Knowledge Tracking Model Based on Learning Process

Yu Zou[1,2], Xiaodong Yan[1,2], Wei Li[1,2]

[1]School of Information Engineering, Minzu University of China, Beijing, China
[2]National Language Resource Monitoring & Research Center Minority Languages Branch, Beijing, China
Email: yanxiaodong@muc.edu.cn

Abstract

Knowledge tracking model has been a research hotspot in the field of educational data mining for a long time. Knowledge tracking can automatically discover students' weak knowledge points, which helps to improve students' self-motivation in learning and realize personalized guidance. The existing KT model has some shortcomings, such as the limitation of the calculation of knowledge growth, and the imperfect forgetting mechanism of the model. To this end, we proposed a new knowledge tracking model based on learning process (LPKT), LPKT applies the idea of Memory Augmented Neural Network(MANN).When we model the learning process of students, two additional important factors are considered. One is to consider the current state of knowledge of the students when updating the dynamic matrix of the neural network, and the other is to improve the forgetting mechanism of the model. In this paper we verified the effectiveness and superiority of LPKT through comparative experiments, and proved that the model can improve the effect of knowledge tracking and make the process of deep knowledge tracking easier to understand.

Keywords

1. Introduction

With the continuous development of Internet education, the use of artificial intelligence technology to promote education has become an inevitable trend [1]. In recent years, intelligent education and online education platforms are becoming more and more popular. How to personalize each student's learning situation and how to make online education system realize intelligent learning

guidance [2] are important research topics in the field of intelligent education at present, and also the basic trend of future education development. The characteristics of knowledge tracking are personalization and automation [3]. Its task is to automatically track the change process of students' knowledge status with time according to the interaction between students and intelligent tutoring system, so as to accurately predict students' mastery of knowledge points and their next presentation. KT task can be formalized as a supervised sequential learning problem: given the historical interaction sequence of students, the probability of students' correct answer to the next exercise can be predicted. The typical knowledge tracking methods are: Bayesian Knowledge Tracking (BKT) [4] and Deep Knowledge Tracking (DKT) [5]. The initial knowledge tracking is based on probability Bayesian knowledge tracking model (BKT), In essence, it is a special case of Hidden Markov Model. It divides the knowledge system into multiple knowledge points, and the students' mastery of each knowledge point is regarded as an implicit state, then the probability distribution of hidden variables is updated according to the students' historical answers. However, the BKT model has the following shortcomings [6]: firstly, it needs to label the data. Secondly, the concept of each knowledge point is expressed separately. BKT cannot capture the correlation between different concepts and cannot effectively represent the complex concept state transition. With the development of deep learning, researchers apply deep learning to the field of knowledge tracking, and propose a deep knowledge tracking model (DKT), which uses Long-Short Term Memory Network (LSTM) [7] for knowledge tracking task. It not only has better prediction performance than BKT model, but also does not need experts to annotate the knowledge points of exercises. However, the DKT model represents the students' mastery of knowledge points with a hidden state, and the hidden state cannot be explained. Therefore, the DKT model cannot output the students' mastery level of each knowledge point in detail [8]. Moreover, LSTM stores all memories in a hidden vector, which makes it difficult for LSTM to accurately remember sequences with more than hundreds of time steps. Memory Augmented Neural Network (MANN) [9] is a solution to the above problems: it allows the network to retain multiple hidden state vectors, which are read and written separately. The representative models of MANN include end-to-end memory network [10], dynamic memory network [11], etc. The latest development in the field of knowledge tracking is the Dynamic Key-Value Memory Network (DKVMN) proposed in 2017 [12]. It draws on the ideas of MANN and combines the advantages of BKT and DKT, and achieves better prediction performance. In addition, DKVMN has several other advantages over LSTM, including avoiding over fitting, fewer parameters, and automatically discovering similar exercises through potential concepts.

2. Deficiency of Dynamic Key Value Memory Network

Although DKVMN has made a breakthrough in the field of knowledge tracking,

which greatly improves the efficiency of knowledge tracking and improves the interpretability of deep knowledge tracking model, there are still several problems as follows:

First of all, there are limitations in the calculation of knowledge growth. In DKVMN, the amount of knowledge growth is calculated by multiplying the learning activity of students' answering questions and a trained embedded matrix, which means that the knowledge growth gained by students after each question answering activity is only related to this activity. However, in fact, from the perspective of human cognitive process, students' knowledge growth in learning should also be related to students' current knowledge status [13]. For the same question answering learning activity, the amount of knowledge growth gained by students with a certain foundation and students who have no contact with this knowledge point is different [14].

Secondly, it relies too much on the forgetting mechanism of the model itself. In DKVMN, the updating process of students' knowledge state is inspired by LSTM forgetting mechanism. Firstly, the "erase" vector is calculated by the hidden layer of a sigmoid activation function, and then the "erase" vector and the student's knowledge growth are used to update the dynamic knowledge state matrix of students [15]. However, according to the research on the forgetting process of human beings by the famous German psychologist Ebbinghaus, the forgetting process of learning is also affected by the current knowledge state of students [16]. In fact, the amount of knowledge forgotten by students who have just started learning should be greater than that after they have learned for a period of time.

Finally, the forgetting mechanism is not considered in the prediction process. DKVMN uses the advantages of large capacity memory of MANN to model the learning process of students. Originally, MANN is widely used in intelligent question answering and machine translation. It stores the learned knowledge in the dynamic matrix by reading a large number of documents. Therefore, the process of intelligent question answering and machine translation is similar to the process of retrieval. However, in the field of knowledge tracking, the prediction process of predicting whether students can correctly answer the next question is not a simple retrieval process. It also needs to consider the students' memory forgetting in learning, which is obviously not considered in DKVMN.

According to the above, we can see that the superiority of the current deep knowledge tracking method is attributed to the deep learning model itself. In essence, to achieve a better knowledge tracking effect, we need to start from human learning cognitive psychology and complete the knowledge tracking process by simulating students' learning and memory process. Therefore, we propose a knowledge tracking model based on learning process (LPKT), it adopts the idea of Memory Augmented Neural Network to model the learning process of students. We make two improvements: one is to consider the current knowledge state of students when updating the dynamic matrix of MANN, the

other is to improve the forgetting mechanism of the model to make the reading and writing process of the model conform to the learning forgetting mechanism of human.

3. Knowledge Tracking Model based on Learning Process

The knowledge tracking model based on learning process (LPKT) aims to complete knowledge tracking by simulating students' learning and memory process. Its structure is shown in **Figure 1**. The static matrix K stores the information of all knowledge points, while the dynamic matrix V stores the knowledge state of students. The process of knowledge tracking in this model is mainly divided into three parts. The attention mechanism is mainly to calculate the number of knowledge points involved in a problem and the proportion of each knowledge point. The other two parts are the key steps of MANN. One is the reading process, which is to observe a period of learning sequence data in the learning system of students in the past period of time. The other is the writing process, that is, given a question answering activity of students, the dynamic matrix V representing students' knowledge state is transformed from state V_{t-1} at time t − 1 to state V_t at time t. I will introduce the three processes in detail with examples.

3.1. Attention Mechanism

The attention mechanism in MANN can be understood as finding the knowledge points involved according to the students' questions in the answering activities. In the answering learning activities, attention will be used in the reading and writing process of MANN. The calculation process of attention includes:
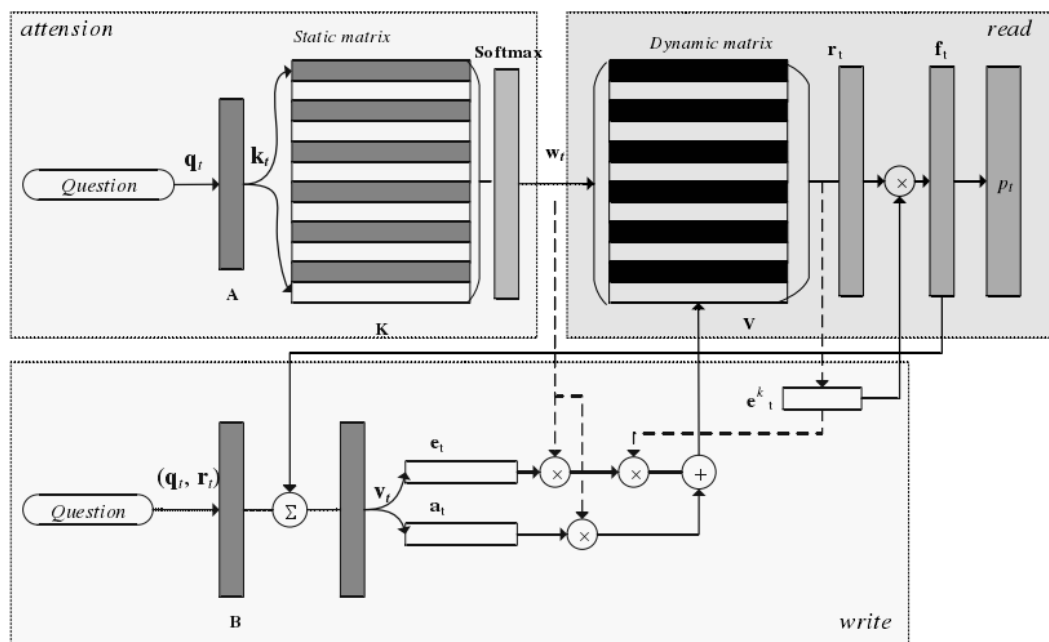

Figure 1. The structure of knowledge tracking model based on learning process (LPKT).

firstly, according to the problem encountered by students at t moment, multiply it with a trained embedded matrix A to obtain vector k_t, and then process k_t through static matrix K to obtain attention vector w_t. The calculation process is as follows:

$$w_t(i) = softmax(k_t^T \bullet M_k(i)) \tag{1}$$

where $K(i)$ represents the vector represented by the i-th row of the matrix K, represents the i-th knowledge point c_i, and $w_t(i)$ represents the attention paid to the i-th knowledge point, that is, how much proportion of the problem involves the knowledge points, and the symbol $\bullet$ represents the inner product operation between vectors.

3.2. Reading Process

The reading process is the prediction process of knowledge tracking. Firstly, according to the attention vector, the students' mastery of the knowledge points involved in the problem is read from the knowledge state matrix of the students. In the DKVMN, this process is calculated as follows:

$$r_t = \sum w_t(i)V_t(i) \tag{2}$$

However, considering the forgetting mechanism in the learning process, we have to carry out two additional steps. First, we calculate the amount of knowledge forgetting of the student according to his knowledge state V_t:

$$e_t^k = sigmod\left(E_e^k V_t + b_e^k\right) \tag{3}$$

Then, referring to the forgetting mechanism of LSTM, according to the forgetting vector, attention vector and input vector, the knowledge state $V_t^{'}$ in accordance with the learning law of students is calculated:

$$V_t^{'} = V_t(i)[1 - w_t(i)e_t^k] \tag{4}$$

So we can modify the formula to:

$$r_t = \sum_i w_t(i)V_t^{'}(i) \tag{5}$$

Then, the knowledge state vector r_t and the input vector k_t are processed by multi-layer perceptron to get vector f_t, which reflects the students' knowledge state and the characteristics of the problem itself, such as the difficulty of the problem, and shows the comprehensive knowledge state of the students for a specific problem:

$$f_t = tanh(W_1^t[r_t, k_t] + b_1) \tag{6}$$

Finally, the vector a is passed through the sigmoid output layer:

$$p_t = sigmod(W_2^T f_t^{'} + b_2) \tag{7}$$

Then we can get the probability that students can correctly answer the question. So far, the reading process of knowledge tracking method based on learning and memory process has been completed.

3.3. Writing Process

The process of writing in MANN is to update students' dynamic knowledge state in knowledge tracking. Firstly, according to the model mechanism of MANN, a question answering activity $x_t = (q_t, a_t)$ is multiplied by another embedded matrix B to obtain vector v_t, and v_t represents the knowledge increment gained by students. Because Ha points out that the knowledge increment of this dependent model is not enough to express the students' gains in the learning process, and proposes that the knowledge state of students should be considered when calculating the knowledge increment of students, so the knowledge increment of students is expressed as v_t':

$$v_t' = [v_t, f_t] \tag{8}$$

After we get the increase of students' knowledge, we update the dynamic matrix V in MANN with the method similar to the "forgetting gate" mechanism in LSTM, which is called "erase" in DKVMN. Generally, the calculation process of "erase" vector to determine the number of forgetting is as follows:

$$e_t = \text{sigmod}(E_e v_t' + b_e) \tag{9}$$

However, from the formula, we can draw a conclusion that for the same student, as long as the amount of knowledge growth is the same, the "erase" vector is also the same, which is obviously contrary to common sense. Moreover, Ha points out that DKVMN model, as a method of calculating forgetting vector, will lead to too much forgetting content. Although Ha gives a regularization method to modify it, however, this correction method is not very explanatory.

According to human cognitive process [17] and the forgetting theory of learning, the forgetting curve of students' memory in the learning process is not only related to the current knowledge increment, but also to the current learning duration of students. In the knowledge tracking model of this paper, the correlation between current and learning duration can be understood as the correlation with the current knowledge state of students. Therefore, combined with the forgetting vector based on the knowledge state of students in formula (3), we use the linear combination of e_t^k and e_t to represent the "erase" vector e_t' in the student's question answering learning activity:

$$e_t' = \lambda_1 e_t + \lambda_2 e_t^k \tag{10}$$

where $\lambda_1 \in (0,1)$, $\lambda_2 \in (0,1)$, $\lambda_1 + \lambda_2 = 1$, the initial values of λ_1 and λ_2 are 0.5.

When the value in the dynamic matrix V is "erased", we calculate the update vector according to the knowledge growth vector, and the calculation process is similar to that of LSTM:

$$\alpha_t = \tanh(W_a v_t' + b_a) \tag{11}$$

Finally, through the process of "erasing" and then updating, the updating process of students' dynamic knowledge state value is as follows:

$$V_t = V_{t-1}[1 - w_t(i)e_t'] + w_t(i)\alpha_t \tag{12}$$

That is to say, after the students' answering behavior at time t, the value of dynamic matrix is transformed from V_{t-1} to V_t.

The optimization goal of our model is to minimize the difference between the predicted value and the actual value of the students' answer results, that is, to minimize the cross entropy of p_t and a_t. So our loss function is:

$$L = -\sum_t (a_t \log p_t + (1-a_t)\log(1-p_t)) \tag{13}$$

And we use the random gradient descent method for training.

4. Experimental Design and Result Analysis

4.1. Dataset

We verify the effectiveness of our method on the data sets of ASSIST Ments 2009 [18] and ASSIST Ments 2015 [19]. These two datasets are from the ASSIST Ments online education platform, reflecting students' learning activities on the platform. Xiong once pointed out that there are problems such as duplicate data in dataset ASSIST Ments 2009, so a new version of dataset "skill builder_ data_ corrected" has been officially released. However, the skill name attribute of this version of data is still empty. In addition, we also found that the sparse learning records of some students are not helpful to our learning process. Therefore, we screened the data and got the statistical results shown in **Table 1**. A total of 315,527 records were obtained, representing 3091 students' answers to 110 different questions. Similarly, in the ASSIST Ments 2015 data set, we collected 628,507 records, which recorded 14,228 students' answers to 100 questions.

4.2. Experimental Design

This paper uses Apache MXNet, an open source deep learning software framework, to implement a knowledge tracking model based on learning and memory process, the experimental results are compared with the current knowledge tracking methods.

Standard DKT model: Piech uses luascripting language to implement DKT model in torch framework. In order to facilitate the experiment, this paper uses Python 3.6 to reimplement DKT model in tensorflow GPU version 1.9.0 framework, and the implementation code refers to public code.

DKVMN model: DKVMN model is currently the best model for knowledge tracking. In this paper, we use the public code provided by Zhang on GitHub.

In addition, in order to compare with DKVMN, the setting of memory

Table 1. Experimental data set information.

Datasets	#Students	#Questions	#Records
ASSIST Ments 2009	3091	110	315,527
ASSIST Ments 2015	14,228	100	628,507

augmented neural network in the two models is referred to Zhang's model setting. As shown in **Table 2**.

We divide the data set into training set, cross validation set and test set. The training set accounts for 60% of the data set, and the cross validation set and test set account for 20% respectively. We use cross entropy as the loss function, use SGD optimization algorithm to train, and set the learning rate to 0.005. We use the area under receiver operating characteristic (ROC) curve to measure the performance of the model. Firstly, as a performance measure, AUC has been widely concerned in the field of machine learning, especially for class imbalance problems. In addition, most of the papers in KT field use AUC as the evaluation index, so our experimental results can be easily compared with those in KT field.

4.3. Experimental Results and Analysis

Our experimental results are shown in **Table 3**.

It can be seen that in ASSIST Ments 2009 data set, the AUC value of LPKT model proposed in this paper (82.35%) is 1.71% and 1.17% higher than that of DKT model (81.18%) and DKVMN model (80.64%), while in ASSIST Ments 2015 data set, the AUC value of LPKT model (73.83%) is 4.62% and 1.15% higher than that of DKT model (69.21%) and DKVMN model (72.68%). From the comparison, the LPKT model in this paper shows better tracking effect. Compared with the DKT method using LSTM, DKVMN and LPKT using MANN have larger capacity to remember more content, so the effect of knowledge tracking is significantly improved. The advantage of LPKT over DKVMN is that

Table 2. Parameter setting of memory augmented neural network.

Datasets	Parameter	Value
ASSIST Ments 2009	Dimension of embedded matrix A	110×50
	Dimension of embedded matrix B	220×200
	Dimension of key matrix K	110×50
	Dimension of value matrix V	110×200
ASSIST Ments 2015	Dimension of embedded matrix A	100×50
	Dimension of embedded matrix B	200×100
	Dimension of key matrix K	100×50
	Dimension of value matrix V	100×100

Table 3. Comparison of AUC values of three KT models.

Model	Datasets	
	ASSIST Ments 2009	ASSIST Ments 2015
DKT	80.64%	69.21%
DKVMN	81.18%	72.68%
LPKT	82.35%	73.83%

its improved forgetting mechanism and calculation method of knowledge growth are more in line with human's learning and memory process, so it shows better knowledge tracking effect.

In order to further analyze the tracking effect of the three knowledge tracking models in the test set, we observe the change process of the test set results of the three knowledge tracking models with the increase of the training iterations of the models. As shown in **Figure 2**, at the beginning of training, the test results of the three knowledge tracking models are relatively close. When the training iterations are more than 20, the knowledge tracking effect of DKT method is close to the limit, while DKVMN model and LPKT model are still growing, which is due to the limitations of LSTM, because both LPKT model and DKVMN model use Mann, their curves are similar. When the training times are less than 20, their curves are very close. However, when the number of training iterations is more than 20, the increase of AUC value of LPKT model is greater than that of DKVMN model under the same training iteration number. This is because the parameter scale of LPKT model is slightly larger than that of DKVMN model, so when the training times are greater than 20, the advantages of LPKT model can be more obvious.

In addition, we also analyze the results of training set and verification set of DKVMN and LPKT in the training process. As shown in **Figure 3**, in the two data sets, the AUC value difference between DKVMN and LPKT is not very
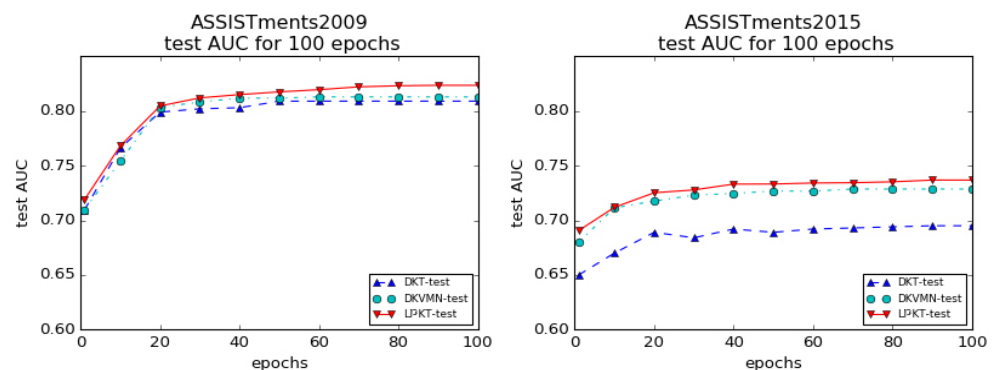

Figure 2. Changes of AUC values of three KT models with training iterations.
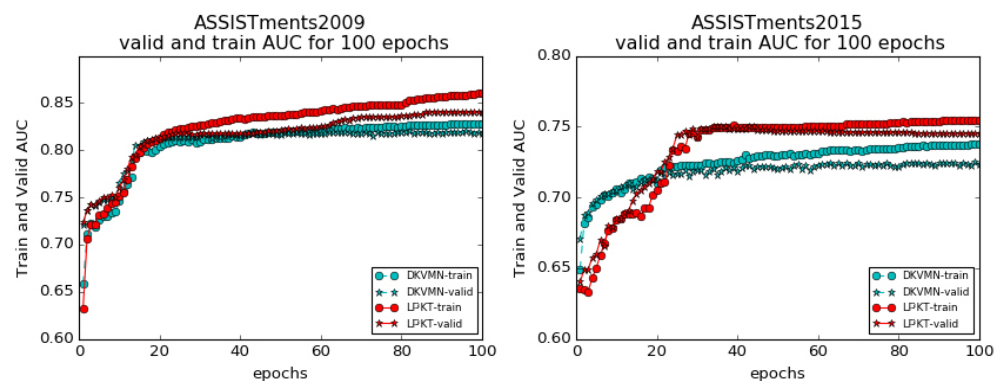

Figure 3. Changes of AUC value of two models on training set and verification set.

large. This means that although the LPKT method increases the parameter scale of MANN, it does not cause over fitting phenomenon to the model, that is, the increase of parameters is scientific and reasonable.

5. Conclusions

Based on the cognitive learning process of human, we proposed a knowledge tracking model based on learning process by improving the forgetting mechanism and knowledge growth mechanism of the existing knowledge tracking model, which makes the knowledge tracking more consistent with the human learning process and enhances the interpretability of the model.

In this paper, the LPKT model is compared with DKVMN model and DKT model on two datasets. The experimental results show that the AUC score of LPKT is significantly higher than that of DKVMN and DKT on the two datasets, and there is no over fitting phenomenon on the premise of increasing the parameter scale. This fully proves the effectiveness and superiority of our model. LPKT can be applied to a variety of online education platforms to help educators to achieve personalized guidance.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Kaplan, A.M. and Haenlein, M. (2016) Higher Education and the Digital Revolution: About MOOCs, SPOCs, Social Media, and the Cookie Monster. *Business Horizons*, **59**, 441-450. https://doi.org/10.1016/j.bushor.2016.03.008

[2] Nwana, H.S. (1990) Intelligent Tutoring Systems: An Overview. *Artificial Intelligence Review*, **4**, 251-277. https://doi.org/10.1007/BF00168958

[3] Baker, R.S.J., Corbett, A.T. and Aleven, V. (2008) More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. *International Conference on Intelligent Tutoring Systems*, Springer, Berlin, Heidelberg, 406-415. https://doi.org/10.1007/978-3-540-69132-7_44

[4] Corbett, A.T. and Anderson, J.R. (1994) Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Model User-Adap Inter*, **4**, 253-278. https://doi.org/10.1007/BF01099821

[5] Piech, C., Bassen, J., Huang, J., *et al.* (2015) Deep Knowledge Tracing. *Advances in neural Information Processing Systems*, 505-513.

[6] Millán, E., Loboda, T. and Pérez-De-La-Cruz, J.L. (2010) Bayesian Networks for Student Model Engineering. *Computers & Education*, **55**, 1663-1683. https://doi.org/10.1016/j.compedu.2010.07.010

[7] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[8] Wang, L., *et al.* (2017) Deep Knowledge Tracing on Programming Exercises. *Proceedings of the Fourth* (2017) *ACM Conference on Learning@ Scale*, 201-204. https://doi.org/10.1145/3051457.3053985

[9] Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D. and Lillicrap, T. (2016) Meta-Learning with Memory-Augmented Neural Networks. *Proceedings of the* 33*rd International Conference on Machine Learning, in PMLR* 48, 1842-1850.

[10] Sukhbaatar, S., Weston, J. and Fergus, R. (2015) End-to-End Memory Networks. *Advances in Neural Information Processing Systems*, 2440-2448.

[11] Kumar, A., Irsoy, O., Ondruska, P., *et al.* (2016) Ask Me Anything: Dynamic Memory Networks for Natural Language Processing. *International Conference on Machine Learning*, 1378-1387.

[12] Zhang, J., Shi, X., King, I. and Yeung. D.-Y. (2017) Dynamic Key-Value Memory Networks for Knowledge Tracing. *Proceedings of the* 26*th International Conference on World Wide Web* (WWW'17). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 765-774. https://doi.org/10.1145/3038912.3052580

[13] Brod, G., Werkle-Bergner, M. and Shing, Y.L. (2013) The Influence of Prior Knowledge on Memory: A Developmental Cognitive Neuroscience Perspective. *Front. Behav. Neurosci.*, **7**, 139. https://doi.org/10.3389/fnbeh.2013.00139

[14] Wandersee, J.H., Mintzes, J.J. and Novak, J.D. (1994) Research on Alternative Conceptions in Science. *Handbook of Research on Science Teaching and Learning*, 177-210.

[15] Graves, A., Wayne, G. and Danihelka, I. (2014) Neural Turing Machines. arXiv preprint arXiv:1410.5401.

[16] Ebbinghaus H. (2013) Memory: A Contribution to Experimental Psychology. *Annals of Neurosciences*, **20**, 155-156. https://doi.org/10.5214/ans.0972.7531.200408

[17] Atkinson, R.C. and Shiffrin, R.M. (1968) Human Memory: A Proposed System and Its Control Processes. *Psychology of Learning and Motivation*, **2**, 89-195. https://doi.org/10.1016/S0079-7421(08)60422-3

[18] Feng, M., Heffernan, N. and Koedinger, K. (2009) Addressing the Assessment Challenge with an Online System That Tutors as It Assesses. *User Model User-Adap Inter*, **19**, 243-266. https://doi.org/10.1007/s11257-009-9063-7

[19] Xiong, X., Zhao, S., Van Inwegen, E.G., *et al.* (2016) Going Deeper with Deep Knowledge Tracing. *International Educational Data Mining Society.*