of model j on the unlabeled process evaluation example a_u and its enhanced process evaluation example $Q(a_u)$. The calculation formula is as follows:

$$V_A = \frac{1}{\omega U}\sum_{u=i}^{\omega U} GH(t_j(g|a_u), t_j(g|Q(a_u))) \tag{6}$$

Assuming the weight factor is represented by ϕ, the final semi-supervised loss calculation formula is as follows:

$$V_R = SS + \varphi V_A \tag{7}$$

In the actual process evaluation dataset, the imbalance of sample categories is a common problem, that is, the number of samples of some categories may be much larger than that of other categories. This imbalance may cause the model to overfit the categories with more samples during training and ignore the categories with fewer samples, thereby affecting the performance of the model.

Based on unsupervised data augmentation methods and a two-stage strategy of process evaluation core viewpoint words, the performance of semi-supervised methods can be further improved. Through this two-stage strategy, samples of all categories can be balanced first, and then consistency training can be performed. This can improve the imbalance of categories to some extent and improve the performance of the model. In process evaluation, core viewpoint words are an important source of information. By incorporating core viewpoint words, the content of process evaluation can be better understood and expressed, thereby improving the understanding and prediction ability of the model. Figure 2 shows the flowchart of the positive and negative example identification model.

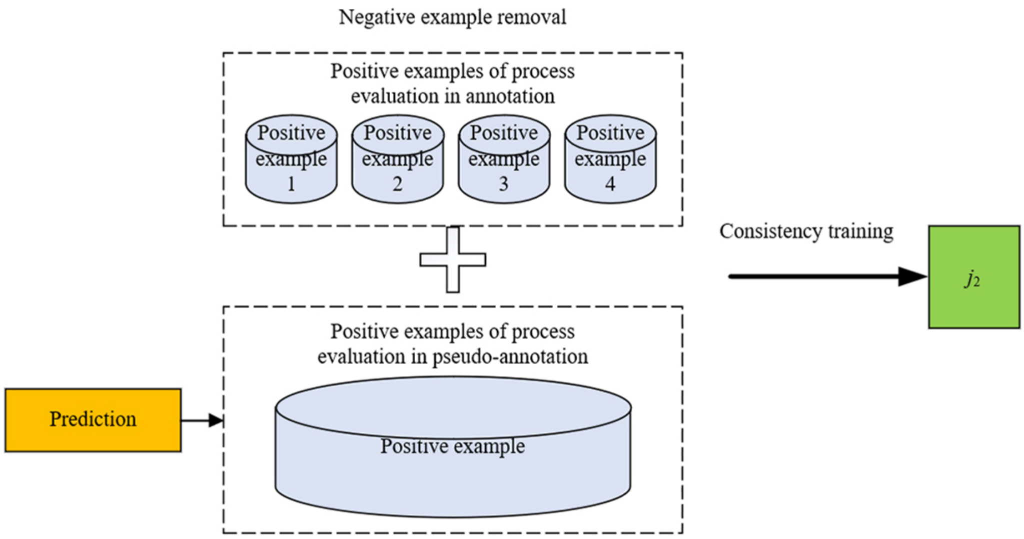

Fig. 2. Flowchart of positive example discrimination model

Text-based process evaluation samples of diversified academic assessment mechanisms in colleges and universities can be divided according to evaluation results and core viewpoint words. In this case, positive examples can be defined as process evaluations that express positive evaluations or positive emotions, while negative examples can be defined as process evaluations that express negative evaluations or negative emotions. This paper provides a possible division method of 4 positive examples and 1 negative example: 1) Positive example 1: Excellent evaluation – This type of evaluation indicates that the student's performance in the assessment process has reached a high level or exceeded expectations. 2) Positive example 2: Significant progress evaluation – This type of evaluation indicates that the student has made obvious progress or improvement in the assessment process. 3) Positive example 3: Active Participation Evaluation – This type of evaluation indicates that the student actively participated in the assessment process and showed a positive attitude and behavior. 4) Positive example 4: Good reviews of courses or teachers – This type of evaluation indicates the student's positive evaluation of courses or teachers. 5) Negative example: Negative evaluation – This type of evaluation indicates that the student's performance in the assessment process did not meet expectations, or indicates dissatisfaction with the course or teacher.

The two core viewpoint words corresponding to each process evaluation example are concatenated in the form of [*SEP*]*ideaword1*[*SEP*]*ideaword2*[*SEP*] behind the corresponding process evaluation example. Then, after word segmentation, each core viewpoint word will be segmented into smaller subwords *TO*. Then all the subwords are input into *BioBERT*, and all the subwords are converted into real-value vectors $d_n \in F^{qp}$. The preprocessed process evaluation example can be expressed as $K = [d_1, ..., d_j, d_{[KH]}, d_{q1}, d_{[KH]}, d_{q2}, d_{[KH]}] \in F^{dj*qp}$. All subwords d_n in K integrate the surrounding subword information through one-dimensional convolution to ensure a richer semantic representation of the process evaluation example.

A weight vector $W_n \in^{Fs*qp}$ is introduced to represent s subword vectors around the subword d_n:

$$w_n = \left[d_{n-(s-1)/2}, ..., d_n, ..., d_{n+(s-1)/2}\right] \tag{8}$$

Furthermore, m convolution kernels will convolve each w_n. Assuming that the element-wise multiplication and then summing operation of two matrices is represented by ⊗, the bias term of convolution is represented by u^{PZX}, and the activation function of convolution is represented by r(). The convolution weight tensor containing all convolution kernels is represented by $z^{PZX} \in F^{qg*s*qp}$, the number of convolution kernels is represented by q^g, and the window size of convolution is represented by s. The weight parameter corresponding to the j-th convolution kernel in z^{PZX} is represented by $z^{PZX}{}_m$. By calculation, m convolution results can be obtained:

$$j_{n,m} = r(Z_m^{PZX} \otimes w_n + u^{PZX}) \tag{9}$$

The representation vector of the entire process evaluation example can be obtained by performing maximum pooling operation on the convolution results of each convolution kernel m on each subword n:

$$e^{PZX} = \underset{m}{MAX}\, j_{n,m} \tag{10}$$

A new convolution layer weight vector $z^{XL} \in F^{qg}{}_{DE*s*q}$ and bias term y^{XL} are defined for the description information of core viewpoint words. Similarly, the maximum pooling operation is performed on the convolution results of each subword to finally obtain the representation vectors e^{XL1} and e^{XL2} of the description information of the core viewpoint words.

This article defines the vector corresponding to the i-th academic performance field of the diversified assessment mechanism of the core viewpoint word entity in the process evaluation example as e_m, and the set of all vectors related to the n-th academic performance field of the core viewpoint word entity as I_n. Assuming the weight parameters and biases in graph convolution are represented by Z_{QZ} and u_{QZ}, and the *ReLU* activation function is represented by $r(\cdot)$, e_n will be updated by convolving in v steps:

In the actual process evaluation dataset, the imbalance of sample categories is a common problem, that is, the number of samples of some categories may be much larger than that of other categories. This imbalance may cause the model to overfit the categories with more samples during training and ignore the categories with fewer samples, thereby affecting the performance of the model.

$$e_n^v = e_n^{v-1} + \sum_{m \in I_n} r(Z_{HID}^{v-1} e_m^{v-1} + u_{HID}^{v-1}) \tag{11}$$

After updating, the diversified assessment mechanism vector of the core viewpoint word entity is obtained by summing all its academic performance field vectors. Assuming the number of academic performance fields contained in the diversified assessment mechanism of the core viewpoint word is represented by J, the weight and bias parameters of the linear layer are represented by Z_{QZ} and u_{QZ}, then:

$$e^{JG} = r(Z_{QZ} \sum_{n}^{J} e_n^V + u_{QZ}) \tag{12}$$

Through this step, the diversified assessment mechanism representation vectors e_{JG} and e_{JG2} of the core viewpoint word entities in the process evaluation example can be obtained.

When e^{PZX}, e^{XL1} and e^{XL1} are obtained, the prediction input vector e^{XL} combining the description information of core viewpoint words can be obtained by concatenating these three vectors:

$$e^{XL} = [e^{PZX}; e^{XL1}; e^{XL2}] \tag{13}$$

Similarly, the two diversified assessment mechanism representation vectors of the core viewpoint word entities will also be concatenated with the representation vector of the process evaluation example to obtain the fusion diversified assessment mechanism information prediction input vector e^{JG}:

$$e^{JG} = [e^{PZX}; e^{JG1}; e^{JG2}] \tag{14}$$

By fusing e^{XL} and e^{JG}, they will eventually be input into the linear prediction layer to obtain the scores k^{XL} and k^{JG} of the process evaluation example being predicted as positive and negative examples. Let $k = [k_0, k_1]$, the score of the process evaluation example being predicted as a negative example is represented by k_0, and the score of the process evaluation example being predicted as a positive example is represented by k_1:

$$k^{XL} = Z^{QF\_XL} e^{XL} \tag{15}$$

$$k^{JGG} = Z^{QF\_JG} e^{JG} \tag{16}$$

The final prediction score can be obtained by:

$$k = k^{XL} + k^{JG} \tag{17}$$

k will pass through the $Soft\max$ function to obtain the probabilities t_{j_1} of the process evaluation example being determined as positive or negative:

$$t_{j_1} = Soft\max(k) \tag{18}$$

Furthermore, using the process evaluation positive and negative example identification model j_1 trained in the first stage, pseudo-labels are given to all unlabeled process evaluation examples. And this type of example is used as unlabeled process evaluation examples in the second stage. Figure 2 shows the flowchart of the positive example discrimination model. Assuming the prediction distribution of j_1 on oy is represented by $T_{j_1}(b \mid o_y)$, the index of the larger category in the prediction distribution is represented by MAX(.), the expression for each batch is given below:

$$O_2 = o_y : y \in (1, \ldots, \omega Y), MAX(t_{j_1}(b \mid o_y)) == 1 \tag{19}$$

The labeled process evaluation examples in the second stage are all labeled positive examples. The expression for each batch is given below:

$$A_2 = \{(c_y, t_y) : y \in (1, \ldots, Y), t_y \in (ME, AD, EF, IN)\} \tag{20}$$

Bringing A_2 and O_2 into the loss function for training can obtain the desired positive example discrimination model j_2.

4 EXPERIMENTAL RESULTS AND ANALYSIS

Figure 3 shows the scores of a student in 5 semesters on 5 evaluation dimensions (Academic Performance, Learning Skills, Participation, Innovation Ability and Teamwork). Overall, the scores of the student's various evaluation dimensions show an obvious upward trend with the progress of the semester. This indicates that the student's learning ability and participation are continuously improving, and innovation ability and teamwork ability are also continuously developing. In particular, the improvement in participation and learning skills is the most significant. This shows that the multi-dimensional assessment mechanism for higher education academic performance has played a very good role in promoting the overall ability of students.

Furthermore, analyze the $F1$ values of association mining under different labeling ratios. From Table 1, it can be seen that the $F1$ values of *SVM*, *RNN*, *GNN* and the model proposed in this paper for association mining under different labeling ratios. The model proposed in this paper reaches an $F1$ value of 0.9744 when the labeling ratio is 0.5, which is the highest among the four models. This is because the model proposed in this paper combines unsupervised data enhancement and consistency training methods, allowing the model to achieve good performance even with few labeled samples. However, when the labeling ratio is 0.7, the performance of the model decreases, because too many labeled samples lead to overfitting of the model. Therefore, for the task of mining the association of process evaluation based on the multi-dimensional assessment mechanism for higher education, the model proposed in this paper can achieve the best performance within a certain range of labeling ratios (especially when the labeling ratio is 0.5). This is because the model uses unsupervised data enhancement and consistency training methods, allowing the model to achieve good performance even with a few labeled samples.

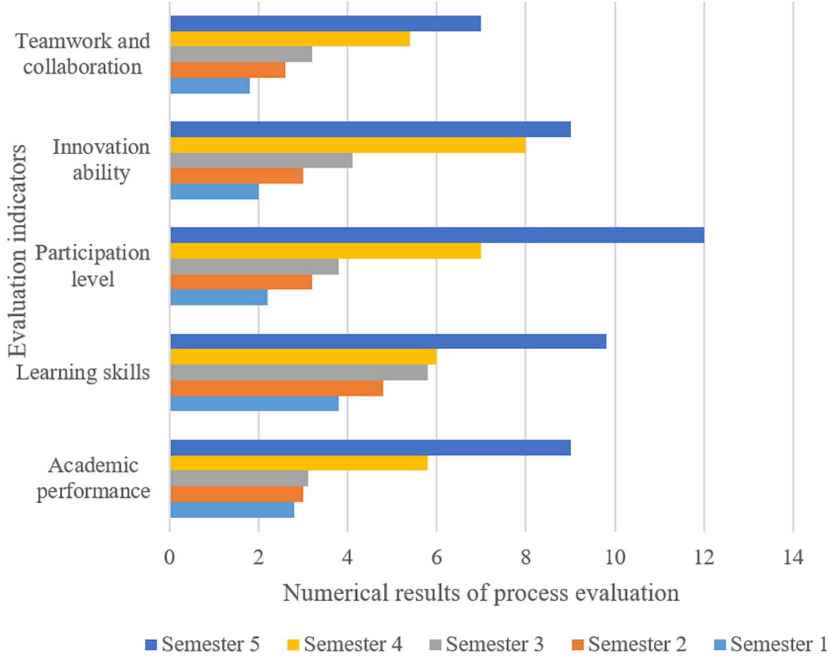

Fig. 3. Logarithmic results of students' process evaluation at different learning stages

Table 1. *F*1 values of association mining under different labeling ratios

Labeling Ratio	*SVM*	*RNN*	*GNN*	This Paper Model
0.1	0.3145	0.4714	0.7153	0.7165
0.2	0.1545	0.7151	0.7115	0.7416
0.3	0.4123	0.8441	0.7613	0.7416
0.4	0.7454	0.7564	0.8411	0.7461
0.5	0.7631	0.7361	0.715	0.9744
0.6	0.751	0.7456	0.7641	0.7646
0.7	0.7616	0.7465	0.8131	0.4156
0.8	0.4641	0.7132	0.8031	0.7416
0.9	0.841	0.4864	0.8	0.4764

Table 2 summarizes the results of ablation experiments. From the given table, it can be seen that when performing different ablation experiments on the model (that is, removing part of the model to study its effect on performance), the performance of each model in different evaluation dimensions (*Advice*, *Effect*, *Int*, *Mechanism*) and comprehensive performance (*F*1, *P*, *R*). Without the positive sample classification model, the performance of the model in each evaluation dimension and the comprehensive performance are lower than the model in this paper, indicating that the positive sample classification plays an important role in improving the performance of the model. Distinguishing different types of positive examples helps the model better understand and distinguish different categories during model training.

The model that only introduces core opinion words has decreased performance in the *Effect* and *Mechanism* dimensions as well as the comprehensive precision and recall, indicating that in addition to core opinion words, other information (such as academic performance field information) also affects the performance of the model. The model that only introduces academic performance field information has lower performance in the *Int* and *Mechanism* dimensions as well as recall than the model in this paper, because only academic performance field information lacks the information of core opinion words, which leads to the inability of the model to fully understand the evaluation content, thereby affecting the performance of the model. The model without two kinds of information has lower performance in all evaluation dimensions and comprehensive performance than the model in this paper, which further proves the importance of core opinion words and academic performance field information in improving the performance of the model.

Table 2. Comparison of model performance under different source sample sets

Sample Set Number	Evaluation Indicators	Model		
		Without Two Kinds of Information	Without Positive Example Classification Model	This Paper Model
1	*Precision*	0.971	0.955	0.978
	Recall	0.745	0.607	0.754
	*F*1	0.862	0.751	0.861
2	*Precision*	0.972	0.978	0.976
	Recall	0.756	0.801	0.751
	*F*1	0.863	0.861	0.818
3	*Precision*	0.903	0.908	0.966
	Recall	0.766	0.787	0.793
	*F*1	0.817	0.863	0.856
4	*Precision*	0.909	0.982	0.981
	Recall	0.767	0.735	0.793
	*F*1	0.864	0.801	0.808
5	*Precision*	0.977	0.971	0.901
	Recall	0.807	0.821	0.786
	*F*1	0.881	0.879	0.801

This paper classifies the data sources of process evaluation indicators. The sample set numbered 1 is mainly based on online learning platform data, the sample set numbered 2 is mainly based on teacher evaluation data, the sample set numbered 3 is mainly based on student self-evaluation data, the sample set numbered 4 is mainly based on peer evaluation data, and the sample set numbered 5 is mainly based on students' scientific research achievements or innovation project data. The table gives the comparison of model performance under different source sample sets, comparing the precision (*Precision*), recall rate (*Recall*) and *F*1 score of the without Two Kinds of Information model, without positive example classification model and the model proposed in this paper. In all data sources, the precision of the model proposed in this paper is better than or equal to the other two models. This shows that in the samples predicted to be positive by this model, the proportion of true positive samples is higher. In terms of recall rate, the performance of the model in this paper is better than or equal to the other two models in sample sets 1, 3, 4 and 5, and is slightly worse in sample set 2. This shows that in all the true positive samples, the proportion of correctly predicted samples in this model is higher, but in the sample source in sample set 2, the recall rate needs to be further improved. In the $F1$ score, the model proposed in this paper performs best in sample sets 1 and 3, and is comparable to the model without positive example classification in sample sets 2, 4 and 5. This shows that in terms of comprehensive performance, the model proposed in this paper is comparable or better than the model without positive example classification.

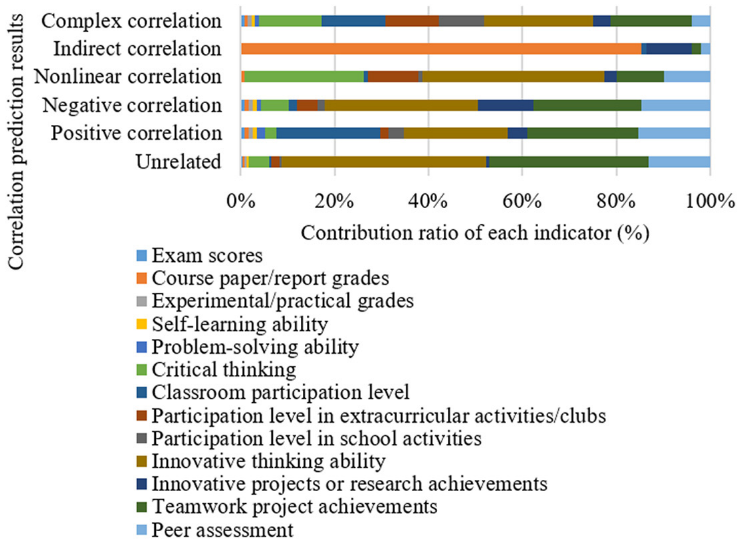

Fig. 4. Contribution ratio of each evaluation indicator to the prediction results of association

Furthermore, the contribution ratio of each evaluation indicator to the prediction results of association is analyzed. Figure 4 provides the contribution ratio of different evaluation indicators in six association predictions (no association, positive association, reverse association, nonlinear association, indirect association and complex association). It can be seen that the contribution ratio of "innovative thinking ability" in no association, positive association, reverse association and nonlinear association prediction is relatively high. This is because innovative thinking ability is a dimension with a wide coverage and far-reaching influence. There are various forms of association with many other indicators. The contribution rate of "team cooperation project achievements" in no association, positive association and reverse association is also high, because team cooperation project achievements can reflect students' performance in team cooperation, which is associated with multiple other dimensions (such as participation, learning skills, etc.). In "nonlinear association", the contribution ratio of critical thinking is the highest, reaching 28%, indicating that improving critical thinking does not always directly lead to improvement of other evaluation indicators, and there is a nonlinear or phased relationship. For "indirect association", the contribution ratio of course papers, reports and innovation projects or scientific research achievements is relatively high. This is because these two indicators reflect students' academic ability and scientific research ability more, which indirectly affect the performance of other indicators. The contribution ratio of each evaluation indicator in "complex association" is relatively high, indicating that the relationship between students' evaluation indicators is usually complex and diverse, requiring a comprehensive consideration of the impact of multiple indicators during analysis. In general, each evaluation indicator has its own unique contribution in different association predictions, indicating that evaluating students' performance requires multi-angle and multi-dimensional examination and analysis, and cannot be oversimplified to only look at one or a few indicators.

5 CONCLUSION

In the process of mining the association of process evaluation, this paper considers various factors such as core opinion words and academic performance field information. This information can help the model better understand and dig out valuable associations. The combination weights of evaluation indicators have an important impact on the results of process evaluation. Different evaluation indicators and different weight settings will lead to differences in evaluation results. Therefore, when setting weights, adjustments need to be made according to the actual situation and objectives. Different data sources of process evaluation indicators have different impacts on model performance. Therefore, in order to obtain most accurate results, it is necessary to consider obtaining data from multiple sources and consider these differences in model training and evaluation. Students may perform differently on different evaluation indicators at different learning stages. Therefore, when conducting process evaluation, students' learning stages need to be considered.

The consistency training and unsupervised data enhancement proposed in this paper are very useful tools that can effectively reduce the model's demand for labeled samples and improve the model's performance, especially in the case of imbalance sample categories. Combined with experiments, by comparing different process evaluation association mining models, it is found that the model proposed in this paper can achieve good performance in most cases, but its performance may be affected when the labeling ratio is too high or too low. This shows that although this model has its advantages, it still needs to be trained with enough labeled data to ensure its generalization ability.

6 ACKNOWLEDGMENT

This work was supported by a grant from Henan Province Educational and Teaching Plan Key Project for 2023 (Grant No.: 2023JKZD25).

7 REFERENCES

[1] Romagnoli, G., Galli, M., Mezzogori, D., Reverberi, D. (2022). A Cooperative and Competitive Serious Game for Operations and Supply Chain Management: Didactical Concept and Final Evaluation. International Journal of Online and Biomedical Engineering, 18(15): 17–30. https://doi.org/10.3991/ijoe.v18i15.35089

[2] Julhadi, Ritonga, M. (2023). Human Resource Management in Islamic Educational Institutions to Improve Competitiveness in Society 5.0 Era. International Journal of Sustainable Development and Planning, 18(2): 611–619. https://doi.org/10.18280/ijsdp.180231

[3] Doukakis, S., Vrahatis, A., Exarchos, T., Hadjinicolaou, M., Vlamos, P., Mouza, C. (2023). Design, Implementation, and Evaluation of Online Bioinformatics and Neuroinformatics Labs. International Journal of Online and Biomedical Engineering, 19(1): 21–37. https://doi.org/10.3991/ijoe.v19i01.34041

[4] Wen, J., Zhao, Y.L. (2022). An Urban and Rural Educational Resource Sharing and Exchange Platform based on Cloud Platform Access Technology. Ingénierie des Systèmes d'Information, 27(3): 515–520. https://doi.org/10.18280/isi.270320

[5] Guraliuk, A., Varava, I., Holovko, S., Shapenko, L., Oleshchenko, V. (2023). Expert Assessment of the Quality of Remote Educational Resources. International Journal of Engineering Pedagogy, 13(1): 34–44. https://doi.org/10.3991/ijep.v13i1.36121

[6] Dhar, J., Jodder, A.K. (2020). An Effective Recommendation System to Forecast the Best Educational Program Using Machine Learning Classification Algorithms. Ingénierie des Systèmes d'Information, 25(5): 559–568. https://doi.org/10.18280/isi.250502

[7] Corzo-Zavaleta, J., Yon-Alva, R., Icho-Yacupoma, J., Principe Somoza, Y., Andrade-Arenas, L., Vargas-Cuentas, N.I. (2023). Hybrid Learning in Times of Pandemic Covid-19: An Experience in a Lima University. International Journal of Engineering Pedagogy, 13(1): 65–81. https://doi.org/10.3991/ijep.v13i1.36393

[8] Ma, J. (2021). Intelligent Decision System of Higher Educational Resource Data under Artificial Intelligence Technology. International Journal of Emerging Technologies in Learning, 16(5): 130–146. https://doi.org/10.3991/ijet.v16i05.20305

[9] Dhar, J., Jodder, A.K. (2020). An Effective Recommendation System to Forecast the Best Educational Program Using Machine Learning Classification Algorithms. Ingénierie des Systèmes d'Information, 25(5): 559–568. https://doi.org/10.18280/isi.250502

[10] Pinto-Santos, A.R., George Reyes, C.E., Cortés-Peña, O.F. (2022). Training and Educational Innovation: An Evaluative Perspective of the Digital Teaching Competence. International Journal of Emerging Technologies in Learning, 17(7): 38–53. https://doi.org/10.3991/ijet.v17i07.28867

[11] Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., Wang, X. (2018). Designing Academic Writing Analytics for Civil Law Student Self-Assessment. International Journal of Artificial Intelligence in Education, 28: 1–28. https://doi.org/10.1007/s40593-016-0121-0

[12] Costa, J.M., Araújo, S., Soliman, M., Sousa, M.J. (2022). Assessment Methods Determining the Higher Education Students' Academic Success. In Advanced Virtual Environments and Education: Third International Workshop, WAVE 2021, Fortaleza, Brazil, March 21–24, 2021, Revised Selected Papers, pp. 77–87. https://doi.org/10.1007/978-3-031-07018-1_7

[13] Patel, S., Patel, A., Patel, M., Patel, U., Patel, M., Solanki, N. (2022). Practicality Assessment of a Frugal Nanoparticle Generator: Opening Doors to Nano Pharmaceuticals for the Researchers and Students of Underprivileged Academic Institutions. Current Nanoscience, 18(2): 237–246. https://doi.org/10.2174/1573413717666210309121405

[14] Baek, J., Lee, J., Choi, W. (2022). Analysis on the Characteristics of Academic Achievement of Middle School Students about 'Composition of Matter': Focusing on the Results of the National Assessment of Educational Achievement (NAEA). Journal of the Korean Chemical Society, Language: Korean, 66(2): 136–149.

[15] Qoiriah, A., Yamasari, Y., Asmunin, Nurhidayat, A.I., Harimurti, R. (2021). Exploring Automatic Assessment-Based Features for Clustering of Students' Academic Performance. In Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020) 12, 125–134. https://doi.org/10.1007/978-3-030-73689-7_13

[16] Cukierman, D.R., McGee Thompson, D., Sun, W. (2019). The Academic Enhancement Program: Assessing Programs Designed to Support Student Success. In Proceedings of the 50th ACM Technical Symposium on Computer Science Education, 686–692. https://doi.org/10.1145/3287324.3287422

[17] Alhassan, A., Zafar, B., Mueen, A. (2020). Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data. International Journal of Advanced Computer Science and Applications, 11(4): 185–194. https://doi.org/10.14569/IJACSA.2020.0110425

[18] Nistor, N., Trăuşan-Matu, Ş., Dascălu, M., Duttweiler, H., Chiru, C., Baltes, B., Smeaton, G. (2015). Finding Student-Centered Open Learning Environments on the Internet: Automated Dialogue Assessment in Academic Virtual Communities of Practice. Computers in Human Behavior, 47: 119–127. https://doi.org/10.1016/j.chb.2014.07.029

[19] Zhang, X., Cheng, H.N. (2019). Mining the Patterns of Graduate Students' Self-Regulated Learning Behaviors in a Negotiated Online Academic Reading Assessment. In Proceedings of the 2019 3rd International Conference on E-Society, E-Education and E-Technology, 109–114. https://doi.org/10.1145/3355966.3355977

[20] Williams, M., Wood, E., Arslantas, F., MacNeil, S. (2021). Examining Chemistry Students' Perceptions Toward Multiple-Choice Assessment Tools That Vary in Feedback and Partial Credit. Canadian Journal of Chemistry, 99(12): 933–941. https://doi.org/10.1139/cjc-2020-0398

[21] MacDonald, C.M., Rozaklis, L. (2017). Assessing the Implementation of Authentic, Client-Facing Student Projects in User Experience (UX) Education: Insights from Multiple Stakeholders. Proceedings of the Association for Information Science and Technology, 54(1): 268–278. https://doi.org/10.1002/pra2.2017.14505401030

8 AUTHOR

Heng Liang graduated from Henan University with a Master of Science degree. He is currently a lecturer in the College of Information Engineering at Xuchang University, where he is engaged in student management. His main research interests include ideological and political education, mental health education, and employment and entrepreneurship education for college students. He has published 11 academic papers and undertaken five research projects of various types.