Towards Multi-Objective Behavior and Knowledge Modeling in Students

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

Sherry Sahebi
Department of Computer Science,
University at Albany - SUNY
Albany, New York, USA
ssahebi@albany.edu

ABSTRACT

Traditional knowledge modeling methods have primarily focused on student knowledge modeling using assessed learning activities, often overlooking the critical interplay between students' knowledge and behavioral preferences. However, students typically interact with multiple types of learning materials, such as questions (assessed), video lectures (non-assessed), and textbooks (non-assessed). We argue that student knowledge can affect their behavioral preferences, and the choice of learning material type can influence their knowledge. In this paper, we address this gap by proposing a novel framework that models student knowledge and behavior as a multi-task learning problem with two objectives. Our dual objectives are to predict student performance and their preferences for selecting different types of learning materials. We utilize the Pareto Multi-Task Learning (MTL) algorithm to effectively handle the complexities of this multi-objective optimization, applying it to two advanced multi-activity knowledge modeling methods, TAMKOT and GMKT, which we refer to as Pareto-TAMKOT and Pareto-GMKT, respectively. We evaluate the framework on one real-world dataset. Our experimental results demonstrate that both Pareto-TAMKOT and Pareto-GMKT improve upon their original models and outperform all baseline models. This underscores the benefits of treating the modeling of student knowledge and behavior as a multi-task learning problem and addresses this multi-objective challenge through the application of Pareto MTL.

CCS CONCEPTS

• **Human-centered computing** → *User models*; • **Computing methodologies** → **Multi-task learning**; • **Computer systems organization** → *Neural networks.*

KEYWORDS

Knowledge tracing, Student behavior, Multi-activity, Multi-task learning, Pareto learning

ACM Reference Format:
Siqian Zhao and Sherry Sahebi. 2024. Towards Multi-Objective Behavior and Knowledge Modeling in Students. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP Adjunct '24), July 01–04, 2024, Cagliari, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3631700.3664880

ACM ISBN 979-8-4007-0466-6/24/07
https://doi.org/10.1145/3631700.3664880

1 INTRODUCTION

Recently, the demand for online education systems has surged, significantly enhancing their importance [35]. A vast amount of data from student interactions, provided by these systems, is beneficial for automatically understanding each student's knowledge and learning behaviors. Traditionally, research has focused only on student knowledge modeling [4, 8, 10, 23], which models student knowledge based on their interaction history with learning materials and aims to predict future performance [13, 16, 22, 24, 26, 29, 31, 31]. However, this often overlooks the modeling of student behavioral preferences. Early student knowledge models mainly addressed only assessed activities like solving questions [4, 5, 10, 17, 19, 20, 23]. Recently, there has been a shift towards multi-activity models that also account for non-assessed activities, such as watching video lectures [1, 6, 30, 33–35]. These multi-activity knowledge models can represent how students learn from both assessed and non-assessed activities. Although these models are crucial for understanding how different activities contribute to knowledge growth, they often neglect the potential insights from behavioral signals, particularly the relationship between student knowledge and their behavioral preferences.

Studies have shown that student knowledge and behavior mutually influence each other [2, 3, 21, 34, 35]. For example, some students repeatedly attempt questions they have already answered correctly to boost their confidence, although this behavior may not effectively enhance their knowledge. Essentially, while students' choice of material is partly driven by their preferences, their knowledge state also determines their choice of material [15, 25]. A student confident about a topic may choose to skip additional materials on the same topic. Moreover, the selection of materials based on student's personal behavior preferences can shape their knowledge acquisition. By choosing topics that align with their interests, students encounter a variety of materials that can enrich their knowledge in diverse ways. Consequently, it is crucial to understand the interplay between students' knowledge and behavioral preferences.

We propose a framework that treats the simultaneous learning of student knowledge and behavior as a multi-task learning problem with dual objectives: (1) predicting student performance, and (2) predicting the types of materials students will interact with. We utilize the **Pareto MTL** algorithm [18], to address this multi-objective optimization challenge. We apply this framework to two transition-aware multi-activity knowledge modeling methods, TAMKOT[35] and GMKT [33], hereafter referred to as **Pareto-TAMKOT** and **Pareto-GMKT**, to evaluate the effectiveness of the proposed framework. Our evaluation is conducted on a real-world dataset. The results of our experiments show that Pareto-TAMKOT and Pareto-GMKT enhance their original models and outperform all other baseline models in both tasks, demonstrating that capturing both student knowledge and behavior can mutually benefit learning in each task. It is worth noting that our approach represents a preliminary exploration of the use of multi-objective optimization for knowledge and behavior modeling, and to the best of our knowledge, this is the first attempt in this direction.

2 PROBLEM FORMULATION

Our goal is not only to predict students' upcoming performance on assessed materials but also to anticipate their selection of future material types. Accordingly, consider an education system that provides two types of learning materials, one assessed (e.g., questions) and one non-assessed (e.g., video lectures). We represent a student's entire trajectory of learning activities as a set of tuples $\langle i_1, d_1 \rangle, \ldots, \langle i_t, d_t \rangle$, where each tuple $\langle i_t, d_t \rangle$ denotes a student's learning activity at time step t. Here, $d_t \in \{0, 1\}$ is a binary indicator to represent the type of material being interacted with at time step t, where 0 signifies the assessed material type, and 1 signifies the non-assessed material type. And, $i_t = \begin{cases} (q_t, r_t) & \text{if } d_t = 0 \\ l_t & \text{if } d_t = 1 \end{cases}$ indicates the specific learning material and, for assessed activities, the corresponding student response at time step t. Specifically, (q_t, r_t) denotes that the student interacted with assessed material q_t at time step t, and their performance is recorded as r_t. Conversely, l_t represents the non-assessed material with which the student interacted at time step t. Eventually, given a student's historical trajectory learning activities, $\{\langle i_1, d_1 \rangle, \ldots, \langle i_t, d_t \rangle\}$, our objective is to predict the material type d_{t+1} that the student will interact with at the next time step $t + 1$, as well as the student's upcoming performance r_{t+1} on the assessed material q_{t+1}, if $d_{t+1} = 0$.

3 METHODOLOGY

As we aim to simultaneously model student knowledge and behavior preferences, which differ from traditional knowledge models focused solely on predicting student performance, we frame this as a multi-task learning problem with two objectives: (1) $\mathcal{L}_r$ for predicting student performance, and (2) $\mathcal{L}_d$ for predicting material type they will choose. Then, we apply a Pareto learning optimization algorithm, specifically Pareto MTL [18], to learn the model and solve this multi-objective problem, thereby finding well-representative solutions for both tasks. An overview of this framework is presented in Figure 1. We first briefly introduce Pareto-Based Multi-Objective Learning.

3.1 Pareto-Based Multi-Objective Learning

Assuming a series of m correlated tasks, characterized by a loss vector: $\min_\theta \mathcal{L}(\theta) = (\mathcal{L}_1(\theta), \mathcal{L}_2(\theta), ..., \mathcal{L}_m(\theta))^\top$, with $\mathcal{L}_i(\theta)$ representing the objective function associated with the i-th task [36]. This scenario represents a multi-objective optimization challenge, where it's not possible to simultaneously optimize all objectives to their fullest extent [18]. Instead, Pareto learning aims to identify a collection of Pareto optimal solutions. These solutions offer a range of optimal trade-offs between the various objectives, as described in [18, 36]: **Pareto dominance**: Let θ^a and θ^b be two solutions, θ^a is said to dominate θ^b ($\theta^a \prec \theta^b$) if and only if $\mathcal{L}_i(\theta^a) \leq \mathcal{L}_i(\theta^b)$, $\forall i \in \{1, ..., m\}$ and $\mathcal{L}_j(\theta^a) < \mathcal{L}_j(\theta^b)$, $\exists j \in \{1, ..., m\}$. **Pareto optimality**: θ^* is a Pareto optimal point and $\mathcal{L}(\theta^*)$ is a Pareto optimal objective vector if it does not exist $\hat{\theta} \prec \theta^*$. The set of all Pareto optimal points is called the Pareto set [18].

3.2 Pareto MTL for Student Knowledge and Behavior Modeling

Suppose there is a model that learns hidden student knowledge and behavioral states based on the historical sequence of learning activities $\{\langle i_1, d_1 \rangle, \ldots, \langle i_t, d_{t-1} \rangle\}$. Predictions for future student performance, p_t, and learning material type, y_t, are calculated using the learned knowledge and behavioral state at time step $t - 1$. The two objective functions for student performance, $\mathcal{L}_r$, and material type prediction, $\mathcal{L}_d$, are determined using a summed binary cross-entropy loss for each time step t, as follows:

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

$$\mathcal{L}_d = -\sum_t (d_t \log y_t + (1 - d_t) \log(1 - y_t)) \tag{2}$$

Here, r_t and d_t represent the actual student response and the type of learning material the student interacts with at time t, respectively. This dual-objective problem could be initially addressed by minimizing a combination of $\mathcal{L}_r$ and $\mathcal{L}_d$, setting a trade-off to balance between the contributions of the student performance objective and the activity-type objective. However, determining how to effectively combine the student performance objective and the activity type objective and establish a proper trade-off among them is a challenging issue [18], and it is time-consuming to experiment with various trade-off values.

Recent developments have introduced strategies to address the multi-objective optimization problem by identifying a single Pareto optimal solution [9, 11, 12, 28, 36]. However, one multi-objective problem can have many optimal trade-offs among its tasks, potentially infinite, and the single solution obtained by this method might not always meet the needs of multi-objective problem practitioners. The Pareto MTL algorithm [18], is designed to identify a collection of representative Pareto optimal solutions at the same time, each offering a different trade-off among tasks. We adopt the Pareto MTL algorithm for training our model, which allows us to effectively solve our dual-objective problem of both student performance and material type tasks. As illustrated in Figure 2, the algorithm employs a series of dividing vectors $\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_m$ to decompose our dual-objective optimization challenge into multiple constrained sub-problems. Each sub-problem represents a trade-off preference, and these sub-problems are solved concurrently. To this end, we obtain a set of well-representative Pareto solutions for the dual-objective problem of predicting student performance and material type, enabling us to select our preferred solution(s) from the set of Pareto optimal solutions.

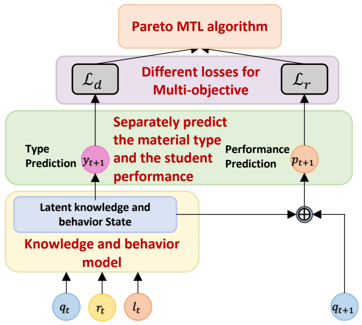

Figure 1: The framework of modeling student knowledge and behavior as dual-objective problem and learning via Perato MLT.

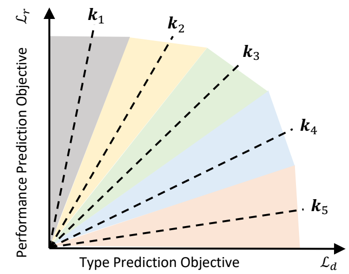

Figure 2: The illustration of Pareto MTL, it finds a set of Pareto solutions through a series of unit dividing vectors k_s.

3.3 Knowledge Model

We utilize multi-activity student knowledge models that handle various types of learning materials to simultaneously model student knowledge and behavior. We implement a Pareto multi-objective knowledge and behavior modeling framework on two models: TAMKOT [35] and GMKT [33]. These are two transition-aware student knowledge models that both learn different knowledge transfer matrices to model the transfer of knowledge across different types of learning activities. Binary indicators are used to represent permutations of transitions between material types from $t-1$ to t, and are used to determine which transfer matrix should be activated for update knowledge. We briefly introduce the key concepts of these models; for the more detailed models, please refer to their original papers in [33, 35].

3.3.1 ***TAMKOT***. The knowledge modeling component of TAMKOT is built upon LSTM [14], where the student's latent knowledge is represented by the LSTM's hidden state, $\boldsymbol{h}_t$. Different transfer matrices are applied to transmit information from $\boldsymbol{h}_{t-1}$ to $\boldsymbol{h}_t$ for each gate and cell of the LSTM. Subsequently, we propose that $\boldsymbol{h}_t$ is used to predict student future performance and material type through two distinct MLPs, as follows:

$$p_{t+1} = \sigma(\boldsymbol{W}_p^T[\boldsymbol{h_t} \oplus \boldsymbol{q_{t+1}}] + b_p) \tag{3}$$

$$y_{t+1} = \sigma(\boldsymbol{W}_y^T\boldsymbol{h_t} + b_y) \tag{4}$$

3.3.2 ***GMKT***. GMKT is designed with a knowledge transfer layer based on memory-augmented neural networks (MANN [31]). It utilizes a static key matrix $\boldsymbol{M}^k$ to represent N latent concept features and a dynamic value matrix $\boldsymbol{M}_t^v$ to track the student's mastery state in concepts. The *erase-followed-by-add* mechanism updates the memory value matrix $\boldsymbol{M}_t^v$. This process involves erasing previous redundant information before adding new information to $\boldsymbol{M}_t^v$, based on different knowledge transfer matrices. A read content $\boldsymbol{c}_{t+1}$ is then retrieved to summarize the student's knowledge state for question q_{t+1}. This summary is obtained using the weighted sum of all memory slots in the value matrix $\boldsymbol{M}_t^v$, calculated using an attention weight vector $\boldsymbol{w}_{t+1}$ that determines the correlation between question q_{t+1} and each of the N latent concepts from $\boldsymbol{M}^k$. Additionally, another read content for learning material type $\boldsymbol{c}_t^o$ summarizes a student's behavior state of material type at each time t, using an attention weight vector $\boldsymbol{w}_t^o$ calculated from $\boldsymbol{M}^k$. The predictions for student performance and material type are then calculated through two distinct MLPs, as follows:

$$p_{t+1} = \sigma(\boldsymbol{W}_p^{\mathrm{T}}(Tanh(\boldsymbol{W}_f^{\mathrm{T}}[\boldsymbol{c}_{t+1} \oplus \boldsymbol{q}_{t+1}] + \boldsymbol{b}_f) + b_p) \tag{5}$$

$$y_{t+1} = \sigma(d_t \cdot \boldsymbol{W}_y^T\boldsymbol{c}_t^o + (1-d_t) \cdot \boldsymbol{W}_y^T\boldsymbol{c}_t^o + b_y) \tag{6}$$

4 EXPERIMENTS

We evaluate the effectiveness of multi-objective behavior and knowledge modeling using a Pareto multi-task algorithm through two sets of experiments. First, we evaluate the predictive capabilities of Perato-TAMKOT and Perato-GMKT in comparison to baseline knowledge modeling methods in terms of student performance Second, we examine how well Perato-TAMKOT and Perato-GMKT predict the types of learning materials students will choose to interact with. Our code and sample data are available at GitHub [1].

4.1 Dataset

We use the **EdNet**[2] [7] dataset to perform the experiments. This dataset is sourced from a multi-platform AI tutoring service named Santa[3], which is designed to help Korean students prepare for the TOEIC[4] English testing. For our research, we use a preprocessed version of the dataset, which utilized questions (assessed) and their associated explanations (non-assessed) as the two learning material types, as detailed in previous studies [33, 35]. To summarize, the dataset includes data from a total of 1,000 students, encompassing 11,249 questions and 8,324 question explanations, along with 200,931 question solving activities and 150,821 explanation review activities.

[1]https://github.com/persai-lab/2024-UMAP-Pareto-GMKT-Pareto-TAMKOT
[2]https://github.com/riiid/ednet
[3]https://www.aitutorsanta.com/
[4]https://www.ets.org/toeic

4.2 Baselines

4.2.1 Performance Prediction Baselines. We compare Pareto-GMKT and Pareto-TAMKOT with six baseline knowledge modeling methods to examine their effectiveness in predicting student performance. This evaluation includes two supervised knowledge modeling methods that focus solely on assessed activities and two multi-activity knowledge modeling methods. Additionally, we extend the two assessed-only supervised modeling methods to also include non-assessed activities. These modified models are denoted by adding '+M' to their original names.

The assessed-only baselines are as follows: **DKT** [24] is the first deep learning method in modeling methods, utilizing recurrent neural networks (RNN) to trace students' knowledge acquisition over time. **AKT** [13] is an attention-based method, that adopts a context-aware strategy using a monotonic attention mechanism to prioritize past student performances that are relevant to the current question. The following are multi-activity baseline models: **DKT+M** [32] is an extension of DKT that incorporates both assessed and non-assessed learning activities. It enhances the original DKT by appending embeddings of non-assessed materials encountered between two assessed activities as additional input features, alongside the original question embeddings. **AKT+M** adapts the AKT framework by incorporating embeddings of non-assessed materials encountered between two assessed activities as an additional feature. It also includes position encoding for each learning material's embedding to enhance the model's context awareness [6]. **TAMKOT** [35] is a transition-aware model that utilizes LSTM [14] technology. It stands out by employing multiple knowledge transfer matrices, which explicitly model the transfer of knowledge across various types of learning activities. **GMKT** [33] is another transition-aware method, that leverages MANN technology and incorporates a Graph Neural Network (GNN) to enhance the modeling of student knowledge through non-assessed learning activities. It is the only existing method that also has an objective on the selection of learning material types, but it does not utilize Pareto learning.

4.2.2 Type Prediction Baselines. To assess the effectiveness of Pareto-TAMKOT and Pareto-GMKT in predicting the types of learning materials, we conduct experiments to compare Pareto-TAMKOT and Pareto-GMKT against four deep sequential baseline models. To facilitate the comparison, we employ learning material embeddings along with the material type as inputs to these baselines and focus on predicting only the upcoming type of material.

The baselines are as follows: **LSTM** [14] is a type of recurrent neural network architecture known for its proficiency in learning long-term dependencies. Its design is particularly effective for tasks that require an understanding of entire data sequences. **MANN** [27] augments neural networks with an external memory component, which facilitates the storage and retrieval of information over long sequences. Such a feature is highly beneficial for tasks that necessitate sustained information retention and manipulation. Additionally, variants of two multi-activity knowledge modeling methods are employed: **TAMOKT** and **GMKT**. For these two baselines, we retain the architecture of modeling knowledge and apply a Multilayer Perceptron (MLP) to the learned hidden behavior states, specifically for only predicting the type of learning material. These two methods bypass the objectives of student performance prediction.

4.3 Experiments Setup

We employ a 5-fold student-stratified cross-validation approach to split the training, testing, and validation datasets [33, 35]. Sequences from 80% of the students constitute the training set, while those from the remaining 20% are used for testing. Additionally, 20% of the students from the training set are allocated as a validation set for hyperparameter tuning. We use five evenly distributed dividing vectors $\{(cos(\frac{k\pi}{10}, sin(\frac{k\pi}{10}))|k = 0, 1, ..., 5\}$ for Pareto MTL optimization. To avoid the potential issue of exploding gradients, we employ the norm clipping. We ensure uniform sequence lengths by truncating or padding them as necessary [24, 30, 33, 35]. The length of these sequences, denoted as L_s, is considered as another hyperparameter and is tuned using the validation set. Sequences longer than L_s are truncated into multiple sequences, while those shorter than L_s are extended using padding with 0s. A coarse-grained grid search is conducted to determine the best hyperparameters.

4.4 Prediction Performance Comparison

Since our experiments with the Ednet dataset involve two types of learning materials, we employ the Area Under the Curve (AUC) metric to evaluate the effectiveness of each model in predicting learning material type. Additionally, since student responses to questions are binary (success or failure), we also use the AUC as the metric to assess the effectiveness of each model in predicting student response. A higher AUC value indicates greater predictive performance. To ensure fair comparisons among different methods, we present the average results across five folds, complete with their confidence intervals, at a significance level of 0.05 for each model. The results of our experiments on student performance predictions and material type predictions are presented in Table 1 and Table 2, respectively.

As mentioned in Section 4.3, for our experiments, we employed five evenly distributed dividing vectors in the Pareto MTL algorithm to identify a well-distributed set of Pareto solutions for our dual-objectives problem. Our experiments showed that both Pareto-TAMKOT and Pareto-GMKT models achieved improvements in predictions for both student performance and material type when the dividing vector was set to $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, which corresponds to the direction of $\frac{\pi}{4}$, as illustrated by the middle vector $\boldsymbol{k}_3$ in Figure 2. Conversely, the optimal prediction performance for each specific task was consistently achieved using the corresponding extreme dividing vectors, such as $(0, 1)$ or $(1, 0)$. Under these settings, the improvement in one task was substantial, while the other task often experienced very limited improvement or even a negative impact. Moreover, altering the dividing vector to other values typically resulted in significant improvements in student performance predictions but only slight or limited enhancements for material type predictions, and vice versa. This underscored the importance of selecting an appropriate dividing vector to achieve balanced performance across both objectives. Due to space limitations, we report results using the $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ dividing vector exclusively in Table 1 and Table 2, which ensures a meaningful comparison that achieves

Table 1: Student performance prediction results (AUC). The best and second-best result 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	0.6786 ± 0.0063
GMKT	<u>0.6819 ± 0.0070</u>
Pareto-TAMKOT	0.6809 ± 0.0063
Pareto-GMKT	**0.6853 ± 0.0071**

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

Methods	AUC
LSTM	0.8768 ± 0.0041
MANN	0.8933 ± 0.0030
TAMKOT	0.8929 ± 0.0042
GMKT	0.8932 ± 0.0046
Pareto-TAMKOT	<u>0.8987 ± 0.0042</u>
Pareto-GMKT	**0.8992 ± 0.0066**

two our objectives that enhance the prediction performance of both student response and material type.

4.4.1 Student Performance Prediction. First, we can see that Pareto-TAMKOT and Pareto-GMKT outperform their counterparts, TAMKOT and GMKT, which do not utilize Pareto MTL, respectively. This demonstrated that incorporating the objective of material type prediction and utilizing Pareto MTL can enhance performance in predicting student responses. However, it was observed that Pareto-TAMKOT does not outperform GMKT, and Pareto-GMKT greatly outperforms Pareto-TAMKOT. These two observations suggested that while applying the Pareto MTL algorithm can improve model optimization for our multi-objective problem of student performance and material type predictions, the inherent strength of the model itself is crucial for accurately learning student knowledge and behavioral preferences. Additionally, the superior performance of GMKT supports the idea that focusing on both student response and learning material type predictions can facilitate student knowledge modeling and improve predictions of student responses. In summary, the strong performance of Pareto-TAMKOT and Pareto-GMKT demonstrated that simultaneously modeling students' knowledge and behaviors, and formulating it as a multi-task learning problem with multiple objectives, optimized using Pareto-MTL, can further enhance the modeling of student knowledge.

4.4.2 Material Type Prediction. Similarly, we observed that Pareto-TAMKOT and Pareto-GMKT outperform all baseline methods in the material type prediction task. Specifically, Pareto-TAMKOT and Pareto-GMKT showed superior performance compared to TAMKOT and GMKT, respectively. This underscored the effectiveness of formulating a multi-objective problem, optimized with the Pareto-MTL algorithm, which improves our understanding of students' learning material behavior preferences. However, when comparing Pareto-TAMKOT to Pareto-GMKT, it is evident that the improvement with Pareto-GMKT is slight. We hypothesized that this is due to the already high performance of all baseline models in the learning material type prediction task, which poses a challenge for significant enhancements; thus, the model itself should have more strength in modeling behavior. The modest improvement of GMKT compared to TAMKOT also supports this observation. Nonetheless, both Pareto-TAMKOT and Pareto-GMKT still managed to enhance material type prediction performance. This again demonstrated that simultaneously modeling students' knowledge and behaviors, and formulating it as a multi-task learning problem with multiple objectives, optimized using Pareto-MTL, can further enhance the modeling of student behavior.

Overall, our results from both the Pareto-TAMKOT and Pareto-GMKT for both student performance and material type prediction, demonstrated that simultaneously modeling student knowledge and tracking their material selection behaviors leads to a deeper mutual understanding of these aspects, ultimately benefiting learning in each task. Consequently, it was evident that framing student performance and learning material type prediction as a multi-objective problem is essential to enhance both tasks. Moreover, applying the Pareto-MTL optimization algorithm proves effective in identifying the optimal solutions for these two tasks. In summary, our approach to addressing the multi-objectives of student performance and material type prediction through Pareto-MTL is crucial for accurately capturing both student knowledge and behaviors related to learning material selection, thereby improving predictions of student performance and material preferences.

5 CONCLUSIONS

We addressed the overlooked relationship between students' knowledge and behavioral preferences by introducing a novel multi-task learning framework with multiple objectives. Utilizing the Pareto MTL algorithm, we applied this framework to two enhanced multi-activity knowledge modeling methods, TAMKOT and GMKT, termed Pareto-TAMKOT and Pareto-GMKT. Our approach outperformed existing models in predicting both student performance and material preferences. This demonstrated the benefits of treating the modeling of student knowledge and behavior as a multi-task learning problem and effectively tackles this multi-objective challenge through the application of Pareto MTL.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Solmaz Abdi, Hassan Khosravi, Shazia Sadiq, and Ali Darvishi. 2021. Open Learner Models for Multi-Activity Educational Systems. In *International Conference on Artificial Intelligence in Education*. Springer, 11–17.

[2] Rakesh Agrawal, Maria Christoforaki, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, and Adith Swaminathan. 2014. Mining videos from the web

for electronic textbooks. In *International Conference on Formal Concept Analysis*. Springer, 219–234.
[3] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, and Krishnaram Kenthapadi. 2011. Enriching textbooks with images. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. 1847–1856.
[4] Hao Cen, Kenneth Koedinger, and Brian Junker. 2006. Learning factors analysis–a general method for cognitive model evaluation and improvement. In *International conference on intelligent tutoring systems*. Springer, 164–175.
[5] Hao Cen, Kenneth Koedinger, and Brian Junker. 2008. Comparing two IRT models for conjunctive skills. In *International Conference on Intelligent Tutoring Systems*. Springer, 796–798.
[6] Youngduck Choi, Youngnam Lee, Junghyun Cho, Jineon Baek, Byungsoo Kim, Yeongmin Cha, Dongmin Shin, Chan Bae, and Jaewe Heo. 2020. Towards an appropriate query, key, and value computation for knowledge tracing. In *Proceedings of the 7th ACM Conference on Learning at Scale*. ACM, New York, NY, USA, 341–344.
[7] Youngduck Choi, Youngnam Lee, Dongmin Shin, Junghyun Cho, Seoyon Park, Seewoo Lee, Jineon Baek, Chan Bae, Byungsoo Kim, and Jaewe Heo. 2020. Ednet: A large-scale hierarchical dataset in education. In *International Conference on Artificial Intelligence in Education*. Springer, 69–73.
[8] 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.
[9] Jean-Antoine Désidéri. 2012. Mutiple-gradient descent algorithm for multiobjective optimization. In *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012)*.
[10] Fritz Drasgow and Charles L Hulin. 1990. Item Response Theory. *Handbook of Industrial and Organizational Psychology* (1990), 577–636.
[11] Jörg Fliege and Benar Fux Svaiter. 2000. Steepest descent methods for multicriteria optimization. *Mathematical methods of operations research* 51 (2000), 479–494.
[12] Jorg Fliege and A Ismael F Vaz. 2016. A method for constrained multiobjective optimization based on SQP techniques. *SIAM Journal on Optimization* 26, 4 (2016), 2091–2119.
[13] Aritra Ghosh, Neil Heffernan, and Andrew S Lan. 2020. Context-Aware Attentive Knowledge Tracing. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, New York, NY, USA, 2330–2339.
[14] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
[15] Xinying Hou, Paulo F Carvalho, and Kenneth R Koedinger. 2021. Drinking Our Own Champagne: Analyzing the Impact of Learning-by-doing Resources in an E-learning Course. In *The 11th Int. Conference on Learning Analytics & Knowledge*.
[16] Michael I Jordan, Michael J Kearns, and Sara A Solla. 1998. *Advances in Neural Information Processing Systems 10: Proceedings of the 1997 Conference*. Vol. 10. Mit Press.
[17] Andrew S Lan, Andrew E Waters, Christoph Studer, and Richard G Baraniuk. 2014. Sparse factor analysis for learning and content analytics. *Journal of Machine Learning Research (JMLR)* 15, 57 (2014), 1959–2008.
[18] Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qing-Fu Zhang, and Sam Kwong. 2019. Pareto multi-task learning. *Advances in neural information processing systems* 32 (2019).
[19] Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su, and Guoping Hu. 2019. Ekt: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering* 33, 1 (2019), 100–115.
[20] Ting Long, Yunfei Liu, Jian Shen, Weinan Zhang, and Yong Yu. 2021. 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*. 173–182.
[21] Amir Shareghi Najar, Antonija Mitrovic, and Bruce M McLaren. 2014. Adaptive support versus alternating worked examples and tutored problems: which leads to better learning?. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 171–182.
[22] Shalini Pandey and George Karypis. 2019. A Self-Attentive Model for Knowledge Tracing. In *Proceedings of the 12th International Conference on Educational Data Mining*. International Educational Data Mining Society, 384–389.
[23] Philip I Pavlik Jr, Hao Cen, and Kenneth R Koedinger. 2009. Performance Factors Analysis–A New Alternative to Knowledge Tracing. *Online Submission* (2009).
[24] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas Guibas, and Jascha Sohl-Dickstein. 2015. Deep Knowledge Tracing. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1* (Montreal, Canada). MIT Press, Cambridge, MA, USA, 505–513.
[25] Cristóbal Romero and Sebastián Ventura. 2010. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 6 (2010), 601–618.
[26] Shaghayegh Sahebi, Yu-Ru Lin, and Peter Brusilovsky. 2016. Tensor Factorization for Student Modeling and Performance Prediction in Unstructured Domain. *International Educational Data Mining Society* (2016).
[27] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. 2016. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*. PMLR, 1842–1850.
[28] Ozan Sener and Vladlen Koltun. 2018. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems* 31 (2018).
[29] Nguyen Thai-Nghe, Tomáš Horváth, and Lars Schmidt-Thieme. 2010. Factorization models for forecasting student performance. In *Educational Data Mining 2011*. Citeseer.
[30] Chunpai Wang, Siqian Zhao, and Shaghayegh Sahebi. 2021. Learning from Non-Assessed Resources: Deep Multi-Type Knowledge Tracing. *International Educational Data Mining Society* (2021).
[31] 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* (Perth, Australia). ACM, New York, NY, USA, 765–774.
[32] Liang Zhang, Xiaolu Xiong, Siyuan Zhao, Anthony Botelho, and Neil T Heffernan. 2017. Incorporating Rich Features into Deep Knowledge Tracing. In *Proceedings of the 4th ACM Conference on Learning at Scale*. ACM, New York, NY, USA, 169–172.
[33] Siqian Zhao and Shaghayegh Sahebi. 2023. Graph-Enhanced Multi-Activity Knowledge Tracing. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 529–546.
[34] Siqian Zhao, Chunpai Wang, and Shaghayegh Sahebi. 2020. Modeling Knowledge Acquisition from Multiple Learning Resource Types. In *Proceedings of The 13th International Conference on Educational Data Mining*. International Educational Data Mining Society, 313–324.
[35] Siqian Zhao, Chunpai Wang, and Shaghayegh Sahebi. 2022. Transition-aware multi-activity knowledge tracing. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 1760–1769.
[36] Eckart Zitzler and Lothar Thiele. 1999. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE transactions on Evolutionary Computation* 3, 4 (1999), 257–271.