Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 20 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i20.44213

PAPER

Construction and Implementation of Curriculum Knowledge Bases by Integrating New Educational Resources

Xiaoqian Zhang(✉)

Education and Teaching Research Center, Shangluo University, Shangluo, China

zxqian@slxy.edu.cn

ABSTRACT

With the development of information technology (IT), the educational field has gradually entered a modernization stage. In this context, it is important to effectively integrate various educational resources to further improve teaching quality. However, existing methods of constructing curriculum knowledge bases face many challenges in several aspects, such as data fusion, knowledge reasoning, and so on. To address these issues, this study conducted in-depth research and practice on the construction of curriculum knowledge bases in the educational modernization field. The main content of this study includes the data fusion and knowledge reasoning of curriculum knowledge bases and the intelligent Q&A framework design of the bases in the educational modernization field. In the data fusion part, an entity alignment method based on semantic similarity was adopted, which effectively solved the problem of semantic relationship fusion between entities. In the knowledge reasoning part, the option-aware network based on attention mechanisms was used to significantly improve the accuracy and efficiency of reasoning. As part of the intelligent Q&A framework design, a sequence-to-sequence machine translation model was adopted, which intelligently queried the knowledge bases successfully and provided customized teaching content, thereby optimizing the learning process. This study provides effective theoretical and practical support for promoting educational modernization.

KEYWORDS

curriculum knowledge base, data fusion, knowledge reasoning, intelligent Q&A framework, educational modernization

1 INTRODUCTION

With the rapid development of information technology (IT), especially the widespread application of technologies such as big data, cloud computing, and artificial intelligence, the educational field has gradually entered a modernization stage.

Zhang, X. (2023). Construction and Implementation of Curriculum Knowledge Bases by Integrating New Educational Resources. *International Journal of Emerging Technologies in Learning (iJET)*, 18(20), pp. 137–150. https://doi.org/10.3991/ijet.v18i20.44213

Article submitted 2023-05-23. Revision uploaded 2023-08-17. Final acceptance 2023-08-20.

In this context, it is an important subject in educational modernization to effectively integrate various educational resources to further improve teaching quality. As a platform for centralized storage, management, and utilization of curriculum knowledge, the curriculum knowledge base provides educators with rich teaching resources and students with a wide range of learning paths [1–4]. However, the construction and application of curriculum knowledge bases currently face many challenges, such as difficulties in data fusion and knowledge reasoning, which require in-depth research on the construction methods of the bases [5–10].

The development of educational modernization has produced a massive amount of educational data, which contains rich educational knowledge. Valuable information needs to be extracted to serve educational decisions and teaching practice. Therefore, the construction method of curriculum knowledge bases studied in this research is not only of theoretical significance but also has practical application value. Through data fusion and knowledge reasoning, the integration ability of educational data and the efficiency of acquiring educational knowledge can be increased, thereby improving teaching methods and optimizing the learning process [11–15]. Through the intelligent Q&A framework, knowledge bases can be intelligently queried, thereby providing customized teaching content and improving the learning experience.

However, there are still some problems with the existing construction methods of curriculum knowledge bases. For example, current data fusion technologies often overlook the semantic relationship between entities, resulting in poor data fusion results [16] [17]. In terms of knowledge reasoning, the lack of effective attention mechanisms leads to low accuracy and efficiency [18–20]. At the same time, the design of an intelligent Q&A framework also faces many challenges. For example, how to effectively correspond questions to answers and how to provide personalized learning suggestions based on the actual situations of students.

This research aimed to study and implement a construction method for curriculum knowledge bases in the field of educational modernization. The main research content included two parts: first, data fusion and knowledge reasoning of curriculum knowledge bases. Data fusion mainly relied on the entity alignment method, which is based on semantic similarity, while knowledge reasoning was achieved through the option-aware network based on attention mechanisms. Second, intelligent Q&A framework design of the bases, taking the use of a sequence-to-sequence machine translation model as the core idea. The main value of this study lies in the efficiency improvement of utilizing educational data, the optimization of educational resource allocation, and the quality and efficiency improvement of educational services by studying and implementing the new construction method of curriculum knowledge bases, thereby further promoting the development of educational modernization.

2 DATA FUSION AND KNOWLEDGE REASONING OF CURRICULUM KNOWLEDGE BASES

2.1 Knowledge fusion

In the field of educational modernization, a curriculum knowledge base is a very important resource that contains various types of educational information, such as textbook content, teaching methods, curriculum arrangement, etc. In different knowledge bases, the information may exist in different forms. For example, a curriculum may be referred to as “advanced mathematics” in one knowledge base and "higher mathematics" in another. Although their names are different, they represent the same curriculum. In addition, different knowledge bases may provide different attributes for the same entity. For example, a curriculum in one knowledge base may only contain basic information, such as the curriculum name, professors, credits, etc., while another knowledge base may contain more information, such as the curriculum syllabus, teaching resources, references, etc. These differences have brought great challenges to data fusion.

Therefore, entity alignment plays a very important role in the integration of curriculum knowledge bases. By aligning entities, the same entities in different knowledge bases can be determined, thereby avoiding data redundancy and duplication. This not only improves the efficiency of data utilization but also reduces the cost of storing and processing data. Entity alignment also helps achieve the interoperability of knowledge bases. When the same entities in different knowledge bases are determined, data can be shared and exchanged between them, thereby improving the efficiency of educational resource utilization and the quality of educational services. Figure 1 shows the process of the entity alignment method.

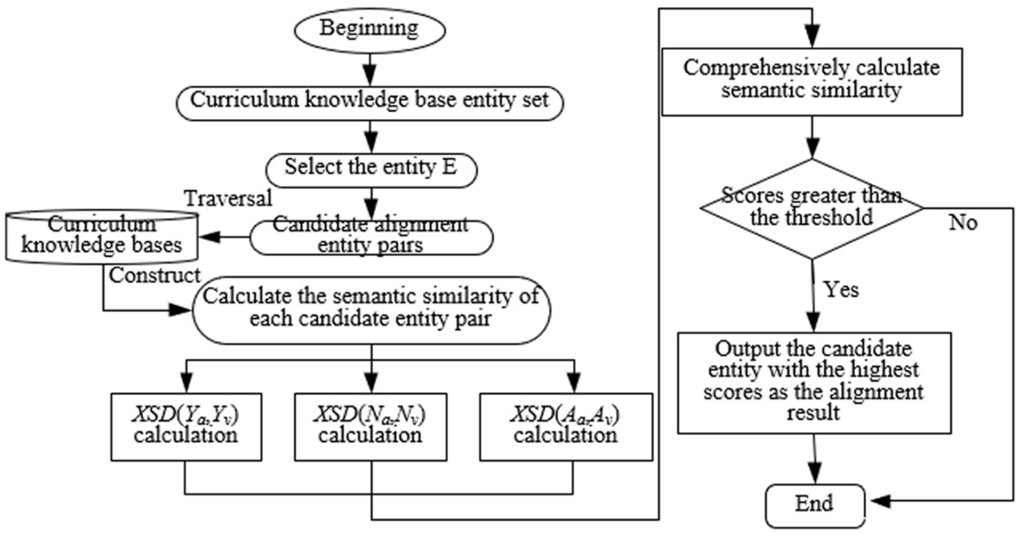

Fig. 1. Process of the entity alignment method

This study achieved entity alignment of curriculum knowledge bases in the educational modernization field by

- Calculating the similarity between entities in the word vector space
- The similarity between the best synonyms
- The semantic similarity between entity description texts

In the word vector space, each word or entity was represented as a vector, which was obtained through the training and learning of a large amount of text data. The similarity between entities was obtained by calculating the cosine similarity between their vectors in the word vector space. Let R_a be the target entity, R_v be the candidate entity, Y_a and Y_v be the context-free word vectors, $COS(Y_a,Y_v)$ be the cosine similarity between Y_a and T_v, and $XSD(Y_a,Y_v)$ be the word-level similarity between Y_a and Y_v. To calculate we use the formula given below:

$$XSD(Y_a,Y_v) = COS(Y_a,Y_v) \tag{1}$$

For each entity, some of its best synonyms were found, i.e., the words being semantically closest to it. The similarity between entities was measured by calculating the similarity of these synonyms in the word vector space. Let N_a be the embedding vector of target entity synonyms obtained using the word vector tool *word2vec*, N_v be the embedding vector of candidate entity synonyms, and $XSD(N_a, N_v)$ be the synonym-level semantic similarity of entities. To calculate we use the formula given below:

$$XSD(N_a, N_v) = COS(N_a, N_v) \tag{2}$$

Each entity usually has some description text, providing more information about the entity. The similarity between entities was measured by calculating the semantic similarity between these description texts. Let $XSD(A_a, A_v)$ be the similarity between descriptive sentences A_a and A_v. A weighted superposition of the above three types of semantic similarity was performed, which obtained comprehensive entity similarity. The calculation formula was as follows:

$$XSD(R_a, R_v) = q_1 * XSD(Y_a, Y_v) + q_2 * XSD(N_a, N_v) + q_1 * XSD(A_a, A_v) \tag{3}$$

This study took the entity pair that had the largest semantic similarity to the target entity in the candidate entity set H_v of curriculum knowledge bases as the quasi-aligned entity. When the similarity scores between the target entity and the quasi-aligned entity exceeded a certain threshold, R_{SC} was output as the entity alignment result, i.e.

$$R_{SC} = \underset{R_v \in H_v}{\operatorname{argmax}} XSD(R_a, R_v) \tag{4}$$

2.2 Knowledge reasoning

In the curriculum knowledge base in the educational modernization field, task texts and evidence knowledge typically contain a large amount of relevant but somewhat different information. After connecting and inputting them into the pre-trained model, the model can understand and process the information in a unified context, thereby improving the efficiency of knowledge fusion. Different tasks or evidence-based knowledge may correspond to different important information points. Based on the needs of tasks and characteristics of evidence knowledge, attention mechanisms can dynamically adjust the level of attention to different parts, thereby achieving refined dynamic weight allocation.

This study, therefore, constructed an intelligent Q&A model for curriculum knowledge bases in the educational modernization field and introduced attention mechanisms. The overall architecture was mainly divided into three modules: the text encoding module, the option-aware attention module, and the answer prediction module. The text encoding module was mainly used to convert the input text information into numerical forms that could be processed by machines, i.e., representation vectors. A lite bidirectional encoder representations from transformers (ALBERT) is a widely used natural language processing pre-training model that captures complex patterns in texts and encodes these patterns into representation vectors. This module used ALBERT for encoding, which obtained representation vectors of text information. As the second module of the model, the option-aware attention module was mainly used to enhance the encoded vectors through attention mechanisms, which achieved knowledge fusion and reasoning. The answer prediction module mainly scored the five representation vectors of contexts through a linear classification layer and a normalization layer and selected the option in the pair with the highest scores as the predicted answer to the question. This module achieved the final decision for Q&A tasks and was a key part of the model's prediction results. Figure 2 shows an example of knowledge reasoning starting from the entity "wick."

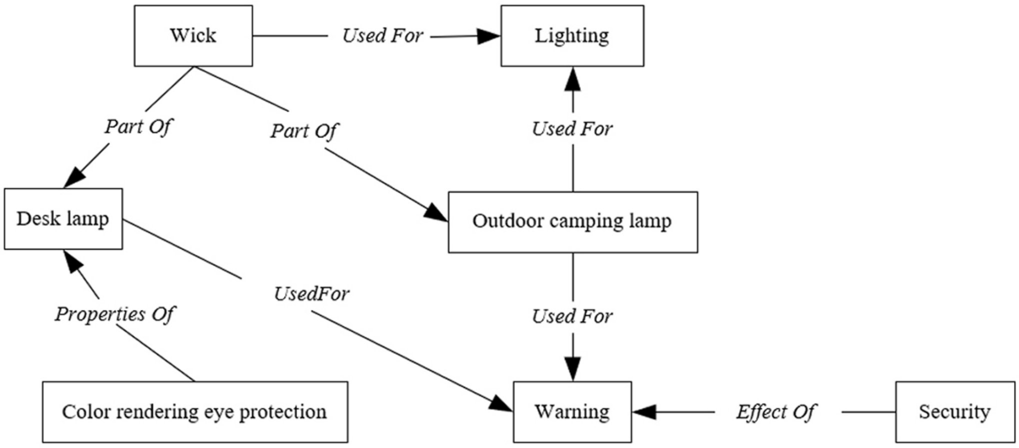

Fig. 2. An example of knowledge reasoning starting from the entity "wick"

This study constructed different attention networks for complex patterns in texts, which processed and represented different information elements under different conditions for different input text sequences. In the case where the encoded text consisted of two parts, i.e., task text and evidence knowledge, a simple attention layer was used to effectively fuse the task text and evidence knowledge, which further extracted the correlation information between the two and highlighted the information that was most critical to the task solution through attention mechanisms, thereby achieving feature enhancement.

Let g_y and g_j be the final hidden layer state vectors of the encoded text in the text encoder and the corresponding hidden layer state vectors of evidence knowledge, respectively. Attention calculations were performed on both vectors, which obtained scores DF_{yj} as follows:

$$DF_{yj} = \text{attention}(g_y, g_j) \tag{5}$$

The state vector g_y was further adjusted based on DF_{yj}. Splicing, linear classification, and normalization operations were performed for the adjusted vector, which obtained the answer prediction. The calculation process was as follows:

$$O(y) = \text{linear}(DF_{yj} g_y, g_j) \tag{6}$$

In the case where the encoded text consisted of three parts, namely, problem text, evidence knowledge, and option text, a multi-step attention strategy was adopted because it involved more information elements with higher complexity compared with the previous situation. The strategy first calculated the weighted combination of option, problem, and evidence knowledge parts in the encoded hidden layer state vector, which obtained the correlation information between the three parts. Then another attention layer was used to calculate these two weighted combinations, which obtained the weighted combination of the final encoding part. Relevant information was further extracted and strengthened through these steps, resulting in richer and more refined text-encoding representations.

In this case, let g_w and g_s be the hidden layer states of the problem and option parts, respectively. Attention calculations were performed for both states, which obtained scores DF_{ws} as follows:

$$DF_{ws} = \text{attention}(g_w, g_s) \tag{7}$$

Similarly, the attention scores DF_{js} of g_j and g_s were calculated as follows:

$$DF_{js} = \text{attention}(g_j, g_s) \tag{8}$$

Finally, weighted combinations of the above scores were performed to achieve the final knowledge embedding representation. Let g_{wjv} be the weighted representation vector of the output, then the answer prediction result was:

$$O(w,s) = \text{linear}(DF_{ws} h_q, DF_{js}, h_s) \tag{9}$$

3 INTELLIGENT Q&A OF CURRICULUM KNOWLEDGE BASES IN THE EDUCATIONAL MODERNIZATION FIELD

An intelligent Q&A task is a way to effectively measure the understanding and reasoning abilities of computers, providing a simple and intuitive way of interacting. Users obtain the necessary information by asking questions without complex operations or profound technical knowledge, which makes the intelligent Q&A task widely applicable and easy to use. Therefore, this study chose it as the main method of knowledge base application.

This study designed two baseline models based on both prediction and planning for Q&A tasks and comprehensively evaluated the effectiveness of knowledge bases. It aimed to test curriculum knowledge bases in the educational modernization field from different perspectives, thereby effectively validating and comparing the performance of these two methods in practical applications. Figure 3 shows the architecture of the intelligent Q&A model.

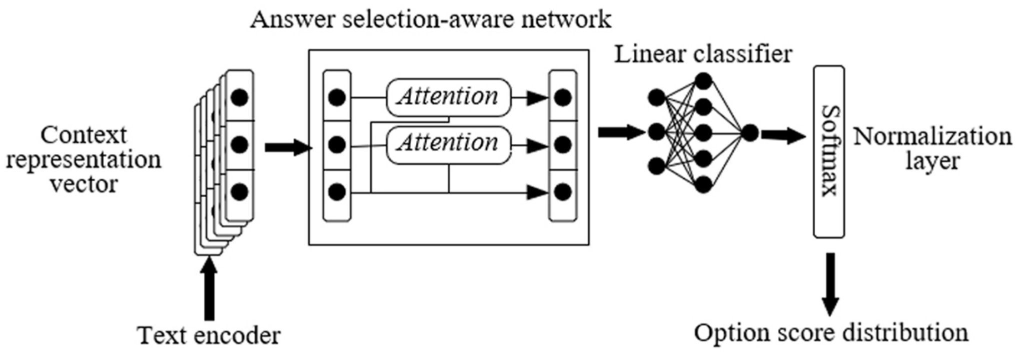

Fig. 3. Architecture of the intelligent Q&A model

The prediction-based model mainly made predictions based on existing knowledge and information. However, the prediction model designed in this study effectively integrated video intention with knowledge sequence prediction to achieve better prediction results. This model mainly included two steps: recognition of entity attributes or connotations and prediction of knowledge sequences. The I3D model was first used to process the input knowledge base entities and recognize the attributes or connotations of the entities. The output of the *I3D* model was input into the *Seq2Seq* model of the bidirectional gate recurrent unit (GRU). The *Seq2Seq* model is a sequence-to-sequence deep-learning model that is widely used in tasks such as machine translation and speech recognition. In the constructed prediction model, the *Seq2Seq* model was mainly used to predict the next knowledge sequences. The bidirectional GRU was used to effectively process sequence data and capture long-distance dependencies in the sequences. Figure 4 shows the architecture of the prediction model.

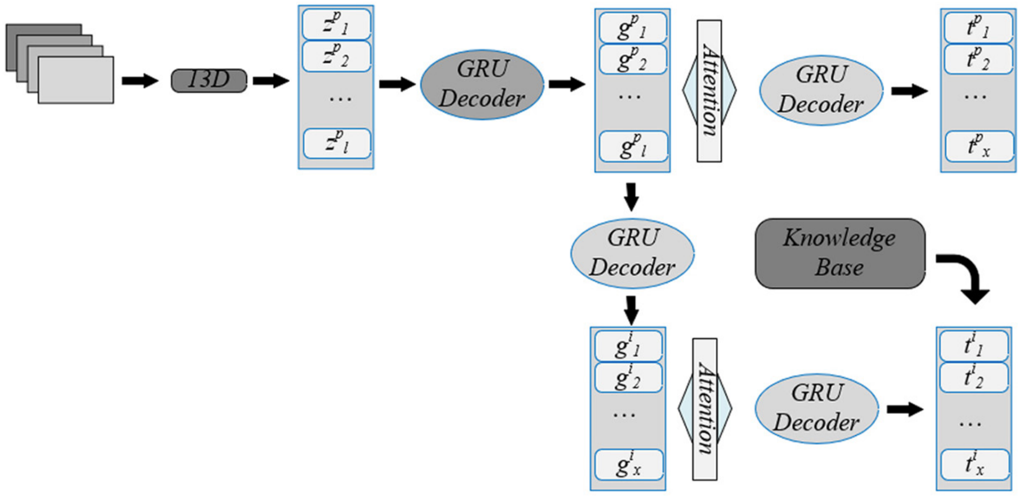

Fig. 4. Architecture of the prediction model

Let $Z^p = (z_1^p,\dots,z_l^p), z_u^p \in E^f$ be the entity attributes or connotative features extracted by the *I3D* model; $G^p = (g_1^p,\dots,g_l^p)$ be the hidden state of the input, which was obtained through a bidirectional GRU encoder; $d_r^p()$ be the encoder used for encoding input sequence features; $D_f^p()$ be the decoder used for decoding hidden features. Then sequence labels were as follows:

$$G^p = d_r^p(Z^p) \tag{10}$$

$$T^p = D_f^p(G^p) \tag{11}$$

Attention mechanisms were not introduced into the GRU decoder, and the hidden state sequence labels of future knowledge sequences were:

$$G^i = d_f^i(g_l^p) \tag{12}$$

The future knowledge flow sequence T^i was decoded, which obtained:

$$T^i = d_f^{it}(G^i) \tag{13}$$

The modified cross-entropy loss function was used as the loss function of the model. Assuming O answers were observed, and B answers were predicted, let $\hat{t}_w$ be the predicted value, and t_w be the true value. The loss function $LOSS_{HS}$ was defined as follows:

$$LOSS_{HS} = \left(1 - \exp\left(-\frac{O}{B}\right)\right)\sum_{w=1}^{B}\exp(-w)M(\hat{t}_w, t_w) \tag{14}$$

Compared with the prediction-based model, the planning-based model emphasizes problem solving based on goals and strategies. The planning model designed in this study adopted a two-stage planning method, which mainly included two stages: knowledge sequence generation and future answer prediction. *UniVL*, which is the existing knowledge entity annotation generation model with the best performance, was first used to map the associated knowledge entities into consecutive knowledge sequences. The *UniVL* model is a deep-learning model that allows for joint learning of videos, audios, and texts to obtain a rich understanding of knowledge entities. Therefore, it was used in this study to generate consecutive knowledge sequences.

Let s^p be the observed knowledge sequence, which was the input of the *UniVL* model. The *I3D* model was used to extract the features of s^p. Let $Y \in E^{b \times f}$ be the encoding matrix of b tokens in the text sequence, and $C \in E^{b \times f}$ be the encoding matrix corresponding to l groups of knowledge entity features corresponding to the knowledge sequence, then the encoding results of the features based on the transformer encoder were as follows:

$$Y = \Omega(s^p) \tag{15}$$

$$C = TF(U3F(a)) \tag{16}$$

Features were further merged based on the cross-modality encoder:

$$L = TF([Y;C]) \tag{17}$$

The encoding feature L, which combined information of two modalities, was further decoded, which obtained:

$$F = TF(L) \tag{18}$$

Then the generated knowledge sequences were input into the *Seq2Seq* language model for predicting future answers. The *Seq2Seq* model is an encoding-decoding structure based on neural networks, which takes sequences as inputs and outputs, and effectively handles tasks, such as text prediction.

$$G_{EN} = EN(z_1, \ldots, z_{m^p}), z_u \in E \tag{19}$$

$$o\left(t_y \middle| t_{<y}, z\right), \ldots, o\left(t_{y+b-1} \middle| t_{<y}, z\right) = DE(t_{<y}, G_{EN}) \tag{20}$$

To reduce prediction errors of the model, this study treated each action phrase as a unique token in the input and output dictionaries, meaning that each independent action phrase was considered as an independent word, thereby effectively avoiding to introduce the bias of specific answers. The loss function was constructed based on the conditional probability loss function, i.e.

$$LOSS = -\sum_{y=1}^{Y} \log o_{\varphi}\left(t_y \middle| t_{<y}, z\right) \tag{21}$$

4 EXPERIMENTAL RESULTS AND ANALYSIS

In the educational modernization field, curriculum knowledge bases can be constructed to cover various disciplines and educational levels, ranging from basic education to higher education, from arts to science, and even comprehensive and interdisciplinary knowledge bases. For the convenience of the experiment, this study set up six curriculum knowledge bases, including *A* -"Basic Mathematics Knowledge Base," *B* -"Physical Principles Knowledge Base," *C* -"Language and Literature Knowledge Base," *D* -"History and Culture Knowledge Base," *E* -"Computer Science Knowledge Base," and *F* -"Biological Science Knowledge Base." The above knowledge bases were designed for the needs of the educational modernization field, which existed independently or were associated with and integrated into specific tasks or studies.

It can be seen from the data in Table 1 that the proposed method in this study has excellent performance in entity alignment tasks in Bases *A-B* and *A-C*. The proposed method in this study is superior to the traditional method using cosine similarity only in *Hit@1* (the proportion of correctly aligned entities ranking in the top 1 among all candidate entities), *Hit@10* (the proportion of correctly aligned entities ranking in the top 10), or *MR* (i.e., *mean rank*, the average ranking of all correctly aligned entities). It was concluded that the entity alignment method based on semantic similarity achieved higher accuracy and efficiency in the entity alignment tasks of curriculum knowledge bases in the educational modernization field and demonstrated its superior performance compared with the traditional cosine similarity method, fully proving its effectiveness in entity alignment tasks.

Table 1. Comparison of entity alignment experimental results based on similarity calculation

Methods	Bases *A-B*			Bases *A-C*		
	Hit@1	*Hit@10*	*MR*	*Hit@1*	*Hit@10*	*MR*
Cosine similarity	81.26	81.32	2567	78.95	78.96	2036
The proposed method in this study	83.48	85.36	42	85.68	84.62	34

Table 2. Comparison of knowledge fusion experimental results

Models	Bases *A-B*		Bases *A-C*		Bases *A-D*	
	Hit@1	*Hit@30*	*Hit@1*	*Hit@30*	*Hit@1*	*Hit@30*
KGFM	19.25	49.25	23.10	51.15	32.01	58.16
DLKFM	22.15	67.62	24.26	71.25	23.52	68.25
SFM	24.26	71.25	27.15	74.22	24.26	73.21
AMFM	23.20	83.21	32.21	81.58	45.26	91.02
The proposed model in this study	25.14	86.32	34.28	82.20	48.26	92.31
Improve	1.89	0.74	1.71	1.75	3.25	1.25

It can be seen from the data in Table 2 that the proposed model in this study has superior performance to other models in the knowledge fusion of Bases *A-B*, *A-C*, and *A-D*. Specifically, in knowledge fusion of Bases *A-B*, the proposed model increases by 1.89% and 0.74%, respectively, compared with the best-performing AMFM model. In the knowledge fusion of Bases *A-C*, the proposed model increases by 1.71% and 1.75%, respectively, in *Hit@1* and *Hit@30*, compared with the AMFM model. For the knowledge fusion of Bases *A-D*, the proposed model increases by 3.25% and 1.25%, respectively, in *Hit@1* and *Hit@30*, compared with the AMFM model. This further validates the effectiveness of the proposed model; that is, better knowledge fusion results can be achieved by learning and mining the semantic information of curriculum knowledge bases.

It can be seen from the data in Figure 5 that the performance of the proposed model shows a trend of continuous improvement as the number of candidate entities increases, which is shown in the *Hit at K* (*H@K*) experimental results in the dataset of Bases *A-E*. The performance of the proposed model surpasses that of other models as the number of candidate entities increases. When the number of candidate entities is 5, the performance of the proposed model exceeds that of the KGFM, DLKFM, and SFM models and is the same as that of the AMFM model. When the number of candidate entities increases to 10, the performance of the proposed model begins to surpass all other models, and this advantage is maintained as the number of candidate entities continues to increase. Overall, these data demonstrate that the proposed model outperforms other models in handling tasks with a large number of candidate entities, which further validates the effectiveness and robustness of the proposed model.

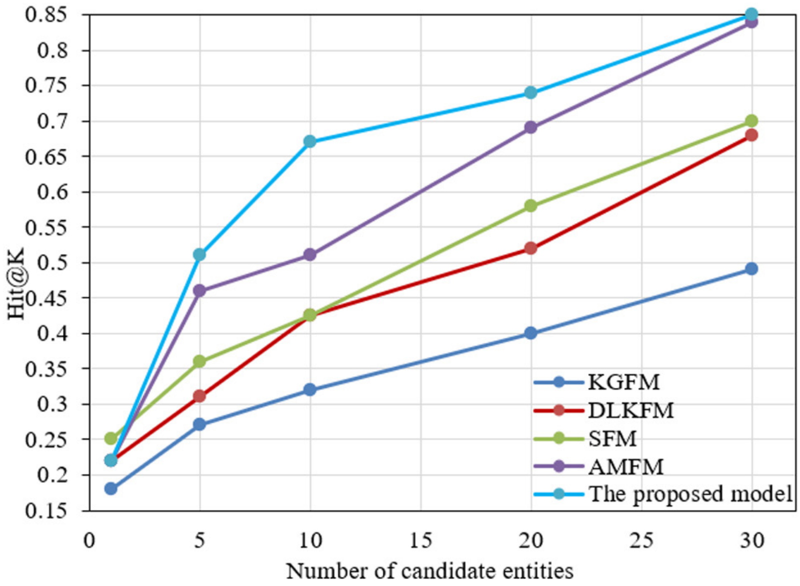

Fig. 5. Variation trend of *H@K* experimental results of the proposed model in the dataset of Bases *A-E*

Table 3. Experimental results of the proposed model in five groups of associated curriculum data

Bases	*Hit@1*	*Hit@10*	*Hit@30*	*MRR*
Bases *A-B*	25.12	46.25	87.12	0.421
Bases *B-A*	36.26	52.03	72.30	0.362
Bases *A-C*	34.06	51.02	88.21	0.442
Bases *C-A*	34.26	45.02	85.21	0.521
Bases *A-D*	48.17	62.03	93.02	0.552
Bases *D-A*	41.15	58.02	84.02	0.536
Bases *A-E*	23.02	62.03	85.03	0.531
Bases *E-A*	35.02	52.03	76.12	0.475
Bases *A-F*	35.06	65.03	81.02	0.445
Bases *F-A*	36.02	52.16	82.31	0.562

The following points can be observed based on the data in Table 3. First, the *Hit@30* index of the proposed model reaches a relatively high level in the combinations of all bases, exceeding 70%, and has the best performance in the combination of Bases *A-D*, reaching 93.02%. This indicates that the proposed model has a very high hit rate for querying entities in the first 30 prediction results, demonstrating good predictive ability. Second, for the *Hit@1* index, i.e., the first prediction result is the correct answer, the combination of Bases *A-D* once again shows the best performance, reaching 48.17%, while the combination of Bases *A-E* performs the worst, with only 23.02%, maybe because of different data characteristics between Base *E* and other bases or the impact of the data scale and quality of Base *E*. For the *Hit@10* index, the combination of Bases *A-F* performs the best, reaching 65.03%, while the combination of Bases *C-A* performs the worst with only 45.02%, indicating that the proposed model demonstrates the best prediction ability in the combination of Bases *A-F* in the top 10 prediction results. For the *Mean Reciprocal Rank* (*MRR*) index, the combination of Bases *B-A* performs the best, reaching 0.362, while the combination of Bases *A-D* performs the worst, reaching 0.552. The lower the *MRR* value, the better the performance of the model. Overall, the proposed model demonstrates good performance in combinations of different bases. However, the specific performance is also influenced by the characteristics and quality of the data between different bases. Nevertheless, the performance of the proposed model is still very impressive on the whole.

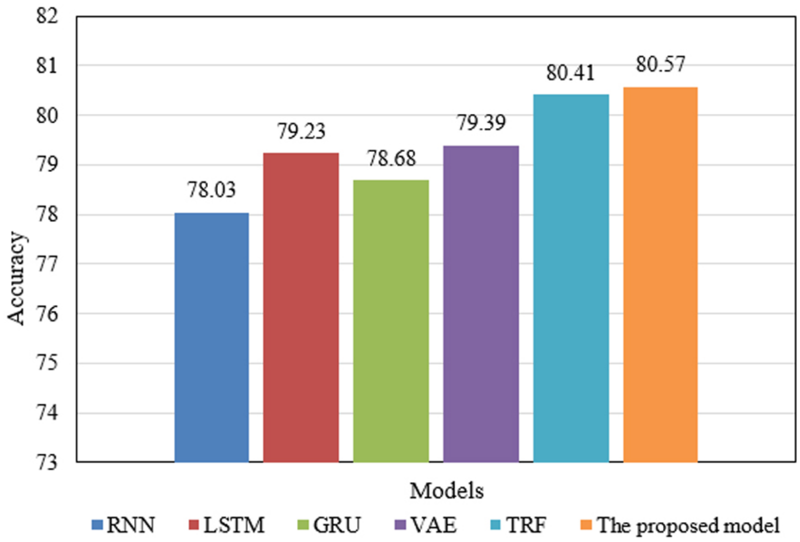

Fig. 6. Histogram comparison of experimental results of intelligent Q&A tasks using different external knowledge base models

The following points can be seen from the data in Figure 6. The proposed model has the highest accuracy of 80.57% in intelligent Q&A tasks, which is higher than that of all other models listed, demonstrating its superiority and effectiveness. In addition to the proposed model, the performance of the transformer (TRF) model is also quite excellent, with an accuracy of 80.41%, second only to the proposed model, maybe because the TRF model handles long sequence data well, especially intelligent Q&A tasks. Among the three traditional recurrent neural network models, recurrent neural network (RNN), long short-term memory (LSTM), and gate recurrent unit (GRU), LSTM performs the best with an accuracy of 79.23%. The variational auto-encoder (VAE) model performs slightly better, achieving an accuracy of 79.39%. Overall, the proposed model performs the best in handling intelligent Q&A tasks, demonstrating the effectiveness and superiority of its design. Other models also have their advantages, but their overall performance does not surpass that of the proposed model.

Table 4. Experimental results of intelligent Q&A tasks using different external knowledge base models

Model	*Seq of 1a (Acc%)*	*Seq of 2a (Acc%)*	*Seq of ≥ 3a (Acc%)*	Total *(Acc%)*
RNN	18.59	14.16	9.71	16.26
LSTM	33.57	15.28	11.24	15.24
GRU	52.29	42.16	25.16	45.11
VAE	37.59	31.12	21.18	33.29
TRF	52.67	36.12	24.29	31.84
The proposed model in this study	69.16	62.25	48.39	66.25
Manual prediction	93.48	91.17	82.15	90.18

It can be seen from Table 4 that the proposed model performs the best among all the models in various situations (i.e., one, two, three, or more action sequences). With an overall accuracy of 66.25%, which again proves the effectiveness and superiority of the baseline models based on both prediction and planning proposed in this study. There is still some gap in accuracy between the proposed model and manual prediction (an accuracy of 90.18%) because humans have strong logical reasoning ability, rich background knowledge, and an understanding of complex contexts and semantics. However, existing models still have some limitations in understanding complex contexts and implementing deep reasoning.

To further improve the accuracy of the proposed model as well as approach or even surpass manual prediction, large-scale corpora can be used for pretraining to learn richer semantic information. At the same time, more efforts can be made, such as introducing external knowledge, utilizing knowledge graphs and other tools, and introducing more background knowledge and contextual information. Overall, although the proposed model has excellent performance, there is still room for improvement. By further optimizing the model and training strategies, it is expected that its performance can approach or even surpass manual prediction.

5 CONCLUSION

This research mainly studied the construction and application of curriculum knowledge bases in the educational modernization field and conducted experiments. Five different models were studied for the fusion of knowledge bases, including KGFM, DLKFM, SFM, AMFM, and the proposed new model in this study. The experimental results showed that the proposed model outperformed other models in terms of fusion performance. For the intelligent Q&A tasks, this study proposed two baseline models based on both prediction and planning and tested them in six different sequence-to-sequence deep-learning models, demonstrating that the accuracy of the proposed model exceeded that of other models. The proposed method was compared with the classic cosine similarity method in the entity alignment experiment, and the results showed that the proposed method significantly improved in the *Hit@1* and *Hit@10* indexes.

Overall, this study achieved significant research results in the construction and application of curriculum knowledge bases in the field of educational modernization. This study plays an important auxiliary role in the Q&A tasks of curriculum knowledge bases by constructing the bases in the educational modernization field and designing excellent baseline models based on prediction and planning. Although there is still a certain gap compared with manual prediction, it is expected that the performance of the model can be further improved by introducing richer background knowledge, adopting more complex model structures, optimizing training strategies, and so on.

6 ACKNOWLEDGMENT

This paper was supported 2021 Special Scientific Research Project approved by Shaanxi Education Department (Grant No.: 21JK0110).

7 REFERENCES

[1] N. Thanachawengsakul and P. Wannapiroon, "Development of a learning ecosystem using digital knowledge engineering through MOOCs knowledge repository system," *International Journal of Engineering Pedagogy*, vol. 11, no. 1, pp. 35–48, 2021. https://doi.org/10.3991/ijep.v11i1.15011

[2] H. Zhang, X. Shen, B. Yi, W. Wang, and Y. Feng, "KGAN: Knowledge grouping aggregation network for course recommendation in MOOCs," *Expert Systems with Applications*, vol. 211, p. 118344, 2023. https://doi.org/10.1016/j.eswa.2022.118344

[3] N. Thanachawengsakul and P. Wannapiroon, "The development of a MOOCs knowledge repository system using a digital knowledge engineering process to enhance digital entrepreneurs' competencies," *International Journal of Engineering Pedagogy*, vol. 11, no. 6, pp. 85–101, 2021. https://doi.org/10.3991/ijep.v11i6.24409

[4] W. Deng, P. Zhu, H. Chen, T. Yuan, and J. Wu, "Knowledge-aware sequence modelling with deep learning for online course recommendation," *Information Processing & Management*, vol. 60, no. 4, p. 103377, 2023. https://doi.org/10.1016/j.ipm.2023.103377

[5] A. J. Atyeh Husain and A. Qasem Al-Shayeb, "Improvement in student achievement through a flipped database management classroom: Shifting from passive traditional to active learning," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 8, pp. 210–221, 2023. https://doi.org/10.3991/ijet.v18i08.37699

[6] J. Xia, "Knowledge-based course map generation for student performance prediction," in *International Conference on Database Systems for Advanced Applications*, Tianjin, China, 2023, pp. 748–752. https://doi.org/10.1007/978-3-031-30678-5_67

[7] P. Petrov, I. Kuyumdzhiev, R. Malkawi, G. Dimitrov, and O. Bychkov, "Database administration practical aspects in providing digitalization of educational services," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 20, pp. 274–282, 2022. https://doi.org/10.3991/ijet.v17i20.32785

[8] H. Guan, "An online education course recommendation method based on knowledge graphs and reinforcement learning," *Journal of Circuits, Systems and Computers*, vol. 32, no. 6, p. 2350099, 2023. https://doi.org/10.1142/S0218126623500998

[9] X. Ma, "The use of knowledge correlation for classification and evaluation of a distance education teaching resource database," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 5, pp. 217–230, 2023. https://doi.org/10.3991/ijet.v18i05.38501

[10] S. D. Bhattacharjee, J. S. A. V. Gokaraju, J. Yuan, and A. Kalwa, "Multi-view knowledge graph for explainable course content recommendation in course discussion posts," in *26th International Conference on Pattern Recognition (ICPR)*, Montreal, QC, Canada, 2022, pp. 2785–2791. https://doi.org/10.1109/ICPR56361.2022.9956098
[11] S. Kryshtanovych, G. Liakhovych, O. Dubrova, H. Kazarian, and G. Zhekalo, "Stages of digital transformation of educational institutions in the system of sustainable development of the region," *International Journal of Sustainable Development and Planning*, vol. 18, no. 2, pp. 565–571, 2023. https://doi.org/10.18280/ijsdp.180226
[12] C. Cassar, "Addressing transformative education and governance through the sustainable development goal 4: A case study," *Opportunities and Challenges in Sustainability*, vol. 1, no. 2, pp. 105–115, 2022. https://doi.org/10.56578/ocs010203
[13] J. Wen and Y. L. Zhao, "An urban and rural educational resource sharing and exchange platform based on cloud platform access technology," *Ingénierie des Systèmes d'Information*, vol. 27, no. 3, pp. 515–520, 2022. https://doi.org/10.18280/isi.270320
[14] H. Du, Q. Huang, C. Zhang, and D. Zhao, "Knowledge-enhanced iterative instruction generation and reasoning for knowledge base question answering," in *CCF International Conference on Natural Language Processing and Chinese Computing*, Guilin, China, 2022, pp. 431–444. https://doi.org/10.1007/978-3-031-17120-8_34
[15] J. Julhadi and M. Ritonga, "Human resource management in Islamic educational institutions to improve competitiveness in society 5.0 era," *International Journal of Sustainable Development and Planning*, vol. 18, no. 2, pp. 611–619, 2023. https://doi.org/10.18280/ijsdp.180231
[16] B. Yang and M. Yang, "Research on enterprise knowledge service based on semantic reasoning and data fusion," *Neural Computing and Applications*, vol. 34, no. 12, pp. 9455–9470, 2022. https://doi.org/10.1007/s00521-021-06382-z
[17] B. Yang and Y. M. Liao, "Research on enterprise risk knowledge graph based on multi-source data fusion," *Neural Computing and Applications*, vol. 34, no. 4, pp. 2569–2582, 2022. https://doi.org/10.1007/s00521-021-05985-w
[18] H. Cui, T. Peng, T. Bao, R. Han, J. Han, and L. Liu, "Stepwise relation prediction with dynamic reasoning network for multi-hop knowledge graph question answering," *Applied Intelligence*, vol. 53, no. 10, pp. 12340–12354, 2023. https://doi.org/10.1007/s10489-022-04127-6
[19] H. Cui, T. Peng, R. Han, J. Han, and L. Liu, "Path-based multi-hop reasoning over knowledge graph for answering questions via adversarial reinforcement learning," *Knowledge-Based Systems*, vol. 276, p. 110760, 2023. https://doi.org/10.1016/j.knosys.2023.110760
[20] J. Yang, L. T. Yang, H. Wang, Y. Gao, Y. Zhao, X. Xie, and Y. Lu, "Representation learning for knowledge fusion and reasoning in cyber-physical-social systems: Survey and perspectives," *Information Fusion*, vol. 90, no. 59–73, 2022. https://doi.org/10.1016/j.inffus.2022.09.003

8 AUTHOR

Xiaoqian Zhang, graduated from Shaanxi Normal University, obtained a Masters' degree in literature, and now works in Shangluo University. Her research interests include education and teaching management, modern and contemporary literature (E-mail: zxqian@slxy.edu.cn; ORCID: https://orcid.org/0009-0005-8693-9936).