Interpretable Personalized Knowledge Tracing and Next Learning Activity Recommendation

Jinjin Zhao
Amazon.com
Seattle, USA
jinjzhao@amazon.com

Shreyansh Bhatt
Amazon.com
Seattle, USA
bhattshr@amazon.com

Candace Thille
Amazon.com
Seattle, USA
cthille@amazon.com

Dawn Zimmaro
Amazon.com
Austin, USA
dzimmaro@amazon.com

Neelesh Gattani
Amazon.com
Seattle, USA
neeleshg@amazon.com

ABSTRACT

Online learning systems that provide actionable and personalized guidance can help learners make better decisions during learning. Bayesian Knowledge Tracing (BKT) extensions [2] and deep learning based approaches have demonstrated improved mastery prediction accuracy compared to the basic BKT model; however, neither set of models provides actionable guidance on learning activities beyond mastery prediction. We propose a novel framework for personalized knowledge tracing with attention mechanism. Our proposed framework incorporates auxiliary learner attributes into knowledge tracing and interprets mastery prediction with the learning attributes. The proposed approach can also provide personalized next best learning activity recommendations. We demonstrate that the accuracy of the proposed approach in mastery prediction is slightly higher compared to deep learning based approaches and that the proposed approach can provide personalized next best learning activity recommendation.

Author Keywords

Knowledge tracing; Personalized knowledge tracing; Recommendation; Attention mechanism

CCS Concepts

•**Applied computing** → **Learning management systems; Interactive learning environments;** •**Human-centered computing** → *User models; User studies;* •**Computer systems organization** → **Neural networks;**

INTRODUCTION AND RELATED WORK

Some online learning systems are moving away from traditional completion and performance metrics to collect and report data in real-time to learners about their predicted knowledge state. As a learner responds to assessment questions, these online learning systems use these question attempts to model how well the learner is mastering the underlying skills associated with a given question. Additionally, these systems can be used to provide an interpretation of a learner's current knowledge state with his/her corresponding attempts and prior knowledge state represented through learner attributes. Further, recommendations on next learning activity can be provided to better support learner's decision making.

Corbett et al. [2] proposed BKT which uses four parameters (p_{slip}, p_{guess}, p_{trans}, p_{init}) to model learner's knowledge state and infer knowledge mastery. Individualized BKT extensions have been proposed to provide personalized knowledge mastery prediction and learning experiences by modeling student-specific features. Corbett et al. [2] also proposed to model student-wise as well as skill-wise four parameters to improve prediction accuracy. Pardos et al. [5] proposed to model 'prior per-student' knowledge by estimating the initial probability of mastery p_{init}. Lee et al. [4] proposed to model per-student BKT parameters, instead of per-student and per-skill BKT parameters, to achieve better accuracy in mastery prediction. Yudelson et al. [12] conducted experiments to analyze the different individualization approaches for different use cases. Khajah et al. [3] proposed to combine latent factor modeling with BKT to predict knowledge state at an individual level with problem difficulty factor and student skill ability factor. Deep learning based knowledge tracing has shown its ability to capture the complex sequential patterns to achieve state-of-the-art accuracy in mastery prediction. Piech et al. [7] proposed Deep knowledge tracing (DKT) to model the knowledge state using recurrent neural network and achieved better prediction accuracy compared to BKT based approaches. DKT with side information proposed by Wang et al. [10] and DKT with question text embedding proposed by Su et al. [8] achieved better prediction accuracy. By adding reconstruction of the interaction, Yeung [11] claimed that prediction accuracy can be further improved. Another set of approaches coupling conventional neural network to external memory bank proposed

by Zhang et al. [13] have also achieved prediction accuracy improvement compared to DKT approaches.

However, the explanation about the mastery prediction provided in the form of four parameters or predefined factors from BKT and its extensions is limited. Deep learning based approached achieved the state-of-the-art accuracy in knowledge state prediction, but it models the relation between the sequential learning activities and the knowledge state implicitly. The relations among huge amount of variables can neither provide explicit explanations of the current knowledge state estimation nor provide actionable suggestions on the next best learning activity to achieve the desired mastery.

We want to model learning attributes in a flexible way in the context of knowledge tracing and also provide reasoning on how learning attributes are impacting the knowledge acquisition. In this work, we propose a personalized knowledge tracing framework, which can incorporate auxiliary learner attributes flexibly and provide the importance of the auxiliary attributes automatically. The framework also provides reasoning on how different attributes are impacting learners' knowledge acquisition. As part of the results, we demonstrate using the personalized knowledge state representation, a mixed representation learnt from interactions and auxiliary attributes, to provide explanations of the current knowledge acquisition and future learning activity recommendations. The knowledge state representation is proposed in our preliminary research result [1]. In the referred paper, the proposed knowledge state representation is the probability of mastery, which is estimated as the ratio of the number of successful future attempts to the total number of future attempts. We also argued in the referred paper that the proposed knowledge state representation is a better approximation of the probability of knowledge mastery.

APPROACH

Framework overview
The overall framework, as shown in Fig 1, consists of learner attributes embedding and personalized knowledge tracing components. The learner attributes embedding module is designed with a simple autoencoder, which can provide the flexibility to incorporate auxiliary learner attributes. Personalized knowledge tracing uses a basic attention module to factor the most important features from learner attributes and interactions to perform mastery prediction.

Learner Attributes Embedding
In Fig1, learner attributes embedding is a process to compute a dense representation of the auxiliary learner attributes. Learner attributes representation e is derived through a vanilla autoencoder, where there are only two fully connected layers as shown in Eq 1.

$$\mathbf{e} = \sigma(\mathbf{Wx} + \mathbf{b}); \mathbf{x}' = \sigma'(\mathbf{W}'\mathbf{e} + \mathbf{b}') \qquad (1)$$

where x is the input attributes, x' is the reconstructed attributes. W/W' and b/b' are the weight and bias matrics in the fully connected layer. σ and σ' is the sigmoid function. With a simple fully connected layer, the importance of raw learning attributes can be traced back with the weight W matrix from the embedded representation e.
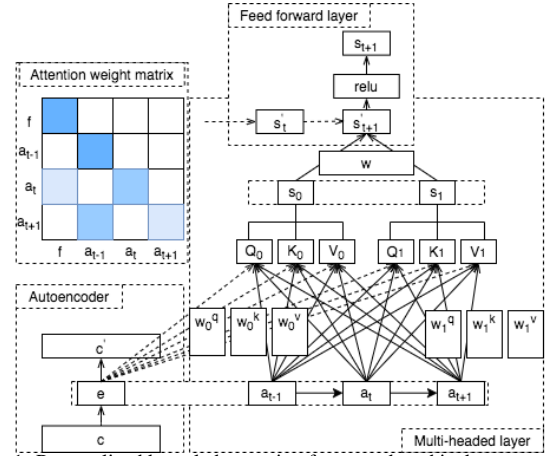

Figure 1: Personalized knowledge tracing framework and its key components

Personalized knowledge tracing
We use attention mechanism [9] as described in equation 2

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{n}}\right) \qquad (2)$$

to model the relation between learning attributes, from both learner background and learner interactions, and the knowledge state. The attention weight matrix shown in Fig1 is the associated attributes to knowledge state weight matrix, which indicates the relative importance of all the attributes to predicting the current state. In the proposed approach, attributes are composed of two parts, auxiliary learner attributes and historical attempts in the current learning experience. The sequential attempt vector is concatenated with the learner representation from auxiliary attributes. Note that the learner specific attribute part of the attribute vector is the same for each timestamp i_t. Intuitively, the knowledge state prediction at t for a learner consists of matching the current learner's activities $i_1, \ldots, i_t$ and attributes with the most similar activities and attributes available from other learners and then providing the prediction based on the matching. The relative importance of the attributes in the context of predicting mastery is derived automatically through model training and can be accessed through the attention matrix. With the attention matrix, we can interpret the knowledge state prediction with its most important learning attributes.

EXPERIMENTS, RESULTS AND APPLICATIONS
In this section, we describe datasets, experiment setup, and results in predicting knowledge state, interpreting knowledge state prediction, and providing a next best learning activity recommendation.

Datasets
We used one open source dataset and two workplace online course datasets to evaluate the proposed approach.

ASSISTment Challenge: We used Assistments data mining competition dataset [6]. It has 102 skills and 942,816 records for 1709 students [6]. Learner attributes used in the experiment includes 'school', 'gender', 'classroom', 'teacher'. In this experiment, we directly concatenate learner attributes to the

Dataset	LSTM		Att		P_Att	
	AUC	ACC	AUC	ACC	AUC	ACC
A_challenge	0.76	0.76	0.78	0.76	0.78	0.77
Course1	0.74	0.79	0.76	0.8	0.77	0.86
Course2	0.8	0.81	0.81	0.81	0.81	0.81

Table 1: Model result comparison

historical attempts rather than using the autoencoder to get the learning attribute embedding.

course 1: We used this dataset from an online learning system, which has 509 students and 207,529 interactions for 28 skills.

course 2: This dataset is also extracted from the same online learning system, which has 315 students and 171496 interactions for 21 skills.

For both course 1 and 2, learner attributes used in the experiment include business organization, office location, and other job related information. An autoencoder process is applied to generate a dense representation of the attributes for both courses.

Evaluation Measures and Experiment Set up
First, we evaluated the proposed knowledge tracing framework based on its ability to predict the knowledge state of a learner. We used accuracy (ACC) and Area Under Curve (AUC) measures for the evaluation. Knowledge state is estimated as the averaged attempt correctness from similar learners who have achieved mastery. We conducted experiments comparing the attention based approach with personalized approach. We also conducted experiments to compare with LSTM based approach.

Experimental Results
Model Comparison
Table 1 shows the comparison results for the LSTM based model, the attention based model without learner attributes (Att), and the personalized attention based model (P_Att) for the knowledge state estimation. As shown in Table 1, P_Att, from both ACC and AUC measurement, the knowledge state estimation accuracy achieves slight improvement or remains the same as LSTM approach. Att and P_Att achieve similar accuracy for all the mentioned course datasets.

Interpretable knowledge state with learning attributes
Learners with different backgrounds learn differently at different learning stages. We want to understand how different learning attributes are associated with knowledge acquisition. With the proposed attention based approach, we can derive the top learning attributes in predicting the learner's current knowledge state. Here are the findings about the impact of modeling auxiliary learner attributes at different learning stages.

- At the early learning stage, when there is not enough information from the learning activities in predicting mastery, learner attributes are the main predictors.

- At later learning stages, learning activities becomes more powerful predictors in predicting knowledge states.

We use results from course 2, an attribute-attributes matrix visualization to demonstrate the interpretation of knowledge
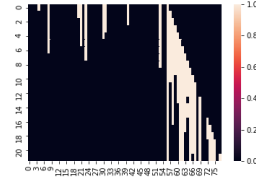

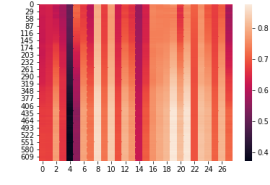Figure 2: Knowledge state interpretation with learning attributes

Figure 3: knowledge state vectorization

state prediction for one learner with their job related attributes and sequential learning activities. In this application, each learner is tagged with 56 profile attributes, which are extracted from a data source outside of the current course. As shown in Figure 2, as the learner starts to interact with the course by answering assessment questions (within the first 8 attempts), the learner attributes have a higher weight than the limited attempts. Within learner attributes, 24th, 8th and 22nd learner attributes have a higher weight than other attributes. As the learner continues interacting and making attempts at assessment questions, the latent sequential learning attributes start to have a higher weight than the learner attributes. Within sequential learning activities, the first attempt (56th in x-axis) shows a high weight in predicting knowledge state along all the attempts, which indicates a high importance of this question in achieving mastery. Interpreting the current knowledge state prediction with available learning attributes can be conducted at any given time. Deriving the key learning attributes for certain learning experiences can help the learning science community better understand the different learning patterns and improve learning experience design.

Effort estimation toward mastery and next best learning activity recommendation
The knowledge gap can be derived from the target knowledge state (the goal) and the current knowledge state. The effort needed towards mastery is defined in terms of unique question and average attempt number. The effort needed to bridge the knowledge gap can be derived based on similar learners' mastery trajectories. Similar learners are defined as the learners who have achieved mastery and have been in the similar knowledge state. The similarity is defined as the cosine similarity on knowledge state representation. The next best learning activity is derived by observing the similar learners' trajectories. The next learning activity the similar learners took when they were in the similar state is considered as one candidate activity. We rank the candidate activities across the Top N similar learners to get the activities to recommend.

Knowledge state is represented as a vector with mastery probabilities for all the knowledge components at any given time. Knowledge state vector is updated after every attempt. Figure 3 illustrates the trajectory of a knowledge state vector, consisting of the mastery probabilities for all the knowledge components, being updated along with learning attempts. The x-axis is the knowledge state vector with 28 knowledge states. The y-axis is the sequential attempts with 611 attempts in total. The value in the heat map is the probability of a learner's knowledge mastery. As shown in Fig 3, for most of the knowledge

components, the probability of knowledge mastery gradually increases along with further attempts.

Result: efforts required towards mastery, next best
learning activity recommendation

while *attempt1 in attempts1* **do**

    1. Calculate Top N similar learners who have achieved mastery:

    if *attempt2 in attempts2* **then**

        similarity score = sim(attempt1, attempt2);
        keep the highest score and the learner;

    end

    2. Calculate the unique question/attempt number required to achieve desired mastery for each learner in Step 1.;

    3. Extract the mean and variance of the unique question/attempt number across learners;

    4. Extract the next learning activity distribution across Top N similar learners;

    5. Rank and get the most popular activity from Step 4. as the recommendation for the next best learning activity;

end

Algorithm 1: algorithm for effort estimation and recommendation

Algorithm 1 demonstrates the pseudocode to derive the efforts estimation towards mastery and the next best learning activity. Attempts1 and attempts2 are the attempts from targeted learner and the rest of the learners, respectively. First, we calculate the Top N similar learners who have achieved mastery and have been in the similar state. Then we derive the remaining efforts towards mastery in terms of unique question and average attempt number from the Top N similar learners. After extracting and sorting the next learning activities, the next best learning activity can be communicated to learners for better decision making.

CONCLUSIONS

In this research we proposed an interpretable personalized knowledge tracing framework. The framework can incorporate auxiliary learner attributes in a flexible way and generate the importance of learning attributes automatically. The interpretation of the current knowledge state can be derived with the relative importance of the learning attributes. The interpretation can help learners and learning designers understand the key factors from learning attributes that are contributing to the current knowledge state and further improve learning experience and environment. Furthermore, these approaches can be used to give learners recommendations about next learning activity based on similarity to other learners who achieved mastery. Additionally, learning designers can be provided

information about which design approaches or question sequences support learners with different attributes.

REFERENCES

[1] Shreyansh Bhatt, Jinjin Zhao, Candace Thille, Dawn Zimmaro, and Neelesh Gattani. 2020. A novel approach for knowledge state representation and prediction. In *Proceedings of the Seventh Annual ACM Conference on Learning at Scale*. ACM.

[2] Albert T Corbett and John R Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4, 4 (1994), 253–278.

[3] Mohammad Khajah, Rowan Wing, Robert Lindsey, and Michael Mozer. 2014. Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In *Educational Data Mining 2014*. Citeseer.

[4] Jung In Lee and Emma Brunskill. 2012. The Impact on Individualizing Student Models on Necessary Practice Opportunities. *International Educational Data Mining Society* (2012).

[5] Zachary A Pardos and Neil T Heffernan. 2010. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 255–266.

[6] Thanaporn Patikorn, Neil T Heffernan, and Ryan S Baker. 2018. ASSISTments Longitudinal Data Mining Competition 2017: A Preface. In *Proceedings of the Workshop on Scientific Findings from the ASSISTments Longitudinal Data Competition, International Conference on Educational Data Mining*.

[7] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. 2015. Deep knowledge tracing. In *Advances in neural information processing systems*. 505–513.

[8] Yu Su, Qingwen Liu, Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Chris Ding, Si Wei, and Guoping Hu. 2018. Exercise-enhanced sequential modeling for student performance prediction. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[10] Zhiwei Wang, Xiaoqin Feng, Jiliang Tang, Gale Yan Huang, and Zitao Liu. 2019. Deep Knowledge Tracing with Side Information. In *International Conference on Artificial Intelligence in Education*. Springer, 303–308.

[11] Chun-Kit Yeung and Dit-Yan Yeung. 2018. Addressing two problems in deep knowledge tracing via prediction-consistent regularization. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 5.

[12] Michael V Yudelson, Kenneth R Koedinger, and Geoffrey J Gordon. 2013. Individualized bayesian knowledge tracing models. In *International conference on artificial intelligence in education*. Springer, 171–180.

[13] Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung. 2017. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 765–774.