Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 1, pp. 341–354, DOI 10.12694/scpe.v25i1.2299

TRADITIONAL CULTURAL NETWORK ONLINE EDUCATION INTEGRATING DEEP LEARNING AND KNOWLEDGE TRACKING

HENG ZHAO* AND ZHIYUAN SUN†

Abstract. With the popularization of the computer network and the development of artificial intelligence technology, the traditional education industry has been reformed. In the past two years, online education has developed rapidly. The combination of the Internet and education enables students to study online at any time, no longer relying on the time and place requirements in traditional education. However, with the rapid development of online education, many problems have gradually emerged. In online education, with a large amount of knowledge and question banks, students are faced with a large number of choices. Therefore, positioning and tracking the knowledge level of students and realizing personalized online education have become the main problems facing the moment. Based on this, this study integrates deep learning and knowledge tracking technology to build a traditional cultural network online education model, aiming at accurately positioning students' knowledge levels and recommending personalized question banks. The experimental results show that the average AUC of the model proposed in this study is 0.781, and the average accuracy rate is 0.886, which is significantly better than other online education models. Through the combination of deep learning and knowledge tracking technology, the research successfully provides a new and efficient model for personalized learning in the field of online education, which is of great guiding value for promoting further innovation and development of online education. In addition, the research also provides practical solution strategies for related fields, which have obvious practical significance and popularization value.

Key words: Online education; Personalized recommendation; Deep learning, Knowledge tracking cost

1. Introduction. As the popularization of the Internet and the growth of artificial intelligence technology, online education has gradually become one of the trends of modern education [1]. With the advantages of the Internet and big data, online education has changed the demand for time and location in traditional education [2]. And it can provide users with rich, real-time updated learning materials, reducing the gap between the education level of different development levels in the region [3]. However, with the gradual development of online education, more problems are gradually revealed. These problems are as follows: teachers can not follow up the learning status of students in real time; students in the knowledge base is difficult to choose; the increase of meaningless learning and other problems. The current online education industry practitioners are plagued by these problems [4]. In the traditional teaching mode, teachers are difficult to rely on manual modeling of learning resources for each student to achieve the screening of learning, and they are also unable to accurately perceive the learning changes of learners. As the advancement of artificial intelligence technology, each learner can easily build a unique learning resource model. However, how the huge database of teaching resources can be accurately recommended to students to achieve their personalized development is still a major problem that needs to be solved [5]. To address this problem, this study proposes a traditional culture network online education model that integrates deep learning (DL) and knowledge tracking (KT), aiming to discover learning resources suitable for learners from a large amount of data through DL. And through KT technology, students' learning data are mined and used for online education, aiming to improve the correct rate of predicting students' correct answers to questions, thus improving the effectiveness of online teaching. In addition, the research innovatively combines knowledge mapping to obtain potential connections between items through the description of semantic associations. The research also expands users' interests through various types of associative relationships, which improves the accuracy and diversity of traditional online education recommendation systems. The research aims to combine modern science and technology to provide more effective aids for online teaching and to help learners learn more efficiently. The research also provides a

*School of Tourism and Hotel Management, The Open University of Shaanxi, Xi'an, 710119, China (hengzhaohz@outlook.com)
†School of Marxism, The Open University of Shaanxi, Xi'an, 710119, China (xuegen2021@163.com)

reference for the application of DL and KT in different fields. This research is composed of five main parts. The first part is an overview of the research as a whole; the second part is a summary of related work at home and abroad; the third part is divided into two subsections; the first one describes the model improvement of KT algorithms based on DL, and the second one describes the construction of KT online education models based on DL; the fourth part is an experimental validation of the performance of the KT models proposed in the study; the fifth part is a summary of this study, and it has been an outlook on the future research.

2. Related works. Recently, research on DL has attracted wide attention in various fields. Shen L et al. believed that the short-term forecast of passenger flow had an important significance to the scheduling of subway systems and other aspects. An improved gravity model with DL was proposed, and a short-term forecasting method was upgraded to balance model interpretability and forecasting accuracy. This method had good performance in the experimental results, which was better than other models [6]. This method provided a more reliable solution for improving the accuracy of short-term forecasting of passenger flow in subway systems. In the data-driven decision support system, Garg et al. designed a new data processing framework based on a DL network by using the combination of density-based clustering and variable-length sequence decoding method. The concept of data association was used to track the occurrence and evolution of such events to understand the operating environment. Simulation experiment denoted that this method could perceive the sensitivity of performance and parameters better than others [7]. The method provided a new reliable reference for DL in data processing. Tang et al. found that the linear defect detection of a photovoltaic module was a key link in the health assessment of photovoltaic devices, but the traditional defect diagnosis was actually inefficient manually. Therefore, by distributing computing tasks among edge devices, edge servers and cloud servers, a defect detection algorithm was proposed based on DL. The effectiveness and accuracy of the method were verified by experiments [8]. The method simplified the task of allocating computation between edge and cloud servers for big data processing. Bernardini et al. introduced a DL framework to predict baryon fields. The new model included two network parts, U-Net and Wasserstein, and combined the universe volume and amplified fluids from the real-life environmental feedback project. The dynamical simulation was used to represent a large range of scales. The experimental results indicated that the accuracy was within 10%, which was in good agreement with the cosmological simulation [9]. The research introduced DL to cosmological tasks, providing a viable solution for baryon field prediction. Liu C et al. developed a hierarchical approach based on DL network embeddings to identify patient subtypes from large-scale patient somatic mutation profiles. A network embedding approach encoded genes on the protein interaction set to construct patient vectors. The high classification accuracy indicated that the web-embedded based patient features were reliable in classifying patients [10]. This study provided a powerful DL approach for personalized cancer treatment.

As online education has become a popular research field, the research on data mining technology in online education has also attracted the attention of scholars, and the KT model has become a key technology for simulating the state of students. Song X et al. believed that there were problems in KT research that considered the single relationship and the lack of interpretability, so they proposed a deep KT framework based on a joint graph convolutional network, which modeled multi-dimensional relationships as graphs for relationship fusion. Experiments proved that the method performed better than others. And the interpretability of learning analysis was proved through a case study [11]. The study provided a more effective solution to the problem of considering a single relationship with a lack of interpretability. Pavlik PI et al. introduced a formal learner modeling method - logical knowledge tracing (LKT), which integrated many existing learner modeling methods. The advantage was that the logistic regression model of choice provided the specification of a symbolic system that could specify many existing models in the literature and many new ones. Experimental outcomes indicated that the approach considering multiple learner model characteristics and learning environments was correct [12]. This study integrated existing learner modeling approaches and provided an easier reference for future research. To alleviate the sparsity problem in the summary KT online learning system, Gan W et al. proposed an enhanced learning model with an attention mechanism for graph representation, and used the model to skill embeddings. Experiments from three datasets verified the superiority and interpretability of this model [13]. This research provided a reliable solution to the sparsity that exists in online learning systems. Huang Y et al. found that it was difficult for students to find suitable exercises from a large number of topics provided by many online systems in programming training. This study found a new model for KT, which added additional information

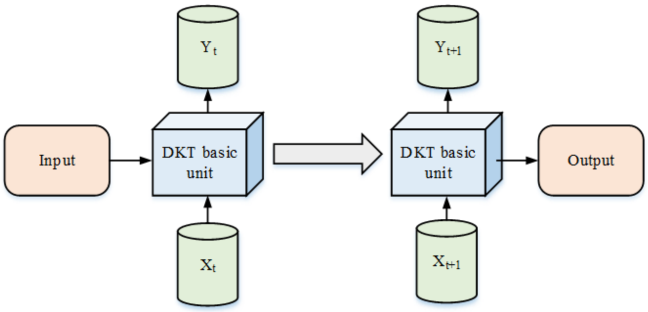


Fig. 3.1: DKT model structure

representing the relationship between exercises to the input data and compressed the input vector to solve the dimensionality problem. Research outcomes proved that the area of the proposed deep KT model was 0.7761, which was better than other KT models and ran faster [14]. This research addressed the dimensionality problem in KT for online teaching and learning systems. Liu S et al. found that the current KT model only attributed the learner's feedback, while ignoring the influence of mental ability factors. Therefore, a new ability-enhanced KT model has been found. It introduced the ability factor into the attribution of feedback, and designed a continuous matrix factorization model. The results proved that the proposed method was better than other KT algorithms in terms of quantitatively evaluated predictive accuracy [15]. This study innovatively introduced the influence of mental ability factors to optimize the KT system more comprehensively.

To sum up, DL is widely used in computer algorithms to improve the accuracy of algorithm models. At the same time, KT models have also been widely used and developed in online education. Therefore, this study innovatively integrates DL and KT, and proposes a new type of traditional cultural online education system. To further improve the interpretability of the deep KT model, the study is based on a dynamic key-value memory neural network, which incorporates two main memory modules for storing knowledge conceptual information and students' state information. In addition, the study recodes the data with the embedding method to avoid the waste of memory resources caused by too high dimensionality. The study aims to provide a more efficient method for accurate recommendation of online learning resources.

3. DL combined with KT online education model.

3.1. Improvement of KT algorithm model based on DL. With the popularity of online education, KT has become one of the best hot research directions. KT uses students' historical learning information to mine students' learning data for online education, with the purpose of improving the correct rate of predicting students' correct answers to questions [16]. The traditional DL-based KT algorithm model, which is named deep knowledge tracing (DKT), is constructed on the long-short-term memory network (LSTM) model. The LSTM model is based on the recurrent neural network (RNN), adding forgetting, input and output gates, so that the network can remember a longer input process [17]. The DKT model is shown in Figure3.1. The input data type of the model is a $0, 1^2N$ sequence, in which N is the number of items. Only one position in the sequence is 1, and the others are 0. For example, if the record "the answer to the first a question is correct", the first element a in the code is 1, and the others are 0. If it is recorded that "the answer to the first a question is wrong", the first element $N + a$ is 1, and the others are 0. The model is an encoder-decoder type model whose output is a sequence of length. During training, the one-hot code $t + 1$ at $q^t + 1$ moment contains both the topic information and the correct and incorrect information. The topic concept scalar at moment corresponds to the prediction information of the next moment in the output sequence. The input gate formula is shown in equation 3.1.

$$i_t = \sigma(W_i[h_t - 1, x_i]) + \sigma b_i \tag{3.1}$$

In equation 3.1, i_t represents the input gate at t moment; W_i means the weight matrix of the input gate; $[h_t - 1, x_t]$ denotes the connection of two vectors into a longer vector; $h_t - 1$ and x_t expresses the output items in the two matrices, respectively; b_i refers to the bias of the input gate; the set item *sigma* means the Sigmoid function. The formula of the forget gate is shown in equation 3.2.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t]) + \sigma(b_f) \tag{3.2}$$

In equation 3.2, f_t stands for the forget gate at t time; Wf refers to the weight matrix of the forget gate; b_f represents the bias term of the forget gate;σ expresses the Sigmoid function. The forget gate memory formula is shown in equation 3.3.

$$\widetilde{C_t} = tanh(W_c \cdot [h_{t-1}, x_t]) + tanh(b_c) \tag{3.3}$$

In equation 3.3,$\widetilde{C_t}$ refers to the forget gate memory at t moment; W_c denotes the weight matrix of the forget gate memory; b_c indicates the bias item of the forget gate memory; *tanh* means the hyperbolic tangent function. The output gate formula is shown in equation 3.4.

$$O_t = \sigma(W_\circ \cdot [h_{t-1}, x_t]) + \sigma(b_\circ) \tag{3.4}$$

In equation 3.4, O_t indicates the output gate at t moment; W_t is the weight matrix of the output gate; $b\circ$ stands for the bias item of the output gate; σ expresses the Sigmoid function. Long memory and short memory are shown in equation 3.5 .

$$\begin{cases} h_t = o_t * tanh(C_t) \\ C_t = f_t * C_{t-1} + i * \widetilde{C_t} \end{cases} \tag{3.5}$$

In equation 3.5, h_t and C_t denote long memory and short memory, respectively. f_t means forget gate, $\widetilde{C_t}$ denote forget gate memory; O_t indicates output gate; *tanh* refers to hyperbolic tangent function. The above formula shows that the output in the model is determined by the long memory, and the long memory is affected by the short memory, so the output is affected by the current input and the long-term input. In the LSTM model, long and short memory vectors are passed between each unit [18]. The improved model of this study is based on the dynamic key value memory neural network with two main memory modules. The memory module composed of M and d vectors stores the knowledge concept information and the student's state information, respectively. M is a parameter that can be set, and the dimensions in the module d can also be different. Another improvement is to re-encode the data using the embedding method. Through the embedding matrix, the length of the data $2N * d$ is reduced to the length of the data $2N$, avoiding the waste of memory resources caused by excessive dimensionality. The improved model is shown in Figure 3.2.

As shown in Figure 3.2, the improved algorithm model mainly consists of the embedding process, the graph process, and the knowledge concept encoding process, which correspond to the blue, green, and yellow labels, respectively. In the figure, the embedding process converts high-dimensional, sparse data (e.g., words, student IDs, topic IDs, etc.) into low-dimensional, continuous vector representations, so that data with similarities are closer to each other in this low-dimensional space. In the education domain, the embedding of students and topics captures their implicit attributes or properties. The purpose of the graph process is to construct a network of relationships between knowledge points or between topics to capture and convey information through graph structures. In each iteration of the DKVMN model, the representation of each node is updated according to its neighboring nodes. In this way, the relationships and hierarchies between knowledge points can be explicitly encoded in the model. The purpose of the knowledge concept encoding process is to encode each knowledge point or concept as a continuous vector that captures the dependencies and complexity between knowledge points. The implementation of knowledge concept encoding first obtains an initial representation of each knowledge point through the embedding process. Then, these representations are refined and optimized through the graph process. Eventually, each knowledge point will have a vector representation that adequately captures the essence of the knowledge and its relationship with other knowledge. The calculation of the weight is processed by the embedding matrix, and the dot product operation is performed on the input data with

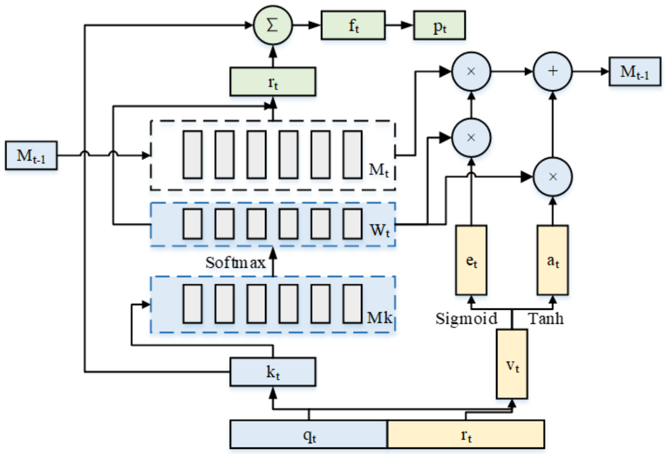


Fig. 3.2: DKVMN model

only the concept of the problem and the vector in the knowledge memory matrix. A vector whose length is equal to the number of vectors in the memory matrix is got, and the weight vector after going through the Softmax layer is got. Softmax is used as an activation function, and its calculation process expression is shown in equation 3.6.

$$w_t(i) = Softmax(k_t^T M^k(i)) \tag{3.6}$$

As shown in the formula, the activation function Softmax will output the sum of the weight vector elements equal to 1. The improved algorithm model uses knowledge memory and state memory modules. The state memory module is used to record the state at a certain moment, and the weight vector is used to weight and add the vectors in the state matrix during reading M . Embedding matrices are enabled to further correlate inputs in global gradient descent optimization. The reading process is shown in equation 3.7.

$$\begin{cases} r_t = \sum_{i=1}^{N} w_t(i) M_t^v(i) \\ f_t = tanh(W_1^T [r_t, k_t] + b_1) \\ p_t = Sigmoid(W_2^T f_t + b_2) \end{cases} \tag{3.7}$$

In equation 3.7, r indicates the product sum of the weight and the state matrix, which is the value obtained after the connection of r and k; p indicates the output of the Sigmoid function. The writing part is to adjust the matrix according to the knowledge concept at t moment and the information of whether the answer is correct. The input information generates vectors that control forgetting and memory after embedding. In the writing, both the control memory vector and the previous vector are affected by the association weight, and this improvement makes the process of writing affected by the relation of knowledge concepts. The process of updating the state matrix is shown in equation 3.8.

$$\begin{cases} e_t = sigmoid(E^T v_t + b_e) \\ \widetilde{M_t^v}(t) = M_{t-1}^v(t)[1 - w_t(i) e_t]) \\ a_t = Tanh(D_{vt}^T + b_a)^T \\ M_t^v(i = \overline{M}_{i-1}^v(i) + w_t(i) a_t \end{cases} \tag{3.8}$$

In the analysis of the improved model, the study find that the three main modules of the model interact with each other, which will complicate the interaction between knowledge concepts and states, thereby reducing

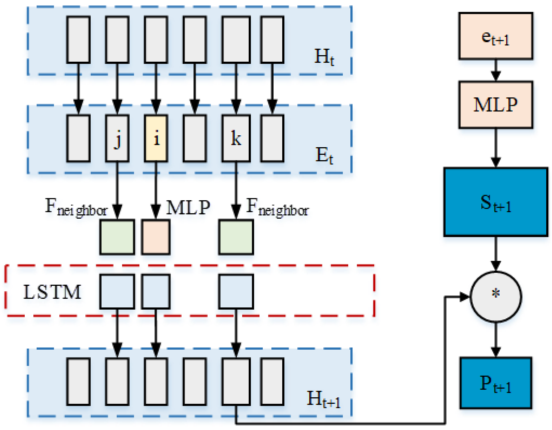


Fig. 3.3: Optimized DL model

the interpretability of knowledge concepts. Therefore, this study introduces deep reinforcement learning for personalized recommendation. The inputs of reinforcement learning include states, actions, and rewards. The key to implementing reinforcement learning is to output the optimal action in the current state to maximize the reward. In reinforcement learning, the long-term rewards obtained through training are optimal, so the summation of rewards is very important. The study introduces a discount factor γ for weighted summation. The expression of weighted summation is shown in equation 3.9.

$$U(S_\circ) = E[\sum_{t=1}^{\infty} \gamma^t R(S_t)] \tag{3.9}$$

In equation 3.9, U stands for the sum of rewards, that is, the average expectation of the weighted sum of rewards for all future action choices. R indicates the reward, that is, the system determines the value of an action S. represents the state, which is the only description of the current environment. This study exploits a recurrence relation when calculating total value to obtain a policy function that maximizes long-term returns. Ideas in reinforcement learning are utilized to divide the long sequence of action decisions, and the mathematical expression is shown in equation 3.10.

$$U^*(S) = max_a R(a) + U^*(S') \tag{3.10}$$

In equation 3.10, S expresses all the best paths; a denotes the current moment selection action; S' means the remaining paths. After optimization, a small part of the end of the optimally selected path can be truncated, and the remaining path will also become the optimal path.

3.2. Construction of KT online education model based on DL. Some studies have introduced two memory modules to store knowledge status and conceptual relationships, increasing the interpretability of the traditional cultural KT [19]. There are some classic theoretical models in the KT problem, including the combination of dynamic key-value storage network for KT and item response theory of classical cognitive diagnosis model. On this basis, some scholars proposed the knowledge query network (KQN) structure, which increases the interaction between knowledge status and knowledge concepts, to clarify the principle of the specific operation of the model, and further enhance the interpretability of the KT model [20]. The optimized DL model is shown in Figure 3.3. In the Figure 3.3, the improved DL model includes three parts. The first part is the embedding process, in which H is a memory matrix, representing the matrix that stores each knowledge state. H contains N independent vectors, in which N is the total number of problem concepts [21]. H_t

denotes each vector dimension, means the time-memory matrix. E_t indicates the result matrix after inputting embedding processing. The moment matrix E_t is a matrix containing $N-e$ dimensional vectors, and the expression is shown in equation 3.11.

$$E_k^t = \begin{cases} x^t E(k=i) \\ E_c(k)(k \neq i) \end{cases} \tag{3.11}$$

As shown in equation 3.11, the second part is the graph process. The input at t time is (q_t, a_t) . q_t denotes the concept of the problem at t time and the adjacent nodes are recorded as j, k . At this time, only the vectors j and k in the memory matrix are updated. The update process goes through a fully connected part and a LSTM network [22]. The result is recorded as the vector i in the memory matrix, as shown in equation 3.12.

$$H_k^{t+1} \begin{cases} RNN(f_{MLP}(h_k^{'t}))(k_i) \\ RNN(f_{neighbor}(h_i^{'t}))(k \neq i) \end{cases} \tag{3.12}$$

In equation 3.12, the core of this part is the function of k vector. In this study, the setting threshold $f_{neighbor}$ of the function is improved to simplify the model. When the moment is t, k is the title.k and n vectors are not considered, only the adjacent vectors need to be updated. It assumes that the concept input at $t+1$ time is e_{t+1}, after the feed-forward network, the output is a code of length S_{t+1}, the expression is shown in equation 3.13

$$S_{t+1} = \sigma(W_1) \cdot (\sigma(W \circ \cdot e_{t+1} + b_\circ)) + b_1) \tag{3.13}$$

The proposed model uses DKT network and the deep reinforcement learning network DQN to solve the problem of large memory requirements and long calculation time of the DKT network. It solves the problem that the DQN network cannot further study the relationship between clustering internal concepts with the help of directed graphs relationship problem [23]. In the application of recommendation algorithms, this study innovatively combines knowledge graph (KG) into an online education system that integrates DL and KT, and establishes a complete traditional cultural online education system [24]. KG obtains potential links between items through the description of semantic associations, expands users' interests through various types of associations, improves the accuracy and diversity of recommendations. KG can also maintain users' historical learning data well, increasing interpretability of recommendation results. In the improved recommendation algorithm, the conceptual calculation of the user's operation on the item is shown in equation 3.14.

$$P_i = \frac{exp(v^t R_i h_i)}{\sum_{(h,t,r) \epsilon S_u^1} exp(v^T R_i h_i)} \tag{3.14}$$

In equation 3.14, v^T indicates the candidate item vector; R_i mean the edge of the i node; h_i represents the feature vector; S_u^1 refers to propagation process set. The user representation is obtained according to the probability, and the calculation process is shown in equation 3.15.

$$\begin{cases} O_u^1 = \sum_{h_{i,r_i,t_i \in S_u^1}} p_i t_i \\ u = o_u^1 + o_u^2 + ...o_u^H \end{cases} \tag{3.15}$$

In equation 3.15, t_i stands for the feature vector of the node's tail node. The similarity between the user and the item is calculated to complete the recommendation at last. The expression is shown in equation 3.16.

$$\widehat{y_{uv}} = \frac{1}{1 + exp(-u^T v)} \tag{3.16}$$

In equation 3.16, $\widehat{y_{uv}}$ indicates the user u 's click probability for the item v . The research and design of the learning system of traditional cultural network online education divides users into students and teachers

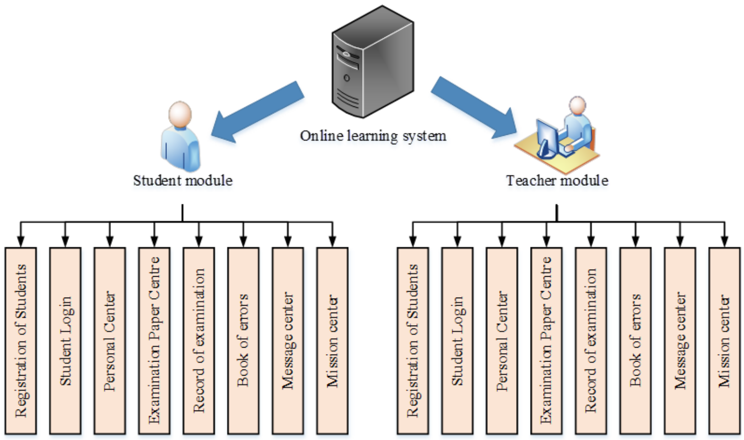


Fig. 3.4: System function overview diagram

Table 3.1: Data set information

Data Set	Algebra_2006_2007	Assistments2009-2010
Number of Logs	52324	58851
Individual Users	511	976
Number of Test Questions	1734	12461
Points of Knowledge	216	110

[25]. Teachers are complex in student management, knowledge management, topic preparation, task release, etc.; students are responsible for completing tasks, taking exams, and answering questions online. An overview of system functions is shown in the Figure 3.4.

In the overall process of traditional cultural online education, the teacher first logs in to the system, checks the students' learning and answering conditions, changes the topics and publishes learning tasks; then the students log in to the system, checks the tasks and notifications, and conducts answering training and knowledge learning. System statistics record students' answers, and input the knowledge structure relationship hidden behind the questions into the model of research design to obtain the sequence of topic recommendations. Finally, students conduct feedback exercises through recommended topics, conduct targeted training, and improve learning efficiency. The overall flow chart of traditional cultural online education is shown in the Figure 3.5.

3.3. Performance comparison results of KT models based on fusion DL. To verify the effect of the knowledge recommendation model based on KT proposed in this study, the experimental data set used algebra_2006_2007 and Assistment2009-2010 data sets that were often used in the field of KT. The algebra_2006_2007 data set was the interaction record between the user and the computer-aided system, and the Assistment2009-2010 data set recorded the practice log of elementary school students' mathematics exercises, including 62955 records. The ratio of the training, test, and verification sets used in the research was 8:1:1. After preprocessing the data, the relevant information of the data is shown in Table 3.1.

The study first explored the impact of different discount factors and different lengths of question sequences on the system, and then established evaluation indicators including area under curve (AUC), accuracy and knowledge mastery fluctuations to verify the improved model's performance proposed in this study. Under different discount factors, the changes in the cumulative rewards of the algorithm model during training iterations

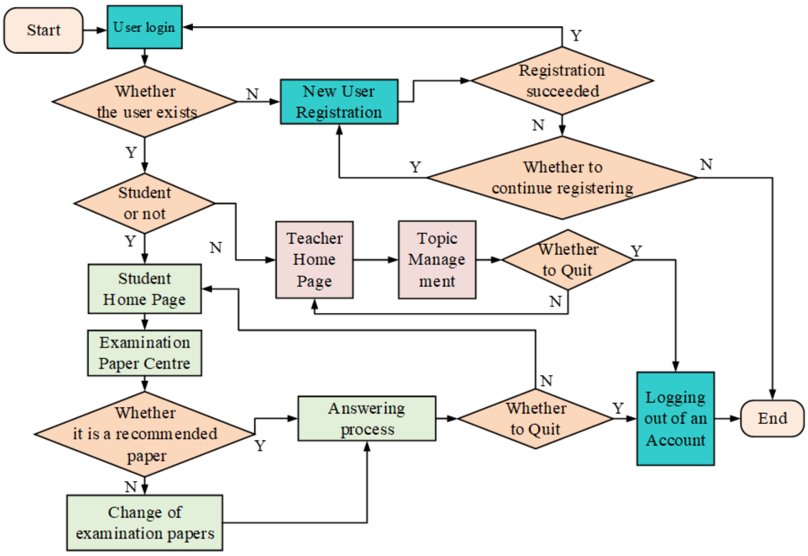
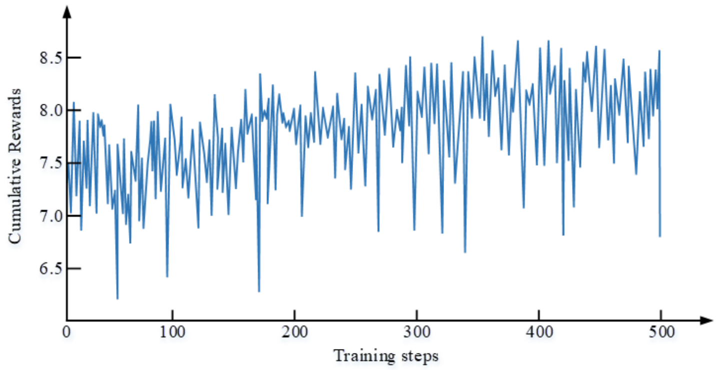


Fig. 3.5: Traditional culture online education flow chart

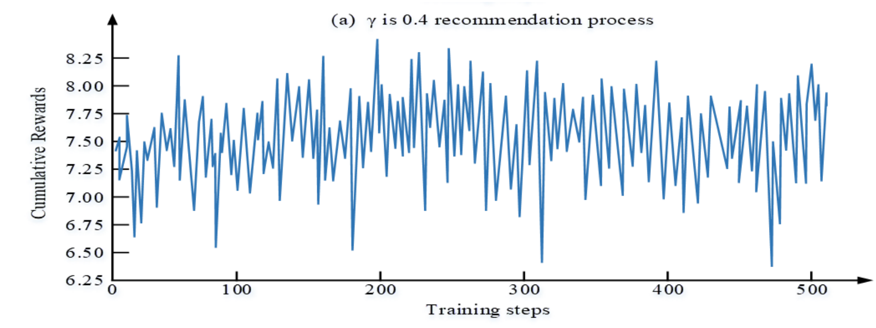
are shown in Figure 3.6.

From Figure 3.6a, a clear trend can be observed with a discount factor set at 0.4. The cumulative reward of the improved algorithm model exhibited a consistent increase as the number of training iterations escalated. This incremental pattern suggested a potential correlation between the training iterations and the model's efficacy. Conversely, Figure 6(b) portrayed a different scenario when the discount factor was zeroed. The cumulative reward oscillated between 6.32 and 8.47 throughout the training iterations. Notably, for up to 500 training iterations, the reward's trajectory appeared relatively static, devoid of any pronounced pattern. A deeper dive into Figure 3.6 revealed an enlightening insight: the model harnessing a 0.4 discount factor outperformed its counterpart, especially as iterations proliferated. This underpinned the assertion that a long-term recommendation strategy held an advantage over a heuristic approach. To further underscore this, it was essential to test the improved model's capability across various topic sequence lengths. Topic sequences of different lengths were recommended to verify the recommendation ability of the improved model. The result is shown in the Figure 3.7.

Figure 3.7 provided an in-depth analysis of the fusion DL and KT recommendation model across varying application settings. Specifically, Figure 3.7a demonstrated the model's performance when recommending 3 out of 5 questions. By approximately the 150th training step, the model verged on full convergence, maintaining a steady trajectory with an overall reward range between 0.38 and 1.53. Meanwhile, Figure 3.7b, which depicted the recommendation of 4 out of 8 questions, revealed a similar pattern of stability by around the 275th training step, having a reward range of 1.78 to 2.25. However, when ventured into more complex scenarios, the dynamics shifted. Figure 3.7c, representing 5 out of 15 recommendations, suggested that the model struggled to converge even after 500 training iterations, with a reward spectrum of 3.05 to 4.03. Similarly, Figure 3.7d which tackled 10 out of a hefty 50 questions confirmed this challenge. Despite 500 training cycles, convergence remained elusive, and rewards oscillated between 6.26 and 8.55. Drawing insights from these findings, a pattern emerged: the model's convergence and stability were inversely proportional to the number of topics to be recommended. For compact topic sets, the model converged swiftly. However, as the breadth of topics expanded, the convergence rate dwindled, culminating in the model's inability to adequately cater to personalized learning needs across a diverse student body. AUC represents the area enclosed by the ROC and the X coordinate, which can better reflect the pros and cons of the classifier. The closer the AUC value was to 1, the better the classifier effect. Common algorithms KT, DKT, DKVMN, GKT and DL-KT algorithm model proposed in this study were



(a) γ is 0.4 recommendation process



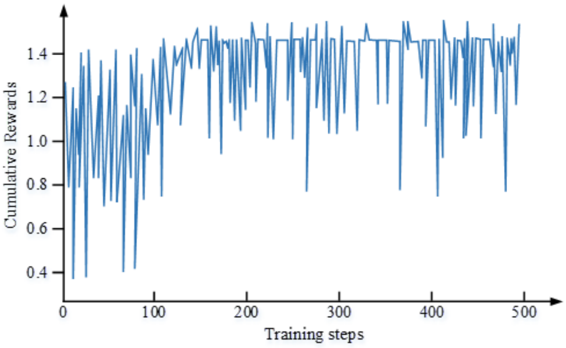
(b) γ is 0 recommendation process

Fig. 3.6: Training performance of the algorithm model under different discount factors

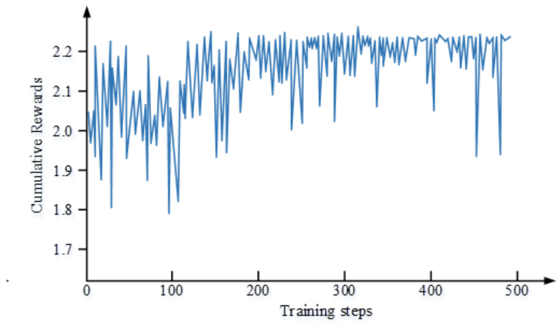
compared. The experimental results of AUC and accuracy are shown in Figure 3.8.

Figure 3.8a shows the AUC comparison of different algorithm models on different training sets. From the figure, the DKT algorithm performed the worst, with an average AUC of 0.762 in the two training sets, followed by the DKVMN, with an average AUC of 0.769, and then the GKT algorithm, with an average AUC of 0.774. The best performance was the algorithm model proposed in this study. The average AUC of the DL-KT algorithm was 0.781. Figure 3.8b shows the comparison of the accuracy of different algorithm models in different training sets. From the figure, the DKVMN performed the worst, with an average accuracy rate of 0.846, followed by the DKT algorithm, with an average accuracy rate of 0.857. The GKT algorithm performed better, with an average accuracy rate of 0.863. The best performer DL-KT was with an average accuracy of 0.886. From the comparison in Figure 8 the DL-KT model proposed in this study performed better than other models in AUC and accuracy. It was verified that the construction mode of the DL-KT model that combined the knowledge display relationship with the potential relationship of knowledge points would make the effect of KT better. The fluctuation of knowledge mastery value (KMV) refers to the fluctuation value between the current knowledge mastery and the previous knowledge mastery after completing the recommended knowledge. KMV is the knowledge fluctuation numerical index. A comparative experiment was carried out on the degree of mastery of knowledge points, and the results are shown in the Figure 3.9.

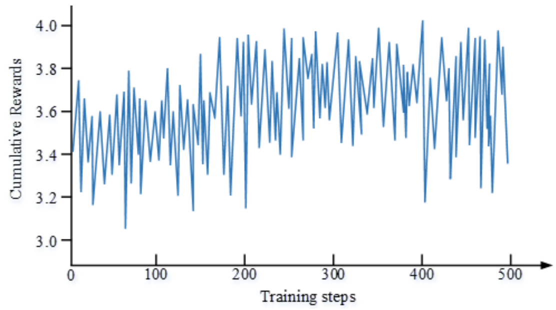
From Figure 3.9, in the Assistment2009-2010 data set, the KMV was the highest when the student's knowledge mastery in the algebra_2006_2007 data set was 0.63. On the whole, the change trend of KMV in the two training sets was the same, and the overall KMV fluctuation range was within 0 to 0.16. The results showed that recommending knowledge within the range of 0.6-0.7 to users improved the user's knowledge



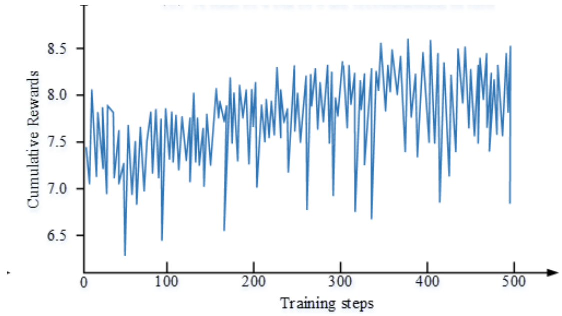
(a) A total of 3 out of 5 are recommended in turn



(b) A total of 4 out of 8 are recommended in turn

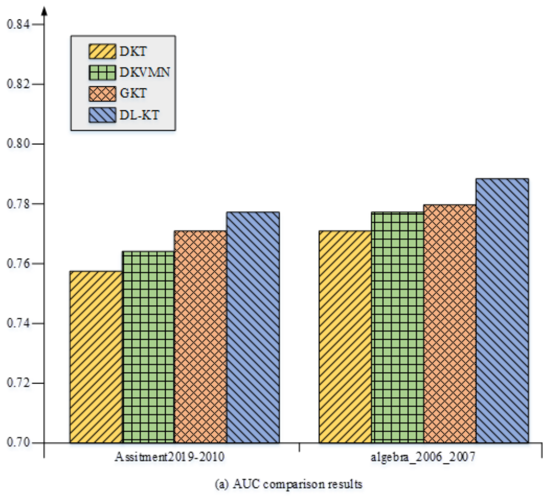


(c) A total of 5 out of 15 are recommended in turn

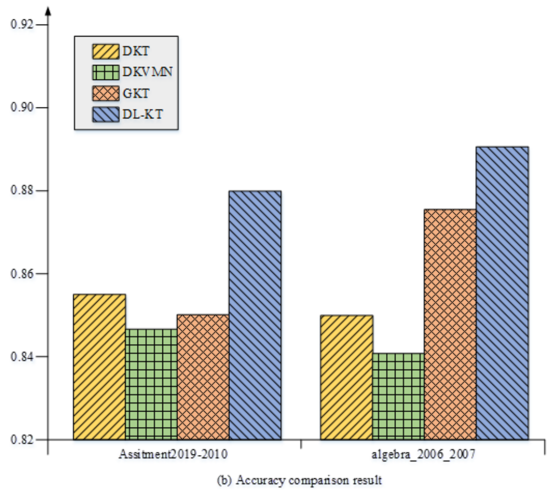


(d) A total of 10 out of 50 are recommended in turn

Fig. 3.7: Model effect performance in different application environments



(a) AUC comparison results



(b) Accuracy comparison result

Fig. 3.8: Comparison of AUC and accuracy of different algorithm models

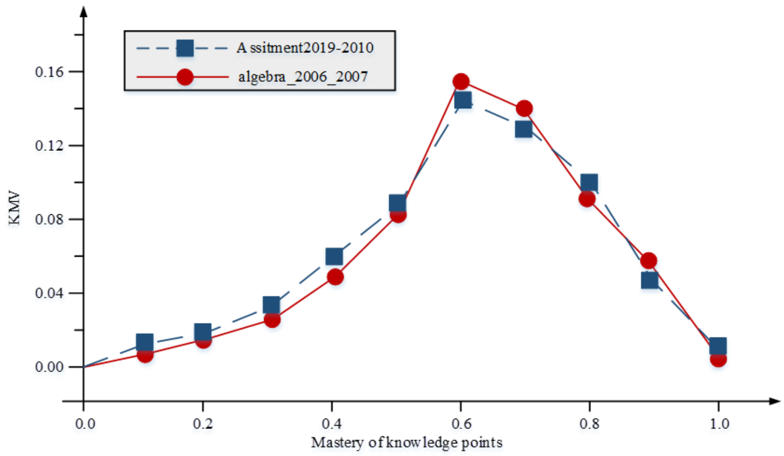
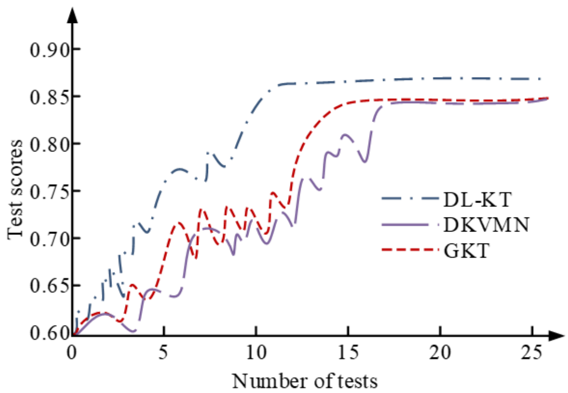
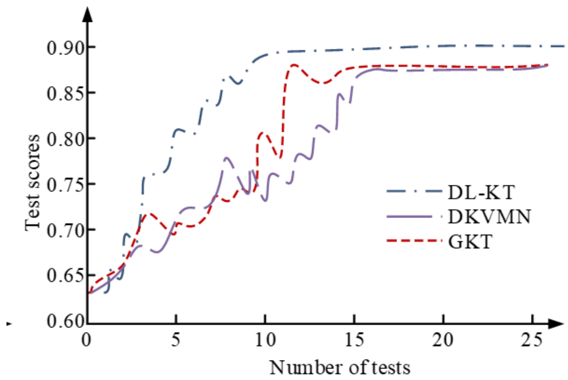


Fig. 3.9: Comparison of knowledge points



(a) Changes in test scores for senior students



(b) Changes in test scores for lower grades students

Fig. 3.10: Comparison of application of different models in mental health education

mastery the fastest, which was obviously better than other algorithm models. It was verified the performance of traditional cultural network online education model proposed in this study that integrated DL and KT. To get the effect of the practical application of DL-KT in online education, the study selected senior and junior students to conduct a comparative experiment, respectively, and there was no significant difference in the performance of each group before the experiment. The results of the effect of different models on students' performance in traditional culture online learning were obtained, as shown in Figure3.10.

Figure 3.10a shows the changes in the grades of the students in the higher grades. With the assistance of the DL-KT model, students' grades increased faster and stabilized at about 86 after the 10th quiz. The difference in students' final grades with the assistance of the DINA and HO-DINA models was not significant. Figure 3.10b shows the changes in students' grades in the lower grades. With the assistance of DL-KT model, students' grades increased at a faster rate and their final grades stabilized around 89, which was better than the DINA and HO-DINA models. From Figure 3.10, the DL-KT model proposed in the study was more effective in improving the online learning performance of students in lower grades and was more suitable for application in online education for students in lower grades.

4. Conclusion. The online learning system is a new learning platform that uses the Internet and artificial intelligence technology to enable students to learn the required knowledge more quickly and conveniently. It is difficult to pass the actual level of students in the current online education system and accurately recommend the knowledge question bank. Therefore this research built an online education model that integrated DL and KT. The experimental outcomes denoted that the algorithm model with a discount factor of 0.4 had better performance, which verified that the long-term recommendation effect used in this study was better than the previous heuristic recommendation effect. The average AUC of the model proposed in this study was 0.781, and the average accuracy rate was 0.886. The results showed that recommending knowledge within the range of 0.6-0.7 to users improved the user's knowledge mastery the fastest, which was obviously better than other online education models. However, the shortcoming of this study was that relatively little research has been conducted on the other key subject in online education, the teacher. Teachers play a crucial role in the learning process of students, especially how to track and adapt to the learning progress and needs of different students in real time. In addition, although the model outperformed other models in some metrics, whether it can be generalized across different educational cultures and curricula still requires further research. In future work, the study plans to delve into the role and function of teachers in online learning systems, especially how to enable real-time tracking of student learning by teachers and how to better integrate teachers and technology to provide personalized learning advice to students. As the fields of technology and education converge further, the study hopes to bring about more far-reaching and broader implications for online learning systems.

Fundings. The research is supported by Scientific Research Program Funded by Education Department of Shaanxi Provincial Government (Program No.22JZ016).

REFERENCES

[1] Dooly, M. & Sadler, R. If you don. *T Improve, What's The Point?' Investigating The Impact Of A 'flipped' Online Exchange In Teacher Education.* **32**, 4-24 (2020)
[2] Adiego, J. & Martín-Cruz, N. Training competences in smart cities: an online program for higher education students. *International Journal Of Sustainability In Higher Education.* **22**, 1630-1645 (2021)
[3] Griffith, P., Doherty, C., Smeltzer, S. & Mariani, B. Education initiatives in cognitive debiasing to improve diagnostic accuracy in student providers: A scoping review. *Journal Of The American Association Of Nurse Practitioners.* **33**, 862-871 (2021)
[4] Root, W. & Rehfeldt, R. Towards a Modern-Day Teaching Machine: The Synthesis of Programmed Instruction and Online Education. *The Psychological Record.* **71**, 85-94 (2021)
[5] Chowdhury, T., Hoque, M., Wanke, P., Raihan, M. & Azad, M. Antecedents of Perceived Service Quality of Online Education During a Pandemic: Configuration Analysis Based on Fuzzy-Set Qualitative Comparative Analysis. *Evaluation Review.* **46**, 235-265 (2022)
[6] Shen, L., Shao, Z., Yu, Y. & Chen, X. Hybrid Approach Combining Modified Gravity Model and Deep Learning for Short-Term Forecasting of Metro Transit Passenger Flows: Transportation Research Record. (2021)
[7] Garg, V. & Wickramarathne, T. Deep Learning Approach for Enhancing Situational Awareness in Surveillance Applications with Ubiquitous High-Dimensional Sensing. *IEEE Journal Of Selected Topics In Signal Processing.* **16**, 869-878 (2022)
[8] Tang, W., Yang, Q., Hu, X. & Yan, W. Deep learning-based linear defects detection system for large-scale photovoltaic plants based on an edge-cloud computing infrastructure. *Solar Energy.* **231**, 527-535 (2022)
[9] Bernardini, M., Feldmann, R., Anglés-Alcázar, D., Boylan-Kolchin, M., Bullock, J., Mayer, L. & Stadel, J. From EMBER to FIRE: predicting high resolution baryon fields from dark matter simulations with deep learning. *Monthly Notices Of The Royal Astronomical Society.* **509**, 1323-1341 (2022)
[10] Liu, C., Han, Z., Zhang, Z., Nussinov, R. & Cheng, F. network-based deep learning methodology for stratification of tumor mutations. *Bioinformatics.* **37**, 82-88 (2021)
[11] Song, X., Li, J., Tang, Y., Zhao, T., Chen, Y. & Guan, Z. Joint Graph Convolutional Network based Deep Knowledge Tracing. *Information Sciences.* **580** pp. 510-523 (2021)
[12] Pavlik, P., Eglington, L. & Harrell-Williams, L. Constrained Framework for Learner Modeling. *Institute Of Electrical And Electronics Engineers (IEEE).* **14**, 624-639 (2021)
[13] Gan, W., Sun, Y. & Sun, Y. Knowledge structure enhanced graph representation learning model for attentive knowledge tracing. *International Journal Of Intelligent Systems.* **37**, 2012-2045 (2022)
[14] Huang, Y. & Cheng, Y. Prediction of Online Judge Practice Passing Rate Based on Knowledge Tracing. 2021. *0.* **38** pp. 003 (0)
[15] Liu, S., Yu, J., Li, Q., Liang, R., Zhang, Y., Shen, X. & Sun, J. Ability boosted knowledge tracing. *Information Sciences: An International Journal.* **596** pp. 567-587 (2022)
[16] Pawan, Y., Prakash, K., B., C., S., H. & C., Y. Particle swarm optimization performance improvement using deep learning techniques. *Multimedia Tools And Applications.* **81**, 27949-27968 (2022)

[17] Lai, Z., Wang, L. & Ling, Q. Recurrent knowledge tracing machine based on the knowledge state of students. *Expert Systems*. **38**, 1-12782 (2021)

[18] Gan, W., Sun, Y., Peng, X. & Sun, Y. Modeling learner's dynamic knowledge construction procedure and cognitive item difficulty for knowledge tracing. *Applied Intelligence*. **50**, 3894-3912 (2020)

[19] Prakash, A., Patro, K., Hammad, M., Tadeusiewicz, R. & Pławiak, P. secured biometric authentication system using ECG signal based on deep learning techniques. *Biocybernetics And Biomedical Engineering*. **42**, 1081-1093 (2022)

[20] Huo, Y., Wong, D., Ni, L., Chao, L. & Zhang, J. Knowledge modeling via contextualized representations for LSTM-based personalized exercise recommendation. *Information Sciences*. **523** pp. 266-278 (2020)

[21] He, J. & Huang, S. Full-length de novo protein structure determination from cryo-EM maps using deep learning. *Bioinformatics*. **37**, 3480-3490 (2021)

[22] Rakhshani, H., Idoumghar, L., Ghambari, S., Lepagnot, J. & Brevilliers, M. On the performance of deep learning for numerical optimization: An application to protein structure prediction. *Applied Soft Computing*. **110**, 7596-15 (2021)

[23] Altaheri, H., Muhammad, G., Alsulaiman, M., Amin, S., Altuwaijri, G., Abdul, W. & Faisal, M. Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review. *Neural Computing And Applications*. **35**, 14681-14722 (2023)

[24] Li, Y., Bao, T., Gao, Z., Shu, X., Zhang, K., Xie, L. & Zhang, Z. new dam structural response estimation paradigm powered by deep learning and transfer learning techniques. *Structural Health Monitoring*. **21**, 770-787 (2022)

[25] Rastogi, K., Bodani, P. & Sharma, S. Automatic building footprint extraction from very high-resolution imagery using deep learning techniques. *Geocarto International*. **37**, 1501-1513 (2022)

Edited by: Mudasir Mohd
Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
Received: May 17, 2023
Accepted: Sep 11, 2023