



HAL
open science

Augmenting Convolution Neural Networks by Utilizing Attention Mechanism for Knowledge Tracing

Meng Zhang, Liang Chang, Tiejuan Liu, Chen Wei

► **To cite this version:**

Meng Zhang, Liang Chang, Tiejuan Liu, Chen Wei. Augmenting Convolution Neural Networks by Utilizing Attention Mechanism for Knowledge Tracing. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.80-86, 10.1007/978-3-031-03948-5_7. hal-04178712

HAL Id: hal-04178712

<https://inria.hal.science/hal-04178712v1>

Submitted on 8 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Augmenting Convolution Neural Networks By Utilizing Attention Mechanism for Knowledge Tracing

Meng Zhang¹, Liang Chang^{1,2}, Tiejuan Liu^{1,2}, Chen Wei¹

¹ Guilin University of Electronic Technology, Guilin 541004, China

² Guangxi Key Laboratory of Trusted Software, Guilin 541004, China
lty205@guet.edu.cn

Abstract. The devastating, ongoing Covid-19 epidemic has led to many students resorting to online education. In order to better guarantee the quality, online education faces severe challenges. There is an important part of online education referred to as Knowledge Tracing (KT). The objective of KT is to estimate students' learning performance using a series of questions. It has garnered widespread attention ever since it was proposed. Recently, an increasing number of research efforts have concentrated on deep learning(DL)-based KT attributing to the huge success over traditional Bayesian-based KT methods. Most existing DL-based KT methods utilize Recurrent Neural Network and its variants, i.e. Long Short-Term Memory(LSTM), Gated Recurrent Unit(GRU) etc. Recurrent neural networks are good at modeling local features, but underperforms at long sequence modeling, so the attention mechanism is introduced to make up for this shortcoming. In this paper, we introduce a DL-based KT model referred to as Convolutional Attention Knowledge Tracing (CAKT) utilizing attention mechanism to augment Convolutional Neural Network(CNN) in order to enhance the ability of modeling longer range dependencies.

Topics: Student Performance Prediction(SPP); Knowledge Tracing(KT); Convolutional Neural Network(CNN); Attention Mechanism(AM).

Contact author: Tiejuan Liu, E-mail:lty205@guet.edu.cn

Augmenting Convolution Neural Networks By Utilizing Attention Mechanism for Knowledge Tracing

Meng Zhang¹, Liang Chang^{1,2}, Tiejuan Liu^{1,2}, Chen Wei¹

¹ Guilin University of Electronic Technology, Guilin 541004, China

² Guangxi Key Laboratory of Trusted Software, Guilin 541004, China
lty205@guet.edu.cn

Abstract. The devastating, ongoing Covid-19 epidemic has led to many students resorting to online education. In order to better guarantee the quality, online education faces severe challenges. There is an important part of online education referred to as Knowledge Tracing (KT). The objective of KT is to estimate students' learning performance using a series of questions. It has garnered widespread attention ever since it was proposed. Recently, an increasing number of research efforts have concentrated on deep learning(DL)-based KT attributing to the huge success over traditional Bayesian-based KT methods. Most existing DL-based KT methods utilize Recurrent Neural Network and its variants, i.e. Long Short-Term Memory(LSTM), Gated Recurrent Unit(GRU) etc. Recurrent neural networks are good at modeling local features, but underperforms at long sequence modeling, so the attention mechanism is introduced to make up for this shortcoming. In this paper, we introduce a DL-based KT model referred to as Convolutional Attention Knowledge Tracing (CAKT) utilizing attention mechanism to augment Convolutional Neural Network(CNN) in order to enhance the ability of modeling longer range dependencies.

Keywords: Student Performance Prediction(SPP); Knowledge Tracing(KT); Convolutional Neural Network(CNN); Attention Mechanism(AM).

1 Introduction

With the continuous development of society, education issues are becoming more and more important, and the issue of teaching students according to their aptitude is particularly important. How to provide different students with training programs that suit their abilities, so that different students can better develop their own potential, is an important issue facing current education. And with the spread of the new coronavirus, more and more students cannot go back to school to study, and there is an urgent need for an online learning system to plan feasible learning plans for students. Therefore, with the continuous development of deep learning, one of the successful tasks of various neural networks is knowledge tracking (KT) [1]. The goal of knowledge tracking is to analyze and evaluate students' knowledge status in the

process of students' learning, so as to predict students' performance in the next time series and give appropriate suggestions[2].

Therefore, we can know that knowledge tracking plays an important role in the online education system. According to recent researches, many scholars have proposed different methods to solve this problem, including Bayesian Knowledge Tracing (BKT)[1], Deep Knowledge Tracing (DKT) [3], Dynamic Key-Value Memory Networks (DKVMN)[4] , Convolutional Knowledge Tracing (CKT) [5], these excellent methods use different models to track the knowledge level of students, by modeling students' practice sequences and other information, although each model has obtained relatively good results , but different models have some potential problems that are not well resolved.

Models utilized in most of these methods are either RNNs or CNNs. CNNs behave well in modeling local features. However, CNNs are inadequate to perform good at long term sequence modeling. Therefore, it is a natural idea to combine a mechanism which models long term sequences well. One of the mechanisms that is fit for this requirement is attention mechanism[6]. In this case, we attempt to combine CNN with attention mechanism. To the best of our knowledge, this is the first paper utilizing attention mechanism with CNN in knowledge tracing area. The model is referred to as Convolutional Attention Knowledge Tracing (CAKT), which handles both short and long student learning sequences better.

2 RELATED WORKS.

2.1 Deep Learning-based Knowledge Tracing

The introduction of deep learning into the field of knowledge tracking is an important turning point. The DKT model [1] proposed in 2015 has received a lot of attention from researchers once it was proposed, and it has improved the accuracy of model prediction to a new level. It analyzes the accuracy of the model through the area under the curve (AUC). After a lot of experiments, it has been proved that its accuracy is much better than the traditional knowledge tracking model.

2.2 Convolutional Neural Network

CNNs have provided a successful example employed in many areas such as computer vision, natural language processing, etc.[8,9]. In 2020, Convolutional Knowledge Tracing(CKT) models individualization of students in a more precise way and shows notable results demonstrating the effectiveness of CKT.

2.3 Attention Mechanism

RNN has the problem of long-term dependence, which makes it unable to better perceive the connection between long-sequence information, and the long-term dependence makes DKT unable to utilize the long-sequence input. The attention

mechanism allows to model the dependencies of input and output sequences regardless of their length in the sequence, and is more suitable for modeling long-distance dependencies. Attention mechanism has proved to be effective in various fields because of its strong modeling ability[10,11]. Recently, a self-attentive model[12] called SAKT(Self-Attentive Knowledge Tracing) has been introduced in KT. SAKT simply utilizes transformer architecture to solve long term modeling problems.

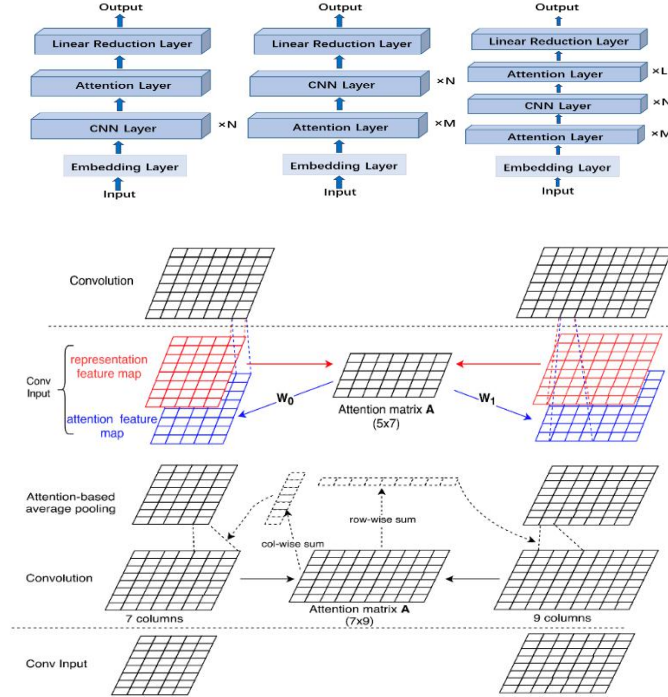


Fig. 1. The architectures of CAKT model.

3 MODEL ARCHITECTURE

3.1 Problem Definition

The knowledge tracking task can be described as a supervised learning task, through the interaction sequence in the process of students doing questions, this sequence contains the questions and other information of the students. where the parameter n is the total number of student interactions in this training, $X = (x_0, x_1, \dots, x_{n-1}, x_n)$, where n is the total number of the student's interactions' times, the task of knowledge tracing is to predict the performance of the student's next interaction x_{n+1} . When interactions are considered as the form of question-answer[12], $X_t = \{q_t, a_t\}$, where q_t is the

question that the student attempts at time step t and a_t indicates the correctness of the student's answer, $a_t \in \{0,1\}$ where 1 indicates a correct answer whilst 0 stands for a wrong one.

3.2 CAKT

The architecture of CAKT model is shown in Fig 1, containing an Embedding Layer, several Convolutional Layers, an Attention Layer, and a Linear Dimension Reduction Layer. The design of the model utilized for both Embedding Layer and Convolutional Layer is similar to that proposed by CKT.

We follow the architecture of attention mechanism in[13], but we slightly modify the architecture in order to fit our task.

$$score_i = h_{attn}^T \sigma(W_{attn} V_i + b_{attn}) \quad (1)$$

$$\alpha_i = \frac{\exp(score_i)}{\sum_{i=1}^n \exp(score_i)} \quad (2)$$

$$Output_{sum} = \sum_{i=1}^n \alpha_i V_i \quad (3)$$

where W_{attn} is the attention weight parameter matrix, V_i is the output of the convolutional layer, h_{attn}^T is the parameter. Attention bias parameter is b_{attn} , and is element-wise multiplication, σ stands for tanh function. α_i is the attention score of V_i . We get the attention output $Output_{sum}$ in formula (3).

In order to reduce the dimension, we add a linear layer at the end of the attention layer. The computation of linear layer is written below:

$$y_t = W_{oy} Output_{sum} + b_y \quad (4)$$

In order to reduce the dimension, we add a linear layer at the end of the attention layer. The computation of linear layer is written below:

$$L = \sum_{t=1}^N (a_t \log y_t + (1 - a_t) \log(1 - y_t)) \quad (5)$$

The goal of model training is to minimize the similarity of students learning sequence predictions. The parameters are learned by minimizing the cross entropy loss between p_t and r_t .

$$\gamma = - \sum_t (r_t \log(p_t) + (1 - r_t) \log(1 - p_t)) \quad (6)$$

4 EXPERIMENTS

4.1 Datasets Description

We use two open-source public datasets and conduct extensive experiments to verify the excellent effect of the CAKT model. The information contained in the two datasets is shown in Table 1 with the information concerning the datasets shown below:

ASSISTments2009: This dataset was gathered by the ASSISTments online tutoring platform during the school year from 2009 to 2010.

ASSISTments2015: The ASSISTments online tutoring platform collected this dataset in 2015, covering response records of students.

Table 1. Statistics of all datasets

Datasets	Statistics		
	<i>Students</i>	<i>Concepts</i>	<i>Records</i>
ASSIST2009	4417	124	328K
ASSIST2015	19,840	100	683K
STATICS	333	1223	189K
Sythetic	4000	50	200K
ASSISTChall	1709	102	943K

4.2 Comparison Methods

We compared CAKT with several classic methods including DKT, DKVMN and CKT. The details of the comparison methods include the following:

DKT utilizes RNN for modeling students' learning performance, we chose LSTM for our implementation because of its stronger modeling capability.

DKVMN Through the dynamic and static network, the knowledge level of students is continuously updated in the process of student training.

CKT models individualization in the student learning process by applying CNNs.

Experimental code We obtained the source code of the DKVMN model and CKT model released by the original author from the open source website, and combined our ideas to better carry out experiments. The original implementation of DKT was programmed by Lua, instead, we used Python to re-implement the DKT model.

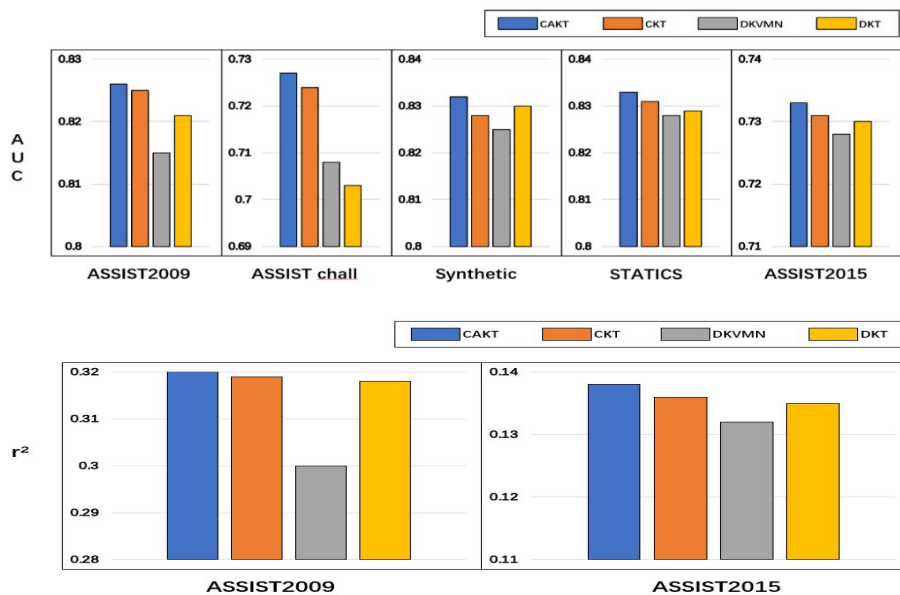


Fig. 2. Results of our model and comparison methods experimented in datasets.

4.3 Experimental Setup

For our experiment, all of the proposed models including data preprocessing, model training and evaluating are implemented by Python, using the deep learning framework Tensorflow on a Linux server with a six-core 2.5GHz Intel Xeon E5-2678 CPU and an NVIDIA GeForce RTX 2080Ti GPU. At each stage of the model, we tune the parameters by randomly initializing the weights and biases of each layer from a Gaussian distribution with zero mean and standard deviation. The size of the amount of data put into the model for each batch is 50. The epoch number is set to 100. To avoid phenomena such as overfitting, we set the dropout parameter of the model to 0.2, the learning rate to 0.03, and the decay every 10 steps to 0.2.

4.4 Conclusions

In order to better judge the performance of CAKT, we adopt the mainstream judgment criteria to analyze the performance of the model. we chose two classic indicators including Area Under Curve (AUC) and coefficient of determination(r^2).

The judgment standard of the experiment is that the higher the index parameters of the model, the higher the prediction accuracy of the model and the better performance. The experimental results of the model can be seen intuitively in the bar graph in Figure 2. From the comparison of the bar graphs, we can intuitively feel that all the indicators of the CAKT model have improved the performance of the previous models. CAKT obtains a test AUC of 82.6% whilst the highest comparison method,

namely, CKT gets 82.4%. The prediction results effectively demonstrate the capabilities of our proposed model.

References

1. A. T. Corbett and J. R. Anderson, "Knowledge Tracing: Modeling the acquisition of procedural knowledge," *User Modeling and User-Adapted Interaction*, vol. 4, pp. 253–278, December 1994.
2. K. Nagatani, Q. Zhang, M. Sato, Y. Y. Chen, F. Chen, and T. Ohkuma, "Augmenting Knowledge Tracing by Considering Forgetting Behavior," *Proc. of the 28th International Conference on World Wide Web*, Apr. 2019, pp. 3101-3107.
3. C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein, "Deep Knowledge Tracing," *Proc. of the 28th International Conference on Neural Information Processing Systems*, Dec. 2015, pp. 505-513.
4. J. Zhang, X. Shi, I. King, and D. Y. Yeung, "Dynamic Key-Value Memory Networks for Knowledge Tracing," *Proc. of the 26th International Conference on World Wide Web*, Apr. 2017, pp.765-774.
5. S. Shen, Q. Liu, E. Chen, H. Wu, Z. Huang, W. Zhao, Y. Su, H. Ma, and S. Wang "Convolutional Knowledge Tracing: Modeling Individualization in Student Learning Process," *Proc. of the ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2020, pp. 1857-1860, doi:10.1145/3397271.3401288
6. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is all you need," *Proc. of the 31st International Conference on Neural Information Processing Systems*, Dec 2017, pp. 6000-6010.
7. I. Bello, B. Zoph, Q. Le, A. Vaswani and J. Shlens, "Attention Augmented Convolutional Networks," 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2019, doi:10.1109/ICCV.2009.00338.
8. A. Krizhevshy, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, Vol. 60, No. 6, May 2017, doi:10.1145/3065386.
9. Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, Oct. 2014, pp. 1746-1751, doi:10.3115/v1/D14-1181.
10. J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho and Y. Bengio, "Attention-based models for speech recognition," *Proc. of the 28th International Conference on Neural Information Processing Systems*, Dec 2015, pp. 577-585.
11. T. Luong, H. Pham, and C. D. Manning, "Effective Approaches to Attention-based Neural Machine Translation," *Proc. of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, Sep. 2015, pp. 1412-1421. doi:10.18653/v1/D15-1166.
12. S. Pandey, and G. Karypis, "A Self-Attentive Model for Knowledge Tracing," *arXiv preprint arXiv: 1907.06837*, 2019.
13. W. Feng, J. Tang, T. X. Liu, S. Zhang and J. Guan, "Understanding Dropouts in MOOCs," *Proc. of the 33rd AAAI Conference on Artificial Intelligence*, 33(01), pp. 517-524, doi:10.1609/aaai.v33i01.3301517.