

# A Student Grouping Method for Massive Online Collaborative Learning

<https://doi.org/10.3991/ijet.v17i03.29429>

Yuan Wang<sup>1</sup>, Qi Wang<sup>2</sup>(✉)

<sup>1</sup> Shijiazhuang University of Applied Technology, Shijiazhuang, China

<sup>2</sup> Department of Postal Communication Management, Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang, China

wangqi@sjzpc.edu.cn

**Abstract**—As a brand-new learning strategy, cooperative learning can greatly improve the quality of classroom teaching and the teamwork and learning efficiency of students. The conventional student grouping models for cooperative learning take less consideration of the complementarity of students' knowledge structures and learning interests, so they are not quite suitable for online teaching, which is provided for a huge number of students and has a large scale of learning resources and great freedom of choice. To this end, this paper studies the student grouping method for massive online collaborative learning. Firstly, the problem of identifying the student's knowledge state was described and characterized, and a diagnosis model of students' knowledge state simulating the cooperative learning process was constructed based on the gated recurrent neural network. The creation of an online collaborative learning scenario that considers both the knowledge state and interests of students was regarded as an NP-hard problem, and then, the enhanced particle swarm optimization algorithm was adopted to realize the grouping of students for massive online collaborative learning. The experimental results verified the effectiveness of the model constructed in this paper.

**Keywords**—cooperative learning, massive online learning, collaborative learning, student grouping

## 1 Introduction

Cooperative learning is a brand-new learning strategy that respects the individual differences among students, cultivates students' self-learning abilities and promotes mutual assistance among students and between teachers and students in the teaching process, and thus it can greatly improve the quality of classroom teaching and the teamwork and learning efficiency of students [1-3]. With the advancement of technologies and the innovation of ideas, various online learning platforms emerge one after another, becoming an important way for students to obtain learning resources and for teachers to carry out information-based teaching [4-7]. However, online teaching involves a

huge number of students, who are difficult to group manually [8-10]. Accurately acquiring the knowledge state of students and optimizing the grouping of massive users for online cooperative learning based on the knowledge state diagnosis results is an effective way to improve the effect of online cooperative learning.

To closely trace the knowledge level of students in the learning process, a number of knowledge tracing methods have been proposed. Chaudhry et al. [11] proposed a recursive knowledge tracking machine, which has enriched the coding of the conventional knowledge tracking machine in terms of time, and also coded learning difficulty and students' abilities, skills and practical history according to their knowledge state. The interactive learning environment facilitates learning by providing student with hints to fill their gaps in conceptual understanding. However, research shows that learners fail to use hints in the best way. Lai et al. [12] proposed a multi-task memory-augmented deep learning model to jointly predict the hint-taking and the knowledge tracing task, which incorporates the past responses as well as the impacts on the hints used on both tasks. The results show that the proposed model is much superior than the past work in terms of hint prediction by at least 12%. In terms of student grouping for online cooperative learning, Chang et al. [13] proposed an online multi-user real-time co-reading system to improve the co-reading performance in collaborative learning. The system can be integrated into a learning management system to support the grouping mechanisms during instruction and thereby enhance students' interaction through co-reading. Chen et al. [14] investigated the impact of the designated leadership role on learners' participation and learning experience in online collaborative learning, and compared six participating behaviours of and collaborative learning experiences between group leaders and members during a 12-week blended course. The research results generally support the influence of designated leadership on learning participation instead of learning experience. Xie et al. [15] examined the relationships between perceived leadership, team cohesion, online participation, self-regulation, and learning results. The data covered surveys and online discussion logs of 171 online undergraduates. Through correlation analysis and structural equation modelling, the research revealed the unique contributions of task leadership and relationship leadership in group collaborative learning.

Conventional diagnosis often evaluates students' knowledge state based on the correct answer rate of preset questions, without considering the dynamic changes in their knowledge state. The conventional student grouping models take less consideration of the complementarity of students' knowledge structures and learning interests, and thus they are not suitable for online teaching, which is provided for a huge number of students and has a large scale of learning resources and great freedom of choice. Therefore, this paper conducts research on a student grouping method for massive online collaborative learning.

## 2 Diagnosis of students' knowledge state in massive online learning

A lot of practice has proved that, in teaching activities, the implementation of cooperative learning can transform students from bystanders to participants, thereby effectively enhancing students' creative abilities, independent learning abilities and teamwork and greatly improving the quality of classroom teaching. In this process, appropriate grouping of students is closely related to the effect of their cooperative learning. A reasonable grouping model for cooperative learning needs to ensure the knowledge structures of the students in each group are complementary. In the case of massive online learning, to accurately obtain the knowledge state characteristics of students from massive online learning behaviour data, research results of artificial intelligence and cognitive diagnosis can be introduced to construct a multi-layer perceptron neural network for diagnosing the knowledge state of students.

The conventional cognitive diagnosis models that rely on manual design can no longer handle the complex dynamic behaviours of cooperative learning. In this paper, the gated recurrent neural network was used to simulate the cooperative learning process of students so as to achieve the accurate identification of their knowledge state. The research problem is described as follows.

### 2.1 Problem description and characterization

Suppose that in an online collaborative learning scenario, there are  $U$  students participating in learning,  $K$  study groups,  $V$  learning tasks, and  $L$  knowledge points, expressed as  $U=\{u_1, u_2, \dots, u_U\}$ ,  $K=\{k_1, k_2, \dots, k_K\}$ ,  $V=\{v_1, v_2, \dots, v_V\}$  and  $L=\{l_1, l_2, \dots, l_L\}$ . Each student  $u$  will choose some learning task  $v$  to do. The task implementation result  $T$  of the student can be represented by a triple  $(u, k, v, t)$ , where  $u \in U$ ,  $v \in V$  and  $t \in [0, 1]$ .  $t$  is the score obtained by student  $u$  on the learning task  $v$ . In addition, let the matrix marked by experts be  $P=\{P_{ij}\}_{N \times L}$ . If the learning task  $v_i$  is associated to some extent with the knowledge point  $l_j$ , then  $P_{ij}$  is 1; and if  $v_i$  is not associated in any way with  $l_j$ , then  $P_{ij}$  is 0. Given a student's task implementation results  $T$  and the matrix  $P$ , the aim of the cognitive diagnosis of the student's knowledge state in this paper can be described as to find out the proficiency of the student in all the required knowledge points by predicting the student's task score. Figure 1 shows the flow chart of a student's knowledge state diagnosis.

The cooperative learning process mainly involves 4 factors, including students, study groups and learning tasks. In order to use the gated recurrent neural network model to simulate and analyse the cooperative learning process, it is necessary to focus on the above factors. The student factor reflects a student's response to learning tasks, and characterizes the student's proficiency in knowledge points. The study group factor reflects the influence of group cooperation on the final performance of the learning tasks, and characterizes the complementarity of the knowledge structures of the students within one group. The learning task factor corresponds to the specific knowledge

points in the diagnosis of students' knowledge state, and characterizes the association between a learning task and the knowledge points.

In view of the above factors, a diagnosis model for students' knowledge state based on the gated recurrent neural network was proposed, in order to simulate the complex process of cooperative learning.

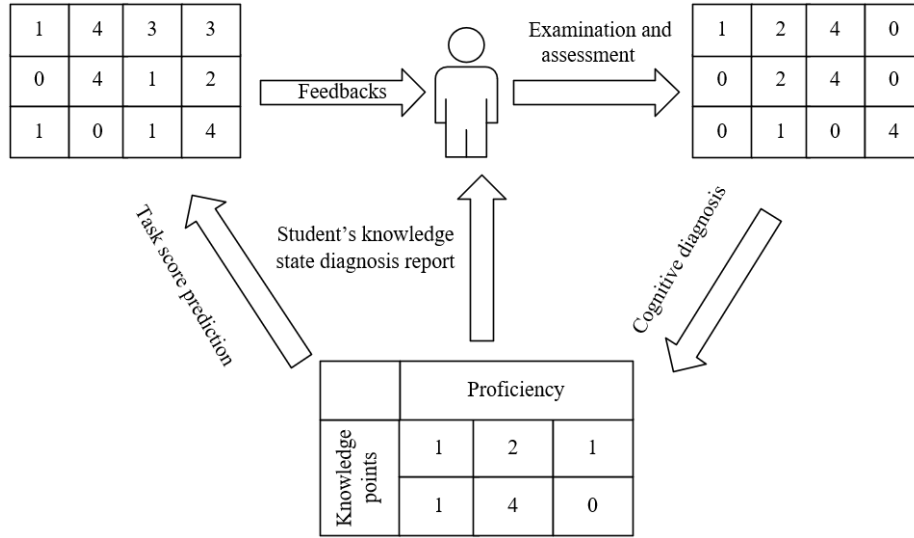


Fig. 1. Flow chart of a student's knowledge state diagnosis

## 2.2 Construction of the knowledge state diagnosis model

Having a basic idea similar to that of the long short-term memory network model, the gated recurrent neural network has a more concise structure. It has two gates - an update gate and a reset gate, which can retain and enhance the original information and thus effectively solve the problem of vanishing gradients that can easily occur in traditional recurrent neural networks. Assuming that the input vector at time  $\varphi$  is represented by  $a_\varphi$ , the update gate  $c_\varphi$  at this time can be expressed as Eq. (1):

$$c_\varphi = \varepsilon(Q_1^{(c)} a_\varphi + S_1^{(c)} f_{\varphi-1}) \quad (1)$$

According to the above equation, the update gate first performs a linear transformation, that is, multiplying the vector  $a_\varphi$  by the weight matrix  $Q_1^{(c)}$ . Then, it performs a linear transformation that multiplies  $f_{\varphi-1}$ , which retains the state at time  $\varphi-1$ , by  $S_1^{(c)}$ . Finally, the results of the two linear transformations are added up and processed by the *sigmoid* activation function, so that the output result of the update gate is greater than 0 and less than 1. The reset gate  $t_\varphi$  at time  $\varphi$  can be expressed as Eq. (2):

$$t_\varphi = \varepsilon(Q_2^{(c)} a_\varphi + S_2^{(c)} f_{\varphi-1}) \quad (2)$$

According to the above equation, the reset gate first performs a linear transformation that multiplies the vector  $a_\varphi$  with the weight matrix  $Q_2^{(c)}$ . Then, it performs a linear transformation that multiplies  $f_{\varphi-1}$ , which retains the state at time  $\varphi-1$ , by  $S_2^{(c)}$ . Finally, the results of the two linear transformations are added up and processed by the *sigmoid* activation function, so that the output result of the reset gate is greater than 0 and less than 1. After all the cooperative learning process information is processed by the update gate and reset gate, the new memory content is finally generated and output. Eq. (3) shows the calculation formula of the memory content:

$$f'_\varphi = \tan\Box(Qa_\varphi + t_\varphi \oplus Sf_{\varphi-1}) \quad (3)$$

The above equation first performs linear transformation, that is, postmultiplication of  $a_\varphi$  by the weight matrix  $Q$ , and the information  $f_{\varphi-1}$  at time  $\varphi-1$  by the weight matrix  $S$ , respectively, and then multiplies the reset gate  $t_\varphi$  by the corresponding elements of  $Sf_{\varphi-1}$ , and further adds up the linear transformation results of  $a_\varphi$  and  $Q$  and of  $t_\varphi$  and  $Sf_{\varphi-1}$ , and processes them with the *tanh* activation function.

Finally, the constructed network model calculates  $f_\varphi$  to retain the information at the current moment and transfer it to the next unit. The memory content  $f'_\varphi$  at the current moment and the memory content  $f_{\varphi-1}$  that needs to be collected at the previous moment are determined by the update gate. The above process can be expressed as the following formula:

$$f_\varphi = c_\varphi \oplus f_{\varphi-1} + (1 - c_\varphi) \oplus f'_\varphi \quad (4)$$

All the cooperative learning process information is output as  $c_\varphi$  through the update gate, with the amount of information transmitted controlled in the above-mentioned way. The information retained at the previous moment can be characterized by the product of  $c_\varphi$  and  $f_{\varphi-1}$ , and the final output of the gated recurrent neural network is the sum of the information retained at the previous moment and that retained at the current moment.

Through the above analysis, it can be seen that the constructed diagnosis model for students' knowledge state based on gated recurrent neural network achieves the storage and screening of cooperative learning process information via the update gate and the reset gate, so that valuable cooperative learning process information will not be eliminated over the long time span of the cooperative learning process. Instead, the constructed model avoids the problem of gradient vanishing or exploding by retaining and transmitting part of the information at a certain probability.

The changes in students' knowledge state in an online collaborative learning scenario are ambiguous, variable and uncertain. To analyze these complex learning psychologies, it is necessary to design the membership function based on records of students' historical learning behaviours, grouping and task performance scores in conjunction with the student, study group, learning task and other factors. The diagnosis model for students' knowledge state based on gated recurrent neural network was therefore constructed to predict the probability of students completing the learning tasks with high quality. The above factors are described in detail below:

In the diagnosis model for students' knowledge state, the vector matrix  $f^u$  describing knowledge proficiency can be used to characterize each individual student, that is, the aforementioned student factor, denoted as  $G^u$ . Suppose the one-hot encoding vector of student is represented by  $a^u$ , and that the trainable matrix is represented by  $TM$ , which satisfy  $TM \in \mathbb{R}^{M \times L}$  and  $a^u \in \{0,1\}^{1 \times M}$ . The student factor  $G^u$  can be obtained by multiplying  $a^u$  by  $TM$ . Eq. (5) shows the corresponding mathematical expression:

$$f^u = SIG(a^u \times TM) \quad (5)$$

In the diagnosis model for students' knowledge state, each study group can be represented by a vector matrix  $o^k$  that describes the complementarity of knowledge structures and teamwork of group members. Assuming that a group has  $\delta$  members, the mathematical expression of the study group factor  $K^u$  is shown below:

$$K^u = \sum_{\delta} a^u \times o^k \quad (6)$$

In the diagnosis model for students' knowledge state, each learning task can be characterized by a vector matrix  $P^v$  about knowledge points, which satisfies  $P^v \in \{0,1\}^{1 \times L}$ . Based on the preset matrix  $P$ , the learning task factor can be expressed as the following formula:

$$P^v = a^v \times P \quad (7)$$

In order to diagnose students' knowledge state more accurately, 4 other factors were introduced as supplement to the study group factor and the learning task factor, which are namely the difficulty of a knowledge point  $f^{ND}$ , the importance of a knowledge point  $f^{ZY}$ , and the difficulty of a learning task  $f^{QF}$ , and the learning efficiency of a study group  $f^{SD}$ .  $f^{ND} \in \{0,1\}^{1 \times M}$  refers to the difficulty of each knowledge point required to be mastered in a learning task;  $f^{QF} \in (0,1)$  means the difficulty of the entire learning task; and  $f^{SD}$  stands for the overall proficiency of the learning group in knowledge. Assuming that the trainable matrices are represented by  $H$ ,  $E$ ,  $J$  and  $R$ , respectively, the mathematical expressions of 4 other factors are shown in Eq. (8):

$$\begin{cases} f^{ND} = SIG(a^v \times H) \\ f^{QF} = SIG(a^v \times E) \\ f^{ZY} = SIG(a^v \times J) \\ f^{SD} = SIG(a^v \times R) \end{cases} \quad (8)$$

The mathematical expression of the diagnosis model for students' knowledge state based on the gated recurrent neural network is given in Eq. (9):

$$a = P^v \times K^u \circ (f^u - f^{ND} + f^{ZY}) \times f^{QF} \times f^{SD} \quad (9)$$

Based on Eq. (9), the diagnosis model for students' knowledge state based on gated recurrent neural network was constructed, as shown in Figure 2. Eq. (10) shows the expression of the loss function between the final output  $b$  of the model and the true label  $ZL$ :

$$SU = -\sum_i(ZL_i \log b_i + (1 - ZL_i)\log(1 - b_i)) \quad (10)$$

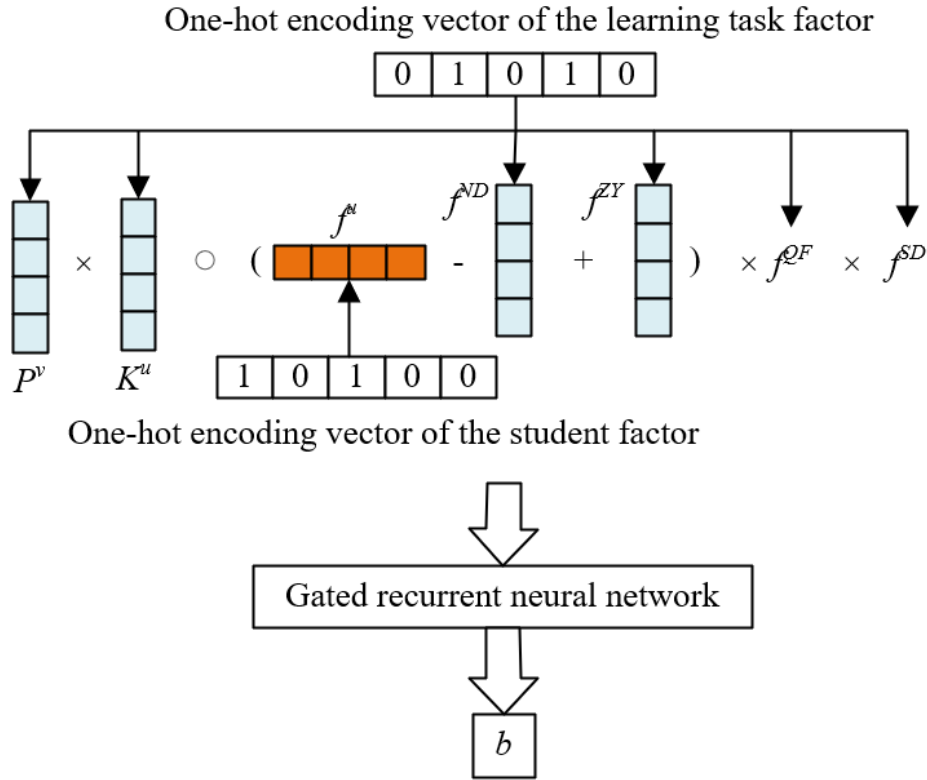


Fig. 2. Diagnosis model for students' knowledge state based on gated recurrent neural network

After training, the value of  $f^i$  is the expected diagnosis result of the student's knowledge state in the cooperative learning process, standing for the student's proficiency in each knowledge point in the cooperative learning process.

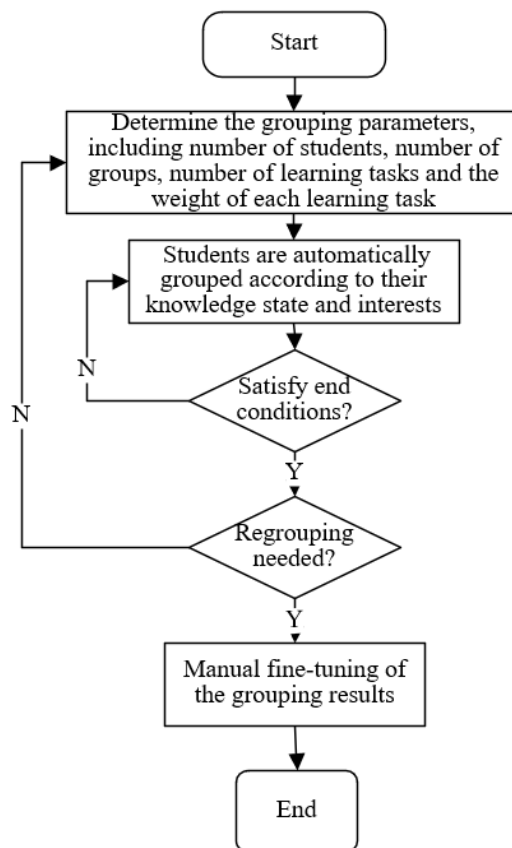
### 3 Student grouping method for massive online collaborative learning

Grouping based on students' knowledge and interests can improve their thinking level and promote discussions and interactions among them in the learning process. The creation of an online collaborative learning scenario that considers both the knowledge and interests of students can be regarded as an NP-hard problem. The existing literatures have applied the particle swarm optimization algorithm to solve different NP-hard problems and achieved good experimental results. In this paper, the enhanced particle swarm optimization algorithm was used to form study groups of cooperative learning effectively and efficiently. The constructed intelligent grouping model for cooperative

learning uses two grouping criteria: the knowledge state of the students in a given learning task and their interest in the given learning task. Figure 3 shows the steps of the constructed grouping model for massive online collaborative learning.

As shown in the figure, in order to quantify the model, the knowledge state and the interest level are characterized by the pre-test score and the learning experience of a student obtained respectively in the previous section. Regarding the optimization of grouping in cooperative learning, Eq. (11) shows the fitness function expression of the intelligent grouping model for cooperative learning:

$$SUF C(W_b) = g + D_1 + D_2 \tag{11}$$



**Fig. 3.** Steps of the student grouping model for massive online collaborative learning

It can be seen from the above formula that the fitness function is composed of three parts:  $g$ ,  $D_1$  and  $D_2$ . Let the average difference between the group with the highest level of understanding and that with the lowest level of understanding for the  $l$ -th learning task be denoted as  $g$ , where students' level of understanding is obtained from the pre-test scores of the knowledge state obtained in the previous section. Assuming that the



weight of the  $k$ -th learning task is denoted as  $\theta_k$ , that the understanding level of the  $a$ -th member in the  $j$ -th group for the  $k$ -th learning task as  $S_{ajk}$ , that the number of learning tasks as  $V$ , that the number of students as  $U$ , and that the number of groups as  $K$ , then there is:

$$g = \frac{\sum_{k=1}^V \left\{ \text{Max}_{1 \leq j \leq K} \{ \sum_{a=1}^U \theta_k S_{ajk} \} - \text{Min}_{1 \leq j \leq K} \{ \sum_{a=1}^U \theta_k S_{ajk} \} \right\}}{UV}, 0 \leq g \leq 1, \quad (12)$$

The maximum difference in understanding between any two study groups on each learning task is calculated according to Eq. (13):

$$D_1 = \frac{\text{Max}_{1 \leq k \leq V} \left\{ \text{Max}_{1 \leq j \leq K} \{ \theta_k S_{ajk} \} \right\} - \text{Min}_{1 \leq j \leq K} \{ \sum_{a=1}^U \theta_k S_{ajk} \}}{U}, 0 \leq D_1 \leq 1, \quad (13)$$

The average number of learning tasks that students in the same group are interested in can be calculated according to Eq. (14):

$$D_2 = \text{Max}_{1 \leq i \leq V} \left\{ 1 - \frac{\sum_{j=1}^K \text{Min}\{1, \sum_{a=1}^U K_{ajk}\}}{K} \right\}, 0 \leq D_2 \leq 1 \quad (14)$$

How interested each student is in different learning tasks can be assessed through the student's learning experience.

The aim of the fitness function constructed is to find the optimal solution of the student's knowledge state and interest for a given learning task, and to minimize the difference between any two study groups. The model velocity function used to characterize the update velocity of each particle position in each iteration can be expressed as Eq. (15):

$$\mu_{i+1} = \eta \times \begin{bmatrix} g_{\mu} \\ D_{1\mu} \\ D_{2\mu} \end{bmatrix} + \lambda_1 \times \text{rand}(\ ) \times \left( \begin{bmatrix} g_w \\ D_{1w} \\ D_{2w} \end{bmatrix} - \begin{bmatrix} g \\ D_1 \\ D_2 \end{bmatrix} \right) + \lambda_2 \times \text{rand}(\ ) \times \left( \begin{bmatrix} g_h \\ D_{1h} \\ D_{2h} \end{bmatrix} - \begin{bmatrix} g \\ D_1 \\ D_2 \end{bmatrix} \right) \quad (15)$$

where, the learning factors that affect the local and global optima in the iterative process are represented by  $\lambda_1$  and  $\lambda_2$ , the inertia weight used to measure individual search and global search is represented by  $\eta$ , and the random number in the interval [0,1] is generated by the  $\text{rand}(\ )$  function. The update function for the position of each particle is shown in Eq. (16):

$$W_{b+1} = \begin{bmatrix} g \\ D_1 \\ D_2 \end{bmatrix} + \begin{bmatrix} g_{e_{i+1}} \\ D_{1e_{i+1}} \\ D_{2e_{i+1}} \end{bmatrix} \quad (16)$$

In order to ensure that the students within one group have the greatest differences in their knowledge state for a given learning task, another objective function expression was introduced in this paper:

$$SUF C_2(W_b) = -1 \times \underset{1 \leq i \leq V}{Min} \left\{ \frac{\sum_{j=1}^K (Max_{1 \leq a \leq U} \{s_{ajk}\} - Min_{1 \leq a \leq U} \{s_{ajk}\})}{K} \right\} \quad (17)$$

It can be seen from the above formula that the objective function is defined as the average value of the greatest differences in the knowledge state of members in each group. In other words, first, the average value of the differences in the knowledge state of all members in the group is calculated, and then the smallest average value of the differences for all learning tasks is selected and maximized. Eq. (18) shows the fitness function at this time:

$$\begin{cases} SUF C(W_b) = g + D_1 + D_2 \\ SUF C_2(W_b) \end{cases} \quad (18)$$

Through the above objective function, the single-objective optimization problem of cooperative learning grouping can be changed to a multi-objective optimization problem with knowledge state and interest as the grouping criteria.

#### 4 Experimental results and analysis

The constructed model for diagnosis of students' knowledge state based on the complementarity of knowledge structures was then applied in actual online teaching as an experiment to evaluate the effectiveness of the model, which specifically involved diagnostic efficiency and diagnostic accuracy. The course selected for the experiment was "Building Power Supply and Distribution". The experiment included 10 real-time diagnostic tests on students' knowledge state including one pre-test and one post-test.

From the perspective of classification and regression, this paper verifies the feasibility of the proposed model with three evaluation indicators - diagnostic accuracy, mean square error and *AUC* value. Figure 4 shows the change curves of the loss function values obtained by the conventional diagnosis model and the proposed model based on the sample sets of online learning behaviour data. It can be seen that the loss function value of the proposed model is higher than that of the conventional one, indicating that the proposed model can better simulate the complex online learning process.

Figure 5 shows the change curves of the error values obtained by the conventional diagnosis model and the proposed one for students' knowledge state based on the sample sets of online learning behaviour data. It can be seen that the error values of both models fell first and then rose, and that the minimum value appeared after the first iteration. The mean square error (MSE) value of the proposed model is generally lower, indicating that the proposed model has better stability. Table 1 shows the diagnosis results of the model after training regarding the knowledge state of students based on the 3 sample sets from different online learning platform. It can be seen that the prediction accuracy of the proposed model for students' knowledge state is significantly improved compared with that of the conventional one, which further verifies the effectiveness of the proposed model.

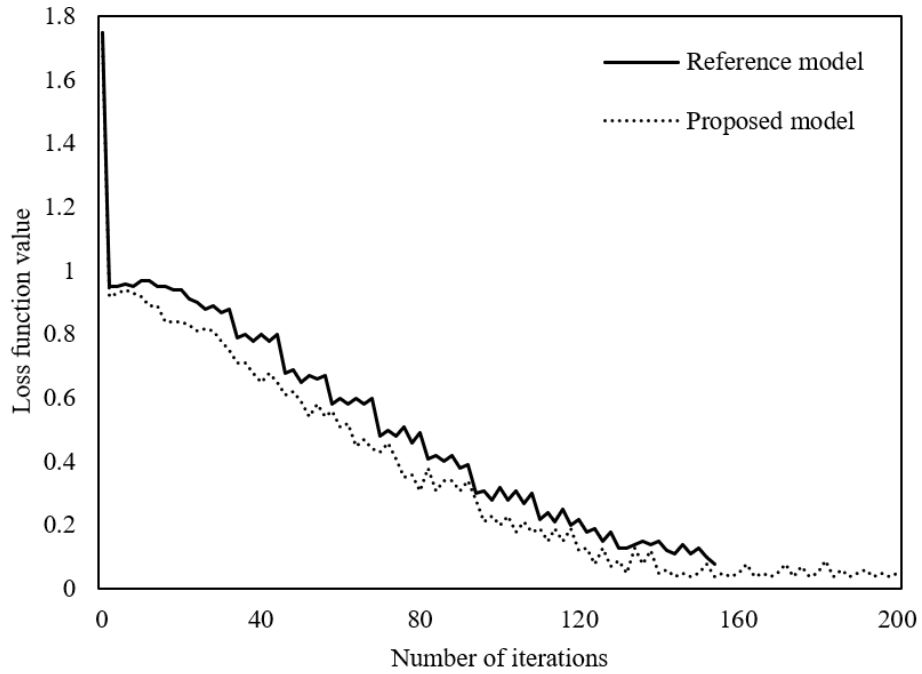


Fig. 4. Loss function curves of different models

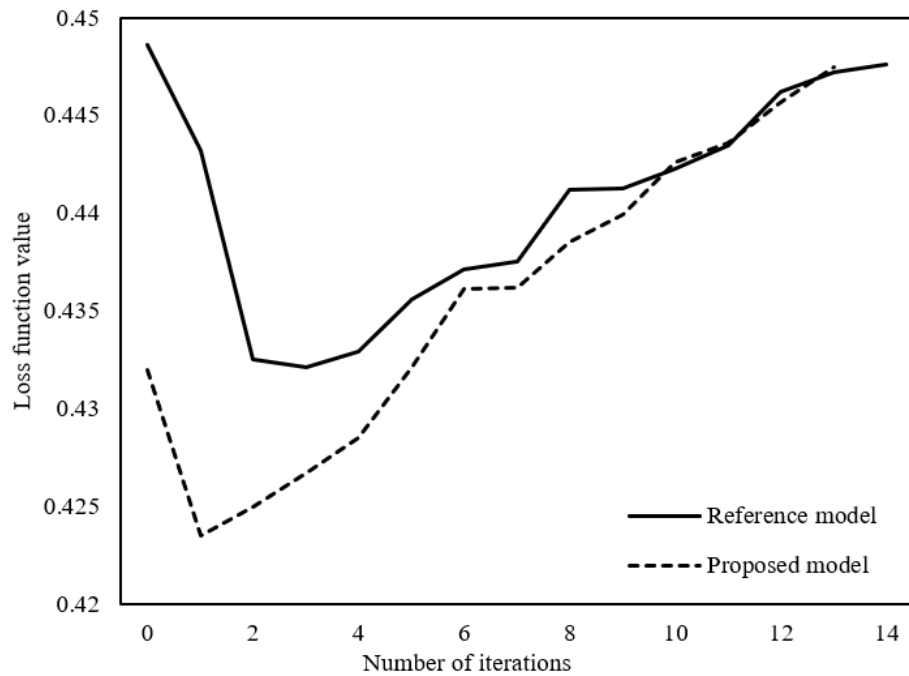


Fig. 5. Error curves of different models

**Table 1.** Prediction results of different students’ knowledge state

Sample set No.		Model No.			
		1	2	3	Proposed model
1	Accuracy	0.765	0.685	0.718	0.732
	MSE	0.451	0.431	0.416	0.420
	AUC value	0.746	0.748	0.801	0.748
2	Accuracy	0.795	0.734	0.719	0.813
	MSE	0.418	0.439	0.428	0.435
	AUC value	0.724	0.692	0.829	0.817
3	Accuracy	0.824	0.624	0.824	0.695
	MSE	0.419	0.382	0.338	0.417
	AUC value	0.736	0.729	0.729	0.727

Before the teaching practice was commenced, a T test based on the T distribution theory was performed on the independent samples of the pre-test results, and the difference between the two average values was calculated to determine whether the difference was significant, so as to find out the knowledge levels of the students in the test group and the control group. Table 2 shows the pre-test results of the students’ knowledge state, where the average scores of the test group and the control group were 78.26 and 79.34, respectively. The significance level of the independent samples in the T test was 0.585, which was greater than 0.05, showing that there was no significant difference in the pre-test results characterizing the knowledge state of the students between the test group and the control group.

**Table 2.** Pre-test results of students’ knowledge state

Group type	Evaluation results	
	Test group	Control group
Number of students	56	51
Mean value	78.26	79.34
Standard error	0.225	0.384
t value	6.28	
Sig	0.585	

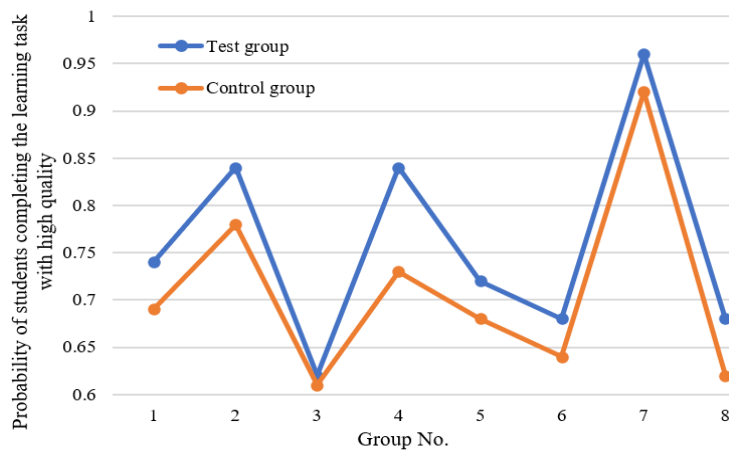
In order to verify whether the grouping method for massive online collaborative learning proposed in this paper can improve the learning effect of students, it is necessary to test whether the pre-test and post-test results of students’ knowledge state satisfy the hypothesis of slope homogeneity. Table 3 shows the post-test results of students’ knowledge state, and the *P* value is 0.662, greater than 0.05, which satisfies the hypothesis of slope homogeneity. It can be seen from Table 3 that the average scores of the test group and the control group were 95.39 and 87.29, respectively, showing that the knowledge state of the students in the test group was significantly better than that of the control group. The significance level of the independent samples in the T test was 0.034, less than 0.05, indicating that the knowledge state of the students was signifi-

cantly different between the test group and the control group. This verifies that the proposed grouping method for massive online collaborative learning can effectively increase the probability of students completing learning tasks with high quality; in other words, it can effectively improve their learning effect.

**Table 3.** Post-test results of students' knowledge state

Group type	Evaluation results	
	Test group	Control group
Number of students	53	49
Mean value	95.39	87.29
Standard error	0.308	0.372
t value	1.62	
Sig	0.034	

In order to find out how satisfied the students participating in the experiment were with the online collaborative learning grouping method in the actual learning process, a questionnaire survey was conducted to investigate three aspects, namely whether the cooperative learning based on the online collaborative learning grouping method proposed in this paper effectively improved the probability of the students completing the learning tasks with high quality, whether it improved their learning quality and efficiency, and whether they liked the teaching model based on the online collaborative learning grouping method. The survey results confirm from the subjective point of view that the students in the test group had a higher probability of completing the learning tasks with high quality than those in the control group. Figure 6 compares the probabilities of students completing the learning tasks with high quality. The survey results are consistent with the post-test results, indicating that the online collaborative learning grouping method has a positive effect on increasing the probability of students completing the learning tasks with high quality.



**Fig. 6.** Comparison of the probabilities of students completing the learning task with high quality

This experiment conducted a comparative analysis of the experimental results of the proposed grouping algorithm and those of the conventional particle swarm algorithm, the exhaust algorithm and the random grouping method. The indicators for comparison included the algorithm efficiency and the quality of the output optimal solution. The running time and fitness values of different grouping algorithms are shown in Table 4. It can be seen that the proposed model was less affected by the number of the students, and its running time and the quality of its output optimal solution were both relatively stable. In particular, when the number of students reached 6000, the proposed model had the best performance compared with the other methods.

**Table 4.** Comparison of the running time and fitness values of different algorithms

Number of students Algorithm No.		100	300	500
1	t	18.26	36.25	572.34
	f	(0.02,-0.85)	(0.03,-0.81)	(0.06,-0.83)
2	t	0.025	0.032	0.041
	f	(0.36,-0.17)	(0.42,-0.21)	(0.37,-0.32)
3	t	1.32	9.85	22.14
	f	(0.03,-0.46)	(0.02,-0.41)	(0.01,-0.36)
Proposed algorithm	t	0.85	11.25	23.16
	f	(0.01,-0.78)	(0.04,-0.76)	(0.06,-0.73)
Number of students Algorithm No.		800	1000	1500
1	t	52.34	/	/
	f	(0.09,-0.88)	(0.06,-0.89)	/
2	t	0.085	0.067	0.027
	f	(0.47,-0.12)	(0.74,-0.26)	(0.83,-0.14)
3	t	56.37	98.35	107.56
	f	(0.03,-0.34)	(0.02,-0.27)	(0.06,-0.25)
Proposed algorithm	t	38.25	71.24	91.64
	f	(0.05,-0.71)	(0.03,-0.67)	(0.01,-0.63)
Number of students Algorithm No.		2500	4000	6000
1	t	/	/	/
	f	/	/	/
2	t	0.038	0.084	0.062
	f	(0.93,-0.57)	(0.95,-0.27)	(0.88,-0.16)
3	t	123.59	189.47	256.14
	f	(0.08,-0.23)	(0.03,-0.17)	(0.09,-0.15)
Proposed algorithm	t	108.37	158.29	178.24
	f	(0.04,-0.61)	(0.02,-0.54)	(0.03,-0.45)

## 5 Conclusions

This paper studied a student grouping method for massive online collaborative learning. In this study, a diagnosis model for students' knowledge state based on gated recurrent neural network was constructed to simulate the cooperative learning process, and the enhanced particle swarm optimization algorithm was adopted to achieve the grouping of students for massive online collaborative learning. Based on the experimental results, the change curves of the loss function values and the error values obtained by the conventional diagnosis model and the proposed model for students' knowledge state from the sample sets of online learning behaviour data were plotted, and the prediction results of different students' knowledge state were given, which verified the effectiveness of the proposed model. The T test was performed on the independent samples of the pre-test results, which proves that the proposed grouping method for massive online collaborative learning can effectively increase the probability of students completing the learning task with high quality, thereby improving their learning effect. After that, the running time and fitness values of different grouping algorithms were compared, which shows that the proposed model has the best performance compared with the other ones.

## 6 References

- [1] Chen, F.S., Ke, H.S., Chen, Y.C. (2020). Online Learning as a Panacea? An Empirical Study to Discuss Problem-Based Cooperative Learning in Taiwan, *International Journal of Emerging Technologies in Learning*, 15(18): 251-259. <https://doi.org/10.3991/ijet.v15i18.15079>
- [2] Rahman, A., Ahmar, A., Rusli, R. (2016). The influence of cooperative learning models on learning outcomes based on students' learning styles. *World Transactions on Engineering and Technology Education*, 14(3): 425-430. <https://doi.org/10.26858/wtetev14i3y2016p6425430>
- [3] Chan, K.K., Zhou, Y.C. (2020). Effects of Cooperative Learning with Dynamic Mathematics Software (DMS) on Learning Inversely Proportional Functions, *International Journal of Emerging Technologies in Learning*, 15(20): 210-225. <https://doi.org/10.3991/ijet.v15i20.14339>
- [4] Liu, J. (2016). Study on the autonomous learning of college English based on online learning platform. *International Journal of Smart Home*, 10(2): 165-174. <https://doi.org/10.14257/ijsh.2016.10.2.16>
- [5] Sun, Y., Chai, R.Q. (2020). An early-warning model for online learners based on user portrait. *Ingénierie des Systèmes d'Information*, 25(4): 535-541. <https://doi.org/10.18280/isi.250418>
- [6] Wang, S.Y. (2021). Online learning behavior analysis based on image emotion recognition. *Traitement du Signal*, 38(3): 865-873. <https://doi.org/10.18280/ts.380333>
- [7] Wen, J., Wei, X.C., He, T., Zhang, S.S. (2020). Regression analysis on the influencing factors of the acceptance of online education platform among college students. *Ingénierie des Systèmes d'Information*, 25(5): 595-600. <https://doi.org/10.18280/isi.250506>
- [8] Jong, B.S., Chen, C.M., Chan, T.Y., Lin, T.W., Hsia, Y.T. (2014). Effect of knowledge complementation grouping strategy for cooperative learning on online performance and

- learning achievement. *Computer Applications in Engineering Education*, 22(3): 541-550. <https://doi.org/10.1002/cae.20579>
- [9] Chen, X., Qin, Z., An, L., Bhanu, B. (2015). Multiperson tracking by online learned grouping model with nonlinear motion context. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(12): 2226-2239. <https://doi.org/10.1109/TCSVT.2015.2511480>
- [10] Zhou, N., Zhang, Z.F., Li, J. (2020). Analysis on course scores of learners of online teaching platforms based on data mining. *Ingénierie des Systèmes d'Information*, 25(5): 609-617. <https://doi.org/10.18280/isi.250508>
- [11] Chaudhry, R., Singh, H., Dogga, P., Saini, S.K. (2018). modeling hint-taking behavior and knowledge state of students with multi-task learning. *International Conference on Educational Data Mining*, Raleigh, NC, pp. 21-31. <https://doi.org/10.29007/dj6b>
- [12] Lai, Z., Wang, L., Ling, Q. (2021). Recurrent knowledge tracing machine based on the knowledge state of students. *Expert Systems*, 38(8): e12782. <https://doi.org/10.1111/exsy.12782>
- [13] Chang, C.T., Tsai, C.Y., Tsai, H.H., Li, Y.J., Yu, P.T. (2020). An online multi-user real-time seamless co-reading system for collaborative group learning. *International Journal of Distance Education Technologies (IJDET)*, 18(4): 51-70. <https://doi.org/10.4018/IJDET.2020100104>
- [14] Chen, Y., Luo, H., Han, X., Zhang, J., Nie, Y. (2020). Does it matter to be a group leader? exploring the impact of assigned leadership on small group online collaborative learning. *2020 International Symposium on Educational Technology (ISET)*, Bangkok, Thailand, pp. 288-292. <https://doi.org/10.1109/ISET49818.2020.00069>
- [15] Xie, K., Hensley, L.C., Law, V., Sun, Z. (2019). Self- regulation as a function of perceived leadership and cohesion in small group online collaborative learning. *British Journal of Educational Technology*, 50(1): 456-468. <https://doi.org/10.1111/bjet.12594>

## 7 Authors

**Yuan Wang**, graduated from Hebei University of Architecture. She has been teaching in Shijiazhuang University of Applied Technology since 2006. Her main research areas are construction electrical, project management and project cost (Email: wangyuan@sjzpt.edu.cn).

**Qi Wang**, graduated from the Hong Kong Baptist University with a master's degree. In August 2007, she entered blue interactive digital marketing company under WPP group and successively served as business analyst and senior media planner. In 2011, she officially joined Shijiazhuang Posts and Telecommunications Technical College and is currently the specialty leading person of network marketing and live e-commerce. Her research areas are communication, new media marketing and live e-commerce.

Article submitted 2022-01-01. Resubmitted 2021-02-05. Final acceptance 2022-02-07. Final version published as submitted by the authors.