An Ensemble Scheme Based on the Optimization of TOPSIS and AdaBoost for In-Class Teaching Quality Evaluation

Junqi Guo[1,2], Aohua Song[1,2(✉)] , Ludi Bai[1], Ziyun Zhao[1,2], and Siyu Zheng[1]

[1] Beijing Normal University, Beijing, China
[2] Engineering Research Center of Intelligent Technology and Educational Application, Ministry of Education, Beijing, China
`aohuasong@mail.bnu.edu.cn`

Abstract. As artificial intelligence has grown, intelligent technology has steadily been used in the classroom. Intelligent in-class evaluation has gained popularity in recent years. In this study, we apply two models: AE-SIS (Analytic Hierarchy Process-Entropy Weight-TOPSIS) and AW-AB (Adjusted Weight in Adaptive Boosting) to evaluate in-class teaching quality. We provide an ensemble scheme for intelligent in-class evaluation that combines the benefits of the two models. We test the current in-class evaluation criteria using classroom datasets for comparison. The outcomes show how great and successful the suggested plan is.

Keywords: in-class teaching · intelligent evaluation · statistical modeling · machine learning

1 Introduction

Education informatization is a breakthrough that stimulates improvements in the outdated educational system while also supporting students in fully developing themselves. It does this by leveraging several information approaches [15,17], including big data and AI methodology. For instance, Malaysia encourages cloud resources while developing innovative curricula, instructional methods, and learning resources [12]. Japan employs a platform for collaborative home-schooling to raise awareness of electronic textbooks and other learning materials [20]. China has started to gradually digitize education with support from the strong policy. Governments must raise the degree of information infrastructure building and set up a top-notch education support system, according to "The 14th Five-Year-Plan of National Informatization Plan" [1] in 2021. Governments should adopt strategic efforts to hasten the digital education transformation and intelligent updates as well as integrate information technology and educational innovations, according to the "Work Highlights of the Ministry of Education" from 2022 [2].

In related research, intelligent education quality evaluation is a crucial component of education informatization. It may evaluate learning outcomes, evaluate teaching quality, and help intelligent systems manage instruction. The observation-based scale method and the questionnaire-based research method, both of which are manual processes, are the two types of in-class teaching quality evaluation techniques that are traditionally used [9]. As a result of the assessor's subjective elements, it is impossible to draw a more unbiased and trustworthy conclusion. Also, it would take more time and effort to manually create questionnaires and observe teaching. Therefore, how to make education quality evaluation more intelligent has become a research hotpot [19].

The Distance Education Center at Beijing Normal University, for instance, has developed a methodology for measuring student involvement in distant learning based on LMS data that considers four factors [10], including online participation and interaction. To gauge student learning, the New Future firm created the Wisroom smart classroom system [18]. The Gradescope platform was developed by the University of California's Department of Computer Science to assist teachers in classifying and revising more than 250 million student assignments [11]. However, most current intelligent systems are analytic techniques such as behavior recognition, and in-class teaching quality models are scarce and have yet to be evaluated. To address the above problem, we propose an ensemble scheme for intelligent in-class evaluation.

2   Background

2.1   Statistical Learning

The Analytic Hierarchy Process and the Entropy Weight Method. By reducing complex problems down into smaller, more manageable components, the Analytic Hierarchy Process (AHP) [13] assists people and organisations in prioritising and making decisions. It entails building a hierarchy of choice criteria and options, then performing pairwise comparisons to ascertain their relative weight. The Entropy Weight Method (EWM) [3,21], a multi-criteria decision analysis technique, uses information theory and entropy to evaluate the relative weight of selection criteria. It works by assessing the degree of diversity or uncertainty among the selection criteria and assigning weights based on the importance of each criterion.

TOPSIS. In the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS), the solutions are ranked according to how closely they resemble the ideal answer.

Denote n as the number of samples, m as the number of features, and x_{ij} as the j_{th} feature of the i_{th} sample. The steps for using the TOPSIS method to perform the calculations are listed as follows:

(1) Regularization. Features need to be classified into four categories: extremely large features (larger is better), extremely small features (smaller is

better), intermediate features (closer to a certain value is better) and interval features (within a certain range is better). All features that are not extremely large must be transformed into the type where larger features are considered preferable.

(2) Normalization. Normalize all the feature series to eliminate the influence of the magnitude.

(3) Calculation of the distance between each sample and the positive ideal solution and negative ideal solution. Denote $X^+ = \{X_1^+, \ldots, X_m^+\}$ as the positive ideal solution, where X_j^+, $j \in \{1, \ldots, m\}$ is the maximum value of j_{th} feature. Denote $X^- = \{X_1^-, \ldots, X_m^-\}$ as the negative ideal solution, where X_j^-, $j \in \{1, \ldots, m\}$ is the minimum value of j_{th} feature. Denote D_i^+ as the distance between i_{th} sample and X^+, D_i^- as the distance between i_{th} sample and X^-. They are calculated as:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} \left(X_j^+ - x_{ij}\right)^2} \tag{1}$$

$$D_i^- = \sqrt{\sum_{j=1}^{m} \left(X_j^- - x_{ij}\right)^2} \tag{2}$$

(4) Calculation of the final score. The final score for each sample is calculated as:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{3}$$

2.2 Ensemble Learning

Ensemble Learning (EL) is a supervised learning algorithm that has gradually become popular [6]. Freund and Schapire [7] proposed the Adaptive Boosting algorithm (AB), which automatically adjusts the weight of each base learner according to the error rate. All base learners are assigned weights according to error rates to obtain an ensemble learner.

3 Method

3.1 The Analytic Hierarchy Process-Entropy Weight-TOPSIS (AE-SIS) Model

Although the TOPSIS technique reveals sample differences clearly and correctly, it tends to equally weight each aspect or relies on subjective experience, which makes it inadequate for assessing the effectiveness of in-class instruction. As a result, appropriate weights must be provided for each model feature. Using the AHP-EW model with the TOPSIS technique results in more appropriate weights for assessing the quality of classroom instruction because the AHP-EW

model incorporates subjective and objective data [16]. The structure of the AE-SIS model is shown in Fig. 1. For different indicators, the specific process of the AE-SIS model is described in Fig. 2.
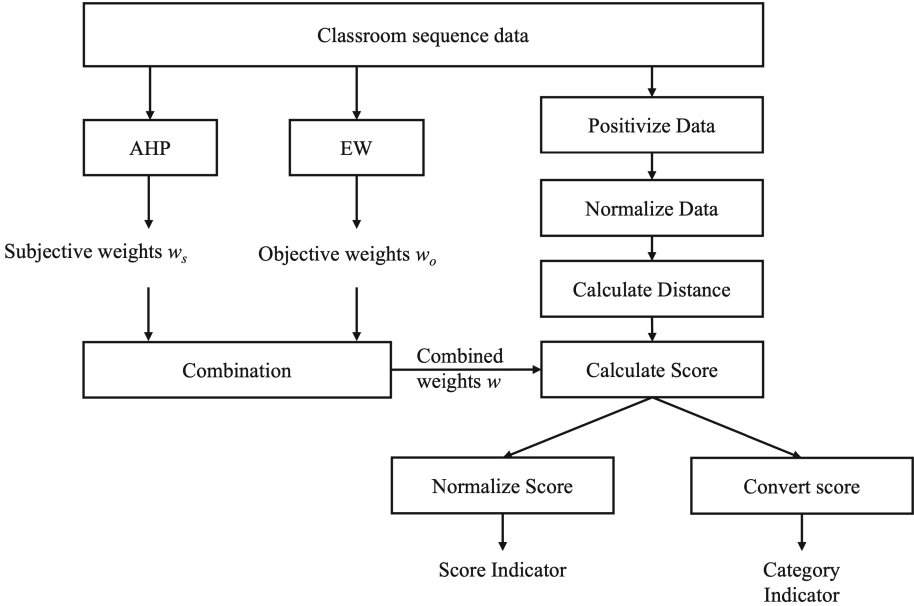

Fig. 1. Framework of the AE-SIS model.

The steps for using the AE-SIS model to evaluate in-class teaching quality are presented as follows:

(1) By the in-class teaching quality evaluation system [8], choose and standardize the features that correspond to the indicators.

(2) Calculate the comprehensive weights of the features. The subjective and objective weights of the features were determined and ranked using the AHP and EWM, respectively. Then, based on the data intensity and order of the objective and subjective weights, the comprehensive weights are decided.

(3) Calculate the scores of the samples. The samples are first evaluated and scored based on TOPSIS. The combined weights derived in the previous step are then introduced in the calculation of the distances to the positive and negative ideal solutions. The updated equation is as follows:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} w_j \left(X_j^+ - x_{ij} \right)^2} \tag{4}$$

$$D_i^- = \sqrt{\sum_{j=1}^{m} w_j \left(X_j^- - x_{ij} \right)^2} \tag{5}$$

where w_j is the comprehensive weight of the j_{th} feature.

(4) Output of the corresponding results. For category indicators, we will compare and analyze the scores of each sample, identify the classification threshold and convert the scores into category results. For the score indicators, the scores need to be normalized so that they meet the requirements.
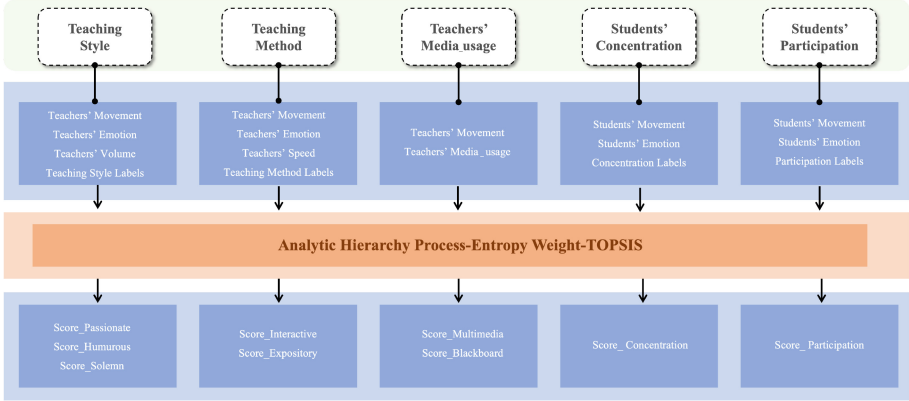

**Fig. 2.** The AE-SIS-based statistical model for in-class teaching evaluation.

3.2 The Adjusted Weight in Adaptive Boosting (AW-AB) Model

Even while Adaptive Boosting excels at solving classification and regression issues, it is highly sensitive to unusual data. The algorithm will give aberrant samples more weight throughout the iterative training phase, which will change the weight of regular samples and thus result in low accuracy. To increase accuracy, we implement a penalty mechanism to lessen the weight of samples with multiple errors. For different indicators, the specific process of the AE-SIS model is described in Fig. 3.

The steps for using the AW-AB model to evaluate in-class teaching quality are described as follows:

Dataset $A = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, where n is the number of the samples, x_i and y_i are the data and label of the i_{th} sample, respectively. h is the base learner, and c is the number of iterations.

(1) Determination of the base learner and dataset. The corresponding features and dataset are determined according to the indicator system.

(2) Initialization of the sample weights as $\frac{1}{n}$.

(3) The base learner h_c is trained using the current weight W_c, and the error e_c is calculated at each iteration. h_c is calculated as:
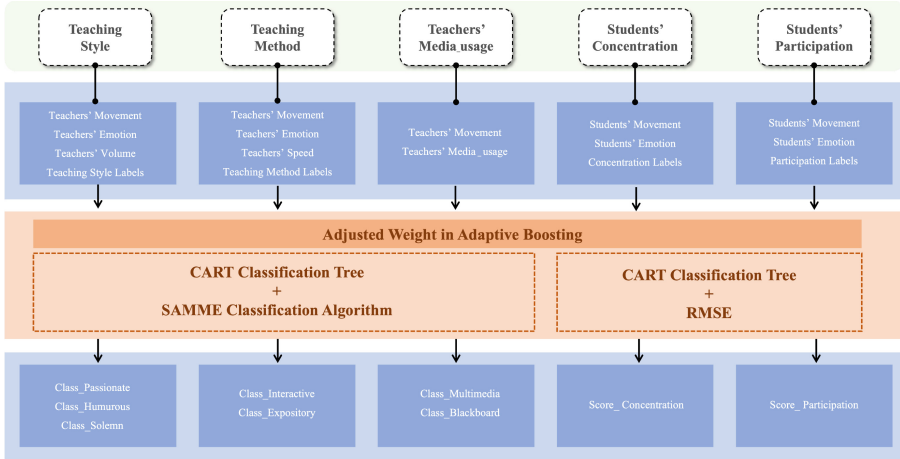
Fig. 3. The AW-AB-based ensemble model for in-class teaching evaluation.

$$h_c = AdaBoost\,(A, W_c) \tag{6}$$

$$e_c = \sum_{i=1}^{N} w_{ci} I\,[h_c(x) \neq y] \tag{7}$$

(4) The weighting factor α_c and normalization factor Z_c are calculated based on e_c. It is calculated as:

$$\alpha_c = \frac{1}{2} \ln \left(\frac{1 - e_c}{e_c} \right) \tag{8}$$

$$Z_c = \sum_{i=1}^{n} W_{ci} e^{-\alpha_c y_i h_c(x_i)} \tag{9}$$

(5) A penalty mechanism is introduced to update the sample weights W_{c+1}. We introduce the weighting threshold th. After a certain iteration, if a sample is assigned weights above this threshold due to a consistently high e_c, the sample will be judged as an anomaly. We then add penalties to reduce the weight of the sample, thus reducing the impact of the sample on the overall model e_c. β is the weight adjustment parameter after the threshold is exceeded. The process is calculated as follows:

(6) Synthetic ensemble learner $H(x)$. The combination of C weak learners into AW-AB ensemble learners is calculated as:

$$H(x) = \mathrm{sign} \left(\sum_{c=1}^{c} \alpha_c h_c(x) \right) \tag{10}$$

Table 1. Formulas for updating weights based on the AW-AB model

| Conditions		Weighting update formula
If $h_c(x_i) = y_i$		$W_{c+1} = \frac{W_c(i)e^{-\alpha_c}}{Z_c}$
If $h_c(x_i) \neq y_i$	If $W_c(x_i) \leq th$	$W_{c+1} = \frac{W_c(i)e^{\alpha_c}}{Z_c}$
	If $W_c(x_i) > th$	$W_{c+1} = \frac{W_c(i)e^{\frac{1-\beta}{2}}}{Z_c}$

4 Experiment

4.1 Dataset

We selected 200 sessions of audio and video data of smart informatization class-rooms in primary and secondary schools at Beijing Normal University. After processing by artificial intelligence algorithms such as object detection, speech recognition, and action recognition [4,5,14], we obtained a total of 200 sets of teacher samples and a total of 300 sets of student samples.
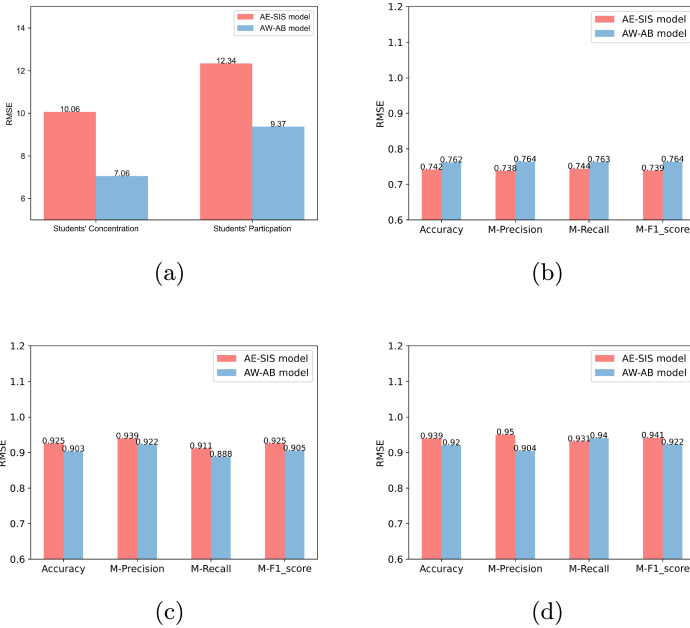
Fig. 4. Performance comparison of students' and teachers' indicators of different models. (a) Performance of students' indicators (b) Performance of teaching style (c) Performance of teaching method (d) Performance of teachers' media_usage.
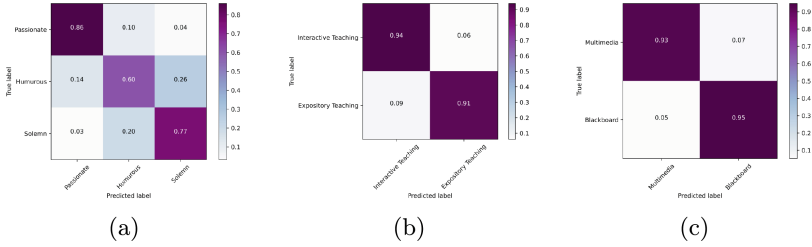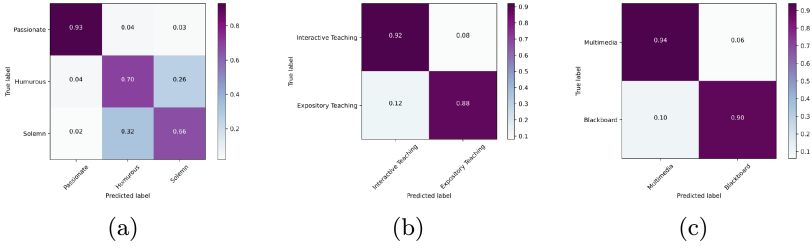
Fig. 5. Confusion matrix for teachers' indicators evaluated by the AE-SIS model. (a) Performance of teaching style (b) Performance of teaching method (c) Performance of teachers' media_usage.

The teacher samples are divided into six categories-movement, emotion, volume, speed, speech text and labels. The student samples are divided into three categories-movement, emotion, and labels. Both teacher data and student emotion data include two categories: laughing or not laughing. Teacher movement data include 9 categories: raising hands, gesturing with both hands, moving around, teaching without gestures, bending over to operate the desktop, holding textbooks, writing on the blackboard, turning over and fingering the multimedia. Student movement data included 5 categories: raising hands, reading, writing, head up, and lying on the table.

Table 2. Performance Comparison of TOPSIS and the AE-SIS model

Indicator	Metric	TOPSIS	AE-SIS
Students' Concentration	RMSE	11.17	10.06
Students' Participation	RMSE	13.99	12.34
Teaching Style	Acc	0.712	0.742
	MP	0.715	0.738
	MR	0.712	0.744
	MF	0.712	0.739
Teaching Method	**Acc**	**0.876**	**0.925**
	MP	**0.878**	**0.925**
	MR	**0.876**	**0.925**
	MF	**0.876**	**0.925**
Teachers' Media_usage	**Acc**	**0.909**	**0.939**
	MP	**0.909**	**0.939**
	MR	**0.909**	**0.939**
	MF	**0.909**	**0.939**

Fig. 6. Confusion matrix for teachers' indicators evaluated by the AE-AB model. (a) Performance of teaching style (b) Performance of teaching method (c) Performance of teachers' media_usage.

Table 3. Performance Comparison of the AB and AW-AB models.

Indicator	Metric	AB	AW-AB
Students' Concentration	**RMSE**	**9.22**	**7.06**
Students' Participation	**RMSE**	**10.66**	**9.37**
Teaching Style	**Acc**	**0.718**	**0.762**
	MP	**0.729**	**0.764**
	MR	**0.718**	**0.763**
	MF	**0.713**	**0.764**
Teaching Method	Acc	0.896	0.903
	MP	0.896	0.904
	MR	0.896	0.904
	MF	0.896	0.903
Teachers' Media_usage	Acc	0.891	0.920
	MP	0.891	0.920
	MR	0.891	0.920
	MF	0.891	0.920

4.2   Results

We drew on the existing evaluation systems as the in-class teaching quality evaluation system for this trial [8]. In this section, we compare the performance of the AE-SIS model, AW-AB model and proposed ensemble model.

Comparison of the AE-SIS and AW-AB Models. Fig. 4(a) shows a comparison of students' indicator evaluations. Figure 4(b)(c)(d) shows a comparison of teachers' indicator evaluations. From the results, we know that two-category classification tasks are simple, that the AE-SIS model fits well, and that the AW-AB model is more suitable for complex tasks such as teaching style classification.
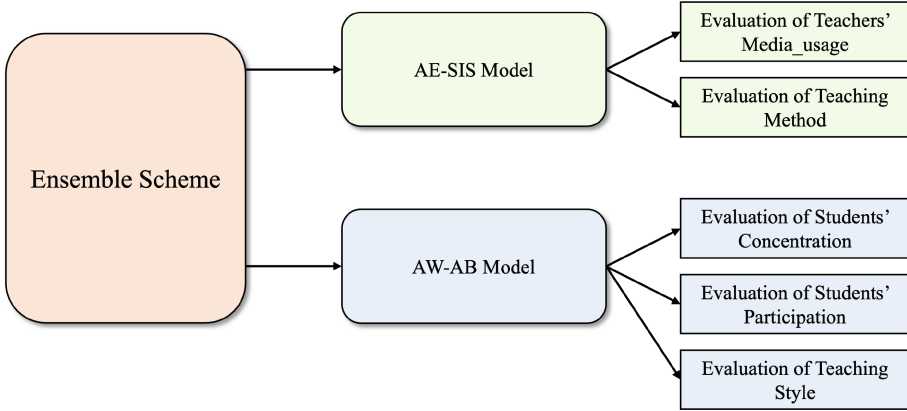
Fig. 7. Framework of the ensemble scheme.

Table 4. Performance Comparison of the Traditional model and Combination model.

Classification	Indicator	Metric	TOPSIS	AdaBoost	Ensemble Scheme
Students' Score Indicators	Students' Concentration	RMSE	11.17	9.22	7.06
	Students' Participation	RMSE	13.99	10.66	9.37
Teachers' Category Indicators	Teaching Style	Acc	0.712	0.718	0.762
		MP	0.715	0.729	0.764
		MR	0.712	0.718	0.763
		MF	0.712	0.713	0.764
	Teaching Method	Acc	0.876	0.896	0.925
		MP	0.878	0.896	0.925
		MR	0.876	0.896	0.925
		MF	0.876	0.896	0.925
	Teachers' Media_usage	Acc	0.909	0.891	0.939
		MP	0.909	0.891	0.939
		MR	0.909	0.891	0.939
		MF	0.909	0.891	0.939

The Results of the AE-SIS Model. The RMSE between the predicted value and the label value of students' concentration is 10.06, and the RMSE of students' participation is 12.34. Figure 5 shows the confusion matrix of the teachers' indicators. Table 2 shows a comparison of the AE-SIS model and TOPSIS method of in-class teaching evaluation. The results show that the proposed AE-SIS model is overall better than the TOPSIS method, which performs well in the classification of teaching methods and teachers' media_usage. The results demonstrate the effectiveness of statistical models in in-class intelligence teaching quality evaluation.

The Results of the AW-AB Model. The RMSE between the predicted value and the label value of students' concentration is 7.058, and the RMSE of students' participation is 9.370. Figure 6 shows a confusion matrix of the teachers'

indicators. Table 3 shows a comparison of the AW-AB model and AB model. The proposed AW-AB model has significant improvements in the classification of teaching style and regression task of students' indicators, is better than the AB model. The results demonstrate the advantages of the proposed model in in-class intelligence teaching quality evaluation.

Comparison of the Ensemble Scheme and Traditional Method. Thus, for different indicators, we choose different suitable models to obtain the combination model, as shown in Fig. 7.

Table 4 shows a comparison of our proposed ensemble scheme and the traditional method:TOPSIS and AdaBoost. The ensemble scheme performs better on the in-class student regression tasks and teacher classification tasks.

5   Conclusion

In this study, we propose an ensemble technique for intelligent in-class evaluation, and we experimentally show how better and more efficient it is. The AE-SIS model outperforms the TOPSIS technique via entropy weighting and analytic hierarchy process feature weighting. The AW-AB model outperforms the AdaBoost method by tweaking basic learners' weight updates to lessen the influence of aberrant samples. In addition to providing a more thorough and varied framework than a single model, the ensemble scheme efficiently addresses the issues with traditional in-class teaching evaluation.

References

1. The 14th five-year-plan of national informatization plan was released. Modern Educ. Technol. **32**(1), 15 (2022)
2. Highlights of the ministry of education 2022. www.moe.gov.cn/jyb_sjzl/moe_164/202202/t20220208_597666.html (2022)
3. Bao, Q., Yuxin, Z., Yuxiao, W., Feng, Y.: Can entropy weight method correctly reflect the distinction of water quality indices? Water Resour. Manage **34**(11), 3667–3674 (2020)
4. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
5. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291–7299 (2017)
6. Dong, X., Yu, Z., Cao, W., Shi, Y., Ma, Q.: A survey on ensemble learning. Front. Comp. Sci. **14**(2), 241–258 (2020)
7. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**(1), 119–139 (1997)
8. Guo, J., Bai, L., Yu, Z., Zhao, Z., Wan, B.: An AI-application-oriented in-class teaching evaluation model by using statistical modeling and ensemble learning. Sensors **21**(1), 241 (2021)

 9. Liu, H., Zhu, Y.: Research on the construction of teaching quality evaluation system. In: J. Phys.: Conf. Series. **1673**, 012055 IOP Publishing (2020)
10. Luan, L., Hong, J.C., Cao, M., Dong, Y., Hou, X.: Exploring the role of online EFL learners' perceived social support in their learning engagement: a structural equation model. Interact. Learn. Environ. 28(9), 1–12 (2020)
11. Singh, A., Karayev, S., Gutowski, K., Abbeel, P.: Gradescope: a fast, flexible, and fair system for scalable assessment of handwritten work. In: Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, pp. 81–88 (2017)
12. Thang, S.M., Hall, C., Murugaiah, P., Azman, H.: Creating and maintaining online communities of practice in Malaysian smart schools: challenging realities. Educ. Action Res. **19**(1), 87–105 (2011)
13. Vaidya, O.S., Kumar, S.: Analytic hierarchy process: an overview of applications. Eur. J. Oper. Res. **169**(1), 1–29 (2006)
14. Van Lancker, D., Kreiman, J., Wickens, T.D.: Familiar voice recognition: patterns and parameters part II: recognition of rate-altered voices. J. Phon. **13**(1), 39–52 (1985)
15. Wang, B., Xing, H.: The application of cloud computing in education informatization. In: 2011 International Conference on Computer Science and Service System (CSSS), pp. 2673–2676. IEEE (2011)
16. Xing, Y., Zhang, Z., Wu, J., Zhao, H., Tu, C., Xue, C.: Evaluation system of distribution network admission to roof distributed photovoltaic based on ahp-ew-topsis. In: 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2), pp. 1349–1352. IEEE (2021)
17. Yan, S., Yang, Y.: Education informatization 2.0 in China: Motivation, framework, and vision. ECNU Rev. Educ. **4**(2), 410–428 (2021)
18. Yu, T.: A case study on the development strategy of TAL education group. J. Innov. Social Sci. Res. ISSN **2591**, 6890 (2020)
19. Zhu, Z.T.: New development of smart education: From flipped classroom to smart classroom and smart learning space. vol. 22, p. 18 (2016)
20. Zhang, W., Li, Z., Okubayashi, T., Jia, R.: Analysis of the educational informatization policies in Japan and significances for China. Modern Educ. Technol. **27**(3), 5–12 (2017)
21. Zhu, Y., Tian, D., Yan, F.: Effectiveness of entropy weight method in decision-making. Math. Problems Eng. **2020**, 0–5 (2020)