Learning Process Oriented Teaching Quality Improvement

Yifan Liu¹, Jing Zhan¹²* and Xue Fan¹

¹ Department of Computer, Beijing University of Technology, Faculty of Information, Beijing, China
² Beijing Key Laboratory of Trustworthy Computing, Beijing, China
Email: zhanjing@bjut.edu.cn

Abstract. In order to get targeted improving strategies for better teaching quality, it is necessary to establish learning data model for specific courses covering the whole learning process, and accurate and stable student performance classification model accordingly. Drawing on existing learning data analysis of online teaching, this paper proposes process oriented learning feature model, student performance classification method, and corresponding improving strategies analysis. With learning data from “Computer Network Fundamentals” course, in which combined online and offline teaching methods are used, learning process oriented feature model are defined using correlation analysis and clustering methods; improved Support Vector Machine (SVM) classification method with grid search optimization are proposed to get optimized student performance classification model, which gains 8% improvement on the accuracy compared to classified SVM model. Based on experiments, the accuracy of optimized classification model is 90%, with 6% false positive rate and 11% false negatives rate, which are reasonable and shows the accuracy and stability of our model. The improving strategies analysis for different students classification are also given, which can provide strong support for teaching quality improvement.

1. Background
Since face-to-face classroom teaching is more conducive to communication between teachers and students, most of the university courses still use the traditional offline teaching and closed-book examination evaluation methods. There is no universal learning process-oriented data model for traditional teaching models. With the continuous deepening of Internet applications, the continuous expansion of online teaching, and the acquisition of data in the learning process have become easier, but how to dig out valuable learning models and improve the quality of teaching is still a problem. This paper draws on existing online learning data analysis and research, and proposes an online and offline comprehensive teaching model quality evaluation model and continuous improvement strategy automatic generation and analysis method for the learning process, so as to achieve the goal of improving teaching quality.

2. Related work
Current research methods for learning process data mainly include machine learning methods such as correlation and regression analysis, support vector machines and neural networks. Macladyen [1] used the learning process data to establish a regression analysis model for performance prediction. Li [2] and Tokan [3] used correlation analysis methods to analyze the significance of each learning behavior data, and find whether there is a significant correlation between the relevant learning behaviors in the learning process of students and the module test scores. Wang [4] conducted correlation and regression
analysis on students' learning social behavior, learning process characteristics and learning quality to predict learning achievements. According to the selection of the influencing factors of the students' different curriculum dynamics in the future, and the use of support vector machines to pre-warn students' performance, some exploratory work has been done for the application of data mining technology in the field of education [5]. Tian [6] based on k-means and SVM methods to predict whether learners can obtain certificates Model. Qiu [7] proposed the LadFG model for course certificate acquisition prediction. Compared with the existing SVM, LRC, FM and other methods, it has a better prediction effect. Brinton [8] predicted student achievement based on time series neural network.

The above work mainly focuses on the feature extraction and modeling analysis of online learning data for multiple courses. The following problems exist: (1) There is a high standardization requirement for learning data, which cannot be directly applied to offline teaching analysis. (2) Due to the lack of a large number of standardized learning data, existing work usually only build a general model for multiple courses, can not effectively analyze a single course; (3) The lack of a correlation between academic performance and learning process, and it is impossible to propose targeted teaching intervention strategies.

3. Feature Model of Learning Process for Online and Offline Integrated Teaching Model
The higher the students’ participation in learning activities, the more active they are, the more they understand the key points, the better their academic performance will be. Referring to the learning behavior classification of learning participation, learning input, and learning interaction [4], this paper defines the following easy-to-measure learning process indicators, as shown in table 1.

| Table 1. Features and indicators of the learning process. |
|-----------------------------------------------------------|
| Learning process features | Feature indicators | Number | Description                        |
|---------------------------|--------------------|--------|------------------------------------|
| Participation             | Absences           | X₁     | Times of absence from class        |
|                           | Visits             | X₂     | Times of visiting online learning platform |
| Commitment                | Preview situation  | X₃     | Times of completing the preview report |
|                           | Task submissions   | X₄     | Times of submitting task           |
|                           | Task completion    | X₅     | Experimental task completion       |
| Interaction               | Reply              | X₆     | Times of replying teacher          |
|                           | Question           | X₇     | Times of questioning teacher       |

The sample data of the learning process data was generated by students (a total of 100 students) of the Information Security in the course of “Computer Network Fundamentals”. The original data are shown in table 2.

| Table 2. Learning process data. |
|---------------------------------|
| Name   | X₁  | X₂  | X₃  | X₄  | X₅  | X₆  | X₇  |
| CY     | 0   | 118 | 0   | 22  | 98  | 2   | 0   |

The z-score normalization method is used to normalize the feature data of the learning process as follows:

\[
X^* = \frac{X - \text{mean}}{\text{std}}
\]  

(1)

Among them, X is the learning process characteristic data of each column, mean is the average value of each column data set, and std is the standard deviation of each column data set.

Make a correlation analysis of the above data. The independent variable X is the learning process
The model training chooses a Gaussian kernel function, which has two hyperparameters: \( C \) and \( \gamma \). The hyperparameter sets selected in this paper are \( C = 1, 16, 64, 20 \) and \( \gamma = 0.001, 0.01, 0.1, 1, 10, 100 \). These two indicators are not representative at present. In summary, the feature classification model is better, but it may still be in an optimized model is improved by 8%.

When \( 0.2 \leq |r| <0.4 \), the two variables can be regarded as low-level correlation; when \( |r| <0.2 \), the two variables can be regarded as basically uncorrelated [9]. The \( r_{x_2,y} \), \( r_{x_3,y} \) in Table 3 are all less than 0.2, so it can be considered that the correlation between the number of online platform visits \( (X_2) \), previews \( (X_3) \) and the quality of learning is small, because the resources on the current learning platform are still not rich enough, and it is mandatory that every student must preview before they can do experiments. Therefore, the data of these two indicators are not representative at present. In summary, the feature vector of the learning process for online and offline integrated teaching modes is \( P = \langle X_1, X_4, X_5, X_6, X_7 \rangle \).

### 4. Optimization of Learning Performance Classification Model Based on Grid Search Strategy for SVM

The classical support vector machine (SVM) can obtain good results with a small number of statistical samples. Therefore, this paper chooses SVM for modeling and optimization of learning quality classification. According to the correlation analysis in Section 3, the learning process feature vector \( P \) is determined. To reduce errors, using the k-means algorithm to cluster each learning process feature and mark the learning quality (test score), the results are shown in Tables 4 and 5.

#### Table 3. Correlation coefficients between variables.

|      | \( X_1 \) | \( X_2 \) | \( X_3 \) | \( X_4 \) | \( X_5 \) | \( X_6 \) | \( X_7 \) |
|------|----------|----------|----------|----------|----------|----------|----------|
| \( Y \) | -0.423   | -0.105   | 0.077    | 0.203    | 0.300    | 0.444    | 0.209    |

When \( 0.2 \leq |r| <0.4 \), the two variables can be regarded as low-level correlation; when \( |r| <0.2 \), the two variables can be regarded as basically uncorrelated [9]. The \( r_{x_2,y} \), \( r_{x_3,y} \) in Table 3 are all less than 0.2, so it can be considered that the correlation between the number of online platform visits \( (X_2) \), previews \( (X_3) \) and the quality of learning is small, because the resources on the current learning platform are still not rich enough, and it is mandatory that every student must preview before they can do experiments. Therefore, the data of these two indicators are not representative at present. In summary, the feature vector of the learning process for online and offline integrated teaching modes is \( P = \langle X_1, X_4, X_5, X_6, X_7 \rangle \).

#### Table 4. Feature clustering of the learning process.

|     | \( X_1 \) | \( X_4 \) | \( X_5 \) | \( X_6 \) | \( X_7 \) |
|-----|----------|----------|----------|----------|----------|
| A   | \( X_1 = 0 \) | \( X_4 \geq 28 \) | \( X_5 \geq 97.3 \) | \( X_6 \geq 6 \) | \( X_7 \geq 5 \) |
| B   | \( X_1 = 1 \) | \( 18 \leq X_4 < 28 \) | \( 89.3 \leq X_5 < 97.3 \) | \( 3 \leq X_6 < 6 \) | \( 2 \leq X_7 < 5 \) |
| C   | \( X_1 > 1 \) | \( X_4 < 18 \) | \( X_5 < 89.3 \) | \( X_6 < 3 \) | \( X_7 < 2 \) |

#### Table 5. Learning quality marks.

| Score (\( Z \)) | Learning level (\( Y \)) |
|-----------------|--------------------------|
| \( Z \geq 85 \) | 1                         |
| 60 \( \leq Z < 85 \) | 0                         |
| \( Z < 60 \) | -1                        |

Then a training sample vector including learning process features and learning quality labels is \( Q = \langle X_1, X_4, X_5, X_6, X_7, Y \rangle \). Randomly select the training sample set \( L = \{ Q_1, Q_2, ..., Q_{10} \} \), the actual proportions of the students of the “good”, “medium” and “poor” categories are 12%, 66% and 22% respectively; use another 50 people to verify the sample set \( M = \{ Q_{11}, Q_{12}, ..., Q_{110} \} \), the actual proportions of the students are 16%, 64% and 20%. The model training chooses a Gaussian kernel function, which has two important hyperparameters, namely gamma and c. In order to obtain an accurate classification model of learning performance from the sample as much as possible, a grid search optimization method is used to optimize the model hyperparameters, the hyperparameter sets selected in this paper are \( C = \{ 0.01, 0.1, 1, 1.2, 2, 4, 8 \} \), \( \Gamma = \{ 0.001, 0.01, 0.1, 1, 1.2, 2, 4, 8 \} \). The algorithm flow is shown in Table 6.

The default parameters of the selected SVM model are: \( C = 0.1 \), \( \Gamma = 1/k \) (k is the number of categories). At this time, the accuracy of the model is 82%. The effect of the SVM model obtained by the algorithm is shown in Table 7 below.

It can be seen that when \( c = 4 \) and \( \gamma = 1 \), the model classification effect is the best. The accuracy was 90%. Compared with the model under the default parameter values, the accuracy of the optimized model is improved by 8%. On the whole, the generalization ability of the optimized classification model is better, but it may still be inaccurate in judging some critical situations, resulting...
in false positives and false negatives.

**Table 6.** Modeling of learning performance classification based on grid search strategy to optimize SVM.

| input: L, C, Gamma | Training sample data set, Hyperparameter C, Gamma set |
|-------------------|-----------------------------------------------------|

| output: model_best, accuracy_best, F_best, Recall_best, Precision_best, gamma_best, c_best |

```plaintext
OptimizedLearningPerformanceClassification ()
Initialize (L, Score, F, Recall, Precision, C, Gamma);
L=load(L);// Load sample set
For c in C do:
    For gamma in Gamma do:
        svm = SVC (gamma, c); // Iterate over each parameter
        svm. train (L);// Training model
        Compute (L, score, f, recall, precision);
        Save(score=>Score, f=>F, recall=>Recall, precision=>Precision, gamma=>Gamma, c=>C);
    End;
End;
For i in Score.length:
    If score > score_best: gamma_best=Gamma[i]; c_best=C[i]
    Else if score==score_best:
        If precision> precision_best: gamma_best=Gamma[i]; c_best=C[i];
        Else if precision== precision_best:
            If recall>recall_best: gamma_best=Gamma[i]; c_best=C[i];
            Else if recall== recall_best:
                If f>f_best: gamma_best=Gamma[i]; c_best=C[i];
                Save (gamma_best, c_best); // Save optimal hyperparameters
        svm=SVC (gamma_best, c_best);
        model_best=svm. train (L);// Use optimal parameters to train the model
        saveBest(model_best, score_best,f_best,recall_best,precision_best,gamma_best,c_best)
    End;
```

**Table 7.** Analysis of grid optimization effect.

| c, Gamma | Accuracy | Precision | Recall | F1-score | Accuracy on M |
|----------|----------|-----------|--------|----------|---------------|
| 0.1, 0.1 | 86%      | 72%       | 85%    | 78%      | 85%           |
| 1, 0.1   | 80%      | 86%       | 88%    | 87%      | 86%           |
| 4, 0.1   | 82%      | 73%       | 83%    | 78%      | 81%           |
| 0.1, 1   | 85%      | 74%       | 85%    | 79%      | 88%           |
| 1, 1     | 88%      | 72%       | 85%    | 78%      | 86%           |
| 4, 1     | 90%      | 94%       | 89%    | 91%      | 92%           |
| 0.1, 4   | 90%      | 71%       | 81%    | 76%      | 82%           |
| 1, 4     | 84%      | 89%       | 92%    | 90%      | 92%           |
| 4, 4     | 86%      | 85%       | 87%    | 86%      | 88%           |

**5. Analysis of Learning Mode and Teaching Quality Improvement Strategies Based on the Characteristics of the Learning Process**

Use the classification model to classify the sample set M. The percentages of the students who was classified as “good”, “medium”, and “poor” were 16%, 64%, and 20%. The false alarm rates of the
three types were 2.32%, 10%, and 2.43%, and the false alarm rates were 4.55%, 10%, and 0. The specific improvement strategy analysis is as follows.

Figure 1 shows the classification results of students with categories of “good”, “medium”, and “poor”.

![Figure 1. Classification results of students with different category.](image)

As can be seen from figure 1, the students of “good” category are all A on indicators X1 and X5; 62.5% of students in indicator X4 are A and B, and 37.5% of students are C; 87.5% and 75% of students perform on indicators X6 and X7 are A or B. One of the 8 classmates misreported, with a score of 82.5 points, but his performance of each indicator is relatively good, so it is classified as "good". Students in this category have a high level of learning participation, homework completion, and learning interaction, with a final score of 87.5% above 85. It can be seen that these students have the abilities to learn. For these students, some more difficult thinking questions can be arranged to help them further improve their understanding and application ability of the course content.

It is known from figure 1b, the indicators of these students are all A or B (6.25%); 59% of the students behave as A or B on the indicator X4, and the remaining 41% are C; all the students in the indicator X5 are A (68.7%) and B; 62.5% of the students are A or B on X6, but on X7 84% of the students are C. Students in this category have a high level of learning participation and completion of homework, but a low degree of interaction. In the end, 93% of the students scored between 60 and 85, and the other two misclassified students scored above 85. Actively discovering problems in the learning process helps deepen students’ understanding of the course. Teachers should increase the interaction process with students during the class.

As can be seen from figure 1c, the performance of this category of students in each indicator is significantly increased. The participation in learning, the number of homework attempts and the interaction of learning are all low. In the end, 90% of the students fail the grade, and the other one has a score of 65, but due to learning The low level of participation is classified as "poor". Students in this category should pay attention to improving their participation in learning, correct their learning attitude, and eliminate absences. Teachers should supervise such students through sign-in and pre-work assignments.

6. Conclusion

This paper draws on existing online teaching and learning data analysis and research, and proposes an online and offline comprehensive teaching mode quality evaluation model and continuous improvement strategy for automatic generation and analysis methods for the learning process. Based on the "Computer Network Fundamentals" course achievement data as the basis, based on correlation analysis and clustering methods to define the online and offline comprehensive teaching model of the learning process feature model; based on support vector machine and grid search optimization strategies to optimize student performance Compared with the classification model with default parameter values, the accuracy of the optimization model is improved by 8%. Based on the analysis of the validation set, the corresponding improvement strategies for different types of learning models are obtained, which is conducive to continuous improvement in the teaching process.
References

[1] Macfadyen L P and Dawson S 2010 Mining LMS data to develop an “early warning system” for educators: A proof of concept Computers & Education 54 (2) 588-599.

[2] Li D M 2017 Application Study of Online Learning Process Analysis in High School Information Technology Flip Classroom Teaching (Sichuan Normal University).

[3] Tokan M, Kopong I and Mbing M 2019 The effect of motivation and learning behaviour on student achievement. South African Journal of Education 39 (1) 1-8.

[4] Wang Y 2018 Analysis and Application of Online Learning Process (Huazhong Normal University).

[5] Liu B P, Fan T C and Yang H 2019 Research on application early warning of students’ achievement based on data mining technology Sichuan Univ.: NatSciEd 56 267-272.

[6] Tian Y, Wen Y, Yi X, et al. 2017 Predicting learning effect by learner’s behavior in MOOCs 18th International Conference on Intelligent Data Engineering and Automated Learning (Cham: Springer).

[7] Qiu J, Tang J, Liu T X, et al. 2016 Modeling and predicting learning behavior in MOOCs ACM International Conference on Web Search & Data Mining.

[8] T Y, Brinton C G, Joe-Wong C, et al. 2017 Behavior-based grade prediction for MOOCs via time series neural networks IEEE Journal of Selected Topics in Signal Processing 5 (11) 716-728.

[9] Schober P, Boer C and Schwarte L A 2018 Correlation coefficients: Appropriate use and interpretation Anesthesia and Analgesia 126 (5) 1763-1768.