

Generative Data Imputation for Sparse Learner Performance Data Using Generative Adversarial Imputation Networks

Liang Zhang , Jionghao Lin , John Sabatini , Diego Zapata-Rivera , Carol Forsyth , Yang Jiang , John Hollander , Xiangen Hu , Arthur C. Graesser

Abstract—As learners engage with Intelligent Tutoring Systems (ITSs) by responding to a series of questions, their performance data, such as correct or incorrect responses, is crucial for assessing and predicting their knowledge states through analysis and modeling. However, data sparsity, often arising from skipped or incomplete responses, poses challenges for accurately assessing learning and delivering personalized instruction. To address this, we propose a generative data imputation method based on Generative Adversarial Imputation Networks (GAIN) to complete missing learning performance data. Our approach employs a three-dimensional (3D) framework structured by learners, questions, and attempts, with an adaptable design along the attempts dimension to manage varying sparsity levels. Enhanced by convolutional neural networks in the input and output layers and optimized with a least squares loss function, our GAIN-based method aligns the input and output shapes with the dimensions of question-attempt matrices across the learners' dimension. Extensive experiments on datasets from three types of ITSs, including AutoTutor Adult Reading Comprehension (ARC), ASSISTments and MATHia, demonstrate that our approach generally outperforms baseline methods, e.g., tensor factorization-based methods and other Generative Adversarial Network (GAN) variants, in imputation accuracy across different setting of maximum attempts. Bayesian Knowledge Tracing (BKT) modeling further validates the imputed data's efficacy by estimating learning parameters, including initial knowledge $P(L_0)$, learning rate $P(T)$, guess rate $P(G)$, and slip rate $P(S)$. Results reveal that the imputed data not only enhances model fit but also closely aligns with the original sparse distributions by capturing underlying learning behaviors, indicating greater reliability in learner assessments. Kullback-Leibler (KL) divergence measurements of all these learning parameters confirm that the imputed data effectively preserve essential learning characteristics, maintaining low divergence.

This paper is currently Under Review for IEEE Transactions on Learning Technologies Journal. Manuscript submitted on March. 14, 2024.

Liang Zhang, John Sabatini, and Arthur C. Graesser are with the Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152, USA (e-mail: lzhang13@memphis.edu, jpsbtini@memphis.edu, art.graesser@gmail.com).

Jionghao Lin is with the Faculty of Education, The University of Hong Kong, Hong Kong, PR China, also with the the Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA, and also with the Centre for Learning Analytics, Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia (e-mail: jionghao@hku.hk).

Diego Zapata-Rivera, Carol Forsyth and Yang Jiang are with Educational Testing Service, Princeton, NJ 08540, USA (e-mail: dzapata@ets.org, cforsyth@ets.org, yjiang002@ets.org).

John Hollander is with Department of Psychology and Counseling, Arkansas State University, Jonesboro, AR 72701 (e-mail: jhollander@astate.edu).

Xiangen Hu is with the Department of Applied Social Sciences, Hong Kong Polytechnic University, Hong Kong, PR China (e-mail: xiangen.hu@polyu.edu.hk).

across parameters for most datasets and attempts. These findings highlight GAIN's potential as a reliable tool for data imputation in ITSs, offering an effective solution for mitigating data sparsity issues and supporting adaptive, individualized instruction in AI-driven education. The improvements in learner modeling accuracy facilitated by GAIN-based imputation ultimately pave the way for more accurate, responsive ITS applications capable of fostering improved learning outcomes.

Index Terms—Learning Performance Data, Data Sparsity, Data Imputation, Generative Adversarial Imputation Network, Intelligent Tutoring System

I. INTRODUCTION

ADVANCEMENTS in AI-driven technologies have significantly enhanced modern education through personalized tutoring and adaptive learning strategies on online platforms [1], [2]. Intelligent Tutoring Systems (ITSs) exemplify this progress by leveraging advanced machine learning and natural language processing models to create interactive learning environments that improve outcomes across domains like literacy [3], mathematics [4], language learning [5], biology [6] and other STEM fields [7]. As human learners interact with ITSs, often through question-and-answer scenarios with immediate responses, their performance data becomes crucial for learner modeling, enabling systems to track progress, predict future performance, and adapt instruction accordingly [8]. Learner models like Bayesian Knowledge Tracing (BKT) and other knowledge tracing variants utilize the learner performance data to uncover learning characteristics, estimate knowledge states and acquisition [9]. However, in real-world scenarios, missing learner performance data is prevalent due to factors, such as learner dropout or disengagement [10], technical issues or incomplete data logging [11], biased sampling within experimental groups [12], and more. These challenges often lead to sparse data, where items (i.e., questions or problems) remain unattempted (e.g., learners may bypass the question, leave it unanswered due to a lack of response initiation, or make no attempt to engage with it), alongside limited learner interactions [13], [14]. As shown in Figure 1, missing performance records can occur along both the attempt and question dimensions during learner-ITS interactions. In the right portion of the figure's two matrices, entries marked with “?” indicate missing data. This data sparsity may arise either randomly or non-randomly over the course of learner-ITS engagement. These data gaps hinder the effectiveness of learner modeling

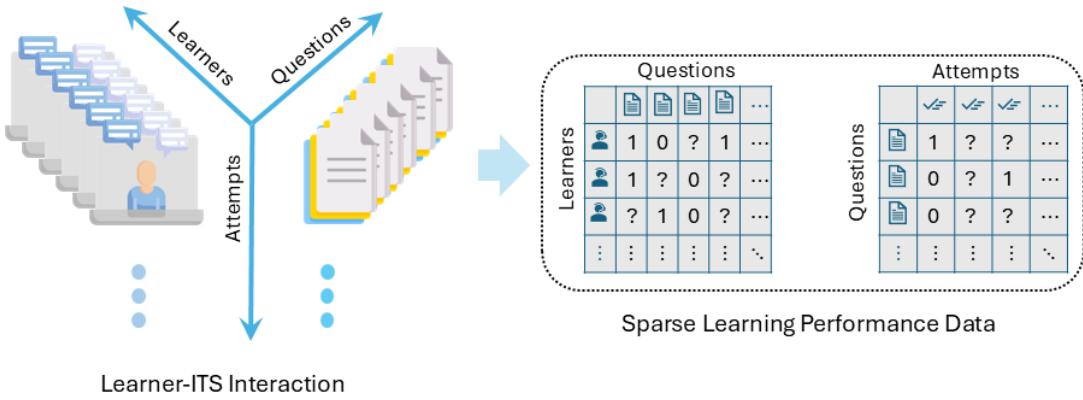


Fig. 1: An illustration for sparse learning performance data in Intelligent Tutoring Systems.

and the adaptive delivery of tailored instruction, ultimately reducing the accuracy and specificity of support provided by ITSs. For instance, sparse learning performance data can lead to biased or overfitted learner models that fail to capture a learner's trajectory or yield misleading predictions about future performance [13]–[16]. Such misrepresentations can have adverse effects, particularly when these models inform instructional decisions or guide personalized learning paths in ITSs. To address these challenges, this study focuses on data imputation techniques for sparse learner's performance data, aiming to enhance the robustness and adaptability of ITSs in delivering personalized learning experiences within AI education.

Various methods have been proposed to address sparse learner performance data. One traditional method, excluding incomplete observations as outliers, may seem straightforward in educational or psychological contexts [11]. However, this method risks discarding valuable information, exacerbating data sparsity, and undermining dataset validity. Alternatively, reverting to real-world experiments can be time-consuming, labor-intensive, and hard to replicate at scale. In light of these challenges, computational data imputation methods have gained prominence, particularly in AI research. Grounded in Rubin's foundational principles [17], these computational imputation methods, such as indicator or mean imputation [18], regression imputation [19], multiple imputation [20] fill missing values based on observed data patterns. While cost-effective and well-established, these methods often simplify the complexities inherent in missing data, risking bias in the resulting models [18], [19], [21], [22]. For instance, indicator or mean imputation can introduce bias by oversimplifying the intricacies of missing data [18], [21], regression imputation frequently falls short of reflecting the complete range of the underlying data structure [19], and multiple imputation struggles with high-dimensional correlations [22]. In the context of human learning data, these challenges are further compounded by its multidimensional nature. Newell and Simon's [23] landmark work conceptualized human learning within a three-dimensional space, comprising the task dimension (different task environments), the performance-learning-development dimension (linking activities to timescales), and the individual-difference dimension (accounting for learner variability), cap-

turing the dynamic and evolving nature of human cognition. Learner responses and problem-solving attempts are sequential, with past performance influencing future actions [24], variation in learner attempts [25], and interdependencies among different knowledge components [26]. These dynamic shifts in knowledge states [9], and the multidimensional cross-effects arise from interactions between learners, tasks, and sequential attempts (e.g., a learner's improved performance on a foundational question enhancing their ability to tackle subsequent, more complex questions, or repeated attempts on a specific task leading to knowledge reinforcement that generalizes to related tasks) [27], introduce temporal and contextual complexities that traditional computational imputation methods struggle to address. Effectively handling sparse learning performance data in ITS contexts remains a significant challenge due to these complexities.

Generative Adversarial Networks (GANs) have demonstrated impressive capabilities in initial image generation [28], as well as in speech and voice recognition [29], multimodal conversations [30], and beyond. By pairing a generator that produces synthetic samples with a discriminator that evaluates their authenticity, GANs leverage game-theoretic principles to learn complex data patterns and distributions, thereby enabling highly effective synthetic data generation [31]. Building on these strengths, GANs have been adapted for data imputation tasks, including image inpainting for 2D images [32], 3D surfaces [33], and time series data [34]. One prominent model, Generative Adversarial Imputation Nets (GAIN), enhances GAN-based imputation by conditioning the generator on observed data and introducing a hint mechanism to help the discriminator detect missing patterns [35], [36], outperforming traditional methods such as MICE (Multiple Imputation by Chained Equations, a statistical method that iteratively models each variable with missing values using regression models) and missForest (a machine learning method that employs random forests to iteratively predict and impute missing data) for clinical datasets [35], [37]. Moreover, GAIN's context-awareness aligns with Rubin's Rules for valid imputations, generating plausible outputs that blend seamlessly with neighboring regions [32]. Despite its success elsewhere, GAIN's potential to impute missing data in sparse learning performance datasets within ITSs remains unexplored. Learn-

ing performance data present unique challenges due to their multidimensional nature, encompassing individual learners, problem-solving attempts, and dynamic interactions across multiple tasks, making conventional imputation methods insufficient. A promising strategy involves adapting GAIN to three-dimensional representations of learners, items (e.g., questions), and temporal factors (e.g., time or attempts) [27], [38], inspired by tensor-based methods in learning engineering [39]–[41]. Such a 3D approach not only captures the multifaceted nature of human learning but also allows GAIN to manage complex data structures, thereby generating more accurate imputed values [27], [38]. Accordingly, this study investigates how GAIN can be refined for imputation in sparse learning performance datasets, optimizing its applicability to educational contexts and potentially improving overall imputation accuracy in ITSs. We are guided by the following two **Research Questions**:

- **RQ1:** How effectively does the GAIN-based method impute sparse learning performance data in ITSs compared to established baselines?
 - **RQ1.1:** How does GAIN perform across different attempts, reflecting varying levels of data sparsity?
 - **RQ1.2:** What is the impact of GAIN on the accuracy of imputed data across various ITS datasets?
- **RQ2:** How do the imputed data align with the original sparse learning performance in ITSs?
 - **RQ2.1:** To what extent does data imputation affect learner modeling?
 - **RQ2.2:** How well do the imputed learning features align with those of the original sparse data?

The impacts of our study are twofold and extend further. First, we aim to advance the accuracy of data imputation in ITSs by applying GAIN within a 3D framework, effectively capturing the multidimensional complexities of learning behaviors that traditional methods often overlook. This advancement will significantly improve learner modeling, predictive analytics, and the ability to provide targeted, data-driven interventions in ITSs. Second, by addressing data sparsity through generative imputation, the study will enhance the adaptive capabilities of ITSs, enabling these systems to deliver more personalized and effective learning experiences for human learners. Beyond these immediate benefits, the integration of generative imputation methods is expected to establish a foundation for scalable and robust AI-driven educational tools, advancing progress tracking, assessment accuracy, and adaptive instructional designs in diverse educational contexts. Our code and results can be found at the following GitHub link: <https://github.com/LiangZhang2017/GenerativeDataImputation>.

II. RELATED WORK

This section reviews related work on data imputation methods in ITSs, identifying key challenges and limitations while also highlighting potential opportunities for advancement.

A. AI-driven Imputation for Sparse Learning Performance in ITSs

Addressing data sparsity in ITSs has become a critical focus in AI-driven education, with various studies exploring imputation techniques to handle incomplete learning performance data. AI-based methods, such as deep learning frameworks, attention mechanisms, and other machine learning approaches, have played a crucial role in modeling the learning process and mitigating the effects of sparse data. For instance, Chen et al. [42] utilized prerequisite concept maps to model logical relationships between knowledge concepts, enhancing knowledge state prediction in sparse data scenarios. Lu et al. [43] extended this by incorporating ordering pairs into concept maps. Pandey et al. [13] employed self-attention mechanisms to predict learner performance by weighting relevant prior answers. Wang et al. [15] integrated question-concept hierarchies into a deep learning framework to better model learner interactions despite data sparsity. However, challenges remain, including the labor-intensive mapping of knowledge concepts [44], limited consideration of temporal learning dynamics [40], and disruption of sequential learning effects critical to knowledge organization [45]. These gaps highlight the need for more robust approaches to addressing data sparsity in ITSs.

B. Tensor-based Imputation for Complex Learning Data

As sparse learner-performance data become increasingly common within ITSs, the need for robust imputation methods has become essential. Tensor-based imputation has emerged as a prominent technique due to its ability to preserve the multidimensional structure of learning data and maintain intrinsic relationships across dimensions such as learners, questions, time or attempts [25], [41]. Tensor-based approaches evolved from two-dimensional matrix factorization techniques initially applied in recommendation systems for adaptive navigation in online learning [46], [47]. Thai-Nghe et al. [48] extended this to three-dimensional tensor factorization, incorporating temporal effects to predict learner performance across dimensions like learners, time, and steps. They further demonstrated its potential for imputing unobserved performance in sparse datasets [40]. Sahebi et al. [41] introduced Feedback-Driven Tensor Factorization (FDTF), integrating sequences of students, quizzes, and attempts, improving knowledge representation and prediction. Later, Doan and Sahebi [49] proposed Ranked-Based Tensor Factorization (RBTF), accounting for concept forgetting and biases to promote positive learning trajectories. Zhao et al. [26] advanced Multi-View Knowledge Modeling (MVKM), applying tensor factorization to analyze multiple materials within a shared latent space while addressing forgetting through rank-based constraints. These and other methods [50]–[53] have enhanced predictive accuracy and addressed data sparsity in educational applications.

These advances highlight the effectiveness of tensor-based representations in predicting and imputing missing data within sparse tensor spaces, preserving the sequential nature of learning events [48], [54]. This approach is crucial for accurately tracing and predicting learner performance, as well as enabling efficient decomposition of interactions across

multiple dimensions [49], [55]. Aligned with Rubin's Rules on interdependence in data imputation and proven effective in recommendation systems, tensor-based techniques can uncover performance similarities and dependencies among learning events. As a result, they hold significant promise for improving imputation in educational data mining, especially when dealing with sparse datasets in ITSs.

C. Generative Data Imputation for Sparse Learning Performance

Recent advances in generative data imputation, utilizing reconstruction mechanisms based on existing data, have achieved significant success in reducing data sparsity and outperforming traditional methods [35], [37]. Morales-Alvarez et al. [56] integrated structured latent spaces with graph neural networks, outperforming baselines such as MICE and missForest on real-world mathematics data from Eedi¹ (an online education platform offering personalized math learning through diagnostic assessments and tailored study plans). Ma et al. [57] used deep generative models to impute multiple-choice question data, handling over 70% missing rates in also Eedi's mathematics dataset. Zhang et al. [38] explored GAN and GPT for data augmentation in adult reading comprehension. Ongoing research continues to uncover further applications of generative imputation in learning engineering and science.

These developments set the stage for applying more advanced models like GAIN, which preserve the multidimensional structure of learning data under a tensor format. By capturing complex relationships across dimensions, integrating GAIN within a tensor-based representation could significantly improve imputation [58]. Inspired by these insights, our study adapts GAIN within a tensor framework to facilitate more effective imputation of sparse learning performance data in ITSs.

III. METHODS

This section presents our proposed data imputation method, specifically designed to address sparse learner performance data by leveraging the hierarchical relationships among learners, questions, and attempts. The framework structures the learning performance data into a 3D tensor where entries represent binary performance outcomes (correct or incorrect). Building on this 3D tensor representation, the method employs GAIN with Convolutional Neural Networks (CNNs) to effectively fill in missing values while preserving multidimensional learning dynamics and improving imputation accuracy across varying levels of data sparsity.

A. 3D Tensor Representation of Sparse Learning Performance Data

Consider an intelligent learning scenario within ITS, where a set of U learners, represented as $\{l_1, l_2, l_3, \dots, l_U\}$, engage with a sequence of N questions, denoted by $\{q_1, q_2, q_3, \dots, q_N\}$. Each question allows up to M attempts, represented by $\{t_1, t_2, t_3, \dots, t_M\}$, for submitting responses

for answers. As learning progresses, learners' performance evolves dynamically based on the sequence of questions and repeated attempts, while the presence of missing data contributes to the sparsity of performance records. The sparse learning performance data can be represented as a 3D tensor $\mathcal{T}_{sparse} \in [\tau_{uij}]^{U \times N \times M}$ or $[0, 1, NaN]^{U \times N \times M}$, where each element τ_{uij} corresponds to the observed performance of the u^{th} learner on the j^{th} question at the i^{th} attempt. Specifically, τ_{uij} takes the value of 1 for correct answer, 0 for incorrect answer, and NaN for unobserved data at a specific question and attempt.

B. The Proposed GAIN-based Imputation Architecture

Building upon the basic GAN method, we demonstrate how GAIN can be adapted to impute tensor-based learning performance data, transforming \mathcal{T}_{sparse} into \mathcal{T}_{dense} .

Consider the \mathcal{T}_{sparse} , representing the learning performance of all learners. This tensor comprises layers along the learner dimension, represented as $\mathcal{T}_{sparse} = (\mathcal{T}_{l_1}, \mathcal{T}_{l_2}, \dots, \mathcal{T}_{l_n})$. Each layer, akin to a single-channel "learner image", is a matrix that encapsulates performance values across different questions and attempts for an individual learner. This is visualized in Figure 2.

For each matrix-based layer $\mathcal{T}_l \in (\mathcal{T}_{l_1}, \mathcal{T}_{l_2}, \dots, \mathcal{T}_{l_n})$, each entry τ_{lij} in the $N \times M$ matrix may include the performance values of 0, 1 or NaN to present the observed data and unobserved data, respectively. One mask matrix $\mathcal{T}_{l_{mask}}$ is supposed to map the observed and unobserved entries within the matrix \mathcal{T}_l , with 1 signifying observed data, and 0 indicates unobserved data. One noise matrix \mathcal{Z} with dimensions matching \mathcal{T}_l , is initialized. These matrices collectively function as inputs to the generator in the GAIN architecture, producing the output $\mathcal{T}_{IG} = G(\mathcal{T}_l, \mathcal{T}_{l_{mask}}, (1 - \mathcal{T}_{l_{mask}}) \odot \mathcal{Z})$ [35]. Here, the \odot denotes as Hadamard product, indicating element-wise multiplication. The imputed matrix $\mathcal{T}_{l_{imputed}} = \mathcal{T}_{l_{mask}} \odot \mathcal{T}_l + (1 - \mathcal{T}_{l_{mask}}) \odot \mathcal{T}_{IG}$, effectively merges observed and generated data to fill in unobserved entries. Particularly, a hint matrix $\mathcal{T}_{l_{hint}}$, also matching the dimensions of \mathcal{T}_l and derived from the mask matrix $\mathcal{T}_{l_{mask}}$, is introduced. It employs a hint rate to specify the conditional probability that a specific entry in $\mathcal{T}_{l_{imputed}}$ can be observed, given both $\mathcal{T}_{l_{imputed}}$ and $\mathcal{T}_{l_{hint}}$. Thereby, the discriminator within the GAIN architecture, formulated as $D(\mathcal{T}_{l_{imputed}}, \mathcal{T}_{l_{hint}})$, evaluate this as probability [35]. We train $D(\cdot)$ to maximize the probability of correctly predicting the $\mathcal{T}_{l_{mask}}$, while the $G(\cdot)$ is trained to minimize the likelihood of $D(\cdot)$ correctly predicting $\mathcal{T}_{l_{mask}}$. So, we introduce the objective function $V(D, G)$ [35]:

$$V(D, G) = E \left[\mathcal{T}_{l_{mask}}^T \log D(\mathcal{T}_{l_{imputed}}, \mathcal{T}_{l_{hint}}) + (1 - \mathcal{T}_{l_{mask}})^T \log(1 - D(\mathcal{T}_{l_{imputed}}, \mathcal{T}_{l_{hint}})) \right] \quad (1)$$

Our proposed imputation architecture incorporates several novel modifications and configurations from the initial GAIN architecture [35]. See below for further details:

- The convolutional layers are employed for both the generator and discriminator, diverging from the original architecture's reliance on dense layers. Five convolutional neural network (CNN) layers [59], excluding the input

¹Eedi Website: <https://eedi.com>

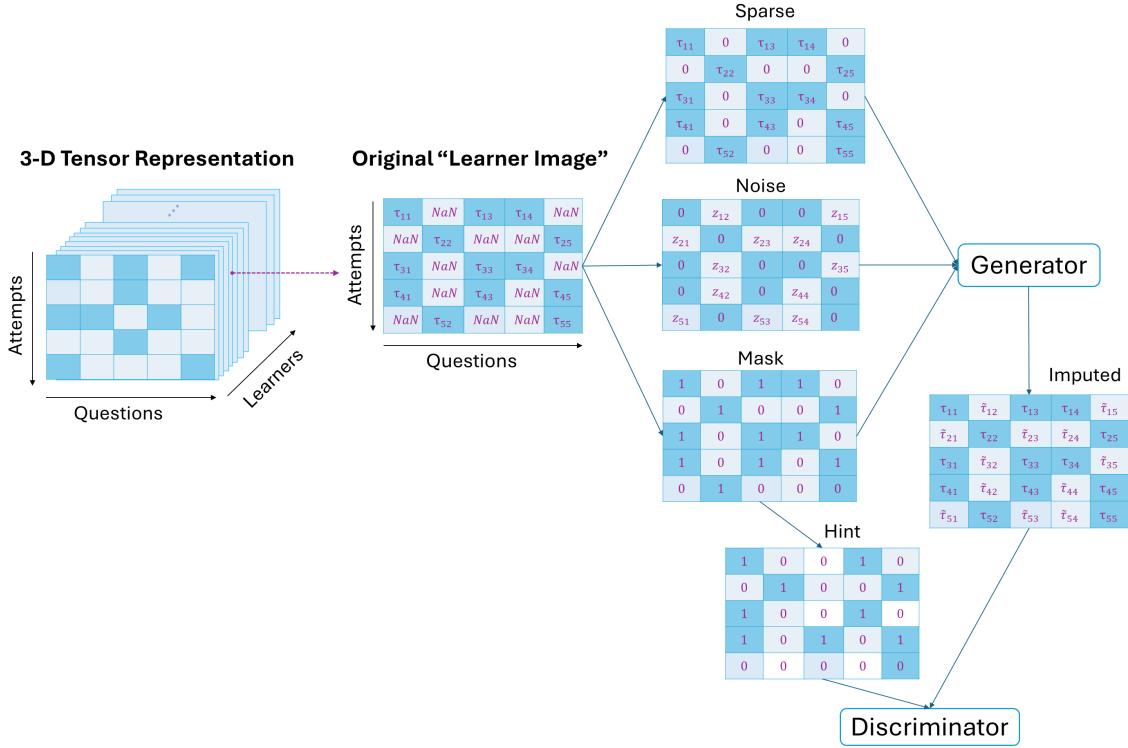


Fig. 2: The proposed GAIN-based imputation architecture for sparse learning performance.

and output layers, with the ReLU activation function are applied to the output of each layer.

- During the iterative training phase, the observed data from \mathcal{T}_l and the corresponding imputed data from \mathcal{T}_{IG} are utilized for optimization via the least square loss function, specifically the Root Mean Square Error (RMSE). This method is chosen to not only ensure enhanced stability and superior quality of the generated data [60], [61] but also align with probability-based predictions of learning performance in peer research on ITSs [9], [24], [62].
- By incorporating a reshape function in the generator's output layer, the shape of generated data \mathcal{T}_{IG} is flexible adjustment to fit the given "learner image" shape, thus accommodating variations across different lesson scenarios without being constrained to a fixed shape, as commonly seen in image-oriented research [31], [63], [64].

The theoretical foundation underlying the inference logic and model assumptions in our proposed generative data imputation approach encompasses the following aspects:

- **Inference Logic.** The entry set within \mathcal{T}_{sparse} can be categorized into two subsets: $\mathcal{T}_{observed}$ for existing values (0 and 1) and $\mathcal{T}_{unobserved}$ for missing ones (NaN). The inference model, formulated as $f_{impute}(\mathcal{T}_{unobserved} | \mathcal{T}_{observed})$, is principle for data imputation, leveraging observed data patterns to impute missing values and predict outcomes [17].
- **Model Assumptions.** Our imputation model operates under several key assumptions within a tensor-based framework: **(a) Probability-based prediction:** Assumes predicted learning performance is a continuous probability between 0 and 1, indicative of knowledge mas-

terly [65]. **(b) Latent domain knowledge relations:** Posits that unobservable latent relationships within the domain knowledge implicitly influence knowledge mastery [66], [67]. **(c) Similarity in learning for individual learners:** Suggests a shared relevance and usefulness of knowledge among learners, aiding in predicting knowledge mastery [39], [48]. **(d) Performance interactions influenced by sequence effects:** Acknowledges that learners' interactions with sequential questions are shaped by priming and recency effects, affecting comprehension and performance [45], [68]. **(e) Maximum attempt assumption:** Defines an empirical maximum number of attempts a learner might require (based on the experimental dataset), highlighting the importance of assessing comprehensive learning states through repeated trials [66].

IV. EXPERIMENTS

In this section, we first introduce the experimental setup, including the evaluation datasets, baselines, and settings. Then, we evaluate the proposed GAIN model and compare it with the baseline models. Finally, we conduct experiments to illustrate the impact of the important learning parameters (derived from BKT) in learning performance.

A. Dataset

To fully evaluate the performance of our revised GAIN model adapted for tensor-based learning data imputation, we utilized three datasets from distinct lessons across differ-

TABLE I: Dataset for the ARC AutoTutor, ASSISTments and MATHia lessons.

Dataset	Lesson Topics	#Learners	#Questions	#Attempts
ARC Lesson 1	Cause and Effect	118	9	9
ARC Lesson 2	Problems and Solution	140	11	5
ASSISTments Lesson 1	Algebra Symbolization Studies	318	64	4
ASSISTments Lesson 2	Skill Builder	392	20	4
MATHia Lesson 1	Scale Drawings	500	28	4
MATHia Lesson 2	Analyzing Models of Two-Step Linear Relationships	500	6	4

ent ITSSs: AutoTutor ARC lessons¹, ASSISTments² and the MATHia³ dataset from mathematics class. As shown in Table I, the AutoTutor ARC lessons include topics such as “Cause and Effect” (ARC Lesson 1) and “Problems and Solution” (ARC Lesson 2), each consisting of 9 to 11 multiple-choice questions designed to test adults’ reading comprehension. This study received ethical approval from the Institutional Review Board (IRB), approval number H15257. The ASSISTments dataset includes lessons on “Algebra Symbolization Studies” (ASSISTments Lesson 1)⁴ and “Skill Builder” (ASSISTments Lesson 2)⁵. The MATHia dataset⁶ covers algebra lessons, specifically “Scale Drawings” (MATHia Lesson 1) and “Analyzing Models of Two-Step Linear Relationships” (MATHia Lesson 2). Table I provides further details, including the total number of learners, questions and attempts.

B. Baselines

Our study compares the GAIN-based imputation method with a range of baseline techniques, including methods from the tensor factorization and GAN series. A detailed description of these baseline approaches is provided below.

Tensor Factorization: The basic tensor factorization factorizes the sparse tensor \mathcal{T}_{sparse} into two components: a learner latent matrix capturing abilities and learning-related features, and a latent tensor representing knowledge during question attempts [25], [38]. A rank-based constraint is used to maintain a generally positive learning trend and accommodate forgetting or slipping [49]. This refined method enhances data imputation within tensor-based structures, providing a robust solution for handling sparse data.

CANDECOMP/PARAFAC Decomposition (CPD): Drawing on the principle of classic CPD [69], [70], the sparse tensor \mathcal{T}_{sparse} is decomposed into three factor tensors that capture learner, attempt and question-related factors in a multidimensional tensor form. A rank-based constraint is additionally applied to enhance the decomposition’s accuracy.

Bayesian Probabilistic Tensor Factorization (BPTF): The BPTF [71] is employed to approximate the sparse tensor $\mathcal{T}_{imputed}$ through the decomposition into a sum of outer

products of three lower-dimensional factor tensors. This approach leverages Bayesian inference for sampling both the factor tensors and the precision of observed entries, effectively enhancing the model’s capacity to manage data sparsity and uncertainty [71], [72].

Generative Adversarial Network (GAN): At one core of the GAN, the “learner image” extracted from \mathcal{T}_{sparse} (depicted in Figure 2), constitutes the base input for the GAN. The GAN architecture includes a generator that simulates data resembling observed entries and a discriminator that assesses the authenticity of this generated data [31]. It uses a consistent CNN layer configuration and least squares loss for optimization.

Information Maximizing Generative Adversarial Nets (InfoGAN): The InfoGAN [63] enhances the traditional GAN framework by integrating the noise with two structured latent variables, allowing for the capture of salient, structured semantic features, such as those relating to learner attributes in ITSSs (e.g., initial learning ability and learning rate). The generator generates imputed $\mathcal{T}_{imputed}$ and decodes latent variables. An auxiliary distribution improves the estimation of these variables’ posterior, boosting mutual information between latent codes and observations and ensuring that the generated outcomes are meaningfully informed.

AmbientGAN: AmbientGAN [73] is used to impute sparse learning performance data by training on partially observed or corrupted data within a GAN framework. It incorporates a dynamically adjusted Gaussian blur in the measurement process, enabling the discriminator to effectively distinguish between real and generated data measurements and accurately infer the original dataset’s true distribution.

C. Experimental Settings and Evaluations

In our experiments for imputing sparse learning performance data, we incorporate several tailored configurations to optimize model training and evaluation. **(a) Cross-Validation:** We employ a five-fold cross-validation strategy, repeated over five cycles, for each model to ensure consistency and reliability of the results. **(b) Varying Attempt Settings:** To assess the stability of the models’ data imputation performance across different levels of data sparsity, we evaluate them under various maximum attempt settings. **(c) Maximum Iterations:** All models are trained for a maximum of 100 iterations to allow adequate learning while monitoring convergence. Training is stopped early if convergence is achieved before reaching the iteration limit, ensuring computational efficiency without compromising the model’s performance. **(d) Learning Rate:** We use a learning rate of either 0.0001 or 0.00001, depending

¹AutoTutor Moodel Website: <https://sites.autotutor.org/>; Adult Literacy and Adult Education Website: <https://adulted.autotutor.org/>

²ASSISTments Website: <https://new.assistments.org/>

³MATHia Website: <https://www.carnegielearning.com/solutions/math/mathia/>

⁴Assistments 2008-2009: <https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=388>

⁵Assistments 2012-2013:<https://sites.google.com/site/assistmentsdata/datasets/2012-13-school-data-with-affect?authuser=0>

⁶MATHia 2019-2020: <https://pslcdatashop.web.cmu.edu/Project?id=720>

on the model, to promote steady progress and convergence during training. **(e) Regularization Techniques:** To prevent overfitting, dropout and batch normalization are integrated into the training process of the GAN-based methods. **(f) Imputation Accuracy Evaluation Metric:** We use the Root Mean Square Error (RMSE) to evaluate the models' performance in data imputation, following previous research [35], [41], [71]. **(g) Mean RMSE and Variability:** The mean RMSE values were obtained over five runs, with standard errors computed to measure variability. **(h) Measuring Sparsity Level:** The sparsity level of the tensor-based learning performance data is computed as the percentage of missing values relative to the total number of elements in the dataset.

D. Evaluating Impacts of Data Imputation on Learning Parameters Using Bayesian Knowledge Tracing

To assess the effectiveness of our proposed data imputation method, we analyze the changes in estimated learning parameters before and after imputation using Bayesian Knowledge Tracing (BKT) [66]. BKT was chosen due to its explainable parameters, which are directly tied to specific learning features in student performance data. Specifically, we focus on four core probability-based parameters in BKT [74]: $P(L_0)$, the probability of initial knowledge, representing the likelihood that the learner has already mastered a skill at the start; $P(T)$, the probability of learning rate, indicating the chance that the learner transitions from an unmastered to a mastered state for a particular skill; $P(G)$, the probability of guess rate, reflecting the likelihood that the learner answers correctly by guessing while still in an unmastered state; $P(S)$, the probability of slip rate, denoting the probability of an incorrect answer despite the learner being in a mastered state. We compare these parameters, derived from the imputed datasets, with those from the original sparse datasets to assess the impact of imputation on the accuracy and reliability of student modeling. The RMSE is also used to evaluate the performance of the BKT modeling, providing a quantitative measure of the impact of our proposed GAIN-based data imputation method. Additionally, significance testing using the one-sided paired test is performed to assess the statistical differences between the original and imputed datasets.

To quantify the changes in BKT parameters before and after imputation, we calculated the Kullback-Leibler (KL) divergence, which measures the difference between two probability distributions, specifically the distributions of BKT parameters derived from the original and imputed datasets. We applied Kernel Density Estimation (KDE) to obtain smoothed probability density functions for each parameter, addressing issues of data sparsity and variance. For each BKT parameter, KDE was used to estimate the continuous probability distributions from both the original and imputed data, and the KL divergence between these two distributions was then computed using the following formula [75]:

$$\text{KL}(P \parallel Q) = \int_{-\infty}^{\infty} P(x) \log \left(\frac{P(x)}{Q(x)} \right) dx \quad (2)$$

where $P(x)$ is the probability density function for learning parameters from the original data and $Q(x)$ is that from

the imputed data. This calculation provides a measure of how much the imputation alters the underlying parameter distributions, thereby evaluating the effectiveness and impact of the GAIN method using the BKT modeling.

V. RESULTS

A. Data Imputation Accuracy

RQ1 investigates how effectively the GAIN-based method imputes sparse learning performance data in ITSs compared to established baselines. This question is examined through the following results. The imputation accuracy of various models (Tensor Factorization, CPD, BPTF, GAN, InfoGAN, AmbientGAN, and GAIN) on sparse learning performance data from six lessons across AutoTutor ARC, ASSISTments, and MATHia is evaluated using RMSE, as shown in Figure 3. The figure also presents RMSE values across different Max Attempt settings (addressing **RQ1.1**), where smaller RMSE values indicate higher imputation accuracy. Error bars in the figure represent standard errors, as previously discussed. As observed, GAIN frequently achieves the lowest RMSE across most lessons, demonstrating superior imputation accuracy, particularly in the ASSISTments and MATHia datasets, where it distinctly outperforms other models (addressing **RQ1.2**). However, an exception occurs in ARC Lesson 1, where GAIN's RMSE is larger than that of BPTF, GAN, and InfoGAN, with longer error bars suggesting greater variance and reduced accuracy in imputation. This divergence likely reflects the unique characteristics of reading comprehension assessment, where questions build upon multiple interrelated skills (decoding, vocabulary knowledge, and inferential reasoning). Unlike mathematical problems where knowledge components are often more discretely defined, reading comprehension questions typically draw upon overlapping cognitive processes. This interconnected nature of reading comprehension skills may challenge GAIN's ability to accurately impute missing values, particularly when learners exhibit uneven skill profiles across different comprehension strategies. The relative higher RMSE values across different Max Attempt highlight challenges in accurately imputing this dataset. This unique case underscores complex data or model interactions that require further investigation. Additionally, while GAN performs well in ARC Lesson 1, it is slightly less robust than GAIN, with CPD and BPTF also showing competitive results.

Despite its overall superior performance, GAIN exhibits greater variance in its results, as reflected by the longer error bars in Figure 3, indicating reduced stability in its data imputation. This heightened variance suggests that while GAIN often achieves superior accuracy, its consistency may be compromised under certain data conditions, potentially requiring additional tuning or pre-processing to enhance stability. In contrast, other baseline models, such as Tensor Factorization and CPD, show lower variance, suggesting they may offer more reliable imputations in specific contexts, even though they do not always achieve the lowest RMSE for optimal imputation accuracy.

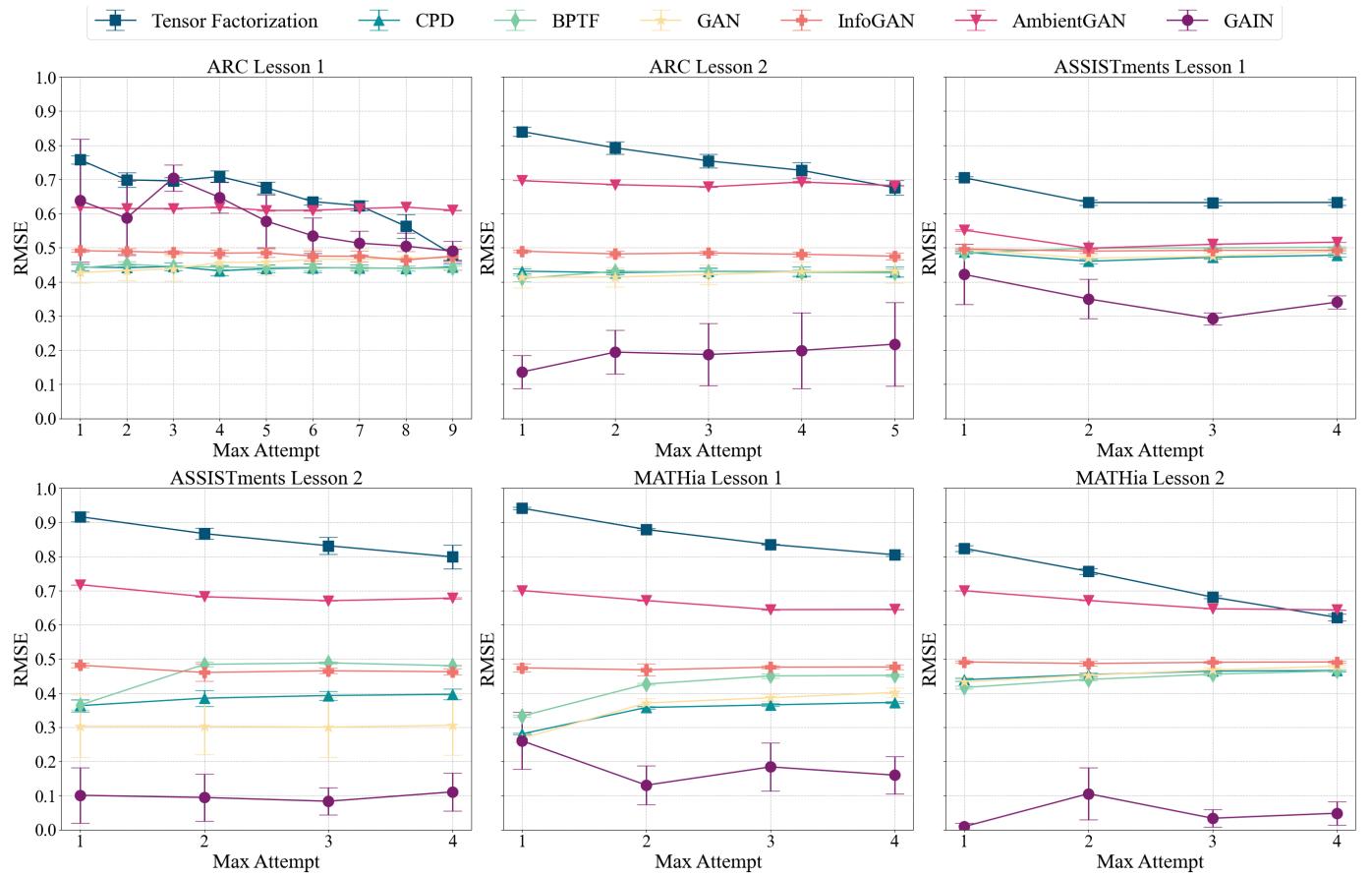


Fig. 3: RMSE comparisons of data imputation models across different attempts for different lessons dataset.

B. Comparative Analysis of BKT Modeling Accuracy Using RMSE: Original Sparse Data vs. Imputed Data

RQ2 examines how the imputed data align with the original sparse learning performance in ITSs, focusing on the impact of GAIN-based data imputation on learner modeling, as exemplified by BKT (**RQ2.1**). This aspect is addressed through the following results. Table II compares the BKT model's performance on original and imputed data across six lessons, highlighting RMSE differences for varying maximum attempt. For each lesson, the table presents RMSE values for the original data ("Original"), the imputed data ("Imputed"), and their difference ("Difference"), calculated as the imputed RMSE minus the original RMSE. A negative difference indicates a decrease in RMSE, reflecting improved BKT modeling accuracy after imputation. Conversely, a positive difference signifies an increase in RMSE, indicating reduced BKT modeling accuracy. For ARC Lesson 1, RMSE initially increases for Max Attempts 1 to 3, indicating a marginal decline in model performance, but decreases from Max Attempt 4 onward, suggesting improved model accuracy at higher attempts. Similarly, in ARC Lesson 2, RMSE increases at lower attempts but decreases at higher attempts. In the ASSISTments datasets, imputed data consistently show reduced RMSE across all attempts, demonstrating enhanced BKT model performance post-imputation. In contrast, for the MATHia datasets, RMSE increases in Lesson 1 for most attempts, while in Lesson 2,

the imputed data consistently result in lower RMSE values. To evaluate the significance of these changes, a paired one-sided t-test was conducted comparing original and imputed RMSE values for each attempt. The t-test produced ($t = -3.2552, p = 0.0014$), confirming that imputation significantly improves BKT model performance (addressing **RQ2.1**). The $t = -3.2552$ indicates a substantial difference between the mean RMSE values of the original and imputed datasets, with the negative sign suggesting that the imputed data generally result in lower RMSE values. Moreover, the $p = 0.0014$, well below the threshold of 0.05, provides strong statistical evidence supporting the significance of these improvements. This statistical evidence highlights the effectiveness of GAIN-based imputation in mitigating the effects of data sparsity and reliably improving model performance across lessons and datasets.

Additionally, a notable observation is the difference in RMSE values from BKT modeling between the original and imputed datasets. These differences, which include both increases and decreases (as shown in Table II), may be attributed to changes in data sparsity driven by the increasing number of attempts in the original dataset. This observation is further verified by the following analysis results, which provide additional insights into **RQ2.1**. See the Figure 4, where the terms "Increased" and "Decreased" in RMSE correspond to positive and negative "Difference" values in Table II, respectively. The

TABLE II: RMSE Comparisons of BKT modeling between the original and imputed data.

Dataset	Type	Max Attempt								
		1	2	3	4	5	6	7	8	9
ARC Lesson 1	Original	0.395	0.400	0.395	0.396	0.396	0.400	0.397	0.399	0.398
	Imputed	0.404	0.421	0.401	0.377	0.360	0.341	0.329	0.316	0.305
	Difference	+0.009	+0.021	+0.006	-0.019	-0.036	-0.059	-0.068	-0.083	-0.093
ARC Lesson 2	Original	0.391	0.390	0.388	0.389	0.390	—	—	—	—
	Imputed	0.423	0.444	0.403	0.370	0.344	—	—	—	—
	Difference	+0.032	+0.054	+0.015	-0.019	-0.046	—	—	—	—
ASSISTments Lesson 1	Original	0.472	0.469	0.471	0.475	—	—	—	—	—
	Imputed	0.364	0.318	0.294	0.280	—	—	—	—	—
	Difference	-0.108	-0.151	-0.177	-0.195	—	—	—	—	—
ASSISTments Lesson 2	Original	0.320	0.398	0.407	0.417	—	—	—	—	—
	Imputed	0.277	0.204	0.172	0.154	—	—	—	—	—
	Difference	-0.043	-0.194	-0.235	-0.263	—	—	—	—	—
MATHia Lesson 1	Original	0.260	0.366	0.410	0.427	—	—	—	—	—
	Imputed	0.228	0.471	0.476	0.454	—	—	—	—	—
	Difference	-0.031	+0.105	+0.066	+0.027	—	—	—	—	—
MATHia Lesson 2	Original	0.408	0.428	0.452	0.466	—	—	—	—	—
	Imputed	0.372	0.388	0.424	0.450	—	—	—	—	—
	Difference	-0.036	-0.040	-0.028	-0.016	—	—	—	—	—

* Note: The positive symbol “+” denotes an increase in the RMSE value, while the negative symbol “-” indicates a decrease. The symbol “—” signifies that the maximum number of attempts was not applicable for that particular dataset.

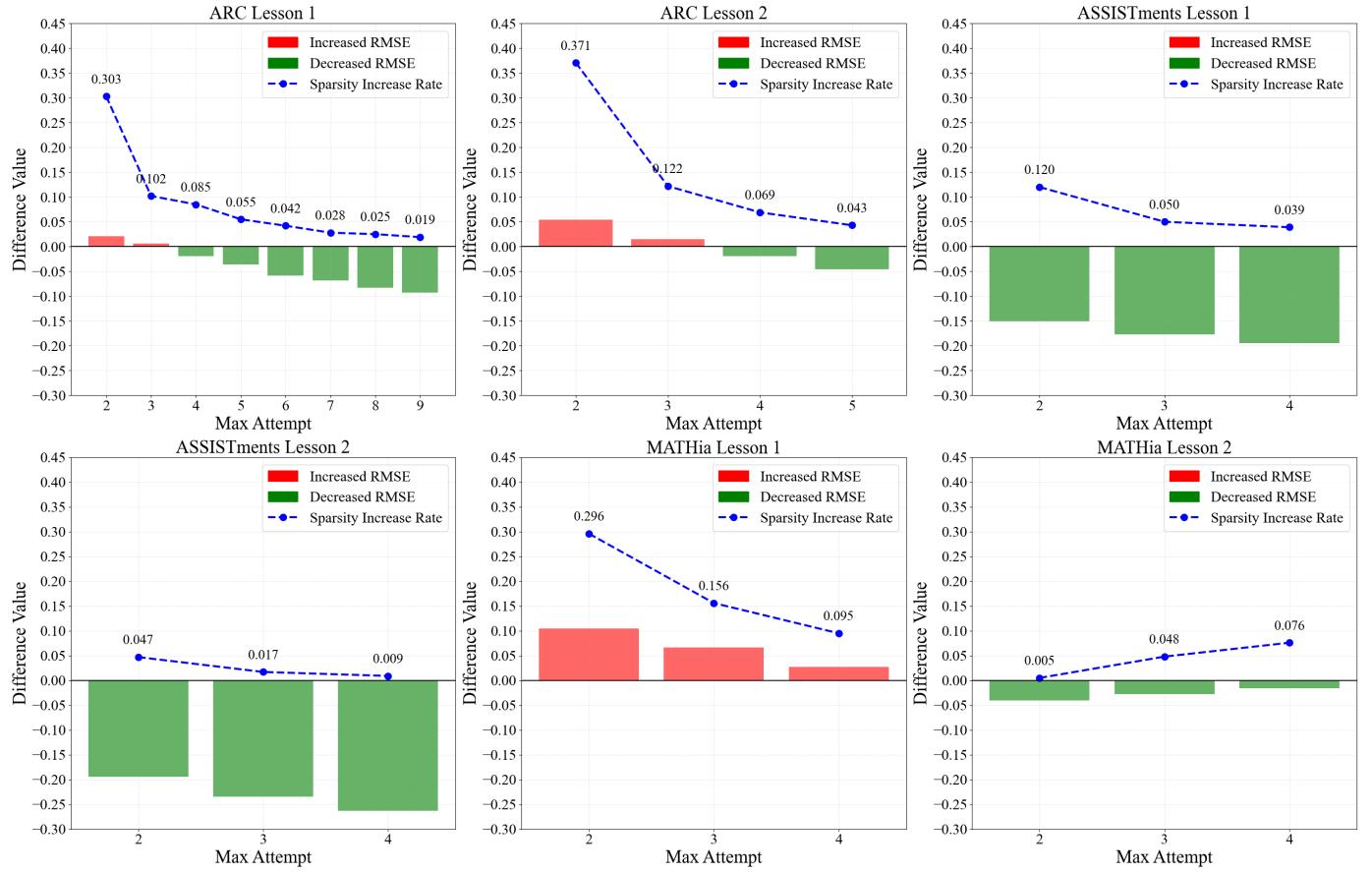


Fig. 4: RMSE changes with varying sparsity increase rates (corresponding to Max Attempts).

sparsity increase rate is calculated by dividing the sparsity level values by their increase with Max Attempt (refer to Appendix A for more details on how sparsity levels change

with increasing Max Attempt, where the slopes of the lines with respect to Max Attempt represent the “sparsity increase rate”). As observed from Figure 4, the sparsity increase rate

generally decreases across Max Attempt in most lessons, with the exception of MATHia Lesson 2, where it shows a gradual increase. In most situations with increased RMSE, the sparsity increase rate tends to be higher, whereas decreased RMSE generally corresponds to lower sparsity increase rates. For example, in ARC Lesson 1, RMSE increases at Max Attempts 2 and 3, where the sparsity increase rates are 0.303 and 0.102, respectively. In contrast, from Max Attempt 4 onward, where RMSE decreases, the sparsity increase rate drops to 0.085 or lower. Similarly, in ARC Lesson 2, RMSE increases at Max Attempts 2 and 3, with sparsity increase rates of 0.371 and 0.122, which are notably higher than those observed at subsequent attempts, where RMSE decreases and the sparsity increase rates drop to 0.069 and 0.043. In MATHia Lesson 1, RMSE increases across Max Attempts 2, 3, and 4, where all sparsity increase rates remain above 0.095. These observations suggest that steeper changes in sparsity levels, which reflect a rapid increase in the proportion of missing data with each additional attempt, make it more challenging for the imputation model to accurately predict missing values, as evidenced by the increased RMSE values. Conversely, when the sparsity increase rate is lower, the model faces fewer abrupt changes in data sparsity, allowing for more accurate imputation and resulting in decreased RMSE values. Additionally, all observed sparsity increase rates are consistently below 0.095 in cases where RMSE decreases, indicating improved GAIN-based imputation. However, it remains unclear whether this value represents a definitive threshold or merely coincides with an underlying threshold that influences the effectiveness of data imputation. Identifying such a threshold, if it exists, presents an intriguing avenue for future research to better understand the relationship between sparsity dynamics and imputation accuracy.

C. Divergence Measurement of Learning Features in Imputed Data from Original Data

The KL divergence measurements for the estimated learning parameters ($P(L_0)$, $P(T)$, $P(G)$, and $P(S)$) distributions for individual questions, derived from BKT modeling, demonstrate the effectiveness of the proposed imputation model in preserving critical learning features and characteristics in the imputed data compared to the original data. These results collectively address **RQ2.2**. As shown in Figure 5, the KL divergence value, calculated using relative entropy, indicate the degree of alignment between question-wise learning parameters in the imputed and original datasets, with ASSISTments Lesson 1 (Max Attempt = 4) serving as an example. The key observations are as follows:

- For $P(L_0)$, the KL divergence values are quite low for most questions (below 1), with a few exceptions (including Q0, Q5, Q19, Q20, and Q43), which exhibit higher divergence (greater than 1). This indicates that while the imputed data generally align well with the original data in modeling initial knowledge, some questions proved more challenging to accurately impute, potentially introducing slight bias in these cases.
- For $P(T)$, the KL divergence values are consistently low, remaining below 1 for all questions. This indicates

that the imputation process had minimal impact on the learning rate parameter, suggesting a strong alignment between the imputed and original data in representing learning rates. The low divergence in $P(T)$ demonstrates the proposed GAIN model's capability in data imputation, effectively preserving the integrity of learning progression estimates across questions.

- For $P(G)$, the KL divergence values are generally low (below 1) across most questions, indicating that the imputation model closely aligns with the original data in modeling guessing behavior. This suggests that our proposed GAIN model effectively captures overall trends in guessing. However, there are notable spikes in divergence values for a few specific questions (e.g., Q11, Q15, Q19 and Q21), where the imputed data differs significantly from the original. These cases suggest that the imputation model may introduce some bias by altering the variability in guessing behavior observed in the sparse data, potentially underestimating guessing tendencies for certain questions.
- For $P(S)$, most questions demonstrate low divergence (below 1), while approximately 15 questions show significant variability with KL divergence values exceeding 1. This indicates that, in most cases, the imputation model closely follows the original data for slip rates. However, for specific questions (e.g., Q1, Q8, Q9, Q11, Q16, etc.), higher divergence suggests that the model may introduce variability in slip characteristics, particularly in cases where the original data was sparse or noisy.

The KL divergence measurements across the four learning parameters ($P(L_0)$, $P(T)$, $P(G)$, and $P(S)$) encompassing all six lessons provide valuable insights into the alignment between imputed and original data. Figure 6 displays the percentages of KL divergence values within the range [0, 1] for each parameter, aggregated across all questions. Using [0, 1] as an acceptable range for smaller divergence values, as suggested in research from other domains [76], [77] (in the absence of specific criteria in learning engineering), provides a practical approach given the unbounded nature of KL divergence. Values closer to 0 indicate a close alignment between the imputed and original data, while values approaching or exceeding 1 suggest greater divergence. The results presented in Figure 6, reveal that most lessons, exhibit high percentages of KL divergence within this range, indicating strong alignment between imputed and original data. This alignment is especially notable for the ASSISTments Lessons, where all parameters consistently show values above 70% for Lesson 1 and Lesson 2, with both ASSISTments Lesson 1 ($P(T)$) and Lesson 2 ($P(L_0)$ and $P(T)$) achieving close to 100% alignment in several parameters. For the ARC Lessons, there is also generally strong alignment, with most parameters showing percentages above 70%. However, ARC Lesson 2 shows more variability, especially in $P(S)$ where the percentage drops to 34.54%, suggesting that the slip parameter in the imputed data diverges more from the original data in this lesson. This variability in ARC Lesson 2 may reflect challenges in capturing certain learning dynamics, such as the tendency to slip, in

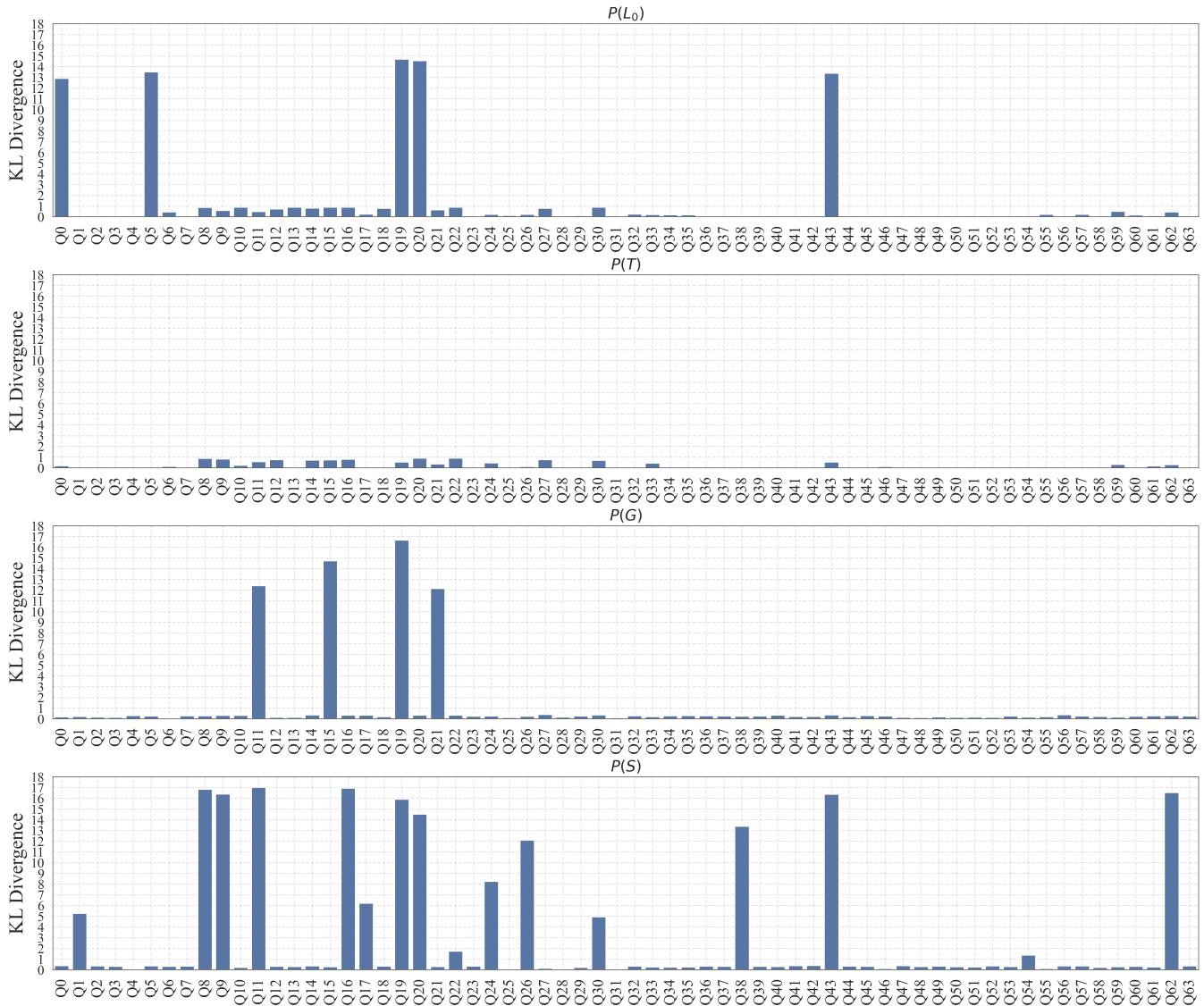


Fig. 5: Comparison of BKT parameters (Original vs. Imputed Learning Performance Data) for ASSISTments Lesson 1 (Max Attempt = 4).

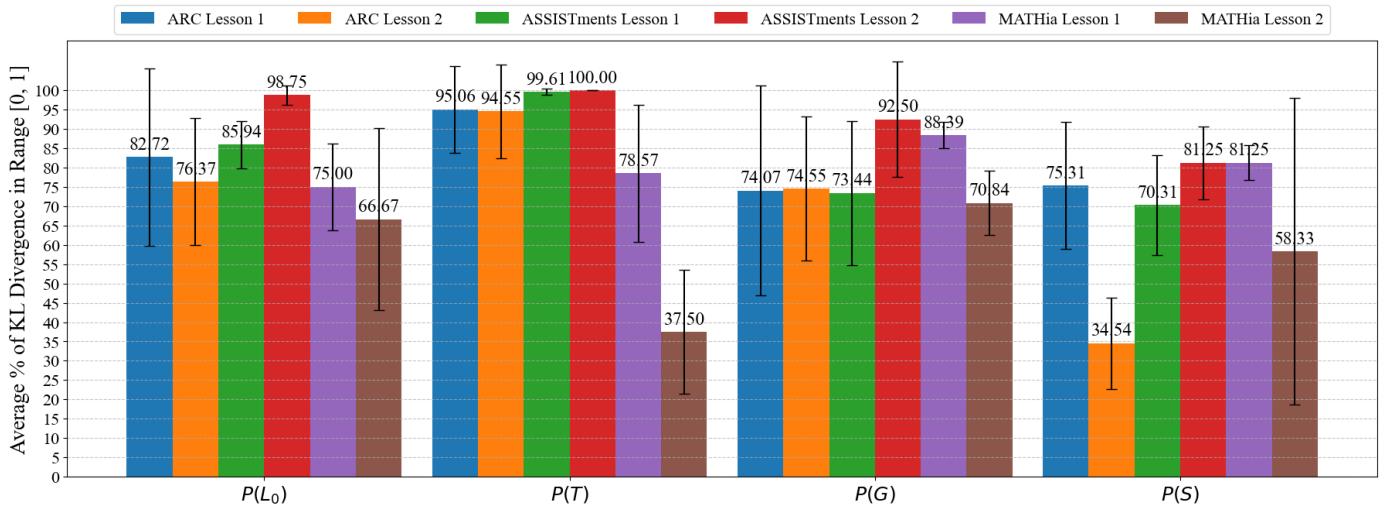


Fig. 6: Percentages of KL divergence values within $[0, 1]$ for learning parameters derived from BKT modeling.

the imputed data. The MATHia Lessons, on the other hand, show a mixed pattern of alignment. MATHia Lesson 1 displays moderate to high alignment, with all parameters falling above 70%. However, MATHia Lesson 2 shows more noticeable divergence across parameters, with $P(L_0)$ below 70%, and $P(T)$ (37.50%) and $P(S)$ (58.33%) falling below 60%. This variability, particularly in MATHia Lesson 2, suggests that the imputed data may have limitations in accurately capturing certain learning behaviors in the MATHia dataset.

In conclusion, the KL divergence analysis results show that the imputation model effectively aligns with the original data across most learning parameters, with minimal divergence observed in $P(L_0)$, $P(T)$, $P(G)$ and $P(S)$ across the majority of questions. While certain parameters, particularly $P(S)$ and $P(T)$, exhibit higher divergence for a subset of questions, these deviations likely reflect the challenges faced by our proposed imputation model in handling sparse data. Overall, these results emphasize the GAIN model's effectiveness in filling missing data while preserving the essential learning characteristics of the original dataset.

VI. DISCUSSION

This study proposes a systematic imputation framework that integrates multidimensional learner modeling with GenAI models, specifically GAIN, to address the critical issue of data sparsity in ITSs. Evaluated using three types of ITS lesson datasets (AutoTutor ARC, ASSISTments, and MATHia), our proposed GAIN model's imputation accuracy generally outperforms baseline methods, including Tensor Factorization, CPD, BPTF, GAN, InfoGAN, and AmbientGAN, thus addressed **RQ1**. Its resilient performance across datasets further highlights its generalizability and adaptability to diverse ITS environments (addressed **RQ1.2**). Moreover, the GAIN model significantly enhances learner modeling accuracy, as evidenced by improved Bayesian Knowledge Tracing (BKT) performance on imputed data compared to original sparse data (directly addressed **RQ2.1**). This improvement is closely linked to changes in sparsity rates, where higher sparsity increase rates present greater challenges for accurate imputation. Additionally, KL divergence analysis of key learning parameters (including initial knowledge $P(L_0)$, learning rate $P(T)$, guess rate $P(G)$) demonstrates close alignment between imputed and original data, preserving critical learning features (answered **RQ2.2**). Collectively, these findings affirm the effectiveness of the GAIN model in mitigating data sparsity while maintaining while preserving the original learning data characteristics.

The 3D tensor-based representation facilitates the estimation of missing performance values for previously unattempted questions and attempts by capturing the complex dynamic interactions and similarities across these dimensions. Within this 3D space, GAIN can impute learner performance by considering the contextual information around specific locations and the similarities among attempts, questions, and even learners, to effectively fill the “learner performance gaps” across the entire 3D structure. This approach represents a successful application of the classic Rubin's imputation rule [78] within the learning engineering context, inferring missing

performance data based on existing and observed learning behaviors for more effective imputation. The efficacy of this 3D framework is also verified with findings from our previous studies [27], [38], [58].

The identification of higher imputation accuracy for sparse learning performance across different types of ITS lessons and varying attempts highlights the compatibility of GAIN with the 3D space and its efficacy in imputing learner performance. In the modeling process, GAIN is adapted to accommodate both the input and output shapes of tensor-based learning performance data. Missing data is imputed through multiple CNN layers in GAIN's generator. The hint matrix, generated based on the sparsity distribution of the original learning performance data, provides conditions for imputing the missing data, while the hint discriminator guides the generator for more accurate imputation. This hint mechanism semantically targets the individual sparsity distribution in the learning data, effectively facilitating data imputation for sparse learning performance data in ITSs. Additionally, the training stage considers the matrix relationships between attempts and questions within the “learner image” and accounts for similarities across individual learners within that group. This multidimensional exploration of learning performance enables effective model training on existing data and facilitates the prediction of likely missing performance data based on patterns learned by the neural networks in the GAIN model.

GAIN's performance in ARC Lesson 1 underscores the broader challenge of applying generative models to datasets characterized by complex cognitive dependencies, such as reading comprehension. The elevated variance and reduced accuracy observed highlight GAIN's limitations in handling overlapping and interdependent skillsets, including decoding, vocabulary, and inferential reasoning. These interconnected skills amplify the difficulty of accurately imputing missing values, particularly for learners with uneven cognitive profiles. This case underscores the critical need for models to go beyond conventional imputation techniques, incorporating mechanisms that account for nuanced learner variability and multidimensional cognitive processes. Exploring advanced architectures, such as hybrid or context-sensitive models, may provide better alignment with these complexities. Ultimately, this exception emphasizes the importance of tailoring imputation strategies to the unique demands of specific domains within ITS applications, particularly those as cognitively intricate as reading comprehension.

To assess the impact of the proposed GAIN-based imputation method on learner modeling, particularly with BKT, we observed a significant improvement in learning modeling accuracy. By reducing data sparsity, the GAIN model provides richer and more complete data, which enhances our understanding of learners' progress across questions and attempts. This additional information from imputed data allows for a more detailed capture of learners' performance patterns, leading to more accurate BKT parameter estimates and a fine-grained understanding of key learning features such as initial knowledge, learning rate, guess rate, and slip rate. As a result, the modeling accuracy for imputed data significantly improves, providing a closer alignment with actual learning

dynamics compared to models based solely on sparse data. Additionally, this improvement in BKT modeling accuracy is closely tied to changes in sparsity rates as the number of Max Attempt increases. As sparsity rates rise sharply with each additional attempt, the proportion of missing data escalates, making it increasingly complex to accurately impute missing values. This pattern suggests that as students engage in more attempts, maintaining high imputation accuracy becomes more challenging due to the growing scarcity of data points necessary for reliable modeling. Implicitly, sparsity rates seem to be linked to information loss in learning performance, with changes in these rates potentially affecting imputation precision and the overall quality of learner modeling. Further research is needed to clarify this relationship and explore strategies to mitigate the effects of high sparsity rates.

A comparative analysis leveraging KL divergence measurements of the distributions of question-wise learning parameters (initial knowledge $P(L_0)$, learning rate $P(T)$, guess rate $P(G)$, and slip rate $P(S)$) between the original sparse and imputed data provides a nuanced understanding of how data imputation affects learning features. This analysis reveals the extent to which the imputed data aligns with the original data, offering insights into the accuracy of the imputation model in preserving the underlying learner behaviors. The generally low KL divergence values, particularly the high percentage within the $[0, 1]$ range, indicate that the imputed data closely mirrors the distributions in the original dataset for all parameters. These results suggest that GAIN-based imputation effectively captures learner dynamics without introducing substantial bias, making it a robust tool for handling missing data in ITS environments. However, some parameters, such as $P(S)$ and $P(T)$, display higher divergence in a few specific lessons, potentially reflecting unique or complex learning patterns that the model does not fully capture. Such divergence may arise in lessons where slipping and learning rates vary significantly among learners, or in cases where data sparsity is more pronounced. Future work could explore further refinement of the imputation process, particularly for parameters sensitive to individual variations or specific instructional contexts, as this extends beyond the current study's scope.

VII. LIMITATIONS

Some misalignment of learning features derived from the imputed dataset may occur, and bias introduced by the imputation process remains a potential concern. Further research is necessary to identify the specific sources of bias resulting from the imputation model and to develop effective methods for mitigating these issues. Additionally, understanding the broader implications of this bias on generalizability and model performance will require more in-depth analysis. The refinement of knowledge components within questions, as well as their connections across different questions, is still an area that needs improvement. This limitation affects the model's ability to accurately track knowledge transfer and interdependencies between various learning tasks. Moreover, the potential utilization of individual levels of attempts, rather than a one-size-fits-all approach in constructing the 3D tensor

of learning performance data, requires further refinement. Although this method offers a theoretically predicted knowledge state in learner modeling for ITSs, the use of equal fixed attempts contributes to increased sparsity levels, necessitating more nuanced methods for managing sparsity and improving precision. Lastly, the varying, and in some cases high, levels of sparsity within the learning performance dataset, as highlighted in Figure 7, can significantly impact modeling and analysis, leading to challenges such as model bias, reduced performance, and complexities in knowledge tracing within ITSs. These limitations underscore the need for future research to address the trade-offs between sparsity management, model accuracy, and scalability in educational settings.

VIII. FUTURE WORKS

Future research can extend generative data imputation on binary numerical values to dialogue-based scenarios, such as dialogue-based ITSs, leveraging GenAI models like Large Language Models (LLMs). These models could enhance the ability to predict learners' responses, making the imputation process more dynamic and context-aware in dialogue interactive environment. In addition to improving computational models, future work should focus on integrating knowledge entities that are actively engaged in tutoring environments. This would allow simulated learner agents to not only model computational behavior but also reflect actual learning processes more accurately. Sequential question generation and recommendation, tailored to learners' knowledge progression, is another promising area of research. Developing systems that can recommend questions based on a learner's current knowledge state, and adjusting in real-time, could significantly improve personalized learning. Finally, addressing the impact of information loss on imputation models is crucial. Future work should explore the effects of spatial data distribution on imputation performance, ensuring that both temporal and spatial patterns are effectively captured to reduce information loss and enhance imputation accuracy.

IX. CONCLUSIONS

In this study, we present a generative data imputation based on GAIN to impute sparse learner performance data in ITSs. By reconstructing the learner performance data into a 3D tensor encompassing learners, questions, and attempts, our method leverages existing data structured in a multidimensional format and adapts along the attempts dimension to accommodate varying levels of sparsity. Enhanced by incorporating convolutional neural networks in the input and output layers and employing a least squares loss function for optimization, our GAIN-based approach aligns the shapes of input and output with the dimensions of the question-attempt matrices along the learners' dimension. The results of our study shows GAIN gains effective in achieving high-accuracy data imputation, as demonstrated through comparisons with other baselines across three types of ITS lessons including AutoTutor ARC, ASSISTments, and MATHia, while maintaining a close alignment with original data patterns across most learning parameters, as verified by low KL divergence. To

assess the impact of data imputation on learning features, we employ BKT modeling to estimate population-level learning parameters. Our analysis of question-wise parameters derived from BKT reveals improved model fitting with imputed data compared to the original sparse data, evidenced by generally low divergence values, suggesting that GAIN preserves essential learning features with minimal bias. However, as sparsity rates increase (especially with additional attempts) the imputation task becomes more complex, and slightly higher divergence in parameters like slip and learning rates in specific lessons indicates areas for refinement to capture unique learning dynamics in high-sparsity conditions. These findings underscore the potential of GAIN-based imputation to enhance learner modeling accuracy, making it a robust tool for supporting adaptive, personalized instruction in ITSs. Future research could explore adaptive modifications within GAIN to improve performance for more complex or individualized learning contexts. Overall, our generative data imputation method effectively overcomes learning performance data sparsity challenges in intelligent systems for AI education.

ACKNOWLEDGMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A200413 to the University of Memphis. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We extend our gratitude to Prof. Mohammed Yeasin from the University of Memphis for his invaluable suggestions on the divergence measurement method employed in this study. Additionally, we sincerely thank Dr. Felix Havugimana from the University of Memphis for his insightful ideas that significantly improved this methodology.

REFERENCES

- [1] K. Crockett, A. Latham, and N. Whitton, "On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees," *International Journal of Human-Computer Studies*, vol. 97, pp. 98–115, 2017.
- [2] G.-J. Hwang, H. Xie, B. W. Wah, and D. Gašević, "Vision, challenges, roles and research issues of artificial intelligence in education," p. 100001, 2020.
- [3] A. C. Graesser, Z. Cai, W. O. Baer, A. M. Olney, X. Hu, M. Reed, and D. Greenberg, "Reading comprehension lessons in autotutor for the center for the study of adult literacy," in *Adaptive educational technologies for literacy instruction*. Routledge, 2016, pp. 288–293.
- [4] N. T. Heffernan and C. L. Heffernan, "The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching," *International Journal of Artificial Intelligence in Education*, vol. 24, pp. 470–497, 2014.
- [5] D. Tafazoli, E. G. María, and C. A. H. Abril, "Intelligent language tutoring system: integrating intelligent computer-assisted language learning into language education," *International Journal of Information and Communication Technology Education (IJICTE)*, vol. 15, no. 3, pp. 60–74, 2019.
- [6] S. D'Mello, A. Olney, C. Williams, and P. Hays, "Gaze tutor: A gaze-reactive intelligent tutoring system," *International Journal of human-computer studies*, vol. 70, no. 5, pp. 377–398, 2012.
- [7] S. Feng, A. J. Magana, and D. Kao, "A systematic review of literature on the effectiveness of intelligent tutoring systems in stem," in *2021 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2021, pp. 1–9.
- [8] L. G. Eglington and P. I. Pavlik Jr, "How to optimize student learning using student models that adapt rapidly to individual differences," *International Journal of Artificial Intelligence in Education*, pp. 1–22, 2022.
- [9] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, "Individualized bayesian knowledge tracing models," in *International conference on artificial intelligence in education*. Springer, 2013, pp. 171–180.
- [10] S. Chen, A. Lippert, G. Shi, Y. Fang, and A. C. Graesser, "Disengagement detection within an intelligent tutoring system." *Grantee Submission*, 2018.
- [11] M. Saarela, "Automatic knowledge discovery from sparse and large-scale educational data: case finland," Ph.D. dissertation, University of Jyväskylä, 2017.
- [12] J. Greer and M. Mark, "Evaluation methods for intelligent tutoring systems revisited," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 387–392, 2016.
- [13] S. Pandey and G. Karypis, "A self-attentive model for knowledge tracing," *arXiv preprint arXiv:1907.06837*, 2019.
- [14] W. Lee, J. Chun, Y. Lee, K. Park, and S. Park, "Contrastive learning for knowledge tracing," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 2330–2338.
- [15] T. Wang, F. Ma, and J. Gao, "Deep hierarchical knowledge tracing," in *Proceedings of the 12th international conference on educational data mining*, 2019.
- [16] X. Wang, S. Zhao, L. Guo, L. Zhu, C. Cui, and L. Xu, "Graphca: Learning from graph counterfactual augmentation for knowledge tracing," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 11, pp. 2108–2123, 2023.
- [17] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.
- [18] A. R. T. Donders, G. J. Van Der Heijden, T. Stijnen, and K. G. Moons, "A gentle introduction to imputation of missing values," *Journal of clinical epidemiology*, vol. 59, no. 10, pp. 1087–1091, 2006.
- [19] Z. Zhang, "Missing data imputation: focusing on single imputation," *Annals of translational medicine*, vol. 4, no. 1, 2016.
- [20] D. B. Rubin, "Multiple imputations in sample surveys-a phenomenological bayesian approach to nonresponse," in *Proceedings of the survey research methods section of the American Statistical Association*, vol. 1. American Statistical Association Alexandria, VA, USA, 1978, pp. 20–34.
- [21] G. E. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning," *Applied artificial intelligence*, vol. 17, no. 5–6, pp. 519–533, 2003.
- [22] S. R. Seaman, J. W. Bartlett, and I. R. White, "Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods," *BMC medical research methodology*, vol. 12, pp. 1–13, 2012.
- [23] A. Newell, H. A. Simon *et al.*, *Human problem solving*. Prentice-hall Englewood Cliffs, NJ, 1972, vol. 104, no. 9.
- [24] P. I. Pavlik, L. G. Eglington, and L. M. Harrell-Williams, "Logistic knowledge tracing: A constrained framework for learner modeling," *IEEE Transactions on Learning Technologies*, vol. 14, no. 5, pp. 624–639, 2021.
- [25] L. Zhang, P. I. Pavlik Jr, X. Hu, J. L. Cockroft, L. Wang, and G. Shi, "Exploring the individual differences in multidimensional evolution of knowledge states of learners," in *International Conference on Human-Computer Interaction*. Springer, 2023, pp. 265–284.
- [26] S. Zhao, C. Wang, and S. Sahebi, "Modeling knowledge acquisition from multiple learning resource types," *arXiv preprint arXiv:2006.13390*, 2020.
- [27] L. Zhang, J. Lin, J. Sabatini, C. Borchers, D. Weitekamp, M. Cao, J. Hollander, X. Hu, and A. C. Graesser, "Data augmentation for sparse multidimensional learning performance data using generative ai," *arXiv preprint arXiv:2409.15631*, 2024.
- [28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [29] Y. Saito, S. Takamichi, and H. Saruwatari, "Statistical parametric speech synthesis incorporating generative adversarial networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 1, pp. 84–96, 2017.
- [30] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "Attngan: Fine-grained text to image generation with attentional generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1316–1324.

- [31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [32] Y. Zeng, Z. Lin, H. Lu, and V. M. Patel, "Cr-fill: Generative image inpainting with auxiliary contextual reconstruction," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 14 164–14 173.
- [33] M. Hernández-Bautista and F. Melero, "3d hole filling using deep learning inpainting," *arXiv preprint arXiv:2407.17896*, 2024.
- [34] Y. Luo, X. Cai, Y. Zhang, J. Xu *et al.*, "Multivariate time series imputation with generative adversarial networks," *Advances in neural information processing systems*, vol. 31, 2018.
- [35] J. Yoon, J. Jordon, and M. Schaar, "Gain: Missing data imputation using generative adversarial nets," in *International conference on machine learning*. PMLR, 2018, pp. 5689–5698.
- [36] Y. Zhang, R. Zhang, and B. Zhao, "A systematic review of generative adversarial imputation network in missing data imputation," *Neural Computing and Applications*, vol. 35, no. 27, pp. 19 685–19 705, 2023.
- [37] W. Dong, D. Y. T. Fong, J.-s. Yoon, E. Y. F. Wan, L. E. Bedford, E. H. M. Tang, and C. L. K. Lam, "Generative adversarial networks for imputing missing data for big data clinical research," *BMC medical research methodology*, vol. 21, pp. 1–10, 2021.
- [38] L. Zhang, J. Lin, C. Borchers, M. Cao, and X. Hu, "3dg: A framework for using generative ai for handling sparse learner performance data from intelligent tutoring systems," *arXiv preprint arXiv:2402.01746*, 2024.
- [39] N. Thai-Nghe, L. Drumond, T. Horváth, A. Nanopoulos, and L. Schmidt-Thieme, "Matrix and tensor factorization for predicting student performance," in *International Conference on Computer Supported Education*, vol. 2. SciTePress, 2011, pp. 69–78.
- [40] N. Thai-Nghe, L. Drumond, T. Horváth, A. Krohn-Grimberghe, A. Nanopoulos, and L. Schmidt-Thieme, "Factorization techniques for predicting student performance," in *Educational recommender systems and technologies: Practices and challenges*. IGI Global, 2012, pp. 129–153.
- [41] 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.
- [42] P. Chen, Y. Lu, V. W. Zheng, and Y. Pian, "Prerequisite-driven deep knowledge tracing," in *2018 IEEE international conference on data mining (ICDM)*. IEEE, 2018, pp. 39–48.
- [43] Y. Lu, P. Chen, Y. Pian, and V. W. Zheng, "Cmkt: Concept map driven knowledge tracing," *IEEE Transactions on Learning Technologies*, vol. 15, no. 4, pp. 467–480, 2022.
- [44] J. D. Novak and A. J. Cañas, "The theory underlying concept maps and how to construct them," *Florida Institute for Human and Machine Cognition*, vol. 1, no. 1, pp. 1–31, 2006.
- [45] C. M. Conway and M. H. Christiansen, "Sequential learning in non-human primates," *Trends in cognitive sciences*, vol. 5, no. 12, pp. 539–546, 2001.
- [46] T. Tang and G. McCalla, "Smart recommendation for an evolving e-learning system: Architecture and experiment," in *International Journal on E-learning*, vol. 4, no. 1. Association for the Advancement of Computing in Education (AACE), 2005, pp. 105–129.
- [47] G. Liang, K. Weinling, and L. Junzhou, "Courseware recommendation in e-learning system," in *Advances in Web Based Learning—ICWL 2006: 5th International Conference, Penang, Malaysia, July 19–21, 2006. Revised Papers 5*. Springer, 2006, pp. 10–24.
- [48] N. Thai-Nghe, T. Horváth, L. Schmidt-Thieme *et al.*, "Factorization models for forecasting student performance," in *EDM*. Eindhoven, 2011, pp. 11–20.
- [49] T.-N. Doan and S. Sahebi, "Rank-based tensor factorization for student performance prediction," in *12th International Conference on Educational Data Mining (EDM)*, 2019.
- [50] S. Sahebi, Y. Huang, and P. Brusilovsky, "Parameterized exercises in java programming: using knowledge structure for performance prediction," in *The second Workshop on AI-supported Education for Computer Science (AIEDCS)*. University of Pittsburgh, 2014, pp. 61–70.
- [51] N. Thai-Nghe and L. Schmidt-Thieme, "Multi-relational factorization models for student modeling in intelligent tutoring systems," in *2015 Seventh international conference on knowledge and systems engineering (KSE)*. IEEE, 2015, pp. 61–66.
- [52] X. Wen, Y.-R. Lin, X. Liu, P. Brusilovsky, and J. BarrÃ-a Pineda, "Iterative discriminant tensor factorization for behavior comparison in massive open online courses," in *The World Wide Web Conference*, 2019, pp. 2068–2079.
- [53] C. Wang, S. Sahebi, S. Zhao, P. Brusilovsky, and L. O. Moraes, "Knowledge tracing for complex problem solving: Granular rank-based tensor factorization," in *Proceedings of the 29th ACM conference on user modeling, adaptation and personalization*, 2021, pp. 179–188.
- [54] L. Luo, L. Xie, and H. Su, "Deep learning with tensor factorization layers for sequential fault diagnosis and industrial process monitoring," *IEEE access*, vol. 8, pp. 105 494–105 506, 2020.
- [55] I. Balažević, C. Allen, and T. M. Hospedales, "Tucker: Tensor factorization for knowledge graph completion," *arXiv preprint arXiv:1901.09590*, 2019.
- [56] P. Morales-Alvarez, W. Gong, A. Lamb, S. Woodhead, S. Peyton Jones, N. Pawłowski, M. Allamanis, and C. Zhang, "Simultaneous missing value imputation and structure learning with groups," *Advances in Neural Information Processing Systems*, vol. 35, pp. 20 011–20 024, 2022.
- [57] C. Ma and C. Zhang, "Identifiable generative models for missing not at random data imputation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 645–27 658, 2021.
- [58] L. Zhang, M. Yeasin, J. Lin, F. Havugimana, and X. Hu, "Generative adversarial networks for imputing sparse learning performance," *arXiv preprint arXiv:2407.18875*, 2024.
- [59] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and J. Jackel, "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems*, vol. 2, 1989.
- [60] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.
- [61] S. Yoon and S. Sull, "Gamin: Generative adversarial multiple imputation network for highly missing data," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8456–8464.
- [62] T. Gervet, K. Koedinger, J. Schneider, T. Mitchell *et al.*, "When is deep learning the best approach to knowledge tracing?" *Journal of Educational Data Mining*, vol. 12, no. 3, pp. 31–54, 2020.
- [63] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.
- [64] Y. Wang, D. Li, X. Li, and M. Yang, "Pc-gain: Pseudo-label conditional generative adversarial imputation networks for incomplete data," *Neural Networks*, vol. 141, pp. 395–403, 2021.
- [65] R. S. d. Baker, A. T. Corbett, and V. Aleven, "More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing," in *Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23–27, 2008 Proceedings 9*. Springer, 2008, pp. 406–415.
- [66] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User modeling and user-adapted interaction*, vol. 4, pp. 253–278, 1994.
- [67] A. Essa, "A possible future for next generation adaptive learning systems," *Smart Learning Environments*, vol. 3, pp. 1–24, 2016.
- [68] M. Ramscar, "Learning and the replicability of priming effects," *Current Opinion in Psychology*, vol. 12, pp. 80–84, 2016.
- [69] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
- [70] R. A. Harshman *et al.*, "Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA working papers in phonetics*, vol. 16, no. 1, p. 84, 1970.
- [71] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, "Temporal collaborative filtering with bayesian probabilistic tensor factorization," in *Proceedings of the 2010 SIAM international conference on data mining*. SIAM, 2010, pp. 211–222.
- [72] H. Morise, S. Oyama, and M. Kurihara, "Bayesian probabilistic tensor factorization for recommendation and rating aggregation with multicriteria evaluation data," *Expert Systems with Applications*, vol. 131, pp. 1–8, 2019.
- [73] A. Bora, E. Price, and A. G. Dimakis, "Ambientgan: Generative models from lossy measurements," in *International conference on learning representations*, 2018.
- [74] Z. A. Pardos and N. T. Heffernan, "Modeling individualization in a bayesian networks implementation of knowledge tracing," in *International conference on user modeling, adaptation, and personalization*. Springer, 2010, pp. 255–266.
- [75] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [76] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

- [77] S. Rustad, P. N. Andono, G. F. Shidik *et al.*, "Digital image steganography survey and investigation (goal, assessment, method, development, and dataset)," *Signal processing*, vol. 206, p. 108908, 2023.
 [78] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*. John Wiley & Sons, 2019, vol. 793.

BIOGRAPHY SECTION



Liang Zhang is a PhD candidate in Computer Engineering at the Department of Electrical and Computer Engineering, University of Memphis. He is a research assistant at the Institute for Intelligent Systems and currently a visiting scholar at Educational Testing Service in Princeton, NJ. He was an intern researcher at the Human-Computer Interaction Institute at Carnegie Mellon University from May to August 2023, and at the machine learning group at NEC Laboratories America from May to August 2024. His main research interests lie

in human-computer interaction, including learning analytics, educational data mining, and Generative AI models, particularly in data imputation and data augmentation within AI-education. He got the Best Full Paper Award at the HCII'24.



Jionghao Lin is an Assistant Professor in the Faculty of Education at the University of Hong Kong. Previously, he served as a Postdoctoral Researcher in the Human-Computer Interaction Institute at Carnegie Mellon University, Pittsburgh, PA, USA (2023–2024). He earned his Ph.D. in Computer Science from Monash University, Clayton, VIC, Australia, in 2023. His research interests include learning science, natural language processing, data mining, and applications of generative artificial intelligence in education. His work has been published in

international journals and conferences and recognized with prestigious awards, including the Best Paper Award at HCII'24, EDM'24, and ICMI'19, and the Best Demo Award at AIED'23.



John Sabatini is a Distinguished Research Professor in the Department of Psychology and the Institute for Intelligent Systems at the University of Memphis. He was formerly a Principal Research Scientist in the Center for Global Research, Research & Development Division at Educational Testing Service in Princeton, NJ. His research interests and expertise are in reading literacy development and disabilities, assessment, cognitive psychology, and educational technology, with a primary focus on adults and adolescents. He has been the principal investigator

of Institute of Education Sciences funded grants to develop pre-K-12 comprehension assessments, as part of the Reading for Understanding initiative, and to adapt those assessments for use in adult education programs, as well as co-PI on a grant project that explores how online collaborative, critical discussions can facilitate the writing of arguments in middle grades students.



Diego Zapata-Rivera is a Distinguished Presidential Appointee at Educational Testing Service Research Institute in Princeton, NJ. He earned a Ph.D. in computer science (with a focus on artificial intelligence in education) from the University of Saskatchewan in 2003. His research at Educational Testing Service has focused on the areas of innovations in communicating assessment results to various audiences and technology-enhanced assessment including work on personalized learning and assessment, conversation-based assessment, caring assessment, and game-based assessment. He has produced more than 150 publications including edited volumes, journal articles, book chapters, and technical papers. Dr. Zapata-Rivera was elected as a member of the International AI in Education Society Executive Committee (2022-2027). He is a Co-PI and research co-director of an NSF AI Institute – the INVITE institute (invite.illinois.edu). He is a member of the Editorial Board of User Modeling and User-Adapted Interaction, Associate Editor for IJAIED, AI for Human Learning and Behavior Change, and former Associate Editor of the IEEE Transactions on Learning Technologies Journal. He is a 2024 IEEE Education Society Distinguished Lecturer.



Carolyn Forsyth is currently a Research Scientist at the Educational Testing Service Research Institute working on novel innovations for digital assessments. She earned her Ph.D. in Experimental Psychology with Cognitive Science Graduate Certification from the University of Memphis in 2014. She has focused her research on understanding and creating environments with natural language conversations for both learning and assessment with a current focus on prompt engineering and development of multi-agent computing systems for generative AI.

She also has developed the methodology of theoretically grounded educational data mining to incorporate principled approaches to understanding students processes during interactions with these systems. She has been a key member serving in various roles for both the International Conference on Artificial Intelligence in Education Conference as well as the International Conference on Educational Data Mining for nearly a decade. Furthermore, Dr. Forsyth has contributed over 100 peer-reviewed publications and presentations on this work.



Yang Jiang a Research Scientist at the Educational Testing Service Research Institute in Princeton, NJ. She received the Ph.D. degree in cognitive science in education from Columbia University, New York, NY, USA, in 2018. Her research focuses on developing and applying AI and data science methods to explore how learners interact with technology-based learning and assessment systems, and how the interactions relate to cognition and learning.



John Hollander is an Assistant Professor in the Department of Psychology and Counseling at Arkansas State University. His research interests include psycholinguistics, computational semantics, and the science of reading.



Xiangen Hu is Chair Professor of Learning Sciences and Technologies at Hong Kong Polytechnic University. He received his Doctor of Philosophy degree in Cognitive Psychology from the University of California, Irvine in 1991 and 1993 respectively. He took up an Assistant Professorship at The University of Memphis in 1993 and was promoted to Associate Professor in 2000 and Professor in the Department of Psychology in 2009. Meanwhile, he has been appointed Prof and Dean of the School of Psychology in the Central China Normal University since 2016 on a visiting basis where he has established good connections with the Chinese mainland. He joined PolyU in Dec 2023 as Director of the Institute for Higher Education Research and Development. His research areas include Mathematical Psychology, Research Design and Statistics, and Cognitive Psychology.



Arthur C. Graesser is a Distinguished University Professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis and is an Honorary Research Fellow in the Department of Education at the University of Oxford. He received his Ph.D. in psychology from the University of California at San Diego. Art's primary research interests are in cognitive science, discourse processing, and the learning sciences. More specific interests include knowledge representation, question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, problem solving, memory, emotions, computational linguistics, artificial intelligence, human-computer interaction, and learning technologies with animated conversational agents.

APPENDIX

APPENDIX A: EXAMPLE PROMPT STRATEGY FOR AUGMENTING SPARSE LEARNING PERFORMANCE LEVERAGING GPT-4

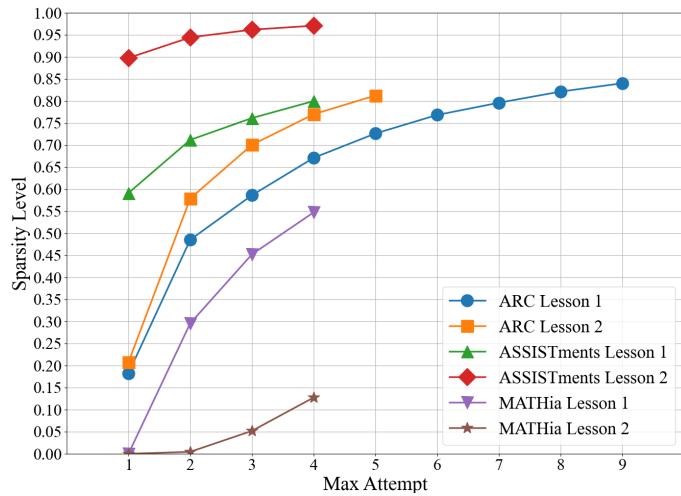


Fig. 7: Sparsity Levels Across Different Lessons and Datasets Based on Max Attempt.

Figure 7 displays the variation in sparsity levels within learning performance data across six lessons, categorized by the maximum number of attempts. Each line represents a different lesson, with data points showing an increase in sparsity as the number of attempts progresses, suggesting a progressive introduction of missing data or non-responses during the learning process. This trend is consistent across all courses, albeit with varying rates of increase. Notably, "ASSISTments Lesson 2" exhibits a gradual ascent, recording the highest sparsity levels across all attempts when compared to other lessons. In contrast, "MATHia Lesson 1" and "MATHia Lesson 2" demonstrate lower initial sparsity levels, with the former experiencing a sharp increase and the latter following a more gradual trajectory as attempts progress. Particularly, "ARC Lesson 1" records the maximum number of attempts observed for this class. The distinct sparsity patterns observed in Figure 7 highlight the heterogeneity of data completeness and the extent of missingness across different lesson datasets. The distinct sparsity patterns underscore the heterogeneity of data completeness and the extent of missingness across different lesson datasets, which impacts the quality of modeling learners' knowledge states and the ITS's ability to provide accurate, personalized instruction.