A Deep Learning-Based Grouped Teaching Strategy for Experimental Training

Shaoxia Mou1,2, Shan Zhang3, Jiawei Chen3, Haiyan Zhai4, Yuandi Zhang5

1 Graduate School Eternal University, University of Perpetual Delta, Las Piñas 6015, Philippines
2 Foreign Language College, Shandong Jianzhu University, Jinan 250101, China
3 Transportation Integrated Administrative Law Enforcement Battalion of Jinan, Jinan 250014, China
4 Business Administration of Jose Rizal University, Mandaluyong 1552, Philippines
5 School of Mathematics and Information Science, Shandong Technology and Business University, Yantai 264000, China

Corresponding Author Email: ydfqly0308@163.com

https://doi.org/10.18280/isi.270216

Received: 8 December 2021
Accepted: 10 March 2022

Keywords:
deeplarning, student grouping, experimental training

ABSTRACT

The diagnosis of students’ pre-class hands-on operation capacity helps to group the students scientifically to facilitate the teaching for experimental training of skill training courses, and benefits the capacity complementation and team spirit cultivation. Practically speaking, the relevant results have rarely been implemented in the practical teaching of skill training courses. No literature has analyzed the correlation between grouped teaching for experimental training, and the key competences of students for skill training courses. To solve the problem, this paper constructs a grouped teaching strategy for experimental training based on deep learning. The key competences of students for skill training courses were evaluated from four aspects, namely, general training objective, thinking training, practical training, and scientific literacy training. The evaluation results were used to diagnose the students’ pre-class hands-on operation capacity, and reasonably group the students participating in experimental training. The deep learning model was called to group students of different majors for experimental training courses. The attention mechanism was added to prevent over fitting, when there are a few samples for capacity diagnosis. The proposed model was proved effective through experiments.

1. INTRODUCTION

Experimental teaching underpins skill training courses. The teaching for experimental training of skill training courses is an essential link of improving students’ skills, abilities, and key competencies [1-6]. It directly relies on group teaching of the students. Grouped teaching for experimental training facilitates students to understand the concepts and teaching rules of skill training courses [7-14]. The diagnosis of students’ pre-class hands-on operation capacity helps to group the students scientifically to facilitate the teaching for experimental training of skill training courses, and benefits the capacity complementation and team spirit cultivation [15-22]. The main difficulty of the diagnosis is how to accurately judge the key competencies of different types of students in the experimental training of skill training courses, and to divide these students into reasonable groups [23, 24].

The traditional experimental teaching of management emphasizes teaching over learning, and stresses knowledge over practice. It no longer meets the development needs of innovative society in the context of new media. In the new media environment, the experimental courses of management are badly in need of collaborative innovation of the teaching system. Drawing on the new media environment, Xia et al. [25] carried out collaborative innovative design of the teaching system for the experimental courses of management, from the aspects of overall design, teaching method, and teaching evaluation. Through practice and application, the experience and shortcomings were summarized, and the improvement measures were presented. Wen et al. [26] put forward a lightweight collaborative experimental teaching method, including computer simulation, teaching guidance, and class reflection. Thirty-three grade 8 students were recruited to explore how the teaching method affects the exploratory scores and scientific knowledge of students. However, further research is needed to verify whether a series of lightweight experiments can improve student performance. Girouard and Kang [27] created an innovative usability training called collaborative learning of usability experiences program for students majoring in human-computer interaction (HCI) of different grades, and detailed the components of the unique experiential training, namely, internship, short-term courses, symposium, and knowledge transfer. In addition, the employment and academic performance of graduates were evaluated, and the five key skills of interns were checked. Kerpen et al. [28] proposed a problem-based lab-centered college course. The course teaching has a high demand for virtual collaboration. The students used various tools to draw a preliminary design plan, which could be realized in future physical prototypes.

The existing literature shows that grouped learning and pre-class hands-on operation capacity are necessary and important to the teaching for experimental training of skill training courses. Practically speaking, the relevant results have rarely been implemented in the practical teaching of skill training courses. No literature has analyzed the correlation between grouped teaching for experimental training, and the key competences of students for skill training courses. To make up
for the gap, this paper constructs a grouped teaching strategy for experimental training based on deep learning. The main contents are as follows: Section 2 evaluates the key competences of students for skill training courses from four aspects, namely, general training objective, thinking training, practical training, and scientific literacy training. The evaluation results can be used to diagnose the students’ pre-class hands-on operation capacity, and reasonably group the students participating in experimental training. Section 3 calls the deep learning model to group students of different majors for experimental training courses, combined meta learning with transfer learning, and introduced the attention mechanism to prevent over fitting, when there are a few samples for capacity diagnosis. The proposed model was proved effective through experiments.

2. CAPACITY DIAGNOSIS

The key competences of students for skill training courses can be evaluated from four aspects: general training objective, thinking training, practical training, and scientific literacy training. The evaluation results can be used to diagnose the students’ pre-class hands-on operation capacity, and reasonably group the students participating in experimental training.

If a student possesses the key competencies of skill training courses, he/she should be able to construct the knowledge system of skill training courses, design the experimental training plan based on the learned basic knowledge or experimental principles, participate in grouped experimental training, and finally acquire valuable experimental conclusions or experience of skill training. The mastery of course knowledge and experimental conclusions measures the degree of realization of the general training objective of the courses. Therefore, the general training objective of skill training courses was further divided into basic knowledge or experimental principles, and valuable experimental conclusions or experience of skill training.

To evaluate the key competences of students for skill training courses, it is necessary to examine the student capacity for problem analysis, reasoning, innovative thinking, and problem solving in the skill training course scenes closely associated with practical application. Grouped discussion can effectively optimize or renovate the ideas. Therefore, the thinking training of skill training courses was further split into problem analysis, reasoning, innovative thinking, and problem solving.

The key competences of students for skill training courses highlight that the students must be able to combine, control, and operate experimental training equipment, and to generate experimental data. The entire flow cannot be completed by one student. Normally, the experiments are carried out collaboratively by several groups of students. Each group is assigned a specific task. Therefore, the practical training of skill training courses was further divided into combination training (grouped), control training (grouped), operation training (grouped), experimental phenomenon observation, data collation, as well as communication and discussion.

The key competences of students for skill training courses also require students to have a good research spirit and learning attitude. Therefore, the scientific literacy training of skill training courses is further divided into the curiosity about training topics, scientific reasoning, and sustained scientific spirit.

3. GROUPING STRATEGY

Based on deep learning model, this paper groups students in different majors for experimental training courses. The students in different majors need to complete different skill training courses, and acquire different key competencies. As a result, there are a few samples for the diagnosis of students’ pre-class hands-on operation capacity.

The small sample set can be improved by data enhancement, transfer learning, or meta learning. This paper combines meta learning with transfer learning, and introduces the attention mechanism to prevent over fitting, when there are a few samples for capacity diagnosis.

As shown in Figure 1, the overall framework of our model consists of a dense residual module, an attention module in global channel and local space, and an attention module between spatial channels. Firstly, the proposed model is trained by the set of original capacity diagnosis samples. The weights of the trained model are finetuned to adapt to the grouping of students for experimental training courses, in the presence of a small sample set.

To overcome vanishing or exploding gradients, this paper introduces a skip connection to the ResNet in the dense residual module of our model. Then, the nonlinear operation can be skipped, once the optimal solution is obtained. Let \( F(a_{k-1}) \) be the residual obtained through nonlinear operation; \( a_{k-1} \) be the optimal solution before executing the nonlinear operation. Then, we have:

\[
 a_k = F_k \left( a_{k-1} \right) + a_{k-1}
\]  

(1)

The DenseNet in the dense residual module adopts a new connection mode to promote the gradient transmission between layers, i.e., the output of each layer is directly transmitted to all subsequent layers. That is, the input of the k-th layer contains the feature maps from the 0-th layer to the k-1-th layer. \( \left[ a_0, a_1, \ldots, a_{k-1} \right] \) be the splicing of the eigenvectors from the 0-th layer to the k-1-th layer. Then, we have:

\[
 a_k = g_k \left( \left[ a_0, a_1, \ldots, a_{k-1} \right] \right)
\]  

(2)

To enhance the feature transfer of capacity diagnosis samples in network training, this paper embeds the dense connection mode of DenseNet to ResNet. Let \( a_{10}, \) and \( a_{11} \) be
the medians of each dense residual block. Taking the residual block structure of ResNet as the basis, the dense residual blocks can be calculated with skip connection as network mapping:

\[
a_i = G(a_{i-1}) + a_{i-1}
\]

\[
G(a_{i-1}) = G([a_{i-1}, a_{i-2}, \ldots , a_{i-10}, a_{i-11}])
\]

(3)

The purpose of the attention module in global channel and local space, and the attention module between spatial channels is to improve the quality of feature representation, i.e., highlight the features of important capacity diagnosis samples, and weaken the features of unimportant samples. Figure 2 illustrates the structure of an attention module.

Let \( A \in \mathbb{R}^{Q \times Q} \) be the eigenvector outputted by the convolutional layer after the input capacity diagnosis sample passes through that layer; \([a_i, a_{i+1}, \ldots , a_{i+l}] \) be channels \( a_i-a_{i+l} \); \( A_i \in \mathbb{R}^{Q \times Q_i} \) be the feature information in the \( i \)-th channel; \( U = [a_i, a_{i+1}, \ldots , a_{i+l}] \) be the filter group for learning; \( u_i \) be the \( i \)-th filter; \( F_i \) be the function of the deep convolutional layer; \( u_i \) be the two-dimensional (2D) spatial convolutional filter. The spatial statistics in a channel can be generated by \( F_i \), and used to generate the output eigenvector \( B_i = B_i = [b_{i1}, b_{i2}, \ldots , b_{ik}] \):

\[
b_i = F_i(a_i, u_i) = a_i * u_i
\]

(4)

Let \( F_i \) be the modeling of the correlation between global channel features; \( \xi \) be the rectified linear unit (ReLU). Suppose parameters \( T_1 \) and \( T_2 \) satisfy \( T_1 \in \mathbb{R}^{l \times l \times Q \times Q_1} \) and \( T_2 \in \mathbb{R}^{l \times l \times Q_1 \times Q_2} \), respectively. Then, the output eigenvector \( B_i \in \mathbb{R}^{Q \times Q_2} \) can be obtained by:

\[
B_i = F_i(A, T) = T_2 * \xi(T_1 * A)
\]

(5)

The compression mechanism and the relationship between global channel indices can be obtained through the dimensionality reduction of channels with \( T_1 \). Let \( F_{\text{hi}} \) be element-by-element multiplication. By sigmoid function, \( B_i \) and \( B_o \) derive the activation value for re-calibration or activating the feature information:

\[
B_{\text{hi}} = F_{\text{hi}}(A \cdot \varepsilon(B_o), \varepsilon(B_i)) = A \cdot \varepsilon(B_o) \cdot \varepsilon(B_{\text{hi}})
\]

(6)

Let \( b \in \mathbb{R}^2 \) be the feature of each capacity diagnosis sample passing through the global average pooling. The channel attention mechanism can be given by:

\[
\theta = \varepsilon(Q \cdot b)
\]

(7)

Let \( Q \) be the \( Z^*Z \) parameter matrix; \( Q \) be the band matrix. The band matrix provides the basis for the learning of channel information attention. The matrix can be expressed as:

\[
\begin{bmatrix}
q_1 & \cdots & q_{n-1} & 0 & 0 & \cdots & 0 \\
0 & q_1 & \cdots & q_{n-1} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & q_1 & \cdots & q_{n-1} & 0 \\
0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & \cdots & \cdots & 0 & 0
\end{bmatrix}
\]

(8)

Formula (25) shows that \( Q \) contains \( n^2 \) parameters. It is only necessary to consider the interaction between \( b_i \) and its 1 adjacent channels. Let \( \Psi_i \) be the set of \( b_i \) and its 1 adjacent channels. Then, the weight of \( b_i \) can be calculated by:

\[
\theta_i = \varepsilon \left( \sum_{j=1}^{n} q_i b_j \right), b_j \in \Psi_i
\]

(9)

To improve the effectiveness of model learning, the learning parameters of all channels should be unified. Thus, formula (26) can be updated as:

\[
\theta_i = \varepsilon \left( \sum_{j=1}^{n} q_i b_j \right), b_j \in \Psi_i
\]

(10)

To keep the learning parameters of all channels consistent, a fast one-dimensional (1D) convolution \( CON_{FD, l} \) can be implemented with a kernel size of \( l \):

\[
\theta = \varepsilon (CON_{FD, l}(b))
\]

(11)

The above formula is the way to compute the attention between channels, and to effectively capture the local information exchange between channels. The computation is of low complexity, involving only \( l \) parameters.

To realize the attention mechanism between channels, it is necessary to determine the kernel size \( l \), and obtain a relatively suitable coverage of adjacent channels. The possible mappings \( \Phi \) between \( l \) and \( Z \) can be nonlinearized:

\[
Z = \Phi(l) = 2^{\alpha \cdot \omega}
\]

(12)

Let \( |x|_{\text{odd}} \) be the odd number the closest to \( x \). If the channel dimension \( Z \) is given, then \( l \) can be adaptively calculated by:

\[
l = Q(Z) = \left\lfloor \frac{\log_2(Z) - \frac{1}{2}}{\alpha} \right\rfloor
\]

(13)

To prevent our model from over fitting, meta learning was integrated with transfer learning, which contains lightweight operations like scaling and translation, to initialize neural network parameters on deep layers. The integration enables meta learning to converge quickly facing a few capacity diagnosis samples, thereby reducing the computing load. Figure 3 shows the flow of the proposed meta-transfer learning model.

During model learning, the first step is to train the capacity diagnosis sample set on multiple task levels. Let \( TT \) be the training task; \( \delta_1 \), \( \delta_2 \) and \( \delta_{1,2} \) be the scaling operation, translation operation, and the combination between the two operations, respectively. Facing a small set of diagnosis samples, the student grouping classifier can be optimized by:

\[
\omega' = \omega - \alpha \nabla_{\omega} \hat{R}(\Pi, \omega, [\delta_{1,2}])
\]

(14)

\[
\delta_1 \text{ and } \delta_2 \text{ can be iteratively updated by:}
\]

\[
\delta_i = \delta_i - \alpha \nabla_{\delta_i} \hat{R}(\Pi, \omega, [\delta_{1,2}]) (i = 1, 2)
\]

(15)

Let \( \Pi \) be the trained feature extractor of capacity diagnosis samples. Suppose the network layer of \( \Pi \) contains \( n \) pairs of weights and thresholds \( \{(Q_{\omega}, \delta_{1,2})\} \). Let \( \odot \) be the point multiplication operation. Then, the combined operation of scaling and translation can be expressed as:

\[
RR(A, Q, \gamma; \delta_{1,2}) = (Q \odot \delta) A + (\gamma + \delta_2)
\]

(16)
The students participating in experimental training achieved relatively good recognition accuracy on mapping, Eigenvector, and our model after 300 iterations. Therefore, our model has a high robustness in the sample set. Therefore, our model has a high robustness in the sample set. Therefore, our model has a high robustness in the sample set. Therefore, our model has a high robustness in the sample set.

4. EXPERIMENTS AND RESULTS ANALYSIS

This paper evaluates the key competences of students for skill training courses from four main aspects: general training objective, thinking training, practical training, and scientific literacy training. The proposed method was verified experimentally on the evaluation dataset of students from different majors and colleges. The results can be applied to diagnose students’ pre-class hands-on operation capacity, and group the students participating in experimental training scientifically. The grouping rule is to allocate the students with different levels of pre-class hands-on operation capacity to the same group.

A comparative experiment of feature extraction effects was designed to verify the effectiveness of our model. Figure 4 reports the experimental results. The proposed model was tested at different number of parameters. In the absence of the attention mechanism, the transfer learning effect was significantly affected by the number of parameters. Our model converged faster and achieved better results than the traditional transfer learning model. As shown in Figure 4, when the sample set of students’ pre-class hands-on operation capacity was very small, the mean classification accuracy on sample set 2 was slightly lower (-5%) than that on sample set 1, for many group features and samples are missing in that sample set. Therefore, our model has a high robustness in the student grouping for experimental training.

To further verify its effectiveness, our model was compared with 9 models, including k-nearest neighbors (KNN), Bayesian, decision tree, random forest, support vector machine (SVM), logistic probability regression, discriminant analysis, neural network, and ensemble boosting. The KNN is only applicable to easily explainable models, sensitive to abnormal values, and susceptible to data imbalance. Bayesian works well on the data with weak correlations between dimensions, and gains popularity in the handling of spam emails. Decision tree can handle different types of data simultaneously, but is vulnerable to attacks. Random forest is not so vulnerable to attacks, and suitable for high accuracy recognition of low-dimensional data. The SVM can withstand attacks effectively. As a linear classifier, logistic probability regression does not apply to strongly correlated features. Discriminant analysis is usually adopted as a dimensionality reduction tool. The neural network is fit for large volumes of data. Ensemble boosting can automatically detect effective features, facilitating the understanding of high-dimensional data.

Table 1 compares the classification accuracies of different models. Obviously, discriminant analysis, neural network, ensemble boosting, and our model achieved relatively good classification results. The best results were realized by our model, because the attention mechanism is suitable for the diagnosis of a few capacity samples.

In addition, the losses of discriminant analysis, neural network, ensemble boosting, and our model after 300 iterations were counted. As shown in Figure 5, our model outperformed the other three models: the loss dropped by 0.6 at the most. The result further confirms that our model can highlight the features of important capacity samples, thanks to the introduction of the attention mechanism. During the training, our model saw a faster decline of the loss function, i.e., a faster convergence, than the other three models. Figure 6 visually displays the grouping effect of experimental training on the sample sets. The effectiveness of our grouped teaching strategy for experimental training was further validated.

Table 1. Classification accuracies of different models

<table>
<thead>
<tr>
<th>KNN</th>
<th>Bayesian</th>
<th>Decision tree</th>
<th>Random forest</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>82.05±0.02</td>
<td>86.58±0.63</td>
<td>81.47±0.1480.85±0.3887.62±0.32</td>
<td></td>
</tr>
<tr>
<td>Dataset 2</td>
<td>85.14±0.75</td>
<td>88.51±0.49</td>
<td>86.89±0.4183.68±1.3581.48±0.74</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Logistic probability regression</th>
<th>Discriminant analysis</th>
<th>Neural network</th>
<th>Ensemble boosting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>86.83±0.53</td>
<td>91.48±0.42</td>
<td>96.52±0.1991.71±0.6293.68±0.47</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>89.28±0.38</td>
<td>97.42±0.24</td>
<td>91.36±0.1896.58±0.3697.42±0.51</td>
</tr>
</tbody>
</table>

Figure 4. Feature extraction effects on different capacity evaluation samples
5. CONCLUSIONS

This paper tries to build a grouped teaching strategy for experimental training based on deep learning. The key competences of students for skill training courses were evaluated from four aspects, namely, general training objective, thinking training, practical training, and scientific literacy training. The evaluation results can be used to diagnose the students’ pre-class hands-on operation capacity, and reasonably group the students participating in experimental training. Based on the deep learning model, the students of different majors were grouped for experimental training courses, and the attention mechanism was introduced to prevent over fitting in the face of a sample set. Then, the feature extraction effects on different capacity evaluation samples were compared experimentally. The results show that our model has a high robustness in the student grouping for experimental training. In addition, our model was compared with 9 models, including KNN, Bayesian, decision tree, random forest, SVM, logistic probability regression, discriminant analysis, neural network, and ensemble boosting. It was observed that discriminant analysis, neural network, ensemble boosting, and our model achieved relatively good classification results. The best results were realized by our model. In addition, the losses of discriminant analysis, neural network, ensemble boosting, and our model after 300 iterations were counted. During the training, our model saw a faster decline of the loss function, i.e., a faster convergence, than the other three models. Finally, the authors displayed the grouping effect of experimental training on the sample sets.

REFERENCES


