Exploring Simultaneous Knowledge and Behavior Tracing

Siqian Zhao
Department of Computer Science
University at Albany - SUNY
Albany, NY 12222, USA
szhao2@albany.edu

Shaghayegh Sahebi
Department of Computer Science
University at Albany - SUNY
Albany, NY 12222, USA
ssahebi@albany.edu

ABSTRACT

Knowledge Tracing (KT) focuses on quantifying student knowledge according to the student's past performance. While KT models focus on modeling student knowledge, they miss the behavioral aspect of learning, such as the types of learning materials that the students choose to learn from. This is mainly because traditional knowledge tracing (KT) models only consider assessed activities, like solving questions. Recently, there has been a growing interest in multi-type KT which considers both assessed and non-assessed activities (like video lectures). Since multi-type KT models include different learning material types, they present a new opportunity to investigate student behavior, as in the choice of the learning material type, along with student knowledge. We argue that student knowledge can affect their behavior, and student interest in learning materials may affect their knowledge. In this paper, we model the relationship between students' knowledge states and their choice of learning activities. To this end, we propose Pareto-TAMKOT which frames the simultaneous learning of student knowledge and behavior as a multi-task learning problem. It employs a transition-aware multi-activity KT method for two objectives: modeling student knowledge and student behavior. Pareto-TAMKOT uses the Pareto Multi-task learning algorithm (Pareto MTL) to solve this multi-objective optimization problem. We evaluate Pareto-TAMKOT on one real-world dataset, demonstrating the benefit of approaching student knowledge and behavior modeling as a multi-task learning problem.

Keywords

knowledge tracing, student behavior tracing, multi-activity, multi-task learning, Pareto learning

1. INTRODUCTION

Knowledge Tracing (KT) [7, 9, 3, 21] is essential for modern education systems as it enables predicting student performance or quantifying student knowledge states, given a large number of students and activities [26, 23, 28, 12, 14, 20, 28, 22]. Traditionally, KT focused on one type of learning activity: assessed activities, such as solving questions [9, 3, 4, 21, 15, 17, 18]. In recent years, interest in multi-type or multi-activity KT has grown [27, 31, 32, 30]. In addition to modeling assessed activities, multi-activity KT models can represent how students learn from non-assessed activities, like watching a video lecture. While these models are instrumental in discovering how a specific activity leads to students' knowledge growth, they miss the opportunity to utilize another signal in student activity data: the relationship between student knowledge and their behavior, particularly, in terms of preference of material type that the student chooses to interact with. Students' choices are influenced partly by their own preferences, but their knowledge state also plays a role in deciding which materials to interact with next [19, 1, 2, 32, 31]. For example, a student might skip some learning materials with similar topics if they feel confident about their understanding of the topic. Alternatively, for a question that the student does not have a clear answer to, they can read the appropriate question hint to learn and assist in determining the answer, rather than repeating the question. Furthermore, student preference for learning materials may affect their knowledge. For example, a student's knowledge may benefit more from practicing a question, while for another student, reading an annotated example may result in a higher knowledge gain. Furthermore, the students may learn more effectively by interacting with the learning materials that they are interested in. Thus, it's important to explore how student knowledge and their choices are interconnected, as this could enhance the assessment of their knowledge as well.

In this paper we introduce Pareto-TAMKOT, framing the simultaneous learning of student knowledge and behavior as a multi-task learning problem. We employ a transition-aware multi-activity KT method (TAMKOT [32]) to concurrently model student knowledge and behavioral preference. Our method targets two objectives: (1) predicting student performance, and (2) predicting the types of materials students will interact with. To address this multi-objective optimization, we apply the Pareto Multi-task learning algorithm (Pareto MTL [16]). We evaluate our method on one publicly available real-world dataset. The results of our experiments demonstrate that Pareto-TAMKOT outperforms all baseline models in predicting both student performance and the types of learning materials. It is worth noting that our work is a preliminary idea in the direction of using multi-

S. Zhao and S. Sahebi. Exploring simultaneous knowledge and behavior tracing. In B. Paaßen and C. D. Epp, editors, *Proceedings of the 17th International Conference on Educational Data Mining*, pages 927–932, Atlanta, Georgia, USA, July 2024. International Educational Data Mining Society.

https://doi.org/10.5281/zenodo.12730001

objective optimization, and to the best of our knowledge, we are the first to undertake this problem.

2. PROBLEM FORMULATION

Our goal is to model student knowledge and trace student behavior preferences of learning materials as they interact with both assessed and non-assessed materials. We use a binary indicator $d_t \in \{0, 1\}$ to represent the type of material being interacted with at t, where 0 represents the assessed material type, and 1 represents the non-assessed type. To represent a student's learning activity at time step t, we use a tuple, $\langle i_t, d_t \rangle$, where $i_t = \begin{cases} (q_t, r_t) & \text{if } d_t = 0 \\ l_t & \text{if } d_t = 1 \end{cases}$, indicates the learning material being interacted with at t. Specifically, (q_t, r_t) represents the student interacted with the assessed material q_t at time step t, with performance r_t, and l_t represents the non-assessed material that the student interacted with at time step t. Eventually, the student's whole trajectory of learning activities is denoted as a sequence of these tuples, $\{\langle i_1, d_1 \rangle, ..., \langle i_t, d_t \rangle\}$. To evaluate the modeling of both student knowledge and behavior preference at the same time, we aim to predict learning material type d_{t+1} at time step $t+1$, as well as student performance r_{t+1} on assessed material q_{t+1}, if $d_{t+1} = 0$, given a student's past learning activity history, $\{\langle i_1, d_1 \rangle, \ldots, \langle i_t, d_t \rangle\}$.

3. METHODOLOGY

Our proposed model, Pareto-TAMKOT, is designed as a sequential multi-activity model to capture both students' knowledge and their behavior preferences for learning material choices. It is further capable of predicting future student performance and the types of learning materials they are likely to interact with next. Pareto-TAMKOT is built based upon the transition-aware multi-activity knowledge transfer method (TAMKOT [32]). Pareto-TAMKOT enhances TAMKOT by adding an additional component in the prediction layer to predict the type of learning materials. Besides the sole objective of predicting student performance in TAMKOT, we propose adding another objective for predicting the type of learning materials. We treat the prediction of student performance and learning material type as a multi-task learning challenge. To solve this dual-objective optimization problem, we employ the Pareto MTL algorithm, which aims to identify a set of well-distributed Pareto optimal solutions for learning model parameters. An overview of Pareto-TAMKOT's architecture is presented in Figure 1.

3.1 Knowledge and Behavior Modeling

Following TAMKOT, we structure Pareto-TAMKOT into three main layers: the embedding layer, the hidden knowledge transfer layer, and the prediction layer.

3.1.1 Embedding Layer

This layer's purpose is to generate an embedding vector for each learning activity $\langle i_t, d_t \rangle$, which serves as input to the hidden knowledge and behavioral preference transfer layer. Assume two types of learning materials: questions and video lectures. For question activity $i_t = (q_t, r_t)$, Pareto-TAMKOT learns two underlying embedding matrices $\boldsymbol{A}^q \in \mathbb{R}^{N^Q \times d_q}$ and $\boldsymbol{A}^r \in \mathbb{R}^{2 \times d_r}$, to map the question q_t and the student response performance r_t into a latent embedding space, producing their respective latent embeddings $\boldsymbol{q}_t \in \mathbb{R}^{d_q}$ and $\boldsymbol{r}_t \in \mathbb{R}^{d_r}$. These embeddings are then concatenated into a single embedding $\boldsymbol{x}_t = [\boldsymbol{q}_t \oplus \boldsymbol{r}_t]$ to represent the activity i_t. For the activity involving video lectures where $i_t = l_t$, Pareto-TAMKOT learns $\boldsymbol{A}^l \in \mathbb{R}^{N^L \times d_l}$ for mapping the video lecture $l_t$to obtain the latent representation $\boldsymbol{l}_t \in \mathbb{R}^{d_l}$, which directly serves as the embedding for the activity i_t.

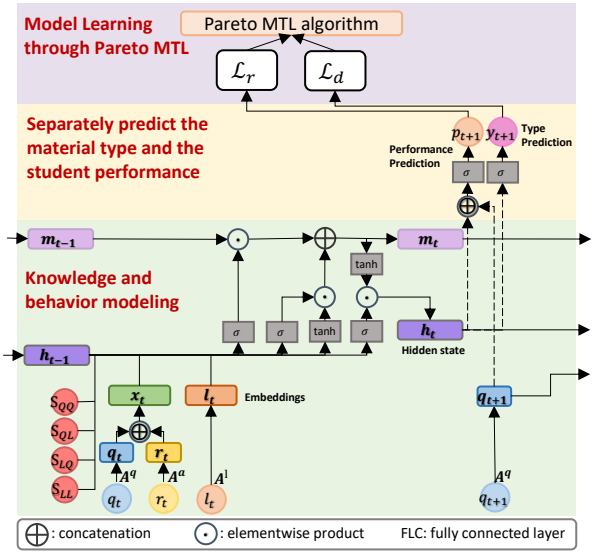

Figure 1: The Pareto-TAMKOT's model architecture.

3.1.2 Hidden Knowledge Transfer Layer

It aims to encapsulate the student's knowledge and behavior state h_t and facilitate the learning of knowledge and behavioral preference transfer as students transition between assessed and non-assessed learning material types. Similar to TAMKOT, our model also consists of a memory cell $\boldsymbol{g}_t$, an input gate $\boldsymbol{i}_t$, an output gate $\boldsymbol{o}_t$, and a forget gate $\boldsymbol{f}_t$. It models the knowledge and behavioral preference transfer patterns during transitions between different types of learning materials by utilizing specific indicators to guide the updates of these gates. We also employ transition-specific weight matrices, denoted by $\boldsymbol{W}$s, to differently update the student's knowledge and behavioral preference state according to the type of learning activity transition. At each timestep t, the knowledge and behavioral preference state $\boldsymbol{h}_t$ is updated using these matrices as follows (due to space limitation, we focus on the input gate equations, other gates have the same structures):

$$\begin{aligned} \boldsymbol{i}_t = \sigma\Big(& (1-d_t) \cdot \boldsymbol{x}_t \boldsymbol{V}_{iQ} + d_t \cdot \boldsymbol{l}_t \boldsymbol{V}_{iL} \\ & + s_{QQ} \cdot \boldsymbol{h}_{t-1} \boldsymbol{W}_{i_{QQ}} + s_{LL} \cdot \boldsymbol{h}_{t-1} \boldsymbol{W}_{i_{LL}} \\ & + s_{QL} \cdot \boldsymbol{h}_{t-1} \boldsymbol{W}_{i_{QL}} + s_{LQ} \cdot \boldsymbol{h}_{t-1} \boldsymbol{W}_{i_{LQ}} + \boldsymbol{b}_i \Big) \end{aligned} \quad (1)$$

$$\boldsymbol{m}_t = \boldsymbol{f}_t \cdot \boldsymbol{m}_{t-1} + \boldsymbol{i}_t \cdot \boldsymbol{g}_t \quad (2)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \cdot tanh(\boldsymbol{m}_t) \quad (3)$$

where S_{**} are the binary indicators that can represent permutations of transitions between learning material types from one time step $t-1$ to the next time step t. Here,

in our model, we assume that $\boldsymbol{h}_t$ can capture information on both student knowledge and their behavioral preferences for choosing learning materials, as our training objectives include predictions on student performance and material type.

3.1.3 *Prediction Layer*

In this layer, we separately predict the type of the next learning material and the student's performance on a given upcoming question q_{t+1} at the next time step $t+1$. This is achieved using two distinct MLPs, each based on the current hidden state $\boldsymbol{h}_t$:

$$p_{t+1} = \sigma(\boldsymbol{W}_p^T[\boldsymbol{h_t} \oplus \boldsymbol{q_{t+1}}] + b_p) \tag{4}$$

$$y_{t+1} = \sigma(\boldsymbol{W}_y^T \boldsymbol{h_t} + b_y) \tag{5}$$

where p_{t+1} denotes the probability of the student correctly solving the upcoming question q_{t+1}, while y_{t+1} indicates the probability of the student interacting with non-assessed materials. $\boldsymbol{W}_*$ refer to weight matrices, and b_* are bias terms.

3.2 Model Learning through Pareto MTL

The objective functions for predicting student performance and material type are calculated using binary cross-entropy losses. This involves comparing the actual and estimated student performance r_t and p_t, as well as the actual material type with the predicted type, d_t and y_t, as follows:

$$\mathcal{L}_r = -\sum_t (r_t \log p_t + (1 - r_t) \log (1 - p_t)) \tag{6}$$

$$\mathcal{L}_d = -\sum_t (d_t \log y_t + (1 - d_t) \log (1 - y_t)) \tag{7}$$

Recent developments have introduced strategies to identify a single Pareto optimal solution that achieves a trade-off compromise between various tasks by treating multi-task learning as an exercise in multi-objective optimization [33, 8, 11, 10, 25]. Then, the Pareto MTL [16] algorithm is noteworthy for its ability to identify a collection of representative Pareto optimal solutions, each offering a different trade-off among the tasks. We adopt the Pareto MTL algorithm for training our model, allowing us to learn model parameters effectively. As shown in Figure 2, it employs a series of dividing vectors $\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_m$ to decompose multi-objective optimization challenge into multiple constrained sub-problems. Each sub-problem represents different trade-off preferences, and these sub-problems are solved concurrently. As a result, Pareto MTL enables us to obtain a set of well-representative Pareto solutions for the multi-objective problem, thus allowing us to select our preferred solution(s) for the tasks of predicting student performance and material type from the set of Pareto optimal solutions.

4. EXPERIMENTS

To assess the effectiveness of Pareto-TAMKOT, we evaluate its capability to predict student performance against baseline KT methods. Additionally, we examine its performance in predicting the type of learning materials students will choose to interact with. Our code and data are available at GitHub [1].

[1] https://github.com/persai-lab/2024-EDM-Pareto-TAMKOT

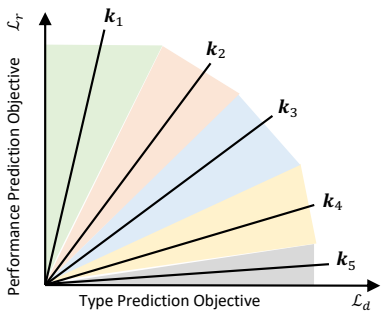

Figure 2: The illustration of Pareto MTL, it finds a set of Pareto solutions through a series of unit dividing vectors $\boldsymbol{k}_s$.

4.1 Dataset

For our experiments, we utilize the **EdNet**[2] [6] dataset, which is publicly available and anonymized, and is exempt from IRB review. The dataset originates from a multi-platform AI tutoring service known as Santa [3], designed to assist Korean students in preparing for the TOEIC [4], a leading English communication skill assessment for the workplace and everyday life. EdNet provides data that captures a range of student learning activities across different material types. In our research, we employ a preprocessed dataset, as introduced in [32, 30], focusing on questions (assessed) and their associated question explanations (non-assessed) as the two types of materials. Each question is a multiple-choice item accompanied by an explanation. Students can choose the questions they wish to practice or follow the platform's recommendations and decide whether to review the explanations while practicing. Eventually, the dataset encompasses data from $1,000$ students, comprising $11,249$ questions and $8,324$ question explanations. It totals $200,931$ question-solving activities and $150,821$ question explanation review activities, with an average student practice length of 352.

4.2 Baselines

4.2.1 *Student Performance Prediction Baselines*

We compare the proposed method with five KT baseline methods to evaluate the proposed method's effectiveness in the student performance prediction task. This includes two supervised assessed-only KT models and one multi-activity KT model. Additionally, we extend the two assessed-only supervised KT models to accommodate both assessed and non-assessed activities. These modified models are indicated by adding "+M" to the original model name. The assessed-only baselines include:

- **DKT** [22] is a pioneer deep learning model for KT, employing recurrent neural networks to capture students' knowledge acquisition over time.
- **AKT** [12] leverages a context-aware approach with a monotonic attention mechanism to focus on past student performances relevant to the current question.

[2] https://github.com/riiid/ednet
[3] https://www.aitutorsanta.com/
[4] https://www.ets.org/toeic

The following are multi-activity baselines:

- **DKT+M** [29] is an extended version of DKT, that integrates both assessed and non-assessed learning activities by appending embeddings of non-assessed materials encountered between two assessed activities as an additional input feature alongside the original question embeddings.
- **AKT+M** modifies AKT by summarizing embeddings of non-assessed materials between two assessed activities as an extra feature and includes position encoding for each learning material's embedding [5].
- **TAMKOT** [32] is a transition-aware model based on LSTM technology. It distinguishes itself by learning multiple knowledge transfer matrices, explicitly modeling the transfer of knowledge across various types of activities.

4.2.2 *Learning Material Type Prediction Baselines*

To evaluate the Pareto-TAMKOT's effectiveness in predicting types of learning materials, we conduct experiments to compare it against three deep sequential baseline models. The baselines include:

- **LSTM** [13] is a type of recurrent neural network architecture that excels at learning long-term dependencies, making it highly suitable for tasks requiring an understanding of entire data sequences.
- **MANN** [24] enhances neural networks with an external memory component, enabling the storage and retrieval of information across long sequences. This feature makes them especially valuable for tasks that demand the retention and manipulation of information over time.

To facilitate this comparison, we employ learning material embeddings along with the material type as inputs to the two baselines mentioned above, focusing only on predicting the upcoming type of material. Additionally, we employ a variant of TAMKOT to predict the type alone as another baseline:

- **TAMOKT** is a multi-activity knowledge method. In this approach, we retain the original architecture of the TAMKOT, leverage the hidden state learned by it, and apply an additional MLP for material type prediction, bypassing the use of the Pareto MTL algorithm to achieve the objective.

4.3 Experiments Setup

To divide the training and testing datasets, we employ a 5-fold student stratified cross-validation approach. In this process, sequences from 80% of the students form the training set, while sequences from the remaining 20% students serve as the testing set. Additionally, for hyperparameter tuning, sequences from 20% of the students in the training set are allocated as a validation set. Our proposed methods are developed using PyTorch [5], and we opt for the Adam optimizer to learn the model parameters. During the Pareto MTL optimization, we employ five evenly distributed preference vectors and initialize these parameters with random values following a Gaussian distribution with a mean of 0 and a standard deviation of 0.2. To prevent the issue of gradient exploding, we apply norm clipping. Consistent with established practices in KT experiments [22], we ensure all sequences are of uniform length by either truncating or padding them. The length of these sequences, denoted as L_s, is determined through hyperparameter tuning using the validation set. Sequences exceeding L_s are truncated into multiple sequences, whereas those shorter than L_s are extended using padding with 0s. The best hyperparameters are identified through a coarse-grained grid search.

Table 1: Student performance prediction results (AUC). The best and second-best results are in boldface and underline.

Methods	AUC
DKT	0.6393 ± 0.01370
AKT	0.63933 ± 0.0104
DKT+M	0.6372 ± 0.0120
AKT+M	0.6404 ± 0.0067
TAMKOT	$\underline{0.6786 \pm 0.0063}$
Pareto-TAMKOT	$\mathbf{0.6809 \pm 0.0063}$

Table 2: Material type prediction results(AUC). The best and second-best results are in boldface and underline.

Methods	AUC
LSTM	0.8768 ± 0.0041
MANN	0.8933 ± 0.0030
TAMKOT	$\underline{0.8929 \pm 0.0042}$
Pareto-TAMKOT	$\mathbf{0.8987 \pm 0.0042}$

4.4 Prediction Performance Comparison

In the EdNet dataset, student responses are binary (success or failure), and there are two types of materials (questions or question explanations). Therefore, we evaluate the effectiveness of each model by calculating the Area Under the Curve (AUC) for both student performance prediction and material type prediction tasks. A higher AUC value indicates superior predictive accuracy. To ensure a fair comparison among methods, we report average results across the five folds, along with their confidence intervals, at a significance level of 0.05 for each method. The results from experiments on the student performance prediction and the learning material type prediction are reported in Table 1 and Table 2, respectively.

During our experiments, we observed that setting the preference vector of Pareto-TAMKOT to $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, corresponding to the direction of $\frac{\pi}{4}$ (as illustrated by the middle vector $\boldsymbol{k}_3$ in Figure 2), leads to improvements in both student performance and material type predictions. On the other hand, when altering the preference vector to other values resulted in significant improvements in student performance predictions but only slight enhancements in predictions for material types, or vice versa. Notably, the optimal prediction performance for each specific task consistently occurred when employing the corresponding extreme dividing vectors, such

[5]https://pytorch.org/

as $(0, 1)$ or $(1, 0)$. Given these observations, and to ensure a meaningful comparison for both predictions for student performance and material type tasks, we decide to exclusively report Pareto-TAMKOT results using the preference vector set at $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ in both Table 1 and Table 2.

4.4.1 Student Performance Prediction

We first observe that Pareto-TAMKOT outperforms all comparison baseline methods in the task of predicting student performance. These results underscore the model's ability to effectively track student knowledge and make accurate predictions about students' future performance. When comparing methods that consider multiple types of activities to those focusing solely on assessed activities, it becomes evident that methods encompassing a variety of activities generally outperform the assessed-only approach, with the exception of DKT+M. This indicates that including non-assessed activities can significantly enhance the understanding of students' knowledge acquisition. Furthermore, the better prediction performance of Pareto-TAMKOT than TAMKOT suggests that simultaneously modeling students' knowledge and behaviors, and addressing this through a multi-task learning approach with performing the Pareto-MTL for optimization, can further refine the modeling of student knowledge.

4.4.2 Learning Material Type Prediction

Likewise, Pareto-TAMKOT surpasses all baseline methods in predicting the type of learning material. This performance underlines the model's adeptness at capturing students' preferences for selecting learning materials and accurately predicting their future learning material choices. Additionally, when comparing our method with TAMKOT, it is evident that Pareto-TAMKOT outperforms TAMKOT. This emphasizes that the concurrent modeling of students' knowledge and behaviors, optimized with the Pareto-MTL algorithm, also can enhance the understanding of students' learning material behavior preferences.

Overall, considering results on both student performance and material type prediction, we can deduce that simultaneously modeling student knowledge while tracking their material selection behaviors leads to a deeper mutual understanding of these aspects. Consequently, the outcomes from the Pareto-TAMKOT experiments underscore the importance of framing this challenge as a multi-task learning problem. Applying the Pareto-MTL optimization algorithm is effective for accurately capturing both student knowledge and behavior regarding learning material selection, thereby enhancing predictions of student performance and material preferences.

5. CONCLUSIONS

Our research presents Pareto-TAMKOT, an initial strategy that addresses the joint learning of student knowledge and behavior through a multi-task learning framework. By deploying the TAMKOT method, we simultaneously model student knowledge and behavior, targeting multiple objectives: (1) predicting student performance, and (2) predicting the types of learning materials students are likely to interact with. Utilizing the Pareto MTL algorithm enables us to adeptly handle this multi-objective optimization challenge. Our experiment results show the benefit of approaching this as a multi-task learning problem. Although our experimental results are based on a single dataset with only limited baseline comparisons. Still, these results are promising, they highlight the potential for future investigations to build upon and extend our understanding in this study.

6. ACKNOWLEDGMENT

This paper is based upon work supported by the National Science Foundation under Grant Number 2047500.

7. REFERENCES

[1] R. Agrawal, M. Christoforaki, S. Gollapudi, A. Kannan, K. Kenthapadi, and A. Swaminathan. Mining videos from the web for electronic textbooks. In *International Conference on Formal Concept Analysis*, pages 219–234. Springer, 2014.

[2] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi. Enriching textbooks with images. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1847–1856, 2011.

[3] H. Cen, K. Koedinger, and B. Junker. Learning factors analysis–a general method for cognitive model evaluation and improvement. In *International conference on intelligent tutoring systems*, pages 164–175. Springer, 2006.

[4] H. Cen, K. Koedinger, and B. Junker. Comparing two irt models for conjunctive skills. In *International Conference on Intelligent Tutoring Systems*, pages 796–798. Springer, 2008.

[5] Y. Choi, Y. Lee, J. Cho, J. Baek, B. Kim, Y. Cha, D. Shin, C. Bae, and J. Heo. Towards an appropriate query, key, and value computation for knowledge tracing. In *Proceedings of the 7th ACM Conference on Learning at Scale*, pages 341–344, New York, NY, USA, 2020. ACM.

[6] Y. Choi, Y. Lee, D. Shin, J. Cho, S. Park, S. Lee, J. Baek, C. Bae, B. Kim, and J. Heo. Ednet: A large-scale hierarchical dataset in education. In *International Conference on Artificial Intelligence in Education*, pages 69–73. Springer, 2020.

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

[8] J.-A. Désidéri. Mutiple-gradient descent algorithm for multiobjective optimization. In *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012)*, 2012.

[9] F. Drasgow and C. L. Hulin. Item response theory. *Handbook of Industrial and Organizational Psychology*, pages 577–636, 1990.

[10] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical methods of operations research*, 51:479–494, 2000.

[11] J. Fliege and A. I. F. Vaz. A method for constrained multiobjective optimization based on sqp techniques. *SIAM Journal on Optimization*, 26(4):2091–2119, 2016.

[12] A. Ghosh, N. Heffernan, and A. S. Lan. Context-aware attentive knowledge tracing. In *Proceedings of the 26th*

ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 2330–2339, New York, NY, USA, 2020. ACM.

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

[14] M. I. Jordan, M. J. Kearns, and S. A. Solla. *Advances in Neural Information Processing Systems 10: Proceedings of the 1997 Conference*, volume 10. Mit Press, 1998.

[15] A. S. Lan, A. E. Waters, C. Studer, and R. G. Baraniuk. Sparse factor analysis for learning and content analytics. *Journal of Machine Learning Research (JMLR)*, 15(57):1959–2008, 2014.

[16] X. Lin, H.-L. Zhen, Z. Li, Q.-F. Zhang, and S. Kwong. Pareto multi-task learning. *Advances in neural information processing systems*, 32, 2019.

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

[18] T. Long, Y. Liu, J. Shen, W. Zhang, and Y. Yu. Tracing knowledge state with individual cognition and acquisition estimation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 173–182, 2021.

[19] A. S. Najar, A. Mitrovic, and B. M. McLaren. Adaptive support versus alternating worked examples and tutored problems: which leads to better learning? In *International Conference on User Modeling, Adaptation, and Personalization*, pages 171–182. Springer, 2014.

[20] S. Pandey and G. Karypis. A self-attentive model for knowledge tracing. In *Proceedings of the 12th International Conference on Educational Data Mining*, pages 384–389. International Educational Data Mining Society, 2019.

[21] P. I. Pavlik Jr, H. Cen, and K. R. Koedinger. Performance factors analysis–a new alternative to knowledge tracing. *Online Submission*, 2009.

[22] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein. Deep knowledge tracing. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, page 505–513, Cambridge, MA, USA, 2015. MIT Press.

[23] S. Sahebi, Y.-R. Lin, and P. Brusilovsky. Tensor factorization for student modeling and performance prediction in unstructured domain. *International Educational Data Mining Society*, 2016.

[24] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850. PMLR, 2016.

[25] O. Sener and V. Koltun. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems*, 31, 2018.

[26] N. Thai-Nghe, T. Horváth, and L. Schmidt-Thieme. Factorization models for forecasting student performance. In *Educational Data Mining 2011*. Citeseer, 2010.

[27] C. Wang, S. Zhao, and S. Sahebi. Learning from non-assessed resources: Deep multi-type knowledge tracing. *International Educational Data Mining Society*, 2021.

[28] J. Zhang, X. Shi, I. King, and D.-Y. Yeung. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th International Conference on World Wide Web*, pages 765–774, New York, NY, USA, 2017. ACM.

[29] L. Zhang, X. Xiong, S. Zhao, A. Botelho, and N. T. Heffernan. Incorporating rich features into deep knowledge tracing. In *Proceedings of the 4th ACM Conference on Learning at Scale*, pages 169–172, New York, NY, USA, 2017. ACM.

[30] S. Zhao and S. Sahebi. Graph-enhanced multi-activity knowledge tracing. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 529–546. Springer, 2023.

[31] S. Zhao, C. Wang, and S. Sahebi. Modeling knowledge acquisition from multiple learning resource types. In *Proceedings of The 13th International Conference on Educational Data Mining*, pages 313–324. International Educational Data Mining Society, 2020.

[32] S. Zhao, C. Wang, and S. Sahebi. Transition-aware multi-activity knowledge tracing. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1760–1769. IEEE, 2022.

[33] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE transactions on Evolutionary Computation*, 3(4):257–271, 1999.