The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)

Student Knowledge Prediction for Teacher-Student Interaction

Seonghun Kim, Woojin Kim, Yeonju Jang, Seongyune Choi, Heeseok Jung, Hyeoncheol Kim
Department of Computer Science & Engineering, Korea University, Seoul, Korea
{ryankim0409,woojinkim1021,spring0425,csyun213,poco2889,harrykim}@korea.ac.kr

Abstract

The constraint in sharing the same physical learning environment with students in distance learning poses difficulties to teachers. A significant teacher-student interaction without observing students' academic status is undesirable in the constructivist view on education. To remedy teachers' hardships in estimating students' knowledge state, we propose a Student Knowledge Prediction Framework that models and explains student's knowledge state for teachers. The knowledge state of a student is modeled to predict the future mastery level on a knowledge concept. The proposed framework is integrated into an e-learning application as a measure of automated feedback. We verified the applicability of the assessment framework through an expert survey. We anticipate that the proposed framework will achieve active teacher-student interaction by informing student knowledge state to teachers in distance learning.

Introduction

The outbreak of the COVID-19 pandemic led to an unprecedented shift of educational paradigm due to the school closures, affecting all stages of public education (Estellés and Fischman 2020). The primary education environment migrated to distance learning, where teachers and students do not share the same physical space (Kaplan and Haenlein 2016). Features of distance learning are asynchronous teacher-student interaction, technology for active communication, and online learning resources to facilitate the learning process.

Asynchronous teacher-student interaction of distance learning disadvantages students since teachers cannot closely monitor students' current knowledge states relative to face-to-face learning. The addressed limitations of distance learning become a barrier to student learning from the perspective of constructivism. Constructivism assumes the learner's construction of knowledge based on his or her past knowledge and experience (Liu and Chen 2010; Hoover 1996). Constructivism regards teacher-student interactions (e.g., teacher's feedback) and the assignment of meaningful and appropriate learning tasks as essential components of learning (Greeno et al. 1996).

Figure 1: A screenshot of a landing page of the Toc-Toc Math Expedition e-learning platform (left). A screenshot of the student analysis page (right). The predicted student test scores and suggestions for learning are generated using our framework.

While the ideal learning of constructivism advocates for active teacher-student interaction, distance learning contradicts constructivism with asynchronous teacher-student interaction. The constraint impedes teachers from thoroughly understanding the knowledge state of students. It may be demanding for teachers to provide suitable assignments depending on the students' knowledge state, which creates difficulty in supporting students with meaningful feedback on their academic performance.

There have been attempts to overcome the downside of distance learning. However, these approaches are limited to stimulating peer-to-peer interaction through online discussions (Bates 2008) or supporting knowledge construction of students through quality online instructions (Ally 2004). These approaches provide measures to remedy distance learning's side effects but do not present student knowledge to teachers. Since an immediate understanding of the student knowledge state is restrained in distance learning due to the limited teacher-student interaction time, it is necessary to provide students' current knowledge state for teachers to apply in their teaching and learning methods.

We propose a Student Knowledge Prediction Framework using Machine Learning models, and an eXplainable Artificial Intelligence (XAI) (Gunning 2017) method that models and explains a student's knowledge state. We used Machine Learning models, the XGBoost (Chen and Guestrin 2016), and the Deep Knowledge Tracing (DKT) (Piech et al. 2015) to model the student's current knowledge state and present the predicted performance of the student to the teacher. We applied the Shapley Value (Lipovetsky and Conklin 2001)

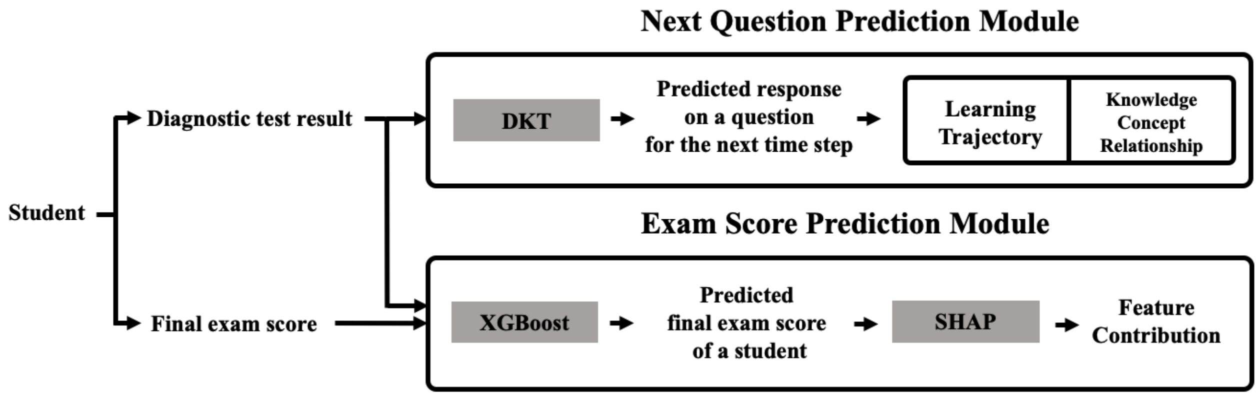

Figure 2: The overall architecture and flow of the Student Knowledge Prediction Framework.

method to explain the prediction in an interpretable form.

The Student Knowledge Prediction Framework has two educational contributions. First, it provides the current knowledge state of a student to the teacher for students' individualized learning. Second, the framework expedites the learning of a student by facilitating teacher-student interaction in distance learning.

The paper's Section 2 describes our proposed framework in detail. Section 3 reports on the training details and performance of the framework. Section 4 discusses the evaluation of the proposed framework. Section 5 presents related works. Finally, we summarize and suggest the educational implications of the framework in section 6.

Student Knowledge Prediction Framework

The Student Knowledge Prediction Framework consists of two modules, the exam score prediction module and the next question prediction module. The input to the framework is the student response on a diagnostic test on a subject's essential skill and the final exam score on the corresponding subject.

The next question prediction module predicts the student's response to a question for the next time step. A student's learning trajectory and a knowledge concept relationship are extracted using the predicted response. The exam score prediction module predicts the final exam score and explains a student's predicted exam score. The overview of our proposed framework is illustrated in Figure 2.

Our framework is integrated into a broader e-learning application, the Toc-Toc Math Expedition[1] by the Korea Foundation for the Advancement of Science and Creativity (KOFAC), which aims for skill mastery of elementary school students in basic mathematics. Screenshots of the e-learning application and student performance analysis generated from our framework are shown in Figure 1. We note in advance that the next question prediction module is not yet integrated into the e-learning application.

[1]www.toctocmath.kr

Data	Grade 1	Grade 2
number of students	259	277
number of exercises	265	257
number of knowledge concepts	7	9
number of interactions	44,310	44,523

Table 1: Overview of the KOFAC data used for training of the framework.

Data

We used the proprietary data provided from KOFAC to train both of our modules. The KOFAC data consists of 259 first grade students and 227 second grade students from 5 different elementary schools in South Korea, gathered from March 2018 to February 2019. The students completed diagnostic tests at the beginning of the academic year[2] on primary mathematics. At the end of each semester, the final exam scores of the students were also collected.

The first grade's basic mathematics knowledge concepts are comparing numbers, ordering numbers, counting numbers, composing numbers, decomposing numbers, addition, and subtraction. For the second grade, the knowledge concepts include comparing numbers, ordering numbers, counting numbers, advanced addition, and subtraction. The advanced addition and subtraction each have three sections, consisting of a total of nine concepts. The summary of the data is provided in Table 1. We also note that we do not own permission to release the data publicly.

Next Question Prediction Module

To serve the application's need for acquiring skill mastery in basic mathematics, the next question prediction module aims to model students' current knowledge state. The objective of skill mastery is to achieve a certain academic level. To identify whether a student achieved a skill, estimating mastery level is necessary than predicting future academic performance. The next question prediction module estimates

[2]Academic year of South Korea begins in March and ends in February.

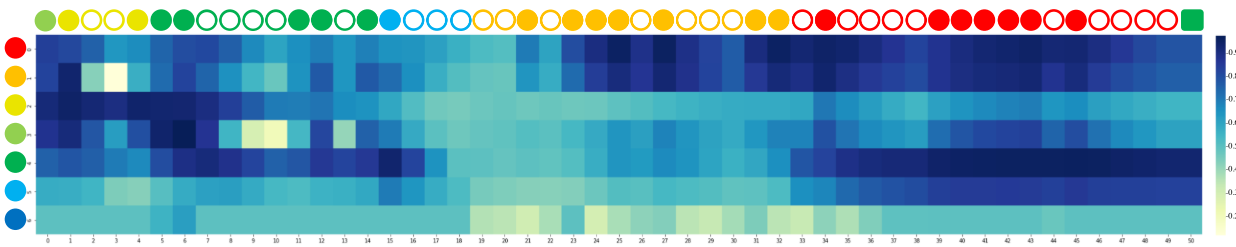

Figure 3: An example of a student's learning trajectory on a sequence of 50 questions, predicted by the DKT for each time step on the KOFAC data. Each row indicates a question, and each cell's color indicates a student's estimated knowledge level of the question. Darker the color, the higher the estimated knowledge level of the student. Best viewed in color.

students' current knowledge state to teachers with a learning trajectory and a knowledge concept relationship.

To model the knowledge state, we mainly adopted the Deep Knowledge Tracing (DKT) network. DKT is a deep learning approach to model the latent knowledge state with the temporal information (Piech et al. 2015). As the student progresses through online coursework, the student produces a series of interactions within the system. Given the student's response to particular coursework, the task of the DKT is to predict the student's response to an exercise question of the next time step. DKT uses Long Short-Term Memory (LSTM) layers (Hochreiter and Schmidhuber 1997) to model the sequence of students' progress and predict the student response of the future step. The DKT implemented in the framework shares the same objective.

Problem Formulation We set the next question prediction module as a supervised sequence prediction task. Given the series of student responses of length t, $\mathbf{x} = \{x_1, x_2, \ldots, x_t\}$, A student response consists of a question and answer at time t, $x_t = (q_t, r_t)$, and the response is binary variable $0, 1 \in \mathbf{r}$. The model outputs a probability $p(r_{t+1} = 1 | q_{t+1}, x_t)$ which is a prediction of the student giving correct answer to the question q_{t+1} in the $t + 1$ time step given the student's past responses.

Learning Trajectory The DKT presents the current knowledge state in the form of a learning trajectory, as shown in Figure 3. The learning trajectory visualizes the probability of giving the correct answer to each of the questions in the next step, given the history of student responses. From the learning trajectory, a student's changing knowledge state is identified and presented to the teacher. The learning trajectory precisely models each student's academic performance, providing an individualized analysis of the student to the teacher.

Knowledge Concept Relationship The next question prediction module provides an individual knowledge state of a student and the global knowledge state of all students through the knowledge concept relationship.

A trained DKT is used to derive relationships between knowledge concepts. The strength of the relationship between questions is the probability of being correct on the jth question given the correct student response on the ith question. The resulting strength of the probability is not dependent on an individual student but is global to all student data that the DKT was trained on.

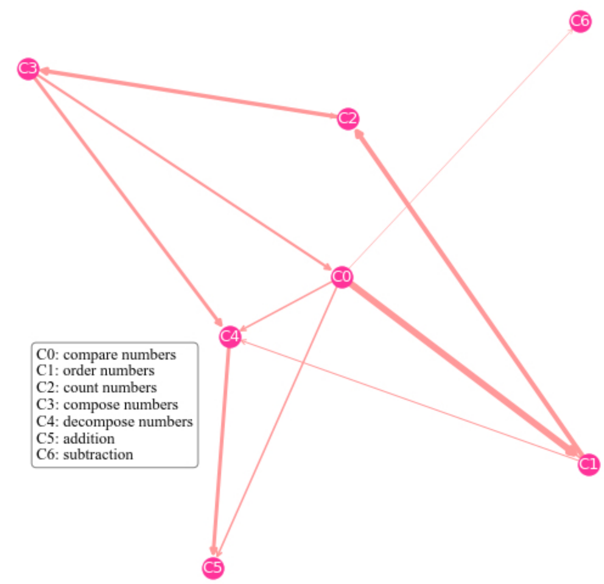

Figure 4: An example of conditional influences between knowledge concepts predicted by the DKT. Relationship strengths below 0.166 are thresholded.

Figure 4 shows the knowledge concept relationship derived from the KOFAC data with the DKT. The weights on the arrows pointing to 'C0' to 'C1' indicate the strength of the probability that students will get the 'C1' question correct given the correct response on 'C0'. The teacher can then assume that students are likely to understand the concept 'C1' better once they master the concept 'C0'.

The next question prediction module models the current knowledge state of an individual student with the learning trajectory and the knowledge-concept relationship's knowledge state.

Exam Score Prediction Module

Once students accomplish skill mastery, we determine the completeness of skill mastery with student evaluation. We

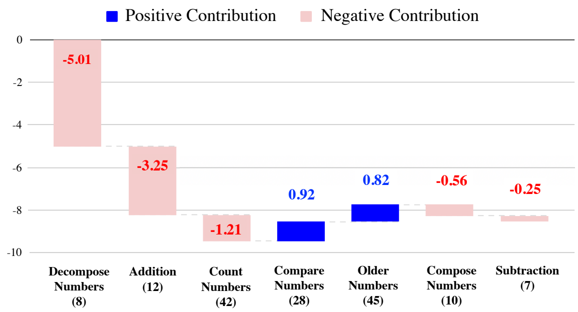

Figure 5: An example of Shapley Values on a first-grade student with a 74.19 predicted score. The numbers under the number sense concepts indicate the actual score for each concept. The true exam score of the student is 74. The concept 'decompose numbers' mostly contributed to lowering the predicted exam score.

consider performance a strong indicator of student learning, and therefore, we set the objective of exam score prediction to model student knowledge. To serve the purpose of the e-learning application to achieve students' skill mastery, the exam prediction module predicts the student's future performance for verifying the completion of the skill mastery.

Problem Formulation The exam score prediction module predicts the final exam score of a student given the results of the diagnostic test. We set the exam score prediction task as a supervised regression problem. The task of this module is formalized as: given the student performance on N number of knowledge concepts $\mathbf{x} = \{x_{c_1}, x_{c_2}, \ldots, x_{c_N}\}$ and the final exam score $\mathbf{y} \in \mathbf{R^1}$, we aim to find a machine learning model $f \in \mathcal{F}$, $\mathcal{F} : \{f|f : \mathbf{R}^N \Rightarrow \mathbf{R^1}\}$. In other words, our objective is to find f that satisfies the $\mathrm{argmin}_{f\in\mathcal{F}} \mathrm{RMSE}(f(\mathbf{x}), \mathbf{y})$.

We used XGBoost to model this task. XGBoost is a tree ensemble algorithm that uses the tree boosting method, which demonstrates state-of-the-art results across diverse machine learning challenges (Chen and Guestrin 2016). The XGBoost is trained on the diagnostic test results and predicts the final exam score. Predicted scores are normalized between the range of 0 and 100.

Shapley Value Explaining the expected performance of a student to the teacher is necessary to employ machine learning models in education. Most machine learning models are inherently black boxes, and they do not reveal the prediction's internal workings. Explaining increases transparency, reliability as well as trust towards the prediction. With the explanation, teachers function as an effective facilitator that helps students in their self-regulated learning by providing personalized learning guidance. In this sense, adopting a local explanation method, which generates an explanation for a single instance, is appropriate in that the focus is on assessing an individual student.

The explanation for the predicted score is produced using Shapley Value. We choose to use Shapley Value for prediction explanation because it is based on solid theory and gives the explanation a reasonable foundation. Further, the Shapley Value guarantees that the prediction is fairly distributed among the features. Initially, Shapley Value determines the order of significance of a player in multiplayer cooperative game theory (Shapley 1953). The Shapley Value indicates the amount of significance each player has over all possible combinations of players. The application of Shapley Value to the regression task extends to finding the magnitude of contribution that each feature has over all possible combinations of features.

In the exam score prediction module, the resulting Shapley Value indicates how the student performance on each of the diagnostic test's knowledge concepts impacts the model's decision to produce the predicted exam score. The knowledge concepts are considered as features and the Shapley Value expresses the amount of the feature contribution to the prediction. Therefore, the Shapley Value is viewed as the degree of strength and weakness of the student on a particular knowledge concept. Teachers may use this information to analyze student performance and provide appropriate feedback to the student.

The exam score prediction module is further utilized as a personalized recommendation measure in the e-learning application. As a student registers to the platform, the student is required to take a one-time diagnostic test. After the student finishes the diagnostic test, the recommended activity is presented to the student based on the Shapley Value. The recommendation decision is based on the assumption that the knowledge concepts with negative Shapley Value are considered academic weaknesses of the student.

An example of the Shapley Value of a student is presented in Figure 5. The student in Figure 5 receives learning advice from the e-learning application based on the resulting Shapley Value. The generated learning advice would suggest students practice more in the order of 'Decompose Numbers', 'Addition', and 'Count Numbers'. The learning advice is then presented to teachers as a written text. An example of learning advice is shown in the right screenshot of Figure 1.

The predicted exam score and the resulting Shapley Value is hidden from students but presented to teachers in the e-learning application. The Shapley Value helps teachers identify the student's academic status and alter learning plans to advance students' overall skill development, such as providing more resources on the specified learning concept.

Experiment

Model Implementation

Data Preprocessing We preprocess the data in two different formats for our two modules. For the next question prediction module, the input to the DKT is one-hot vectors of the student response on a series of questions. We set the maximum number of sequences to the maximum question number of both data.

For the exam score prediction module, the input to the XGBoost is the summed scores of each knowledge concept of the diagnostic test, and the target is the final exam score of the second semester. Min-Max normalization is applied

Data	RMSE
Grade 1	6.48
Grade 2	13.81

Table 2: Performance of the XGBoost

Data	Loss	ACC	AUC
Grade 1	0.0741	98.10	91.10
Grade 2	0.1521	95.02	90.23

Table 3: Performance of the DKT

to all scores. We split the data into the train and test set with an 8:2 ratio.

Training Details We train and evaluate our model in a single machine with three GeForce GTX 1080 Ti GPUs. We did not use GPUs when training the XGBoost. We present the final result of the models from a single run of training.

For the exam score prediction module, we use the XGBoost with 1,000 estimators. The selected evaluation metric is the root mean squared error (RMSE). The learning rate for XGBoost is set to 0.01. We used XGBoost library for implementation (Chen and Guestrin 2016). The Shapley Value for this module is implemented using the SHAP framework (Lundberg and Lee 2017).

For the next question prediction module, we use the DKT with a single LSTM layer with 100 hidden dimensions. The DKT is implemented using the PyTorch library and trained with the distributed data parallelism method. We train the DKT with ten epochs and with a batch size of 64. We use Noam optimizer with the learning rate of 0.001, and a warmup step of 4,000 (Vaswani et al. 2017). The network is trained with the binary cross-entropy loss function. Evaluation metrics for the DKT are accuracy and Area under the ROC Curve (AUC).

Model Performance

The performance of the implemented XGBoost is presented in Table 2 and DKT is presented in Table 3. XGBoost achieves 6.48 RMSE on the first-grade student data and 13.81 RMSE on the second-grade student data. Our implementation of the DKT achieves an AUC of 91.10% on the first-grade student data and 90.23% on the second-grade student data. Considering that our RMSE of XGBoost is small and the AUC of DKT is over 90% out of 100%, our models achieve adequate performance for providing teachers with meaningful feedback.

Expert Evaluation

Participants

We surveyed nine experts in mathematics education in South Korea to evaluate the Student Knowledge Prediction Framework's efficacy. Since the e-learning application was in a closed beta test, we selected the expert survey as our evaluation method. All nine experts have master's or doctoral degrees in Education. Six of the nine experts are working as elementary school teachers, two are professors of educational psychology, and the remaining one is a mathematics education researcher. All of the experts have more than ten years of experience in the field.

Method

We investigate whether both modules of our framework fulfill the function of informing the student knowledge state to teachers and help with teacher-student interaction. The questions in the survey are:

1. The Shapley Value from the exam score prediction module assists students to enhance their mathematical skills.
2. The Shapley Value is helpful in understanding the student's current knowledge state.
3. The Shapley Value is reliable.
4. The learning trajectory is reliable.
5. The concept map is helpful in providing feedback to students.

We created a simulated student and provided arbitrarily sampled instances for examination. We asked the experts to respond to the usability and reliability of Shapley Values, learning trajectories, and the related concept map. We also provide how each output is generated and how to interpret the output to assist with the expert's understanding. The experts examined the instances and the above questions on a Likert scale from 1 to 5.

We also consider observations of the experts. We collected short-answer responses from the nine experts on their opinions in using and applying the three outputs of the framework to teaching methods.

The questions for short-answer responses are:

1. How can teachers use the Shapley Values in their teaching method?
2. Are learning trajectories reliable and why?
3. How can teachers use the knowledge concept relationship for their teaching method?

We specifically investigate the reliability of learning trajectories, that there have been addressed limitations of DKT (Ding and Larson 2019). The collected responses are translated from Korean to English by the authors.

Survey Result

The expert responses on both modules are summarized in Figure 6. The label of each column represents the index number of questions. Overall, we received positive feedback on the two modules for educational purposes.

Most of the experts were positive towards the role of the exam score prediction module, recognizing that it helps students enhance their mathematical skills and helps understand students' knowledge state. On the survey question 1 and 2, seven of the experts responded 'Strongly Agree'. Based on the survey responses, experts view that the Shapley Values provide meaningful information on students' knowledge state, supporting the teacher-student interaction.

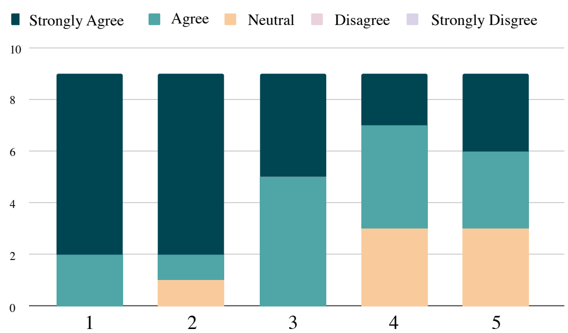

Figure 6: Expert responses on the usability and reliability of the framework.

For survey questions 3, 4, and 5, the experts exhibited a decrease in their certainty towards reliability. On the Shapley Values' reliability, five experts responded 'Agree' and on the reliability of the learning trajectory, three experts responded 'Neutral', and four experts responded 'Agree'.

We assume that the increased uncertainty originates from the lack of a baseline to evaluate our framework's reliability. No framework implemented machine learning models for practical use in public mathematics education in South Korea to the best of our knowledge. The experts also do not have prior knowledge of the students that they examined during the survey. The instances are randomly sampled from a simulated student, which contributed to growing uncertainty towards the reliability.

Survey Comments on Application to Teaching Method

On the application of Shapley Value, we asked how teachers can implement the Shapley Value to their teaching methods. Expert A addressed the local explanation nature of using Shapley Values to estimate the knowledge state of an individual student: *"... able to predict individual student's learning performance and identify academic weaknesses"*. Expert B reflected on the framework's usability: *"...provides student's relative strengths and weaknesses reliably and straightforwardly"*.

Expert responses on the reliability of the learning trajectories confirmed the reliability or were uncertain in an application for student assessment. Expert C mentioned *"...convenient to identify student's performance."* and expert D commented, *"the student's exercise progress corresponds to the graph"*. On the other hand, expert E pointed out the limitation of DKT that *"despite that learning trajectories are generally reliable, the change in a question's probability looks like they are affected by other exercises. It is difficult to identify the cause of the change"*. Expert F also commented *"it is insufficient to evaluate a student only based on their correct and incorrect responses"*.

Although our framework does not aim to evaluate or assess students, the expert's comments provide meaningful insight into the future research direction on evaluating student performance with machine learning models.

On the question of applying knowledge concept relationship to teaching methods, expert F responded that the knowledge concept relationship might be applied as *"1) learning material recommendation algorithm, especially for math subjects where the hierarchy among concepts are distinct, 2) development of curriculum and 3) help teachers to determine what knowledge concepts that a student struggles on"*. Expert G suggested a potential application *"...may be employed to reconstruct curriculum"* and expert G also addressed *"...is useful when creating a new curriculum"*.

Based on the expert's comments, using Shapley Values, learning trajectory, and the knowledge concept relationship allows for the student knowledge estimation of teachers and suggests a new approach to education, contributing to teacher-student interaction in distance learning.

Related Work

Deep Knowledge Tracing

Knowledge Tracing (KT) is an attempt to model the changing students' knowledge state during skill acquisition (Corbett and Anderson 1994). The knowledge tracing model acts as a tutor that tracks the current mastery level or knowledge state of a student on knowledge concepts in probability. The estimated knowledge state is beneficial for students' individualized learning that teachers can effectively instruct appropriate topics depending on the student's needs to improve academic performance.

Bayesian Knowledge Tracing (BKT) is an approach to model students' knowledge state in a Hidden Markov Model (HMM) with Bayesian inference (Yudelson, Koedinger, and Gordon 2013). However, BKT suffers from the limitations of assuming binary representation of student skill and independence between the student's skill.

Recently, deep learning models have been actively applied to model the student knowledge state and show promising performance. The first work to apply a deep learning model to KT is Deep Knowledge Tracing, which uses the LSTM network to predict a student's response for the next time step (Piech et al. 2015). Other works following the deep learning approach include the use of Memory Networks (Zhang et al. 2017), and incorporation of Attention mechanisms (Pandey and Karypis 2019).

These approaches demonstrate superior performance but come with a limitation on explainability. To apply the deep learning models in the domain of education, students' comprehension of the model's prediction is crucial because they use the model's prediction as a form of feedback for their self-reflection in learning. (Conati, Porayska-Pomsta, and Mavrikis 2018). Recent attempts that address explainability augments a deep learning model with Item Response Theory (Yeung 2019) or applies Layer-wise Relevance Propagation (Lu et al. 2020).

eXplainable Artificial Intelligence

Many machine learning models are black-boxes, which the internal workings of the model are opaque and obscure to humans. As machine learning models are embodied in society, understanding and trusting these opaque models' prediction is imperative. Especially for the sensitive domains such as healthcare, military, and finance, the lack of explanation of the model's prediction becomes a restriction to apply machine learning models in the process of decision-making (Varshney 2016). Explaining the prediction increases transparency, reliability, and trust towards the model in a real-world application.

The scope of explaining an opaque model is divided into interpreting the entire model behavior or comprehending a single prediction (Adadi and Berrada 2018). Explaining a specific decision of a model indicates an explanation is generated locally. Prominent local interpretability approaches use saliency masks (Xu et al. 2015), class activations (Zhou et al. 2016), game theory (Lipovetsky and Conklin 2001) and backpropagation (Bach et al. 2015) to explain a prediction on an input. The Shapley Value (Lundberg and Lee 2017) employs game theory to draw the comparative importance of variables to the model. This method has been applied to the domains of healthcare (Lundberg et al. 2018) and the environment (Stojić et al. 2019).

Teacher-Student Interaction of Constructivism

The proposed Student Knowledge Prediction Framework has its base on constructivism learning theory (Kanselaar 2002). Constructivism learning theory articulates how students construct knowledge during a process of learning.

Cognitive constructivism by Jean Piaget regards a student as a being who actively acquires knowledge rather than passively accepting knowledge from the environment (Piaget and Cook 1952). This notion opposes the traditional learning theory, where knowledge is delivered from a teacher to students. Cognitive constructivism also emphasizes students' current knowledge as a prominent element for the active construction of knowledge (Liu and Chen 2010). The process of using current knowledge to construct new knowledge is considered as a critical aspect of cognitive constructivism (Hoover 1996).

In cognitive constructivism, teachers play a central role in the active knowledge construction of their students. Constructivist teachers guide through examining the competence of their student's current knowledge state (Hoover 1996). Teachers should aid students' active learning through authentic tasks, experiences, collaboration, and assessment (Christie 2005). In the same sense, constructivism also considers teachers' role as examining the student's knowledge and assisting them to apply the newly acquired knowledge. (Mvududu and Thiel-Burgess 2012).

Social constructivism is strongly influenced by the theories of Vygotsky (Vygotsky 1980). The social constructivism suggest that knowledge is first constructed with social learning of environments, then it is embodied and utilized by students (Bruning, Schraw, and Ronning 1999). Social constructivism emphasizes the process of sharing individual perspectives through collaboration elaboration and encouraging students to construct knowledge together (Van Meter and Stevens 2000).

Conclusion

We propose a Student Knowledge Prediction Framework that predicts and explains students' current knowledge state for teachers in distance learning. The framework consists of two modules, the exam score prediction module and the next question prediction module, which both present modeling of a student's current knowledge state. We used DKT and XGBoost to model the knowledge state and utilized Shapley Value to explain the prediction. Our implementation of the model demonstrates adequate performance and received positive evaluations from the expert survey.

The proposed Student Knowledge Prediction Framework has two educational implications that reflect the constructivism's emphasis on the teacher's role in the teacher-student interaction in distance learning.

First, the framework operates as a pedagogical tool for the significant learning of students. The assessment framework's main educational functionality is providing students' knowledge state to teachers in distance learning. With the framework's current knowledge, teachers may adjust their teaching methods to provide personalized academic guidance tailored to the student's knowledge state.

Second, we adopt the constructivist view of ideal learning, emphasizing teacher-student interaction in distance learning. Presenting the individual knowledge state allows teachers to estimate students' current mastery level, which serves as a basis for teachers to connect with students through tailored feedback. The framework proposes an opportunity to expand AI application that contributes to teaching and improves student learning.

The framework will be implemented in public mathematics education and used by 700,000 elementary school students of South Korea. We plan to investigate whether our framework effectively enhances students' academic performance and interest in mathematics for future work. We hope to advance our framework by implementing memory networks with an explainability feature. Further, we aim to explore implementing a reinforcement learning approach for dynamic and adaptive student feedback in the e-learning platform.

Acknowledgements

This work was supported by Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2020-0-00368, A Neural-Symbolic Model for Knowledge Acquisition and Inference Techniques).

References

Adadi, A.; and Berrada, M. 2018. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6: 52138–52160.

Ally, M. 2004. Foundations of educational theory for online learning. *Theory and practice of online learning* 2: 15–44.

Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; and Samek, W. 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one* 10(7): e0130140.

Bates, T. 2008. Online learning tools and technologies, Retrieved September 10, 2020. *URL: https://www. tonybates. ca/wp* .

Bruning, R. H.; Schraw, G. J.; and Ronning, R. R. 1999. *Cognitive psychology and instruction*. ERIC.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Christie, A. 2005. Constructivism and its implications for educators. *Retrieved February* 19: 2014.

Conati, C.; Porayska-Pomsta, K.; and Mavrikis, M. 2018. AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling. *arXiv preprint arXiv:1807.00154* .

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

Ding, X.; and Larson, E. C. 2019. Why Deep Knowledge Tracing Has Less Depth than Anticipated. *International Educational Data Mining Society* .

Estellés, M.; and Fischman, G. E. 2020. Imagining a Post-COVID-19 Global Citizenship Education. *Práxis Educativa* 15: 1–14.

Greeno, J. G.; Collins, A. M.; Resnick, L. B.; et al. 1996. Cognition and learning. *Handbook of educational psychology* 77: 15–46.

Gunning, D. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* 2: 2.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Hoover, W. A. 1996. The practice implications of constructivism. *SEDL Letter* 9(3): 1–2.

Kanselaar, G. 2002. Constructivism and socio-constructivism. *Constructivism and socio-constructivism* 1–7.

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(4): 441–450.

Lipovetsky, S.; and Conklin, M. 2001. Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry* 17(4): 319–330.

Liu, C. C.; and Chen, I. J. 2010. Evolution of constructivism. *Contemporary issues in education research* 3(4): 63–66.

Lu, Y.; Wang, D.; Meng, Q.; and Chen, P. 2020. Towards Interpretable Deep Learning Models for Knowledge Tracing. *arXiv preprint arXiv:2005.06139* .

Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, 4765–4774.

Lundberg, S. M.; Nair, B.; Vavilala, M. S.; Horibe, M.; Eisses, M. J.; Adams, T.; Liston, D. E.; Low, D. K.-W.; Newman, S.-F.; Kim, J.; et al. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering* 2(10): 749–760.

Mvududu, N.; and Thiel-Burgess, J. 2012. Constructivism in practice: The case for English language learners. *International Journal of Education* 4(3): 108.

Pandey, S.; and Karypis, G. 2019. A Self-Attentive model for Knowledge Tracing. *arXiv preprint arXiv:1907.06837* .

Piaget, J.; and Cook, M. 1952. *The origins of intelligence in children*, volume 8. International Universities Press New York.

Piech, C.; Bassen, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L. J.; and Sohl-Dickstein, J. 2015. Deep knowledge tracing. In *Advances in neural information processing systems*, 505–513.

Shapley, L. S. 1953. A value for n-person games. *Contributions to the Theory of Games* 2(28): 307–317.

Stojić, A.; Stanić, N.; Vuković, G.; Stanišić, S.; Perišić, M.; Šoštarić, A.; and Lazić, L. 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Science of The Total Environment* 653: 140–147.

Van Meter, P.; and Stevens, R. J. 2000. The role of theory in the study of peer collaboration. *The Journal of Experimental Education* 69(1): 113–127.

Varshney, K. R. 2016. Engineering safety in machine learning. In *2016 Information Theory and Applications Workshop (ITA)*, 1–5. IEEE.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008.

Vygotsky, L. S. 1980. *Mind in society: The development of higher psychological processes*. Harvard university press.

Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; and Bengio, Y. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, 2048–2057.

Yeung, C.-K. 2019. Deep-irt: Make deep learning based knowledge tracing explainable using item response theory. *arXiv preprint arXiv:1904.11738* .

Yudelson, M. V.; Koedinger, K. R.; and Gordon, G. J. 2013. Individualized bayesian knowledge tracing models. In *International conference on artificial intelligence in education*, 171–180. Springer.

Zhang, J.; Shi, X.; King, I.; and Yeung, D.-Y. 2017. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web*, 765–774.

Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; and Torralba, A. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2921–2929.