Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Prerequisite-driven Q-matrix Refinement for Learner Knowledge Assessment: A Case Study in the Online Learning Context

Wenbin Gan[a*], Yuan Sun[b]
[a]*National Institute of Information and Communications Technology, Tokyo, Japan*
[b] *National Institute of Informatics, Tokyo, Japan*
*wenbingan@nict.go.jp

Abstract: The ever-growing abundance of learning traces in the online learning platforms promises unique insights into the learner knowledge assessment (LKA), a fundamental personalized-tutoring technique for enabling various further adaptive tutoring services in these platforms. Precise assessment of learner knowledge requires the fine-grained Q-matrix, which is generally designed by experts to map the items to skills in the domain. Due to the subjective tendency, some misspecifications may degrade the performance of LKA. Some efforts have been made to refine the small-scale Q-matrix; however, it is difficult to extend the scalability and apply these methods to the large-scale online learning context with numerous items and massive skills. Moreover, the existing LKA models employ flexible deep learning models that excel at this task, but the adequacy of LKA is still challenged by the representation capability of the models on the quite sparse item-skill graph and the learners' exercise data. To overcome these issues, in this paper we propose a prerequisite-driven Q-matrix refinement framework for learner knowledge assessment (PQRLKA) in online context. We infer the prerequisites from learners' response data and use it to refine the expert-defined Q-matrix, which enables the interpretability and the scalability to apply it to the large-scale online learning context. Based on the refined Q-matrix, we propose a Metapath2Vec enhanced convolutional representation method to obtain the comprehensive representations of the items with rich information and feed them to the PQRLKA model to finally assess the learners' knowledge. Experiments conducted on three real-world datasets demonstrate the capability of our model to infer the prerequisites for Q-matrix refinement, and also its superiority for the LKA task.

Keywords: Learner Knowledge Assessment, Prerequisite Inference, Q-matrix Refinement, Graph Representation Learning

1. Introduction

The popularity of online learning has increased in recent years, with increasingly many intelligent tutoring systems (ITSs) becoming available to learners (Hasanov, Laine, &Chung, 2019; Gan, Sun, & Sun, 2020; Gan, Sun, & Ye, 2019a, 2019b). The application of techniques from artificial intelligence and cognitive psychology has brought increasing intelligence to these systems (Hasanov et al., 2019), and has led to a very wide range of advantages, e.g., enabling the knowledge-acquisition through a series of individualized learning activities that accommodate learners with different needs and knowledge proficiencies (Gan, Sun, Peng, & Sun, 2020; Q. Liu, Shen, Huang, Chen, & Zheng, 2021). A key technique underlying these adaptive tutoring services is LKA (Gan, Sun, Peng, & Sun, 2020; Gan, Sun, & Sun, 2022b, 2022a).

LKA assesses learner knowledge in the granularity of a set of specific skills, also called knowledge concepts, based on the learners' exercising logs in the systems, thereby providing the detailed information about their strength and weakness. Detailed assessment of skills requires a fine-grained mapping of items to skills, i.e., the Q-matrix (M. Desmarais, Beheshti, & Xu, 2014). A binary Q-matrix indicates how the skills are combined to correctly answer the items, hence bridging the latent skill mastery patterns of learners and their explicit responses. However, building the Q-matrix for a domain is a non-trivial task (Desmarais, Xu, & Beheshti, 2015; Desmarais & Naceur, 2013), which requires the involvement of domain experts. It is widely recognized that this subjective process may consist of some misspecifications (Chiu, 2013; Desmarais et al., 2014), which may negatively degrade the estimation of learner knowledge (Chen, 2017; Jiang et al., 2021). To alleviate the subjective bias in manually-defined Q-matrix, existing researches in the data mining and cognitive diagnostics fields have proposed various methods to infer the Q-matrix in a data-driven manner (J. Liu, Xu, & Ying, 2012), further refine the existing Q-matrix (M. Desmarais et al., 2014; Desmarais et al., 2015; Chiu, 2013; Kang, Yang, & Zeng, 2019), and validate the Q-matrix by combining with the cognitive diagnostic models (CMD) (Ma & de la Torre, 2020; de la Torre & Chiu, 2016). They also verified that in most of the cases, the refined Q-matrix has better model fit on the data than the original expert-labeled Q-matrix (M. C. Desmarais & Naceur, 2013; Matsuda, Furukawa, Bier, & Faloutsos, 2015). However, most of the efforts have been devoted to the scenarios of small tests with quite limited number of skills and items. Even for the famous fraction-subtraction dataset with only twenty items and eight skills, the obtained Q-matrix is still controversial (Kang et al., 2019). It is difficult, and sometimes impractical, to extend the scalability and apply these methods to the large-scale online learning context with numerous items and massive skills.

Moreover, most of the existing research refine the Q-matrix by considering a set of Q-matrices in the latent space and determining the one with superior model fitness on the data using some fit strategies (Kang et al., 2019). Typical studies are the maxDiff (De La Torre, 2008), minRSS (Chiu, 2013), and the ALS (M. C. Desmarais & Naceur, 2013). These researches showed good performance on the learner data, but they generally ignore the explainable structural and inherent information among the questions and skills. Actually, the interdependencies between the skills in a domain have long been acknowledged as prerequisites in both cognitive and education science (P. Chen, Lu, Zheng, & Pian, 2018). Correspondingly, a question with a specific skill also requires the prerequisite skills. These prerequisites between pedagogical concepts can be represented as a knowledge graph (KG) (Q. Liu et al., 2019), which contains additional structural information about the learning domain, and provides the potential for further refining the Q-matrix. Nevertheless, the prerequisites specifying the structure of skills has rarely been applied in the refinement of Q-matrix because obtaining the KG of a domain is labor-intensive (Pan, Li, Li, & Tang, 2017; Zhang & King, 2016; Nakagawa, Iwasawa, & Matsuo, 2019). Fortunately, the ever-growing abundance of learning traces in the online learning platforms further enhances our capacity to obtain the prerequisites (KG) directly from the data through data mining techniques. Based on this idea, we infer the prerequisites between skills from the data and use them to refine the Q-matrix, and apply the refined Q-matrix to assess the knowledge proficiency of learners. We try to balance the interpretability of Q-matrix refinement and the performance in applying it to the LKA task.

To assess learner knowledge in the online learning systems, plenty of researches have been conducted leveraging the compelling attribute of deep learning methods, such as the deep knowledge tracing (DKT) (Piech et al., 2015) and Dynamic key-value memory networks for KT (DKVMN) (Zhang, Shi, King, & Yeung, 2017), and excelled at this task. However, most of these models used the underlying skills as input (also termed as skill-level KT models) and did not distinguish questions containing the same skills. The loss of distinctive information related to the questions may lead to imprecise inferences of the learners' knowledge states (Ghosh, Heffernan, & Lan, 2020; Gan et al., 2022b). To enrich the question representation, graph KT methods, such as the graph-based interaction KT (GIKT) (Y. Yang et al., 2020) and graph for KT (GKT) (Nakagawa et al., 2019), have been proposed to learn the embedding of questions from the question-skill graph. However, they ignore the prerequisite information of the skills in the domain, and the sparseness of the original question-skill graph and the learner data greatly limit the capability of the graph learning methods. To overcome this sparseness, in this paper we refine the Q-matrix and propose a graph representation learning model to enrich the embedding learning from the prerequisite-enhanced question–skill graph and integrate it into the LKA.

To summarize, in this paper we refine the Q-matrix for learner knowledge assessment in online learning context and propose a prerequisite-driven Q-matrix refinement framework for LKA (PQRLKA). Specifically, we first explore eight methods to infer the prerequisites from learners' response data in the online learning systems and use it to refine the expert-defined Q-matrix to eliminate the potential subjective tendency of experts in designing the Q-matrix. This Q-matrix refinement method offers two advantages over the existing methods: first, it leverages the additional structural information about the learning domain to refine the Q-matrix rather than use the model fit indices to brutally change some entries from zeros to ones, thus enabling the interpretability of this process.

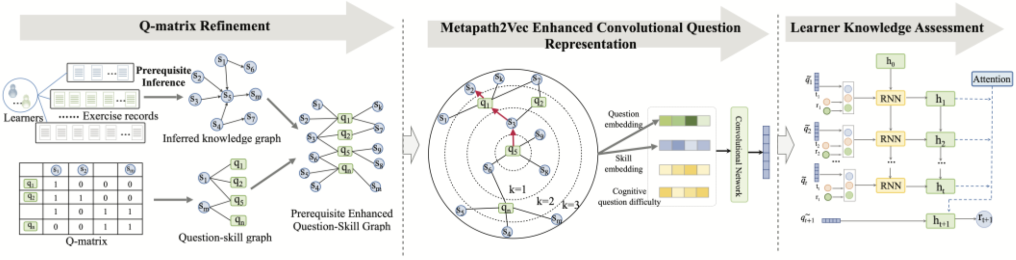

Figure 1: The framework of the proposed method

Second, it refines the Q-matrix by applying the inferred prerequisites in a date-driven manner without relying on the specific cognitive diagnosis models (i.e., model-independent), and hence has good scalability. Based on the refined Q-matrix, we propose a graph representation learning model for LKA to verify its effectiveness in the online learning scenarios. A Metapath2Vec (Dong, Chawla, & Swami, 2017) enhanced convolutional representation method is then proposed to obtain the comprehensive representations of the attempted questions with rich information. These representations for the learners' exercising sequences are fed into the PQRLKA model, which considers the long-term dependencies using an attention mechanism, to finally predict the learners' performance on new problems. The main contributions are listed below.

- We explore eight methods that automatically discover the domain prerequisites from learner response data and leverage it to refine the expert-defined Q-matrix.
- We propose the PQRLKA framework to assess the learner knowledge with a Metapath2Vec enhanced convolutional representation method for comprehensive question representations with rich information.
- We integrate the Q-matrix refinement with the task of learner knowledge assessment, joint these two tasks together and leverage the former to benefit the later.
- We conduct comprehensive experimental evaluations on three real-world datasets. The results demonstrate the superiority and effectiveness of our method in Q-matrix refinement and learner knowledge assessment in the online learning context.

2. Proposed Method

This section introduces the proposed PQRLKA framework for assessing learner knowledge, as shown in Figure 1. The framework proceeds the LKA task using three modules: Q-matrix refinement, Metapath2Vec enhanced convolutional question representation, and learner knowledge state evolution. It first builds a KG for the learning dom ain by inferring the prerequisite relations from the learner response data, and then uses it to refine the expert-defined Q-matrix, thus generating a refined Q-matrix, which can be naturally represented as a prerequisite enhanced question–skill graph. Based on this enhanced graph, it conducts the question representation learning based on a Metapath2Vec enhanced convolutional neural network. To incorporate more distinctive information, A convolutional representation method is then proposed to fuse the question and skill embeddings with the question difficulty, thus obtaining the comprehensive representations of questions (see Figure 3). Compared with the existing methods, the proposed question representation learning method incorporates both the distinctive information of the specific question and the various relations between questions and skills, a factor making it more superior for the later assessment of learner knowledge. These representations are then fed into the attentive LKA network for predicting learner performance. It is worth noting that we conduct the Q-matrix refinement and the learner knowledge assessment together.

2.1 Q-matrix Refinement

2.1.1 Prerequiite Inference from Data

In practical educational scenarios, there always exists a topological order (prerequisites) among the skills in a domain, because skills are taught and learned in sequence. Here we infer the prerequisites

Table 1: *Contingency table for a pair of skills s_i and s_j*

	s_j master	s_j not master	total
s_i master	a	b	a+b
s_i not master	c	d	c+d
total	a+c	b+d	a+b+c+d

Table 2: *Evaluation indices for obtaining KG using the contingency*

Evaluation Index	Formula
Cohen's Kappa	$R_{i,j}^{Kappa} = 2(ad-bc)/\{(a+b)(b+d)+(a+c)(c+d)\}$
Adjusted Kappa	$R_{i,j}^{Kappa'} = 2(ad-bc)/\{(a+c)(c+d)\}$
Phi coefficient	$R_{i,j}^{Phi} = (ad-bc)/\sqrt{(a+b)(b+d)(a+c)(c+d)}$
Yule coefficient	$R_{i,j}^{Yule} = (ad-bc)/(ad+bc)$
Ochiai coefficient	$R_{i,j}^{Ochiai} = a/\sqrt{(a+b)(a+c)}$
Sokal coefficient	$R_{i,j}^{Sokal} = (a+d)/(a+b+c+d)$
Jaccard coefficient	$R_{i,j}^{Jaccard} = a/(a+b+c)$

from the order of the learners' mastery of skills, which is explicitly represented by the exercising performance data. We first build a skill relation matrix $R \in R^{|S|\times|S|}$, in which entry $R_{i,j}$ represents the prerequisite relation $s_i \rightarrowtail s_j$ between skill s_i and s_j. Inspired by the definition of question similarity in previous methods (Tong et al., 2020; Pel´anek, 2019), we explore the underlying KG using eight methods: Skill Transition, Cohen's Kappa, Adjusted Kappa, Phi coefficient, Yule coefficient, Ochiai coefficient, Sokal coefficient, Jaccard coefficient, as described below.

Skill Transition: The skill-transition matrix R contains the transitions of different skills. Its entries are $R_{i,j}^{SK} = \frac{n_{i,j}}{\sum_{k=1}^{|\kappa|} n_{i,k}}$, where $n_{i,j}$ denotes the number of times in which skill s_j is trained immediately after training skill s_i.

To further leverage the impact of the learners' performance of one skill on the performance of another, we summarized the learners' performance on skill pair s_i and s_j in a contingency table (see Table 1). As mentioned above, we interpreted the learners' correct or incorrect responses as mastery indicators of the underlying skills of the given questions. It is worth noting that in Table 1, the question requiring skill s_i occurs before the question requiring s_j in the learning sequence. Based on the contingency table, we discovered the KG using the evaluation indices (Pel´anek, 2019) in Table 2, which measure the agreement of the prerequisite relation between a pair of skills.

As the KG is always a unidirectional graph, we simplified it by a suitable strategy. We also imposed a threshold that controlled the sparsity of the relations in KG. The final skill relation matrix was denoted as R^w, $w \in$ *{SK, Kappa, Kappa', Phi, Yule, Ochiai, Sokal, Jaccard}*. The elements along the diagonal of R^w were set to one.

$$\begin{cases} R_{i,j}^{w} = max(R_{i,j}^{w}, R_{j,i}^{w}), R_{j,i}^{w} = 0, \ if\ R_{i,j}^{w} \geq R_{j,i}^{w} \\ R_{j,i}^{w} = max(R_{i,j}^{w}, R_{j,i}^{w}), R_{i,j}^{w} = 0, \ otherwise \end{cases} \tag{1}$$

$$R_{i,j}^{w} = \begin{cases} 1, \ if\ R_{i,j}^{w} \geq threshold \\ 0, \ otherwise \end{cases} \tag{2}$$

2.1.2 Refine the Expert-defined Q-matrix

After obtaining the inferred skill relation matrix $R^w \in R^{|S|\times|S|}$ of the KG that includes all the prerequisites among skills, we refine the original expert-defined q-matrix $O \in R^{|Q|\times|S|}$. The matrix representation $\hat{O} \in R^{|Q|\times|S|}$ of the refined new Q-matrix is obtained as follows:

$$\hat{O} = O(R^w)^T \tag{3}$$

$$\hat{O}_{i,j} = \begin{cases} 1, \ if\ \hat{O}_{i,j} \geq 1 \\ 0, \ otherwise \end{cases} \tag{4}$$

In Eq. (3), the transpose of the skill relation matrix R^W accounts for the skills that are prerequisite to the skills of the current question. In other words, a question requiring a specific skill is also related to the prerequisite skills. The procedure of the refinement can be shown in Figure 2.

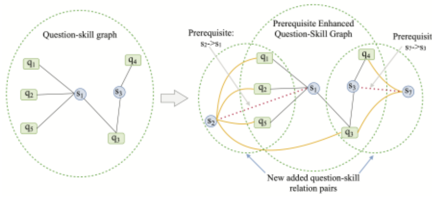

Figure 2: The procedure of Q-matrix refinement.

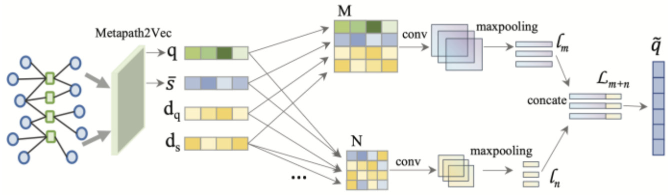

Figure 3: Use of Metapath2Vec enhanced convolutional question representation

2.2 *Learner Knowledge Assessment Based on Refined Q-matrix*

2.2.1 *Metapath2Vec Enhanced Convolutional Question Representation*

The matrix $\widehat{O}$ can be naturally represented as a graph, hereafter denoted as the KG-enhanced question--skill graph. This graph includes not only the multi-hop relations between questions and skills, but also the prerequisite relations among the skills. The embeddings in this graph were obtained using the adapted Metapath2Vec method (Dong et al., 2017).

The adapted Metapath2Vec method proceeds in two main steps: meta-path generation and skip-gram-based embedding learning. A meta-path is a sequence of nodes following the edges in the graph. To assure that all questions and skills appear in the final embeddings, we generate the meta-paths from the KG-enhanced question--skill graph using a question--skill--question (QSQ) pattern, in which every meta-path begins with a question node followed by a skill node and then by a question node; for example, ρ: $q_1 \xrightarrow{\hat{O}_{1,1}} s_1 \xrightarrow{\hat{O}_{2,1}} q_2 \xrightarrow{\hat{O}_{2,2}} s_2 \xrightarrow{\hat{O}_{3,2}} q_3$. Setting two hyper-parameters---the path length $\wp$ and number of paths $\aleph$---for each question node, we generated all meta-paths on the graph. The heterogeneous skip-gram is leveraged to learn the node embeddings. The interested reader can refer to (Dong et al., 2017) for the original Metapath2Vec method. Finally, we obtain all embeddings (q for questions and s for skills) with the same dimension d of question and skill nodes in the graph, which also enclose the relation information.

To distinctively represent the questions, we fuse the various feature representations through convolutional operations to obtain the comprehensive question embeddings with additional distinctive information, i.e., the skill information and the cognitive difficulty information (Gan, Sun, Ye, Fan, & Sun, 2019). The cognitive difficulty d_q and d_s (represented as vectors) is calculated following prior work in (Gan, Sun, Peng, & Sun, 2020). Following (Y. Liu et al., 2020) and (S. Yang, Zhu, Hou, & Lu, 2020) that learn the high-order latent patterns through feature interactions and convolution operations (rather than directly concatenating the features), we map and fuse the separate features and their interactions using convolution operations, as shown in Figure 3. For questions containing multiple skills, we represent the skill as the average skill embedding $\bar{s} = Avg(s_1, \ldots, s_q)$. Fusing the above-obtained features, we generate the linear information M and quadratic information N for question q.

$$M = [q, \bar{s}, d_q, d_s] \in R^{4\times d}, \quad N = [\langle M_i, M_j \rangle] \in R^{4\times 4}, \tag{5}$$

where $\langle . \rangle$ represents the interactions of two vectors obtained by the inner product. We then apply the two-dimensional convolution operation with eight kernels of size 2×2 on both M and N, and maxpooling on each feature map to obtain $l_m \in R^{1\times(d-1)\times 8}$ and $l_n \in R^{1\times 3\times 8}$. These two parts are then concatenated into eight longer vectors $l_{m+n} \in R^{1\times(d+2)\times 8}$ including the convolutions from the separate features and their interactions. Inspired by the multi-head mechanism in the transformer model (Vaswani et al., 2017), we concatenate and linearly transform the eight vectors and hence obtain the final question representation $\tilde{q} \in R^{1\times d'}$.

$$\tilde{q} = Concat(l^1_{m+n}, \ldots, l^8_{m+n})W^O \tag{6}$$

where $W^O \in R^{((d+2)\times 8)\times d'}$ is the parameter that transforms the convolution results into a vector.

2.1.2 *Learner Knowledge State Evolution*

Since the dynamic evolution of learner knowledge in the online learning systems, the learner exercising sequences are fed into an attentive KT framework that predicts the learner performance, as shown in the right part of Figure 1.

The log data of each interaction in the exercising sequences consists of a tuple representing the question, the correctness, and the elapsed time. Look-up operations are performed on an embedding matrix $E_r \in R^{2\times d'}$, in which row vector r_t contains the incorrectness or correctness of the responses. The elapsed time et strongly evidences a student's proficiency in knowledge and skills (Shin et al., 2020). This time is converted to seconds and capped at 500 seconds. A d'-dimensional latent embedding vector for et_k is computed as $t_k = et_k W_{et} + b_{et}$, where W_{et} and b_{et} are learnable vectors. The interaction embedding is obtained as $x_t = Concat(\tilde{q}, t_t, r_t)$.

Table 3: *Statistics of the three datasets used in this study*

Dataset	Assist0910	Assist1213	Ednet
# of learners	3,002	22,339	5,000
# of questions	17,705	52,825	12,372
# of skills	123	265	188
# of interaction records	277,540	2,672,532	347,864
# skills per item	1.20	1.00	2.28
# questions per skill	172.33	199.34	149.78
# of attempted items per learner	92.45	119.64	69.57

The sequence data of the learners' exercising process are modeled using LSTM (Piech et al., 2015) within the KT framework, The hidden knowledge state h_t of a learner at step t is updated based on the current input and the previous state as: $h_t = LSTM(x_t, h_{t-1}; \theta)$.

We then employ an attention mechanism that accounts for the impact of previous attempts on the current attempt. A new question will likely be strongly affected by similar questions or questions requiring the same skillset as the new question. To describe these effects, we assume that the learner-knowledge state in the current step is the weighted sum of the aggregated states in the previous steps. The weights are based on the correlations:

$$h_{t+1} = \sum_{i=1}^{t} \alpha_{i,t+1} h_i. \tag{7}$$

The attention $\alpha_{i,t+1}$ was calculated using a combination of the shared skill-based attention and the question similarity-based attention. The learner performance at step $t+1$ can be predicted from the question representation $\widetilde{q_{t+1}}$ and the current knowledge state h_{t+1} as follows:

$$s_{t+1} = tanh(W_s[\widetilde{q_{t+1}}, h_{t+1}]) + b_s, \quad p_{t+1} = \sigma\left(W_p s_{t+1} + b_p\right), \tag{8}$$

where W_* and b_* are parameters in the fully connected layer and the sigmoid activation layer, respectively. Finally, we optimized our model by the cross-entropy loss; specifically, we minimized the following objective function between the true answer l_t and the predicted performance p_{t+1} at each interaction: $\mathcal{L} = -\sum_t \left(r_{t+1} log\, p_{t+1} + (1 - r_{t+1}) log(1 - p_{t+1})\right)$.

3. Experiments

3.1 Datasets

E xperiments were performed on three well-established datasets, namely, Assist0910[1], Assist1213[2], and EdNet[3]. All datasets are public real-world datasets containing the temporal interaction records between learners and real computer-aided tutoring systems. For the EdNet dataset, following (Y. Yang et al., 2020), we randomly selected 5000 students who answered 12,372 questions requiring 188 skills, thus obtaining 347,866 interaction logs. As done in existing studies (Vie & Kashima, 2019; Y. Yang et al.,
2020), we eliminated noise by deleting users with fewer than 10 interaction entries and questions with NaN skills from the three datasets. Table 3 summarizes the statistics of the datasets.

3.2 Compared Model

As part of our model evaluation, we competed the model against several state-of-the-art skill- and question-based KT models: BKT (Corbett & Anderson, 1994), DKT (Piech et al., 2015), DKVMN (Zhang et al., 2017), KTM (Vie & Kashima, 2019), DKT-Q (a variant of DKT that replaces the skills

[1] Assist0910: https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data/skill-builder-data-2009-2010
[2] Assist1213: https://sites.google.com/site/assistmentsdata/home/2012-13-school-data-with-affect
[3] EdNet: http://bit.ly/ednet_kt1

embedding with a one-hot encoding of questions as the input), DKT-Q&S (a variant of DKT that inputs both the questions and skill representations to the DKT), DKT-CQE (a variant of DKT that inputs our convolutional question embeddings to the DKT), GIKT (Y. Yang et al., 2020), DAS3H (Choffin et al., 2019), RKTM (Lai et al., 2021), and AKT (Ghosh et al., 2020).

Among the baseline models, the former three are skill-based models, in which the LKA is based on the skills contained in the questions. The latter eight and the proposed model are question-based models that account for the distinctive question information.

3.3 Setup and Implementation

Before conducting the experiments, we extracted 20% of the sequences in the dataset as the test set and retained the remaining 80% as the training set. To embed the nodes in the graph using Metapath2Vec, we set the length of all meta-paths as $\wp = 7$ and the number of paths as $\aleph = 100$ for each question node in the graph. The embedding dimension d of the skill and question representations was set to 128. The final dimension of the convolutional question representation was $d' = 256$. The size of the hidden layers of the LSTM was set to 256. The other hyperparameters were set through grid searching. The model was optimized using Adam optimization of the learning rate on a case-by-case basis in the three datasets. The norm clipping threshold and batch size were maintained at 10 and 64, respectively. Similar to the existing models, the sequence length of the model input was fixed at 200. The proposed model was implemented using TensorFlow. The other baselines were implemented with their best parameter settings, as specified in the original works. As the evaluation metric, we selected the area under the receiver operating characteristic (ROC) curve (AUC), which is widely used in existing studies.

4. Results and Analysis

4.1 Performance Prediction

The different models were evaluated by their performances in predicting the future learner scores from the estimated knowledge state. Table 4 presents the AUC results of all models on the three datasets. Our model outperformed the other models on all three datasets. Specifically, the AUC scores of the PQRLKA model were 0.8242, 0.7851, and 0.7754 on the Assist0910, Assist1213 and EdNet datasets, respectively, 3.97%, 1.39%, and 2.25%, respectively, above those of the state-of-the-art AKT model. Similarly, to the original DKT model, our model processes the time-series data using a recurrent neural network framework, but achieved 8.3%, 5.95%, and 8.65% higher AUCs than the DKT model on the Assist0910, Assist1213 and EdNet datasets, respectively.

The skill-based BKT model was the worst performer among the models because it tracks the mastery of each skill separately, without considering a contextual trial sequence of all skills. KTM framework, which incorporates several traditional models, typically obtained similar AUC scores to those of DKT and DKVMN. DKT extended with various input-question embeddings (DKT-Q, DKT-Q&S, and DKT-CQE) demonstrated noticeable performance differences. DKT-Q using the one-hot encoding of question representations performed much worse than the original DKT model, owing to the sparsity of question interactions in these datasets. DKT-Q&S and DKT-CQE decidedly outperformed the original DKT and DKT-Q models, consistent with our intuition that each question contains distinctive information even when it requires the same skills as one or more other questions in the dataset. Therefore, incorporating the distinctive question and skill information into the question representations can improve the model performance. The comparison between DKT and DKT-CQE also shows the effectiveness of the proposed convolutional question representation. DAS3H and RKTM are built on the KTM model by considering additional rich information and show better results than the original KTM model. Here we do not present the results of RKTM on the EdNet as the code for implementing RKTM is not published by the author. Our PQRLKA model outperforms GIKT and AKT, which also take the well-designed question embedding as input. This result validates the effectiveness of the Q-matrix refinement for the task of LKA.

	Model	ASSIST0910	ASSIST1213	EdNet
Skill-based Model	BKT	0.6571	0.6204	0.6027
	DKT	0.7412	0.7256	0.6889
	DKVMN	0.7559	0.7247	0.6921
Question-based Model	KTM	0.7582	0.7212	0.6899
	DKT-Q	0.7306	0.7158	0.6812
	DKT-Q&S	0.7616	0.7389	0.7235
	DKT-CQE	0.7998	0.7686	0.7523
	GIKT	0.7845	0.7712	0.7529
	DAS3H	0.789	0.741	0.731
	RKTM	0.7650	0.7690	–
	AKT	0.8152	0.7720	0. 7658
	PQRLKA	**0.8242***	**0.7851***	**0.7754***

Table 4: *Comparisons of the AUC results of different models on the three datasets*

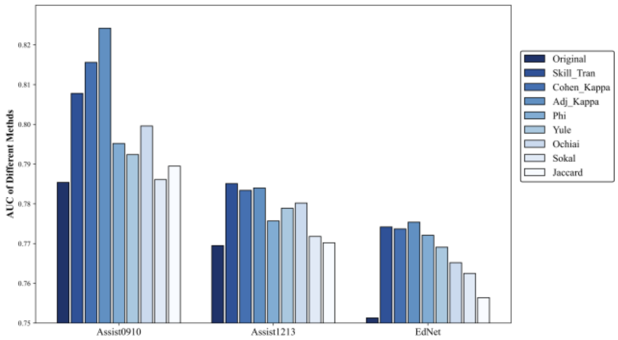

Figure 4: KS inferred from the learner response data

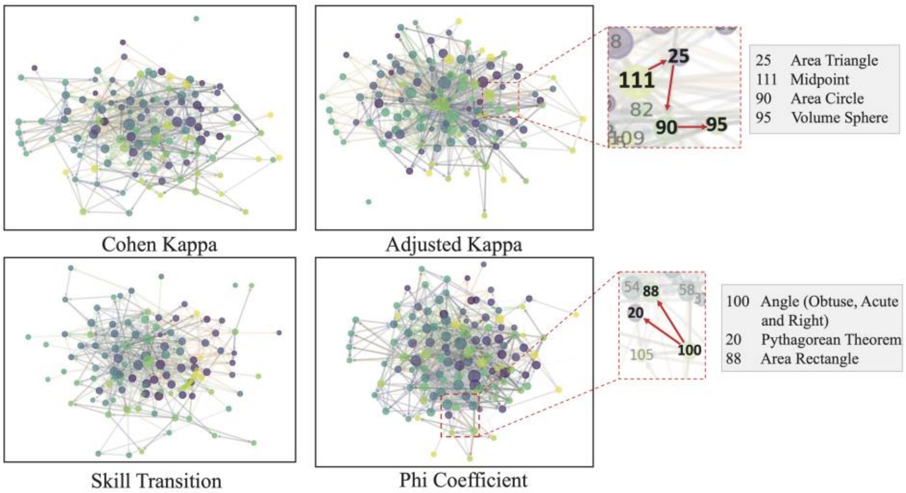

Figure 5: Visualization of the directed KG graphs generated by four methods on Assist0910

4.2 Model Analysis

4.2.1 Comparison of Prerequisite Inference by Different Methods

Figure 4 compares the AUC results of the predictions of nine methods on the three datasets. Here the "original" method represents the embedding learning on the original question--skill relation graph and the other eight methods are based on the prerequisite-enhanced graph. As evidenced in the figure, the eight KG enhanced methods generally outperformed the "original" method on all three datasets, validating the effectiveness of the KG enhanced graph in KT tasks. The adjusted Kappa yielded the best performance on both Assist0910 and EdNet, whereas the skill transaction method performed best on Assist1213.

4.2.2 Visualization of Inferred Prerequisites

Figure 5 illustrates the KGs generated from the prerequisites using four methods, inferred from the learner response data in Assist0910. The right-hand side of this figure enlarges a part of the graphs to show their local connections. The nodes and edges in the KG form dense graphs with similar structures, showing several interconnected nodes. These graphs also show some interesting properties. In the adjust-Kappa graph, four nodes (25, 111, 90, and 95) were locally interconnected and revealed a perfect ordering of the skills (prerequisite and post-requisite relations) in the geometry. The local connections in the Phi coefficient graph also presented reasonable relations among the three skills. These results confirm that our KG discovery methods can infer prerequisite skill pairs from the ordering of learners' mastery of skills.

5. Conclusions

In this paper we proposed a prerequisite-driven Q-matrix refinement framework for learner knowledge assessment in online learning context. We first explored eight methods to infer the prerequisites from learners' response data in the online learning systems and used it to refine the expert-defined Q-matrix to eliminate the potential subjective tendency of experts in designing the Q-matrix. This refinement leveraged the additional structural information (prerequisites) to allow the interpretability of the optimized Q-matrix while enabling the scalability to apply to the large-scale online learning context. Based on the refined Q-matrix, we proposed a Metapath2Vec enhanced convolutional representation method to obtain the comprehensive representations of the attempted items with rich information. These representations for the learners' exercising sequences are fed into the PQRLKA model, which considers the long-term dependencies using an attention mechanism, to finally predict the learners' performance on new items. Extensive experiments conducted on three real-world datasets demonstrated the capability and interpretability of our model to infer the prerequisites from the learning data, and the better performance of the embedding representation on the prerequisite enhanced graph, thus contributing to the superiority of the proposed model for learner knowledge assessment and validating its potential applicability to real online learning environments.

In future work, we intend to embed KG learning into our model in an end-to-end manner. The KG will then be automatically learned in the training process rather than computed from the learner performance data. Such automation can potentially expand the KG to new skills in the domain, for example, enlarging the KG of mathematics from the primary school to high school level.

Acknowledgements

This work is supported by JSPS KAKENHI with Grant Number 18K11597, 20H01720 and the grant-in-Aid for JSPS Fellows 20J15339.

REFERENCES

Baker, F. B. (2001). The basics of item response theory. ERIC.

Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In Its (pp. 164–175).

Chen, C. (2017). Q-matrix optimization for cognitive diagnostic assessment (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Chen, P., Lu, Y., Zheng, V. W., & Pian, Y. (2018). Prerequisite-driven deep knowledge tracing. In 2018 ieee international conference on data mining (ICDM) (pp. 39–48).

Chiu, C.-Y. (2013). Statistical refinement of the q-matrix in cognitive diagnosis. Applied Psychological Measurement, 37 (8), 598–618.

Choffin, B., Popineau, F., Bourda, Y., & Vie, J.-J. (2019). Das3h: Modeling student learning and forgetting for optimally scheduling distributed practice of skills. Proceedings of the 12th EDM (pp. 29–38).

Choi, Y., Lee, Y., Shin, D., Cho, J., Park, S., Lee, S., Heo, J. (2020). Ednet: A large-scale hierarchical dataset in education. In International conference on artificial intelligence in education (pp. 69–73).

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4 (4), 253–278.

De La Torre, J. (2008). An empirically based method of q-matrix validation for the dina model: Development and applications. Journal of educational measurement, 45 (4), 343–362.
De La Torre, J. (2009). Dina model and parameter estimation: A didactic. Journal of educational and behavioral statistics, 34 (1), 115–130.
Desmarais, M., Beheshti, B., & Xu, P. (2014). The refinement of a q-matrix: Assessing methods to validate tasks to skills mapping. In Educational data mining 2014.
Desmarais, M. C. (2012). Mapping question items to skills with non-negative matrix factorization. ACM SIGKDD Explorations Newsletter, 13 (2), 30–36.
Desmarais, M. C., & Naceur, R. (2013). A matrix factorization method for mapping items to skills and for enhancing expert-based q-matrices. In AIED (pp. 441–450).
Desmarais, M. C., Xu, P., & Beheshti, B. (2015). Combining techniques to refine item to skills q-matrices with a partition tree. International Educational Data Mining Society.
Dong, Y., Chawla, N. V., & Swami, A. (2017). metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD (pp. 135–144).
Gan, W., Sun, Y., Peng, X., & Sun, Y. (2020). Modeling learner's dynamic knowledge construction procedure and cognitive item difficulty for knowledge tracing. Applied Intelligence, 50 (11), 3894–3912.
Gan, W., Sun, Y., & Sun, Y. (2020). Knowledge interaction enhanced knowledge tracing for learner performance prediction. In 2020 7th international conference on behavioural and social computing (BESC) (pp. 1–6).
Gan, W., Sun, Y., & Sun, Y. (2022a). Knowledge interaction enhanced sequential modeling for interpretable learner knowledge diagnosis in intelligent education systems. Neurocomputing, 488, 36-53.
Gan, W., Sun, Y., & Sun, Y. (2022b). Knowledge structure enhanced graph representation learning model for attentive knowledge tracing. International Journal of Intelligent Systems, 37 (3), 2012–2045.
Gan, W., Sun, Y., Ye, S., Fan, Y., & Sun, Y. (2019a). Field-aware knowledge tracing machine by modelling students' dynamic learning procedure and item difficulty. In 2019 ICDMW (pp. 1045–1046).
Gan W., Sun Y., Ye S., et al. (2019b). AI-tutor: Generating tailored remedial questions and answers based on cognitive diagnostic assessment, In 6th international conference on behavioral, economic and socio-cultural computing (BESC). IEEE, 2019: 1-6.
Ghosh, A., Heffernan, N., & Lan, A. S. (2020). Context-aware attentive knowledge tracing. In Proceedings of the 26th ACM SIGKDD (pp. 2330–2339).
Hasanov, A., Laine, T. H., & Chung, T.-S. (2019). A survey of adaptive context-aware learning environments. Journal of Ambient Intelligence and Smart Environments, 11 (5), 403–428.
Jiang, C., Gan, W., Su, G., et al. (2021). Improving Knowledge Tracing through Embedding based on Metapath. In Proceedings of the 29th ICCE (pp. 11–20).
Lai, Z., Wang, L., & Ling, Q. (2021). Recurrent knowledge tracing machine based on the knowledge state of students. Expert Systems, e12782.
Liu, Q., Shen, S., Huang, Z., Chen, E., & Zheng, Y. (2021). A survey of knowledge tracing. arXiv preprint arXiv:2105.15106.
Liu, Q., Tong, S., Liu, C., Zhao, H., Chen, E., Ma, H., & Wang, S. (2019). Exploiting cognitive structure for adaptive learning. In Proceedings of the 25th ACM SIGKDD (pp. 627–635).
Liu, Y., Yang, Y., Chen, X., Shen, J., Zhang, H., & Yu, Y. (2020). Improving knowledge tracing via pre-training question embeddings. arXiv preprint arXiv:2012.05031.
Ma, W., & de la Torre, J. (2020). An empirical q-matrix validation method for the sequential generalized DINA model. British Journal of Mathematical and Statistical Psychology, 73 (1), 142–163.
Matsuda, N., Furukawa, T., Bier, N., & Faloutsos, C. (2015). Machine beats experts: Automatic discovery of skill models for data-driven online course refinement. International Educational Data Mining Society.
Minn, S., Desmarais, M. C., & Fu, S. (2016). Refinement of a q-matrix with an ensemble technique based on multi-label classification algorithms. In EC-TEL (pp. 165–178).
Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019). Graph-based knowledge tracing: modeling student proficiency using graph neural network. In 2019 IEEE/WIC/ACM WI (pp. 156–163).
Pan, L., Li, C., Li, J., & Tang, J. (2017). Prerequisite relation learning for concepts in MOOCs. In Proceedings of the 55th ACL (pp. 1447–1456).
Pandey, S., & Srivastava, J. (2020). RKT: Relation-aware self-attention for knowledge tracing. In Proceedings of the 29th ACM international conference on information & knowledge management (pp. 1205–1214).
Pavlik Jr, P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis–a new alternative to knowledge tracing. Online Submission.
Pel´anek, R. (2019). Measuring similarity of educational items: An overview. IEEE Transactions on Learning Technologies, 13 (2), 354–366.
Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In Advances in neural information processing systems (pp. 505–513).

Shin, D., Shim, Y., Yu, H., Lee, S., Kim, B., & Choi, Y. (2020). Saint+: Integrating temporal features for ednet correctness prediction. arXiv preprint arXiv:2010.12042.

Tong, H., Zhou, Y., & Wang, Z. (2020). HGKT: Introducing problem schema with hierarchical exercise graph for knowledge tracing. arXiv preprint arXiv:2006.16915.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

Vie, J.-J., & Kashima, H. (2019). Knowledge tracing machines: Factorization machines for knowledge tracing. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, pp. 750–757).

Yang, S., Zhu, M., Hou, J., & Lu, X. (2020). Deep knowledge tracing with convolutions. arXiv preprint arXiv:2008.01169.

Yang, Y., Shen, J., Qu, Y., Liu, Y., Wang, K., Zhu, Y., Yu, Y. (2020). Gikt: A graph-based interaction model for knowledge tracing. arXiv preprint arXiv:2009.05991.

Zhang, J., Shi, X., King, I., & Yeung, D.-Y. (2017). Dynamic key-value memory networks for knowledge tracing. In Proceedings of the 26th www (pp. 765–774).