Rodrigo, M. M. T. et al. (Eds.) (2021). Proceedings of the 29th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Improving Knowledge Tracing through Embedding based on Metapath

Chong JIANG[a], Wenbin GAN[b], Guiping SU[a], Yuan SUN[b] & Yi SUN[a*]
[a]*University of Chinese Academy of Science, China*
[b]*National Institute of Informatics, Japan*
*sunyi@ucas.ac.cn

Abstract: The goal of knowledge tracing (KT) is to track students' knowledge status and predict their future performance based on their learning logs. Although many researches have been devoted to exploiting the input information, they do not strictly distinguish between questions and the involved skills when taking the learning logs as input, and hence leading to performance degradation due to the fact that the inherent relations between skills and questions are not fully utilized. To solve this issue, we propose an embedding pre-training method based on metapath by explicitly considering the relations between skills and questions in the domain. Specifically, we construct a heterogeneous graph composed of skills and questions, and obtain the meaningful embeddings of nodes using the metapath2vec method, hence the explicit relation information can be embedded in the dense representation of skills and questions while still maintaining their own characteristics. Adopting these pre-trained embeddings to existing models, experiments on three public real-world datasets demonstrate that our method achieves the new state-of-the-art performance, with at least 1% absolute AUC improvement.

Keywords: Knowledge Tracing, matapath, network embedding

1. Introduction

Knowledge tracing is a fundamental task in intelligent tutoring systems to provide adaptive services to learners. The main aim of KT is to track students' knowledge status based on their learning records over time and predict their future performance accordingly.

At present, considerable progress has been achieved on this task due to the prevalence of online education. The existing models can be divided into two categories: traditional method and deep learning-based method. Bayesian Knowledge Tracing (BKT) (Corbett & Anderson, 1994) is a typical model in traditional method, which infers the evolution of a student's skill mastery using a Markov model. Deep Knowledge Tracing (DKT) (Piech, Spencer, & Sohl-Dickstein, 2015) is the first model to apply deep learning method into KT task, and has made a breakthrough.

For the KT task, the number of questions in a tutoring system generally far exceeds the number of skills. A skill may relate to many questions and a question may also correspond to more than one skill. To simplify the modeling process and improve the prediction efficiency, KT task deals with the model input in a unified way: KT models in both the traditional and deep learning categories conduct the KT process and make predictions based on the skills instead of the questions themselves. They assume that skill mastery can potentially reflect the possibility that a student can correctly solve a question incorporating that skill to some extent, hence the input to the KT models is actually skill tags alternatively. Moreover, each question is assumed to associated with only one skill. For a question containing multiple skills, a new skill is generated to represent the combination of the multiple skills.

Therefore, the relations between skills and questions, and their own characteristics are neglected, which can cause two important issues: the first one is that the tracking of students' knowledge states only stays at skill-level, and cannot truly reflect the actual ability of students to correctly solve the related questions; the second one is that the neglected relation and characteristic information are essential in predicting students' performance, and tracing students' knowledge states without considering this information will potentially cause the performance degradation. As shown in Figure 1, question q_1 is related to skill s_1, question q_2 is related to skill s_1 and skill s_2, question q_3 and question

q_4 share the same skill s_2. The main deep models track students' states at skill-level, when taking the input to the models, the question q_1 is replaced by skill s_1, the question q_3 and question q_4 are both replaced by skill s_2 though they have different difficulties. As for the question q_2, it is usually replaced by a new skill tag s_3, which means the combination of skill s_1 and skill s_2. Although the existing models perform well in the skill-level prediction, the problem of incorporating the more informative embedding representation by considering the relations between skills and questions, and their own characteristics, into the KT models to achieve more precise prediction of student performance still remains under-explored.

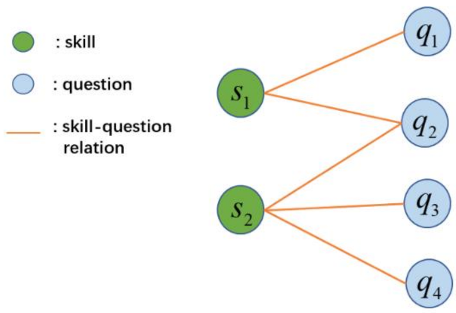

Figure 1. A Simple Example of Question-skill Relation.

In this paper, we take a further step towards exploiting the relations between skills and questions and obtaining the meaningful embedding of questions and skills together. Inspired by PEBG method (Liu, Y., Yang, Y & Yu, 2020), which learns a low-dimensional embedding for each question on the side information and the side information includes question difficulty and three kinds of relations contained in a bipartite graph between questions and skills, we abstract the relations between questions and skills as a heterogeneous bipartite graph considering that a skill may relates to many questions and a question may corresponds to more than one skill, and construct a series of meta-path to sample the whole graph. Deep embedding representation of each node in the graph is then obtained by not only using the own information, but also aggregating the information from connected skills and questions by utilizing the relations of skill-question in the graph. Finally, we incorporate the pre-training embeddings with the existing deep KT models to test its performance.

The contributions of this paper can be summarized as follows:

- By training the embeddings of questions and skills, we can track students' knowledge states at both skill- and question-level.
- We propose a pre-training approach to learn the pre-training embedding of skills and questions by leveraging the explicit relations between skills and questions in the heterogeneous bipartite graph, and incorporate the embeddings into the KT process.
- The extensive experiments on three benchmark datasets show that our model outperforms the existing state-of-the-art models, with at least 1% absolute AUC improvement.

2. Related Work

2.1 Knowledge Tracing

Existing KT models can be divided into two categories: traditional models and deep models, and deep models have shown overwhelmingly better performance than the traditional models. In this paper, we only focus on the deep KT models. Deep knowledge tracking (DKT) model (Piech, Spencer, & Sohl-Dickstein, 2015) is a milestone work to apply deep learning model in KT task, it uses a recurrent neural network (RNN) to model learners' learning process by taking the one-hot embedding representation of exercises and interactions. The other researchers subsequently have proposed various models to improve the DKT, the improvements are mainly in two aspects.

The first category improves the DKT model by designing different kinds of neural network structures to model students' learning process. Dynamic Key Value Memory Networks (DKVMN) (Zhang & Yeung, 2017) introduces Memory-Augmented Neural Networks (MANN) to solve the KT task and abstractly extracts the knowledge states of students. Graph-Based Knowledge Tracing (GKT) (Nakagawa & Matsuo, 2019) introduces the graph neural network (GNN) to solve the KT task which reformulates the KT task as a time series node-level classification problem in GNN.

The second category improves the DKT model by taking different embedding representations as input. The one-hot embedding representation of DKT only contains extremely sparse skill information and neglects the special characteristics of problems. Moreover, this representation can't reflect the relationships between skill and question. To solve these issues, Exercise-Aware Knowledge Tracing (EKT) (Liu, Q & Hu, 2019) trains the students' exercise records and the corresponding text content into a new embedding as the input of the network. Context-Aware Attentive Knowledge Tracing (AKT) (Ghosh & Lan, 2020) uses the transformer to train the embeddings of questions and skills based on the prior knowledge. Deep Knowledge with Convolutions (CKT) (Yang & Lu, 2020) uses three-dimensional convolution to aggregate the answer information of students in a period of time, and then trains the corresponding embedding as the input of LSTM network. KTM-DLF (Gan & Sun, 2020) model students' learning and forgetting behaviors by taking account of their memory decay and the benefits of attempts when an item can involve multiple KCs.

In addition to the above progress, the application of graph neural network and dense embedding to KT task also achieves good results. GIKT (Yang & Yu, 2020) uses a Graph Convolutional Network to aggregate the relationship between question and skill and obtain the skill embeddings as final input to RNN models. Following this line, the PEBG proposes a pre-training framework to get the embeddings of question. This paper also applies graph neural network and pre-trained dense embedding in KT task, however, different from these models, it considers the relations between skills and questions as a heterogeneous information graph and uses the metapath based method to learning the dense embeddings of skills and problems, and hence it can track students' knowledge states at both skill- and question-level, and obtains better performance.

2.2 Network Embedding

Network embedding (Chen & Skiena, 2018) aims to learn the low dimensional potential representation of nodes in the network. The learned feature representation can be used as the feature of various graph-based tasks, such as classification, clustering, link prediction and visualization. DeepWalk (Perozzi & Skiena, 2014) was proposed as the first network embedding method using deep learning. DeepWalk treats nodes in graph as words and generating short random walks as sentences. Then, neural language models such as Skip-gram (Mikolov & Dean, 2013) can be applied on these random walks to obtain network embedding. On the basis of DeepWalk, node2vec (Grover & Leskovec, 2016) adopts a biased wandering strategy which contains DFS and BFS. The network embedding method mentioned above is mainly used in isomorphic graphs. For heterogeneous graphs, it has different node types and edge types. Metapath2vec (Dong, Chawla & Swami, 2017) uses random walks based on meta path to construct the heterogeneous neighborhood of each node, and then uses skip gram model to complete the node embedding.

3. Problem Definition

3.1 Definition 1(Knowledge Tracing)

The KT task can be formally defined as follows: given a student's exercising records over a period of time $X_t = (x_1,x_2,\ldots,x_t)$, the KT model predicts the student's performance at the next moment x_{t+1}. Each interaction is composed of a question and the label indicating the correctness of student answer, hence x_t can be expressed as a pair of (q_t,a_t). The ordered pair indicates that the student has answered question q_t at time t and the correctness is a_t. The KT task aims to predict the probability $P(a_{t+1}=1|q_{t+1},X_t)$ of the student answering question q_{t+1} correctly.

3.2 *Definition 2(Question-Skill Relation Graph)*

A skill s_i may relate to many questions such as $\{q_1,\dots,q_m\}$ and a question q_i may contain many skills such as $\{s_1,\dots,s_n\}$. Here, we abstract a heterogeneous bipartite graph $G = (S,Q,R)$, where $R = [r_{ij}] \in \{0,1\}^{|S|\times|Q|}$ is a binary adjacency matrix and S means the set of all skills and the Q means the set of all questions. If question q_j contains skill s_i , there is edge between them in the graph G, and the entry in adjacency matrix $r_{ij} = 1$; otherwise $r_{ij} = 0$. In our model, the edge is bidirectional as shown in Figure 2.

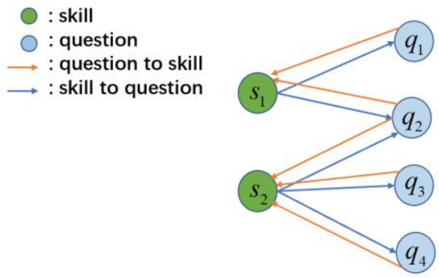

Figure 2. The Bidirectional Question-skill Relation Graph.

3.3 *Definition 3(Meta-path)*

Meta-path is a path containing a sequence of relations defined between objects of different types. The specific form is

$$A_1 \overset{R_1}{\rightarrow} A_2 \overset{R_2}{\rightarrow} \dots \overset{R_l}{\rightarrow} A_{l+1} \tag{1}$$

It represents a compound relation between node types A_1 and A_{l+1}.The probability that the walker will take one step specified in the path is:

$$P(v_{t+1}|v_t,\rho) = \begin{cases} \frac{1}{|N_{t+1}(v_t)|}, & (v_t,v_{t+1}) \in E, \phi(v_{t+1}) = A_{t+1} \\ 0, & otherwise \end{cases} \tag{2}$$

Where $N_t(v)$ denotes v 's neighborhood with the t^{th} type of nodes and $\phi(v_{t+1})$ indicates the node v_{t+1} is in type of A_{l+1}. The constructed Meta-path is usually required to be symmetric based on the types to facilitate expansion. For a given Graph $G=(V,E,T)$ and a node v, we learn the embedding by maximizing the probability:

$$\arg\max_{\theta} \sum_{v\in V} \sum_{t\in T_V} \sum_{c_t\in N_t(v)} \log p(c_t \mid v;\theta) \tag{3}$$

where c_t is the context of node v and $p(c_t|v;\theta)$ is commonly defined as a Softmax function. In our question-skill heterogeneous bipartite graph, we only have two node types and two relation types.

4. Method

This section introduces our proposed method. Metapath2vec is applied to learn question and skill embeddings aggregated on the question-skill relation graph, and then a recurrent layer is used to model the sequential change of knowledge state. Then, we design an interaction module for the final prediction. The process of learning meta-path embedding representation of skill and question is shown in Figure 3. The recurrent layer and the interaction module for modeling student performance are shown in Figure 4.

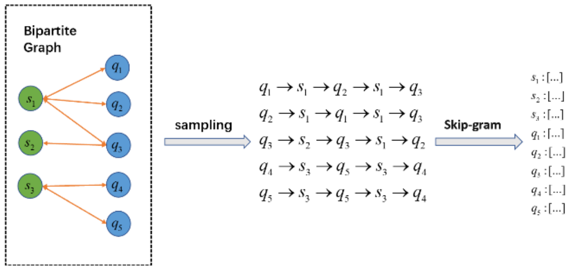

Figure 3. The process of Metapath based Embeddings.

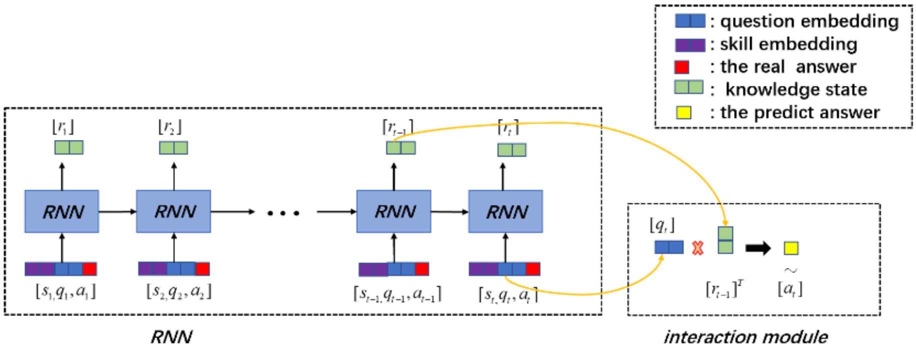

Figure 4. The Recurrent Layer and the Interaction Module.

4.1 Pre-training Question and Skill Embedding

As shown in Figure 3, a heterogeneous bipartite graph is constructed based on the relationship between skill and question in the dataset. We use a two-way arrow to show the connection between skill and question. These arrows represent the "question to skill" and "skill to question" relations in Figure 2. In the following, we use send to represents "question to skill" and back to represents "skill to question". To predict whether a student can answer a question correctly, we need to understand the correlation between the question and the student's knowledge state. This requires a careful consideration of the relationship between question and skill when coding them.

After constructing the relation graph of question and skill, a sampling process is conducted in the heterogeneous bipartite graph based on Metapath. Here we define a Metapath with five nodes, which is defined as follows:

$$Q \xrightarrow{send} S \xrightarrow{back} Q \xrightarrow{send} S \xrightarrow{back} Q \quad (4)$$

where Q means "question" and S means "skill". By constructing such Metapath, we can aggregate "question", "skill" and their relationship in a path at the same time. When we sample the path from the graph, we also set condition: the number of sampling paths must be equal to the number of questions, which requires that the starting node of each path is a different question node. As shown in Figure 3, there are five question nodes in the figure, and we get five paths by sampling. The advantage of setting this condition is that we can start from each question node when sampling, hence we can collect the information from the whole graph.

Finally, the unified d dimension embeddings of all the questions and skills can be obtained by training the sampled paths using metapath2vec (Dong, Chawla & Swami, 2017).

4.2 Metapath-Based Knowledge Tracing

We use DKT model as the baseline model and extend it with the pre-trained embeddings of questions and skills to formulate our model. In our model, we use combination of skill embedding and question embedding to replace one-hot embedding in DKT to trace students' knowledge state and use question embedding to directly predict the probability of students correctly answering the next question.

In order to model the learning process of students, we use the RNN to process the exercise sequence, as shown in Figure 4. The input features x_t to our model is the combination of the skill embeddings and question embeddings attached with the students' answer at time t, hence the dimension of final input features is $2*d+1$. In addition, we can also use skill embeddings or question embeddings alone to track students' knowledge status, and the dimension of input features is $d+1$.

$$x_t = [s_t, q_t, a_t] \tag{5}$$

The formula of knowledge tracing process in RNN is as follows:

$$h_t = \tanh(W_{hx} x_t + W_{hh} h_{t-1} + b_h) \tag{6}$$

$$r_t = \sigma(W_{yh} h_t + b_y) \tag{7}$$

Where the x_t, h_t, r_t represents the input features, the hidden state and output state. The output state r_t represents a students' current knowledge state. The prediction result on the correctness of the student on the next question is obtained by interacting the output state in the current step with the question embedding at the next moment.

$$\tilde{a}_t = q_t \odot r_{t-1}^T \tag{8}$$

To optimize our model, we update the parameters in our model using gradient descent by minimizing the Cross Entropy Loss between the predicted probability of answering correctly and the true label of the student's answer:

$$\mathcal{L} = -\sum_t (a_t \log \tilde{a}_t + (1-a_t)\log(1-\tilde{a}_t)) \tag{9}$$

5. Experiments

We conduct several experiments to investigate the performance of our model. We first compare the performance of our model with five baselines on three public datasets. Then we conduct some ablation studies to investigate the effectiveness of our proposed model.

5.1 Datasets

We conduct experiments on three public datasets: ASSISTments2009, ASSISTments2012 and EdNet. The statistical information of these datasets is reported in Table 1.

Table 1. *Dataset Statistics*

	ASSIST09	ASSIST12	EdNet
#students	3,852	27,485	5,000
#questions	17,737	53,065	12,150
#skills	167	265	1774
#exercises	282,619	2,709,436	676,974
question pre skill	107	200	6.849
skill pre question	1.197	1.000	2.280
attempts per question	15	51	56
attempts per skill	1,692	10,224	381

ASSISTments2009 was collected from ASSISTments online education platform during the school year 2009-2010. For this dataset, we remove records without skills and scaffolding problems. Similar to other knowledge tracking methods, we also remove users with less than three records from the original dataset. In ASSISTments2009 dataset, it has 3852 students with 167 skills, 17,737 questions and 282,619 exercises.

ASSISTments2012 was collected from ASSISTments online education platform during the school year 2012-2013. ASSISTments2012 is different from ASSISTments2009 in that it has one characteristic that needs special emphasis. In ASSISTments2012 dataset, each question is only related to one skill, but one skill still corresponds to several questions. After the same data processing as ASSIST09, it has 2,709,436 exercises with 27,485 students, 265 skills and 53,065 questions.

EdNet was collected by (Choi, Lee & Heo,2020). In EdNet dataset, we only choose part of it. We choose EdNet-KT1 which has 676,974 exercises with 5,000 students, 1,774 skills and 12,150 questions after consistent processing with other data sets.

5.2 Baselines

We compare with five models, as follows:

BKT (Corbett & Anderson,1994) is a Bayesian model defined by initial knowledge, learning rate, slip and guess parameters. It models the knowledge state of the skill as a binary variable.

DKT (Piech, Spencer & Sohl-Dickstein, 2015) is the first method that introduces deep learning into knowledge tracing task. It uses recurrent neural network to model knowledge state of students.

DKVMN (Zhang, Shi & Yeung, 2017) introduces Memory-Augmented Neural Networks (MANN) to solve the KT task and abstractly extracts the knowledge states of students.

PEBG (Liu, Yang & Yu, 2020) uses the bipartite graph of question-skill relations to obtain question embeddings, which provides plentiful relation information.

AKTHE (Zhang, Du & Sun, 2020) capture the relevance of historical data to the current state by using attention mechanism.

5.3 Implementation Details

We implement all the compared methods using Pytorch. In the Metapath2vec, we use heterogeneous skip-gram model to get the embeddings and set the window size as 5, the batch as 128, the number of epoch as 10, the learning rate as 0.01. The Pre-traing embeddings of skill and question are both 10 dimensions.

In the implementation of RNN, a RNN with only one hidden layer is used and the max length of the RNN is set to 50. The batch size is set to 64 and the max epoch is 50, the learning rate is set to 0.001.

Table 2. *The AUC Results over Three Datasets*

	ASSIST09	ASSIST12	EdNet
BKT	0.6571	0.6204	0.6027
DKT	0.7561	0.7286	0.6822
DKVMN	0.7550	0.7283	0.6967
PEBG+DKT	0.8287	0.7665	0.7765
AKTHE	0.8310	0.7782	0.7757
OURS	0.8432	0.8047	0.7881

5.4 *Overall Analysis*

We use the area under the curve (AUC) as an evaluation metric to evaluate the performance. The higher the AUC, the better the model performances. The result is shown in Table 2.

From the results we observe that our MKT model achieves the highest performance over three datasets, which verifies the effectiveness of our model. To be specific, our MKT model achieves at least 2% higher results than other baselines on ASSIST09. The main reason why our model performs well is that the input embedding contains more information than the previous model and the ASSIST09 dataset is a typical bipartite graph which nodes interacting each other. As the figure 5 shows, after visualizing the question embedding through t-sne, we find that although there are many questions, we can find that the question embedding with the same skill tag is more intensive in the visualization graph, which reflects the effectiveness of specific meta path extraction information to a certain extent.

So, even in EdNet, our model also gets best result. As for ASSIST12, its question only relates to one skill, our model may not be able to fully capture the relationship between skill and skill, which has also been proved in next ablation study, but our model still performs better than other models.

Compared with PEBG, which also uses the pre-training embedding in the KT task, our model achieves better results with significantly fewer dimension in the embedding process (we only use 10 dimensions while PEBG use 128 dimensions). Moreover, PEBG also adds the similarity between skill and skill, and the similarity between question and question, better results may be further achieved by integrating the similarity calculation in the PEBG into our model.

Compared with AKTHE, which capture the relevance of historical data to the current state by using attention mechanism, our model gets the information between skills and questions by interpretable meta-path and get better performance.

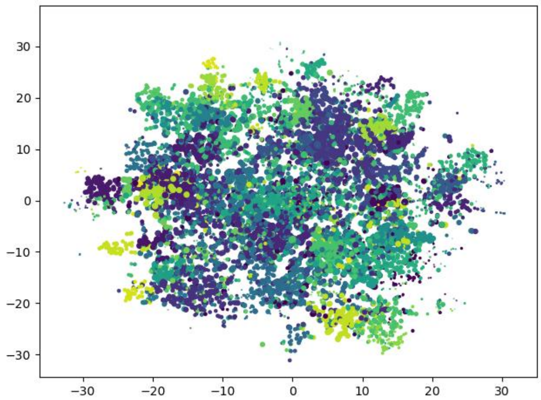

Figure 5. Visualization of Question Embedding (ASSIST09).

5.5 *Ablation Study*

In this section, we conduct some ablation studies to investigate the effectiveness of our proposed model. We use the single skill embeddings or question embeddings to replace the combination of the skill embeddings and question embeddings. Our experiment shows that the combination of skill and question improves the performance of our model.

As the result shows, the single question embeddings get the best performance in ASSIST12 dataset, but the gap between it and skill-question embedding is very small. After analyzing this dataset, we find that this is related to the characteristics of datasets. The dataset ASSIST12 is different with other two datasets because its question is only related to one skill. Hence in our model, the question embedding can fully contain the information of other questions with the same skill in the meta-path. The MKT-question model does not perform as well as MKT-skill in other datasets where questions generally contain multiple skills, but performs better than MKT-skill in ASSIST12.

Table 3. *The AUC Results over Three Datasets*

	ASSIST09	ASSIST12	EdNet
MKT-S	0.8410	0.7950	0.7815

MKT-Q	0.8260	**0.8055**	0.7784
MKT-S-Q	**0.8432**	0.8047	**0.7881**

6. Conclusion

In this paper, we have proposed a Metapath-based interaction model named MKT, which improves the KT by formulating the question-skill relations as a bipartite graph and introducing Metapath2vec to learn low dimensional embeddings of questions and skills for knowledge tracing. Experiments on three datasets show that MKT significantly improves the performance of existing deep KT models at both skill-level and question-level. Besides, ablation study shows the effectiveness of combination of skill and question for the task of KT, which provides an explanation of its high performance.

Acknowledgements

This work is supported by JSPS KAKENHI Grant Number 18K11597.We would also like to thank four anonymous reviewers for their comments.

References

Chen, H., Perozzi, B., Al-Rfou, R., & Skiena, S. (2018). A tutorial on network embeddings. *arXiv preprint arXiv:1808.02590*.

Choi, Y., Lee, Y., Shin, D., Cho, J., Park, S., Lee, S., ... & Heo, J. (2020, July). Ednet: A large-scale hierarchical dataset in education. In *International Conference on Artificial Intelligence in Education* (pp. 69-73). Springer, Cham.

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.

Dong, Y., Chawla, N. V., & Swami, A. (2017, August). metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 135-144).

Ghosh, A., Heffernan, N., & Lan, A. S. (2020, August). Context-aware attentive knowledge tracing. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2330-2339).

Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).

Liu, Q., Huang, Z., Yin, Y., Chen, E., Xiong, H., Su, Y., & Hu, G. (2019). Ekt: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering*, *33*(1), 100-115.

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, November). Knowledge Interaction Enhanced Knowledge Tracing for Learner Performance Prediction. In *2020 7th International Conference on Behavioural and Social Computing (BESC)* (pp. 1-6). IEEE.

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*.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019, October). Graph-based knowledge tracing: modeling student proficiency using graph neural network. In *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 156-163). IEEE.

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 701-710).

Piech, C., Spencer, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *arXiv preprint arXiv:1506.05908*.

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, April). Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web* (pp. 765-774).

Zhang, N., Du, Y., Deng, K., Li, L., Shen, J., & Sun, G. (2020, August). Attention-Based Knowledge Tracing with Heterogeneous Information Network Embedding. In *International Conference on Knowledge Science, Engineering and Management* (pp. 95-103). Springer, Cham.