

Received September 8, 2021, accepted November 11, 2021, date of publication November 16, 2021, date of current version November 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3128578

PAKES: A Reinforcement Learning-Based Personalized Adaptability Knowledge Extraction Strategy for Adaptive Learning Systems

MUHAMMAD ZUBAIR ISLAM¹, (Student Member, IEEE), RASHID ALI¹, (Member, IEEE), AMIR HAIDER¹, MD ZAHIDUL ISLAM², AND HYUNG SEOK KIM¹, (Member, IEEE)

¹Department of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, Republic of Korea

²School of Computing, Mathematics, and Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia

Corresponding authors: Amir Haider (amirhaider@sejong.ac.kr) and Hyung Seok Kim (hyungkim@sejong.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Government of Korea [Ministry of Science and ICT (MSIT)] under Grant 2019R1A4A1023746 and Grant 2019R1F1A1060799, and in part by the Strengthening Research and Development Capability Program of Sejong University.

ABSTRACT Advancements in adaptive educational technologies, specifically the adaptive learning system, have made it possible to automatically optimize the sequencing of the pedagogical instructions according to the needs of individual learners. The crux of such systems lies in the instructional sequencing policy, which recommends personalized learning material based on the learning experiences of the learner to maximize their learning outcomes. However, limited available information such as cognitive, affective states, and competence levels of the learners ongoing knowledge points servers critical challenges to optimizing individual-specific pedagogical instructions in real-time. Moreover, making such decisions policy for every learner with a unique knowledge profile demands a trade-off between learner current knowledge and curiosity to learn next knowledge point. To address these challenges, this paper proposes a personalized adaptability knowledge extraction strategy (PAKES) using cognitive diagnosis and reinforcement learning (RL). We apply the general diagnostic model to track the current knowledge state of the learners. Subsequently, an RL-based Q-learning algorithm is employed to recommend optimal pedagogical instructions for individuals to meet their learning objectives while maintaining equilibrium among the learner-control and teaching trajectories. The results indicate that the learning analytics of the proposed framework can fairly deliver the optimal pedagogical paths for the learners based upon their learning profiles. A 62% learning progress score was achieved with the pedagogical paths recommended by the PAKES, showing a 20% improvement compared to the baseline model.

INDEX TERMS Adaptability recommendation, adaptive learning system, knowledge acquisition, Markov model, general diagnostic models, educational technology.

I. INTRODUCTION

In recent years, educational technologies have gained immediate attention by transforming traditional classroom settings to an online learning environment such as Khan Academy, massive online open courses, and knetwon. The main advantage of these learning systems over classroom learning is the adaptive guideline during the learning trajectories of the learners. Adaptive learning indicates to assist learners with personalized instructional sequencing to maximize learning

progress. The learning system that provides adaptiveness to the learners to learn as per their learning needs is called the adaptive learning system (ALS). Implemented through adaptive algorithms, ALSs lead to personalized instructional sequencing for individual learning attributes [1]. Contrary to the traditional “one-size-fits-all” teaching methodology where a single tutor teaches hundreds of students in a class with the same pedagogical action [2], [3], the ALS focuses on providing a personalized learning experience at an individual level. Currently, ALSs are employed in multiple educational institutes worldwide to train students. For instance, Cognitive Tutor mathematics courses [2] have been used to

The associate editor coordinating the review of this manuscript and approving it for publication was James Harland.

teach around 60,000 students annually, and a similar system called ASSISTment [3] has been employed to train 6000 students annually at various schools. The ALSs use academic information metrics, such as proficiency level, learning time, resulting grades, and response time to uncover hidden patterns about the learner and recommend productive learning activities to optimize the learning outcomes [4].

The core component of ALSs is a pedagogical strategy with various kinds of tracking models to represent the hidden knowledge state of the learner and to deliver new learning actions effectively. Pedagogical strategies are described as instructional sequencing policies to decide what to recommend in the next step as per ongoing available information of both the learners and the learning material. In recent literature, a variety of approaches have been introduced to design these pedagogical strategies for ALSs by employing rule-based [5], machine learning (ML)-based [6], and reinforcement learning (RL)-based method [7], [8]. In a rule-based approach, a domain expert defines a teaching sequence and corresponding learning actions, and the learner acquires knowledge by following them. The drawbacks associated with such an approach for designing strategies are its high cost and domain dependence. The ML models for ALSs have proven to be a significant development for handling many students with diversity in a dynamic environment [9], [10]. The ML models are capable of delivering personalized learning actions to the learners by analyzing their learning experiences. However, ML algorithms, such as the recurrent neural network and long short-term memory, have some limitations during designing strategies and implementation for the ALSs, as they require an enormous amount of knowledge records and high computational overhead. The significant advances in RL models play a vital role in the development of intelligent pedagogical strategies because they have the capability of optimizing the learning gains of the students within an uncertain environment [11]–[13]. However, it still remains an open issue to empower ALSs to make recommendations by giving the right to the learners to explore and review their knowledge structure during the learning session. Additionally, how to optimize the learning sequence with limited knowledge about learner's cognitive and affective states while estimating the learner's ongoing competence level in a specific knowledge point.

One of the emerging research interests in RL-aided ALSs is to incorporate psychological theories along with RL models to tackle the existing challenges [14], [15]. Moreover, ALSs with such hybrid approaches comprise various modules for adaptive teachings, such as the assessment model, learning model, and student model. The task of the assessment model is to question the learners and evaluate their performance to predict their knowledge levels. The learning model receives predictions from the assessment model and develops a relationship between the learning material and assessment results. Thereafter, the learning material is delivered to the learners in accordance with their learning abilities to optimize learning engagement. Finally, the student model is updated to

reflect the current knowledge levels of the learners. A learning model that makes possible task recommendations for the best learning path was proposed in [16].

To enhance the learning outcomes by recommending the next activity, the learning model integrates the learning material with the learners' proficiency levels. Similarly, the studies in [17] and [18] on competency-based recommendation systems cover hybrid approaches to combine data-driven RL models with psychometric assessment models. Such hybrid approaches can efficiently track the learner's knowledge (performance, experience, and degree of achievement) and deliver the best learning items. In these recommendation systems, the assessment model deduces the latent knowledge states of the learners, and RL models play an essential role in making the best decision for personalized task selection. However, the problems of designing an adaptive strategy for ALSs that involves the dynamic nature of the environment and customizing the learning path for the learner's performance remain open. Numerous studies on the potential of ALSs highlight that a robust relationship between student learning and the learning material is required to develop an optimal pedagogical strategy [19]. There is a need to deal with the critical challenge for ALSs to design an optimal strategy that tracks the learners' ongoing knowledge proficiency in an online fashion and recommends the learning material as per the learner's psychological needs while considering limited cognitive and affective state information.

In this paper, a personalized adaptability knowledge extraction strategy (PAKES) is proposed for adaptive assessment and personalization in ALSs to address the aforementioned open research issues. The term adaptability in learning analytics and educational systems refers to providing learners with a choice to personalize their learning materials according to their learning styles, leading toward more learner control [20]. We summarize the contributions of this paper as follows:

- To detect the learners' effective knowledge states and recommend suitable learning materials to achieve intrinsic cognitive learning improvements. The proposed framework incorporates cognitive psychology and RL-induced policies for optimal learning path recommendations using the Markov decision process (MDP).
- For the proposed system, a general cognitive model is used to predict the hidden knowledge states of the learners to provide ongoing personalized expectations during learning sessions to enhance the learning progress even when restricted information is provided.
- To deliver the best instructional sequencing corresponding to the learners' exclusive knowledge attributes estimation, an RL-based Q-learning algorithm is employed.
- To improve the learners' engagement and effective education, the adaptability concept has been utilized to achieve a balance between learner control and system teaching control.
- Experiments reveal that the proposed system PAKES outperforms the traditional approaches by making it

possible for each learner to learn at their own pace, choice, and intrinsic desire for learning even there is a new student with limited information. Hence, it indicates that the above-average learners can learn as per their desire without wasting time, while below-average learners can continuously work to reach a higher level with motivation.

The rest of the paper is organized as follows: Section II presents an overview of the empirical study of the related work. The proposed PAKES algorithm, along with the systematic mechanism, is described in Section III. Section IV demonstrates the experiment analysis to investigate the performance of the proposed framework. Finally, Section V concludes this article and discusses possible future avenues.

II. RELATED RESEARCH WORK

This section presents an overview of the relevant empirical studies on ALSs in artificial intelligence in education. Recent work addressing ML techniques, including RL approaches and content related to cognitive assessment models, are briefly discussed.

In the literature, several approaches to design personalized ALSs have been presented that aim to recommend learning paths to improve the overall learning outcomes of the learners. A comprehensive study on ALSs was presented in [21], where the authors classified the ALSs into five clusters based on RL approaches. The authors designate that RL-induced instructional sequencing proves successful when a combination of learning science and cognitive psychology theories is employed. Such a classification suggests integrating data-driven and theory-driven methods in ALSs for optimal learning instructions.

One study [22] reported an adaptive teaching algorithm for ALSs called the plain vanilla strategy. The strategy infers the mastery and nonmastery knowledge level of the learners for suitable learning path recommendations. The plain vanilla system encodes the understanding of the students for an inquiry to the probability of success to predict whether they have mastered specific knowledge components. The goal of the plain vanilla strategy is to optimize the objective function that estimates the entire gain of the learning process using a multi-armed bandit (MAB) framework with the Gittins index approach. Extensive research has been conducted to employ MAB for ALSs, including the contextual bandit [23] and adversarial bandit [24]–[28]. The Gittins index is one of the most primitive methods to understand the MAB problem and only proves promising when no assessment error exists and the skills are independent of each other in the ALS [29]. An adaptive oracle strategy was proposed in [29], where the information of the transition model is considered to be unknown for the learner, similar to a real-world practice. In particular, an RL model is employed to design optimal pedagogical policies rather than the Gittins index. The oracle strategy uses a dynamic cognitive diagnostic model and a Q-learning algorithm to track the hidden knowledge states

of the learners and estimate the effectiveness of the learning material on knowledge and skill, respectively. However, a critical limitation in the oracle strategy is not considering earlier stopping with the fixed time horizon, which does not make it feasible for fast learners [30]. Therefore, fast learners must wait for the entire trajectory, leading to a source of frustration for the learners. Moreover, the assessment design in the oracle strategy merely uses one point of information to estimate the knowledge profiles of the learners.

Several novel instructional strategies for the ALS that incorporate deep RL approaches with cognitive assessment models have been proposed [30], [31]. The deep Q-network (DQN) recommendation strategy uses a deep Q-learning algorithm to design the optimal learning paths. In particular, a feed-forward neural network was employed to generate an adaptive learning design by approximating the objective function values to optimize sequential policies [30]. The work in [31] further extended the DQN recommendation strategy presented in [30] with a policy network and introduces a new predictive model for inferring the learner's learning curiosity. The system stores student learning experiences and then employs a feed-forward neural network to predict a curiosity reward. The key issue with such a curiosity-driven recommendation strategy (CDRS) is that it requires a specific amount of the learning experience data of students to make an accurate recommendation that is only possible during a long session of interaction with the ALS. During initialization, a lack of a sufficient amount of data in the CDRS leads to low-performance decisions. This discourages the learners during their interaction with the system, thus leading to the abandonment of the system by the learners [32]. Various studies demonstrate that accuracy in the estimation of latent knowledge states affects the performance of the recommended strategies [33], [34].

In ALSs, designing adequate learning material is insufficient for the best learning path recommendations [35]. The best assessment model is required to evaluate individual-specific knowledge levels and needs [36], [37]. An excellent assessment model serves as an integral part of the ALS by inferring the hidden learning capabilities of the students within specific knowledge points. Therefore, it is vital to develop an active instructional policy for ALSs that can be applied to accommodate the learners' needs on knowledge points over time. The main challenge faced in tracking the knowledge competencies of the learners is that a predictive model requires the right amount of information to assess the potential knowledge states of the learners. Most of the currently available ALSs use large datasets containing thousands of learners, which makes it challenging to integrate into a realm where restricted information is obtainable. To address this problem, we propose a pedagogical strategy called PAKES that employs a general diagnostic model online to make precise predictions within a finite time and with limited information. After providing an assessment, it captures the below-average knowledge competencies of the learners and recommends the best learning material to promote the

overall progress during the learning process. The goal of the PAKES is to enable the ALSs to be robust in any event, even when less information is accessible.

III. PROPOSED FRAMEWORK

A. SYSTEM MODEL

The PAKES aims to assess the multiple competence levels of the learners in specific knowledge components and provides personalized pedagogical instructions to individual learners. To achieve this goal, PAKES deals with three significant challenges: (1) the knowledge attributes of the learners change over time as they progress, (2) the same evaluation method is not applicable for each learner under study, and (3) capturing the essential information of the learner under study from the binary responses (yes or no). This section presents the proposed strategy to optimize the learning schedules by quantifying the latent attributes of the learners. The proposed framework integrates an online learning approach and a cognitive measurement model with adaptive psychometric tests to deal with the aforementioned challenges.

In an ALS, we assume there is a learning course with knowledge component K and latent concept knowledge states $s(t)$ of the learner. The total states based on the K knowledge components can be given as $S = \{s_1(t), s_2(t), s_3(t), \dots, s_K(t)\}$, within a finite learning time step $t \in [0, T]$. The hidden knowledge states of the learner are unobservable directly, whereas the performance of the learner corresponding to the knowledge states is observable. Therefore, a knowledge state of the learner is represented as a sequence of competence levels given as $C_{s(t)} = (c_{s(t)}^1, c_{s(t)}^2, c_{s(t)}^3, \dots, c_{s(t)}^n)$, where n indicates the total number of competencies involved in this state. These competencies represent the set of skills and proficiency levels of the learner. With this central fact, the first step is to estimate the knowledge state by modeling the degree of proficiency of the learners. The proposed system conducts adaptive testing to uncover the knowledge states and captures the necessary knowledge concepts. A general multidimensional adaptive (GenMA) model [38] is used as an assessment model to reveal the learning states of the learner by forecasting learner performance. This assessment model uses task difficulty with the required competencies in the corresponding knowledge components as an adaptive test. During adaptive testing, if the learner answers the question correctly, $c_{s(t)}^n$ is equal to 1; otherwise, it is 0. The adaptive assessment module returns a list of passed and failed competencies $\hat{C}_{s(t)} = \{C_{s(t)}^P, C_{s(t)}^F\}$, which represent the estimated ability levels of the learner concerning the corresponding knowledge state. Moreover, $C_{s(t)}^P$ refers to the total passed competencies, and $C_{s(t)}^F$ represents the failed competencies of the learner. Table 1 presents a summary of the notation used in this paper. Fig. 1 represents the overview of the learning path recommendation process of the proposed system for adaptive learning. The learner moves into the next succeeding state after learning the recommended learning material proposed through personalized

TABLE 1. Summary of notations.

Notation	Definition
K	Set of knowledge components
S	Set of learners state space
A	Set of action space
$s(t)$	Student knowledge state s at time t , $s(t) \in S$
$a(t)$	Learning action a at time t , $a(t) \in A$
$r(t)$	Reward r at time t
$\pi(a_t s_t)$	Policy for state $s(t)$ with action $a(t)$
$V^\pi(s)$	Expected long-term return for state s under policy π
$V^{\pi^*}(s)$	Expected long-term return under optimal policy π^*
$C_{s(t)}$	Set of competence levels
$\hat{C}_{s(t)}$	List of estimated competencies
$C_{s(t)}^P$	Total number of pass competencies at state $s(t)$
$C_{s(t)}^F$	Total number of fail competencies at state $s(t)$
$\Pr(x_{ij})$	Success probability, for learner i with item j
Φ	IRT logistic function
B	Total involved knowledge competencies during test
β_i	Main learning ability of the learner i
θ_{ib}	Learner i ability for knowledge competence b
q_{jb}	Q-matrix cell entry, for question q with $b \in B$
d_{jb}	Difficulty of question q for $b \in B$
α	Learning rate
γ	Discount factor
$Q(a_t, a_t)$	Q-value for state $s(t)$ at action $a(t)$
L_{ps}	Learning progress score
$W_{a(t)}$	Learning weights, for action $a(t)$
$L_{s(t)}^s$	Learning achievement in state $s(t)$ with action $a(t)$

learning actions (PLAs), which is depicted as follows:

$$s_i(t) \xrightarrow{PLA(a_t)} s_j(t+1),$$

where $s_i(t)$ refers to the learner's i^{th} current state at time t , and $s_j(t+1)$ indicates the j^{th} next latent knowledge state. Based on the assessment results of the GenMA model, the proposed system ranks the failed competencies by the learners starting with the most failed competence on top. The adaptability recommendation strategy takes suitable action from the action set to recommend the learning material corresponding to the current understanding of the individual learners.

In this paper, we formalize the adaptability recommendation learning process as an MDP [39], where the adaptability recommendation strategy works as an intelligent agent. The adaptability agent interacts with the learners as a sequential decision-making procedure to optimize the pedagogical strategy. According to the MDP, the successor state $s_j(t+1)$ of the learner depends only on the current state $s_i(t)$ rather than the entire prior learning history $\{s_1, s_2, s_3, \dots, s_{(i-1)}\}$. In this stochastic process, when a learner interacts with the system during learning episodes, the learning records of the learner are generated as a sequence of estimated competencies, actions, and rewards $\{\hat{C}_1, a(1), r(1), \hat{C}_2, a(2), r(2), \dots, C_{(T-1)}, a(T-1), r(T-1), \hat{C}_T, a(T), r(T)\}$. In each episode, the proposed system takes action $a(t)$ to recommend the learning materials corresponding to the instructional policy $\pi(a_t|s_t) = P[A = a_t|S = s_t]$. The learners select the material of their own choice to master a specific task, and the agent receives a

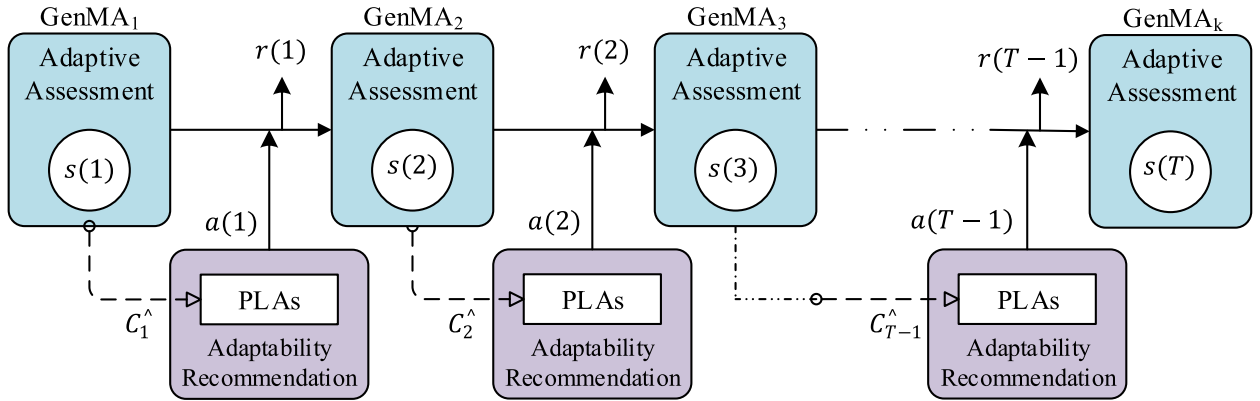


FIGURE 1. Framework for the personalized adaptability-recommendation system: (1) adaptive assessment model GenMA estimates the knowledge state $s_j(t)$ by modeling degree of competencies of the learner as $\hat{C}_s(t)$; (2) following the policy π the system recommends learning material by taking action $a(t)$ from personalized learning action space; and (3) after receiving reward $r(t)$ and taking learning actions system transits the learner into next state $s_j(t+1)$.

reward $r(t)$. The goal is to find the policy value, $V^\pi(s) = \mathbb{E}_\pi[\sum_{t=1}^{T-1} r(t)]$ of the recommendation for the unknown state transition model for individual learners and then maximize it using the optimal value $V^{\pi^*}(s) = \max_\pi V^\pi$ [40], [41]. To accomplish this, we integrated a GenMA model with the Q-learning algorithm to predict the success rate and maximize the learning gain over time in an uncertain environment. Moreover, the proposed strategy works in an online fashion and does not depend on students' previous learning experience data to represent knowledge proficiency. Overall, the goal of the proposed system is to recommend the optimal personalized action to the learners given the current estimated knowledge state to increase their learning outcomes on the following adaptive test. The adaptability recommendation strategy aims to empower the learners to improve knowledge competencies while using ALS.

B. Q-LEARNING TECHNIQUE

For a long time, the Q-learning algorithm and its variants have been proven to be robust RL approaches to solving the MDP problem in the ML community [42]–[46]. In addition, Q-learning is an off-policy technique and applies to running on any scheme in the MDP framework. In the Q-learning method, a parametric mechanism is used to approximate the Q-function of the current control strategy. Then, the equivalence principle is applied to improve the procedure, from which the optimal policy is selected for the action network. The Q-learning algorithm approximates the value function, and it is defined as follows:

$$Q^N(s_t, a_t) = Q^o(s_t, a_t) + \alpha(r_t + \gamma \times \max_a Q^o(s_{t+1}, a) - Q^o(s_t, a_t)), \quad (1)$$

where $Q^o(s_t, a_t)$ is the old value, and $Q^N(s_t, a_t)$ is the newly calculated value in the Q-table. Moreover $\alpha \in (0 < \alpha \leq 1)$ is a nonnegative step-size coefficient called the learning rate, r_t is the immediate reward after applying action a_t in state

s_t , and γ is the discount factor used to measure the significance of the future reward. Finally, $Q'(s_{t+1}, a)$ is used to estimate the optimal value for the successor state. The Q-table is similar to the lookup table, where rows represent states, and columns indicate actions. Moreover, it is used to store state-action pair values called Q-values and updates the Q-values after training.

C. GENERAL MULTIDIMENSIONAL ADAPTIVE ASSESSMENT MODEL

The GenMA model is a hybrid adaptive assessment model that incorporates general diagnostics [47] and multidimensional item response theory model [48], [49]. In [38], the authors proposed the GenMA by using a general diagnostic model for partial credit data, which is defined as follow:

$$Pr(x_{ij}|i \text{ selects } j) = \Phi(\beta_i + \sum_{b=1}^B \theta_{ib} q_{jb} d_{jb}) \quad (2)$$

where $Pr(x_{ij})$ is the probability of the proficiency of learner i , $\Phi : z \rightarrow 1/(1+e^{-z})$ is the logistic function, β_i represents the aptitude of learner i , and B is the total number of knowledge competencies included in the adaptive test. Moreover, θ_{ib} is the ability of learner i for the required competence b , and q_{jb} is the cell (j, b) entry in the Q-matrix. Table 2 represents the Q-matrix and provides the description of the knowledge competencies corresponding to Tatsuoka's fraction subtraction [50] dataset, and hierarchical knowledge map of the leading online learning platform, the Khan Academy.¹ The involved competencies in the pedagogical testing are represented by *TRUE*, whereas *FALSE* indicates their absence. In this paper, an expert-defined Q-matrix with some contribution to extending the number of skills ($C = 16$) was used, as depicted in Table 2. The parametric value d_{jb} is the difficulty of the test question j for competence b , which is well-calibrated using historical data and the Metropolis-Hastings Robbins-Monro (MH-RM) algorithm [51]. The

¹<https://www.khanacademy.org/exercisedashboard>

TABLE 2. The q-matrix corresponding to khan academy knowledge map simple addition and Tatsuoka (1984) fraction subtraction dataset with 20 questions and 16 knowledge competencies.

		Knowledge Competencies (<i>T</i> = True, <i>F</i> = False)															
		<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>	<i>C9</i>	<i>C10</i>	<i>C11</i>	<i>C12</i>	<i>C13</i>	<i>C14</i>	<i>C15</i>	<i>C16</i>
<i>Q1</i>		<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>
<i>Q2</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q3</i>		<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q4</i>		<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q5</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>
<i>Q6</i>		<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q7</i>		<i>T</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q8</i>		<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q9</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>
<i>Q10</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>
<i>Q11</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q12</i>		<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>
<i>Q13</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q14</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q15</i>		<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q16</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q17</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q18</i>		<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>F</i>
<i>Q19</i>		<i>T</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>
<i>Q20</i>		<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>

Description of knowledge competencies
C1 : Counting objects 1
C2 : Making 5
C3 : Counting objects 5
C4 : Add within 5
C5 : Add within 10
C6 : Subtract within 5
C7 : Subtract within 10
C8 : Making small numbers in different ways
C9 : Convert a whole number to a fraction
C10 : Separate a whole number from a fraction
C11 : Simplify before subtracting
C12 : Find a common denominator
C13 : Borrow from whole number part
C14 : Column borrow to subtract the second numerator from the first
C15 : Subtract numerators
C16 : Reduce answers to simplest form

comprehensive overview of the GenMA model at the parametric level is depicted in Table 3.

D. SYSTEM MECHANISM FLOWCHART

Fig. 2 presents an overview of the systematic implementation of the adaptive test generation and evaluation of the learner performance using the GenMA model. Furthermore, to cope with the failed competencies of the learner and to provide choice control for different learning activities, the adaptive test is updated based on the present performance. In the proposed mechanism, the learners go through the adaptive benchmark assessment method, where their responses to questions are encoded to the success rate. Each item in the adaptive test belongs to a specific knowledge component that predicts the learners’ knowledge competencies during the learning sessions that they endeavor to master. The assessment module asks the next question using several types of input parametric information, such as the difficulty of the item, knowledge components, and learner’s knowledge attributes. Afterward, based on the achievements, the system promotes the learner to a pedagogical decision-process module to recommend learning resources. Based on the ongoing evaluation, the system identifies the most unsatisfactory skill for each learner. To improve the overall learning experience, the system then suggests learning material that corresponds to the learners’ understanding, needs, and style. The Q-table is updated corresponding to the knowledge state of the learner, which represents the individual learning characteristics. The core components of the proposed system are the GenMA model and adaptability recommendation modules. Contrary to the ordinary adaptive tests in the GenMA model, the maximum number of tasks asked during the test is defined in advance to avoid the frustration of the learners. In PAKES, a total of eight questions ($j = 8$) were chosen during the adaptive testing using the GenMA model to obtain interactive

Algorithm 1 Q-Learning Algorithm for PAKES Mechanism

```

Input: Initialization of knowledge profile state ( $S_t$ )
action set  $A$ , learning rate  $\alpha$ , discount factor  $\gamma$ ,
exploration probability  $\epsilon$ 
Output: Optimal Q function value
Initialization of weight:  $W_{a(t)}$ 
Initialization of Q-table with zeros
while learning do
  for  $t = 0, \dots, T - 1$  do
    Estimate competencies  $\hat{C}_{s(t)} = \{C_{s(t)}^P, C_{s(t)}^F\}$ 
    using assessment model GenMA
    Select an action  $a_t \leftarrow \operatorname{argmax} q(s_t, a_t)$  with
    probability  $1 - \epsilon$  and otherwise explore using
    system estimation
    Recommend learning material ( $a_t$ ) as per
    estimated failed competencies  $C_{s(t)}^F$ 
    Learners select material ( $a_t$ ) as per their desire:
    Adaptability selection
    Transit into next state:
     $s(t + 1) = 1 - (1 - s(t)) \odot \exp\{-\xi \cdot W_{a(t)} \odot L_{a(t)}^{s(t)}\}$ 
    Calculate reward  $r(t)$ :
    The qualifying failed competencies becomes
    adaptability reward  $r(t) = L_{a(t)}^{s(t)}$ 
    Calculate Q-value
     $Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_t + \gamma$ 
       $\times \max_a Q(s_{t+1}, a) - Q(s_t, a_t))$ 
    Update Q-table
  end
end

```

information for the prediction rate. The pseudocode of the PAKES framework for the selection and recommendation of the optimal material is depicted in Algorithm 1.

TABLE 3. A detailed description of the GenMA parameters corresponding to the q-matrix simple addition and fraction subtraction test question (Q) along with competencies (C).

	C1	C2	C3	C4	C5	C6	C7	C8
Q1	011q11d11			014q14d14		016q16d16	017q17d17	
Q2		012q22d22		014q24d24			017q27d27	
Q3			013q33d33	014q34d34			017q37d37	
Q4		012q42d42	013q43d43		015q45d45		017q47d47	
Q5		012q52d52	014q54d54				017q57d57	018q58d58
Q6							017q67d67	
Q7	011q71d71	012q72d72					017q77d77	018q78d78
Q8							017q87d87	
Q9		012q92d92						
Q10		012q102d102			015q105d105		017q107d107	018q108d108
Q11		012q112d112			015q115d115		017q117d117	
Q12							017q127d127	018q128d128
Q13		012q132d132		014q134d134	015q135d135		017q137d137	
Q14		012q142d142					017q147d147	018q148d148
Q15	011q151d151						017q157d157	
Q16		012q162d162					017q167d167	
Q17		012q172d172			015q175d175		017q177d177	018q178d178
Q18		012q182d182			015q185d185	016q186d186	017q187d187	
Q19	011q191d191	012q182d182	013q193d193		015q195d195		017q197d197	
Q20		012q202d202	013q203d203		015q205d205		017q207d207	018q208d208
	C9	C10	C11	C12	C13	C14	C15	C16
Q1				0112q112d112		0114q114d114	0115q115d115	
Q2				0112q212d212			0115q215d215	
Q3				0112q312d312			0115q315d315	
Q4		0110q410d410	0111q411d411		0113q413d413		0115q415d415	
Q5		0110q510d510		0112q512d512			0115q515d515	0116q516d516
Q6							0115q615d615	
Q7	019q79d79	0110q710d710					0115q715d715	
Q8							0115q815d815	
Q9		0110q910d910						
Q10		0110q1010d1010			0113q1013d1013		0115q1015d1015	0116q1016d1016
Q11		0110q1110d1110			0113q1113d1113		0115q1115d1115	
Q12							0115q1215d1215	0116q1216d1216
Q13		0110q1310d1310		0112q1312d1312	0113q1313d1313		0115q1315d1315	
Q14		0110q1410d1410					0115q1415d1415	
Q15	019q159d159						0115q1515d1515	
Q16		0110q1610d1610					0115q1615d1615	
Q17		0110q1710d1710			0113q1713d1713		0115q1715d1715	
Q18		0110q1810d1810			0113q1813d1813	0114q1814d1814	0115q1815d1815	
Q19	019q199d199	0110q1910d1910	0111q1911d1911		0113q1913d1913		0115q1915d1915	
Q20		0110q2010d2010	0111q2011d2011		0113q2013d2013		0115q2015d2015	

IV. PERFORMANCE EVALUATION AND RESULTS

A. SIMULATION SETUP

We have developed our simulation environment in Python and R programming language on a computer equipped with Intel Core i7 processor, 16GB running memory, CPU frequency 3.6GHz 64bit Ubuntu 18.04.5 operating system. The GenMA model is developed using R based mirt package, the actor-critic algorithm is build using the PyTorch framework, and the Q-learning algorithm is built upon Python. As we mentioned above in III-C we utilized the Tatsuoka’s fraction subtraction [50] dataset with some addition in skill set. The dataset includes dichotomous responses of 536 school students over 20 questions, and 5-fold cross-validation was performed as suggested by authors in [51]. Table 4 enlist the parameters and their corresponding values used in the simulation environment.

B. BASELINE MODEL

A robust adaptive learning model [31] was used as a baseline to compare the effectiveness of the learning path

TABLE 4. Simulation parameters.

Parameters	Values
Simulation environment	Python, R
Learning rate (α)	0.1, 0.7, 0.8, 0.9
Discount factor (γ)	0.1, 0.7, 0.8, 0.9
Exploitation/exploration factor (ϵ)	0.5
Finite time episodes (T)	40
No. of epochs	1–1000
Action space (A)	20
State space (S)	40
Set of competencies $C_{s(t)}$ in each state	16
Test questions (j) in GenMA	8

recommendations made by the PAKES. In [31], the authors proposed a CDRS using an RL technique with cognitive diagnostic models. The CDRS follows the three-parameter logistic (3PL) model from the multidimensional item response theory to measure the knowledge state of the learner at the initial stage. In the later stage, it applies a data-driven RL actor-critic model [52] to recommend personalized learning material. For the CDRS, a predictive model is developed

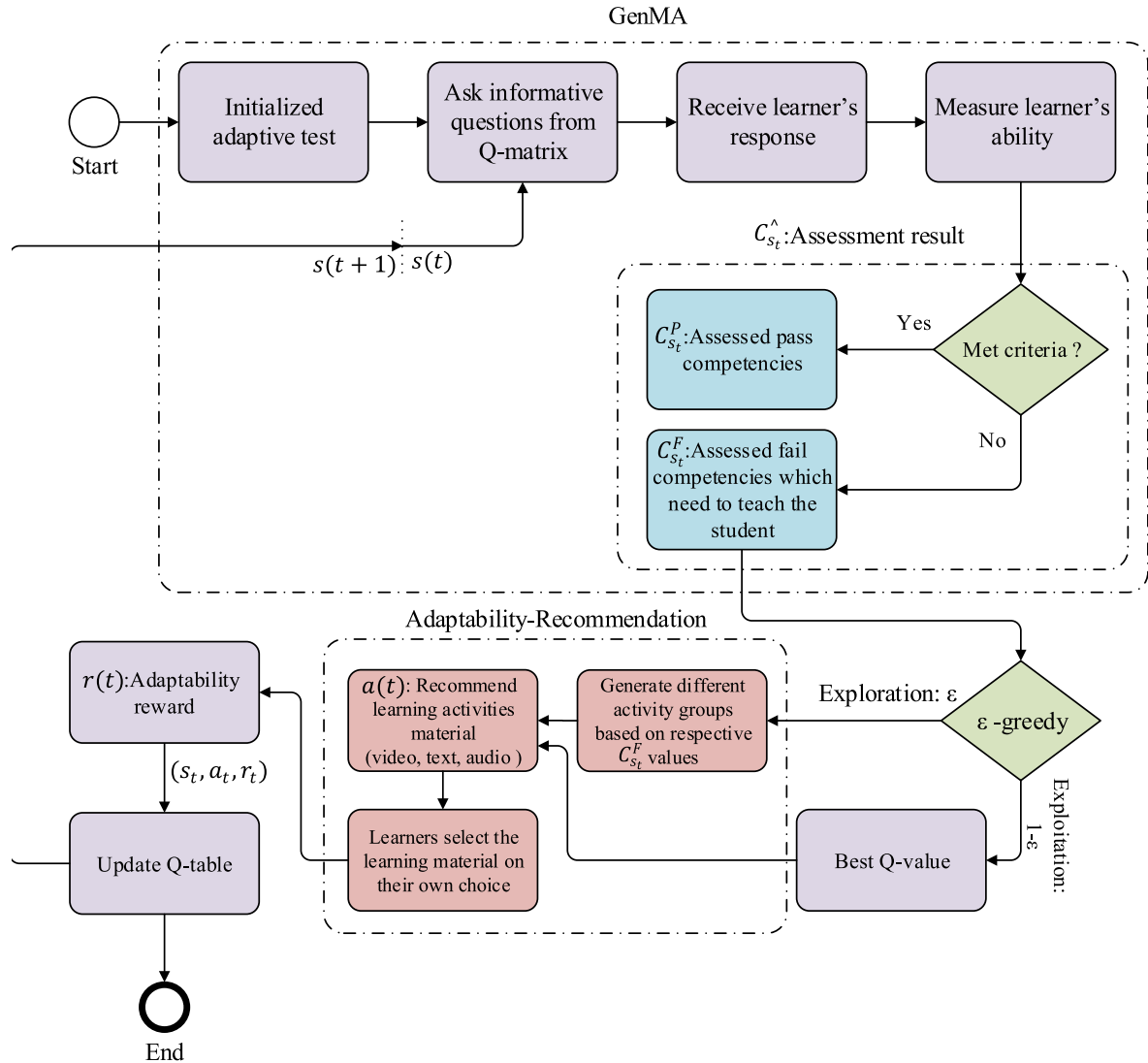


FIGURE 2. The flowchart illustrates the general flow of the entire system dynamics with the crux of the personalization, assessment model GenMA and adaptability-recommendation.

using a feed-forward neural network to return a prediction error. This error serves as a curiosity reward of the learner to the corresponding knowledge state. Moreover, this curiosity reward is used by the RL model to approximate the objective function and design an optimal recommendation strategy. After specific iterations of the trajectories of the learner’s interaction data, the CDRS follows the learner’s knowledge attributes and recommends adaptive material to the individual.

C. EVALUATION METRICS

A simple final exam prototype has been employed to measure the efficacy of the proposed framework for the ALS. The evaluation setup adopted in this work proceeds identically to the one defined in [31]. A simple promotion of knowledge competencies is employed as an evaluation metric to compare the learning path recommendations. A similar evaluation strategy is employed by several educational institutes to evaluate the

performance of the students [53], [54]. The performance metrics to evaluate the learning rate of the students is given as follows:

$$L_{ps} = 100 \times s(T), \tag{3}$$

where $s(T)$ refers to the final state of the learner after interaction with ALSs, and L_{ps} indicates the learning progress score of the learner in this state. The L_{ps} is employed to quantify the performance gains offered by the PAKES over the baseline model. The effectiveness of the proposed framework is elaborated using a combination of empirical graphs and parametric analysis.

D. EXPERIMENT RESULTS

1) PARAMETER EVALUATION

It is important to discuss the role of the transition model employed in the proposed framework, which specifies the learning progress of the learners through their learning

TABLE 5. Learning material and training weights for tutoring individual competencies.

Learning material		Training weight to the corresponding competencies																Competencies to be trained by learning material
		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	
1	W_1	0.06	0.00	0.00	0.09	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.09	0.00	0.02	0.01	0.00	$C1$
2	W_2	0.00	0.06	0.00	0.11	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.01	0.00	$C2$
3	W_3	0.00	0.00	0.06	0.11	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.01	0.00	$C3$
4	W_4	0.00	0.06	0.02	0.00	0.02	0.00	0.02	0.00	0.00	0.06	0.02	0.00	0.02	0.00	0.02	0.00	$C8$
5	W_5	0.00	0.06	0.00	0.03	0.00	0.00	0.02	0.01	0.00	0.06	0.00	0.03	0.00	0.00	0.02	0.01	$C1, C3$
6	W_6	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	$C2, C4$
7	W_7	0.09	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.09	0.02	0.00	0.00	0.00	0.00	0.01	0.00	$C2, C5$
8	W_8	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.06	$C2, C6$
9	W_9	0.04	0.05	0.01	0.00	0.00	0.00	0.02	0.00	0.04	0.05	0.01	0.00	0.00	0.00	0.02	0.00	$C2, C7$
10	W_{10}	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	$C9, C10$
11	W_{11}	0.00	0.06	0.00	0.00	0.03	0.00	0.02	0.01	0.00	0.06	0.00	0.00	0.03	0.00	0.02	0.01	$C5, C6, C7$
12	W_{12}	0.00	0.05	0.00	0.00	0.04	0.02	0.01	0.08	0.00	0.05	0.00	0.00	0.04	0.02	0.01	0.00	$C6, C7, C9$
13	W_{13}	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.06	$C1, C3, C12$
14	W_{14}	0.05	0.03	0.00	0.30	0.10	0.00	0.00	0.00	0.05	0.03	0.00	0.30	0.10	0.00	0.00	0.00	$C1, C3, C7, C13$
15	W_{15}	0.11	0.05	0.00	0.00	0.00	0.00	0.01	0.05	0.11	0.05	0.00	0.00	0.00	0.00	0.01	0.05	$C1, C2, C14$
16	W_{16}	0.30	0.80	0.0	0.00	0.00	0.00	0.01	0.06	0.30	0.80	0.00	0.00	0.00	0.00	0.01	0.00	$C1, C2, C15$
17	W_{17}	0.00	0.05	0.00	0.00	0.03	0.00	0.02	0.00	0.00	0.05	0.00	0.00	0.03	0.00	0.02	0.00	$C4, C5, C8, C9$
18	W_{18}	0.00	0.02	0.00	0.00	0.06	0.00	0.04	0.00	0.00	0.02	0.00	0.00	0.06	0.00	0.04	0.00	$C1, C3, C7, C11$
19	W_{19}	0.08	0.02	0.02	0.01	0.01	0.00	0.00	0.00	0.06	0.02	0.02	0.01	0.01	0.00	0.00	0.00	$C1, C2, C6, C10$
20	W_{20}	0.08	0.00	0.02	0.04	0.00	0.02	0.04	0.09	0.00	0.00	0.02	0.04	0.00	0.02	0.04	0.00	$C1, C3, C7, C16$

experiences during each episode. Learning materials that correspond to the action space and their respective weights ($W_{a(t)}$) for all knowledge competencies are defined in Table 5. The transition model employed in this work follows the guidelines described in [31] with a slight variation in the evaluation parameters. The employed transition model considers the current knowledge state $s(t)$ and the probability of success to predict the next knowledge state $s(t+1)$. The probability of success is defined as the learner’s ability to acquire the next knowledge state. The proposed framework maximizes the probability of success for each learner by identifying which failed competencies the learner should address first by taking the corresponding adaptability action $a(t)$. The output of the transition model is written as follows:

$$s(t + 1) = 1 - (1 - s(t)) \odot \exp\{-\xi \cdot W_{a(t)} \odot L_{a(t)}^{s(t)}\}, \quad (4)$$

where ξ denotes the chi-squared distribution of X_2^2 , a random variable with m degrees of freedom. The value of m specifies the type of learners with different learning profiles and is used in various transition models for individual learners [30], [31]. In this work, $m = 2$ was employed. In addition, $L_{a(t)}^{s(t)}$ indicates the acquired failed competencies of the learner at state $s(t)$ after the learning adaptability recommendation $a(t)$. From (4), it can be inferred that, at the initial stage, the learners would feel comfortable and make quick progress, as their knowledge state is low and they are assigned with easy initial tasks. However, it becomes more challenging for the learners in the later stage to progress, as their learning proficiency is increases, and they struggle to show significant improvement in learning outcomes [55]. When the learner transits from one knowledge state to the next, the progress in the learning proficiency is indicated by (ΔLP).

To show the efficacy of the proposed framework, we consider a learner with an initial knowledge state $s(1) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ along with a sequence of competencies ($\hat{C}_{s(1)} = 16$), as estimated by the GenMA model. The learners gain knowledge of their failed and unsatisfactory competencies within finite time episodes ($T = 40$). The Q-matrix depicted in Table 2 is employed to instruct the learners based on their personalized educational proficiency. An action space A with a total of 20 actions, constituting 20 learning materials was created to train the learners. Among the 20 learning materials, four materials were linked with one correlated knowledge competence, six materials were linked with two competencies, six materials were linked with three competencies, and the remaining four materials relate to four competencies. Learning actions with knowledge competencies and corresponding k -dimensional training weights ($W_{a(t)}$) are depicted in Table 5. The episodic learning progress of the learners after interacting with the proposed framework is calculated using (4). The proposed PAKES aims to analyze the learners’ educational experiences and predict the competencies state $\hat{C}_{s(t)}$ by employing the GenMA model with test questions ($j = 8$), and it delivers the adaptability actions.

We study the performance of the proposed framework under different parameter settings. Two weighted tuning parameters (α, γ) were used to adjust the learning of the recommending agent in the Q-learning algorithm, as defined in (2). The values of α and γ , lying between a range of 0.0 to 1.0, were described in the proposed framework in terms of the learning progress. The parameter α represents the learning of policies recommended by the proposed framework. The value of learning rate α depends on the learner’s progress in knowledge attributes from the learning activities. It was employed

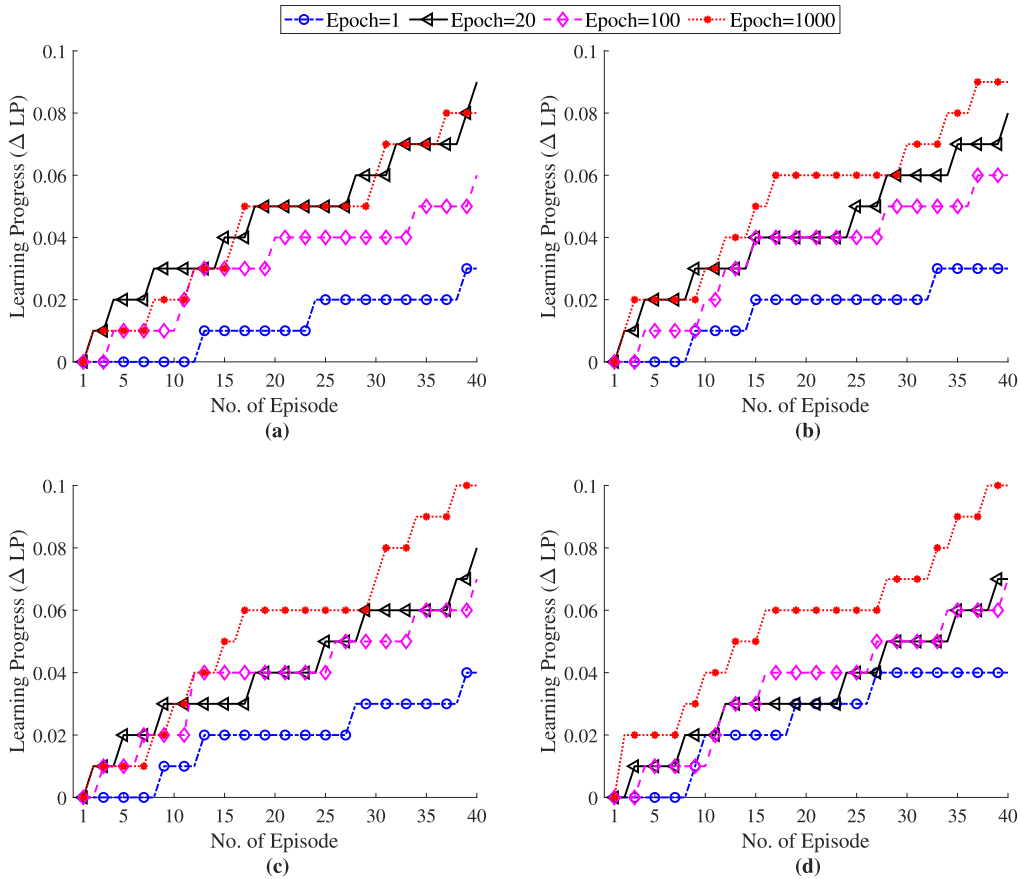


FIGURE 3. System progress with different learning rate values; (a) learning rate = 0.1, (b) learning rate = 0.7, (c) learning rate = 0.8, (d) learning rate = 0.9.

to determine how much change in the Q-value of the states ($s_{t+1} - s_t$) is useful to accept. A lower learning rate and more change in the Q-value indicates an easy task for a learner to progress. The learning rate value $\alpha = 0$ indicates no learning for the recommending agent, and $\alpha > 0$ implies improvement in recommendation policy. In practice, the learning rate values from 0.1 to 0.9 were adopted. The value $\alpha = 1$ directs the recommending agent to ignore the prior knowledge and just consider the current information, which specifies that no exploration was done based on prior knowledge. The proposed framework was simulated by varying the learning rate with the conventionally used values of 0.7 to 0.9. To show the effect of a lower learning rate, $\alpha = 0.1$ was also investigated. The parameter γ indicates the importance of the future reward for the learner. For the proposed framework, $\gamma = 0$ was employed, which designates that the reward is based only on the ongoing performance of the learners.

The performance of the proposed PAKES in terms of episodic learning is depicted in Fig. 3. The system performance is evaluated using the above-mentioned parameters. To analyze the effect of variation in α and the number of training epochs on the system performance, the proposed framework was simulated for $\alpha = 0.1, 0.7$ to 0.9 , and 1 to 1000 training epochs. The trend with an increase in α and the number of training epochs can be perceived

comprehensively for the episodic learning progress of the learner, where the learning progress of the learners increases from 0 to 0.1 as the number of episodes T increases from 1 to 40. The episodic learning of the proposed system for $\alpha = 0.1$ to 0.9 is presented in Fig. 3(a)-(d), respectively. In Fig. 3(a) with $\alpha = 0.1$, the value of the learning progress ΔLP increases with an increase in the number of episodes. In addition, the increase in the number of training epochs results in an increased value of learning progress ΔLP after completion of 40 episodes. However, by employing a lower α value, the system is prone to errors, as the proposed framework seems to make random recommendations for learning. In particular, at episode $T = 3$ to 40 , the recommended learning path at 20 epochs exhibits a higher learning progress ΔLP value than 100 epochs. This trend is perceived as an error, because more training epochs should result in a better learning rate for the learners. A similar trend can be observed for $T = 3$ to 15 and at $T = 29$, where (ΔLP) is higher at 20 epochs, when compared to 1000 epochs. For a higher value of $\alpha = 0.7$ to 0.9 , the results indicate that the proposed framework learning path recommendations suggest the best learning activities to the learners and tend to push them towards their learning targets. Fig. 3(b) and 3(c) with $\alpha = 0.7$ and 0.8 illustrate the improvement in the recommendation policies at different epochs. In particular, the proposed

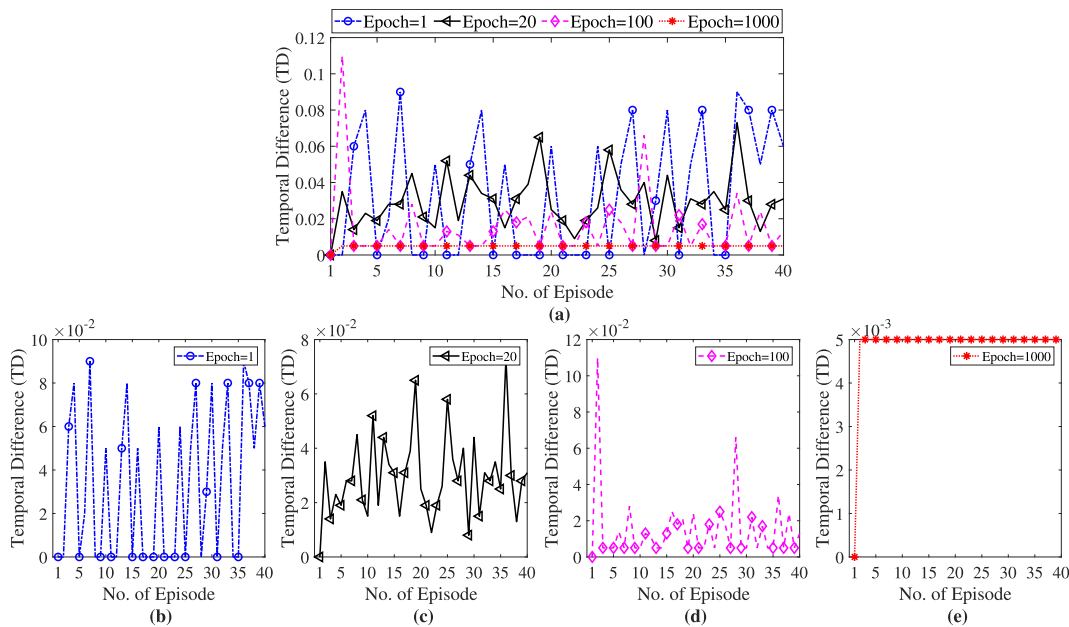


FIGURE 4. Comparing temporal difference error for parameter value $\alpha = 0.1$ and number of epochs; (a) Epoch = 1, 20, 100, 1000 (b) Epoch = 1 (c) Epoch = 20 (d) Epoch = 100 (e) Epoch = 1000.

framework after training at 20, 100, and 1000 epochs shows less randomness and fast convergence. However, a similar trend of errors is dominant, especially for the results for 20 and 100 epochs, where lower epoch numbers exhibit better learning progress.

To observe the effect of the increased learning rate on the error, Fig. 3(d) illustrates the episodic learning progress of the learners at $\alpha = 0.9$. The analysis depicted in Fig. 3(d) indicates that, with a higher value of the learning rate α , the proposed framework fine-tunes its policies for the adaptability of the learning path recommendations. This analysis demonstrates that, at the very beginning, the proposed framework attempts to learn the hidden attributes of the learner’s knowledge state to recommend an optimal learning path. This is the fundamental logic that, at one epoch, the proposed framework does not recommend the best learning path. The proposed framework performs better with $\Delta LP = 0.04$ for episode $T = 30$ at one epoch compared to other results with different learning rate α values, such as $\Delta LP = 0.03 - 0.03$, for episode $T = 30$. Fig. 3(d) shows that the proposed framework with $\alpha = 0.9$ shows the minimum error compared to the results in Fig. 3(b) and Fig. 3(c) with $\alpha = 0.7$ and 0.8 . In summary, the results demonstrate that the proposed framework makes the best learning path recommendations with a higher learning rate $\alpha = 0.9$ and number of training epoch = 1000. These results also illustrate that, during a long learning session, the proposed framework allows the learner to explore and review more learning concepts for overall learning progress.

2) LEARNING CONVERGENCES

In this section, we present the analysis of the temporal difference (TD) error made by the proposed framework during

interaction with the learners over time to study the convergence and further confirm the value of the employed learning rate α . The target is to achieve good convergence with a lower number of training epochs so that the computational overhead during the online phase can be reduced. The TD error is defined in (2) and computed as follows:

$$\Delta Q = TD(s, a) = r_t + \gamma \times \max_a Q'(s_{t+1}, a) - Q^o(s_t, a_t), \quad (5)$$

where $r_t + \gamma \times \max_a Q'(s_{t+1}, a)$ indicates the TD target of the newly calculated Q-value, whereas $Q^o(s_t, a_t)$ is the prior recorded Q-value for the state s_t in Q-table. Moreover, ΔQ is known as the change in the Q-value and is estimated by subtracting the previous value from the target value.

To observe the effect of a lower value $\alpha = 0.1$ on the error rate, the TD errors values for the proposed framework are summarized in Fig. 4(a). This analysis identifies the reasons for the lower convergence rate for the proposed framework with a lower learning rate $\alpha = 0.1$ value as the number of training epochs increases. Fig. 4(a) demonstrates that, with an increase in training epochs, the TD error tends to decrease over the episodic learning progress. The higher epoch values exhibit lower TD error values compared to the lower epoch values. To elaborate on this decrement in the TD error, Fig. 4(b)-(e) represents the TD error for the training epoch values of 1, 20, 100 and 1000, respectively. The TD error exhibits a decreasing trend from 9×10^{-2} to 5×10^{-3} , as the training epochs increase from 1 to 1000. However, this reduction in the TD error values is still not significant, resulting in a higher learning progress ΔLP at a lower epoch value, particularly for epoch values 20 and 100, as shown in Fig. 3(a), where the TD error has similar values.

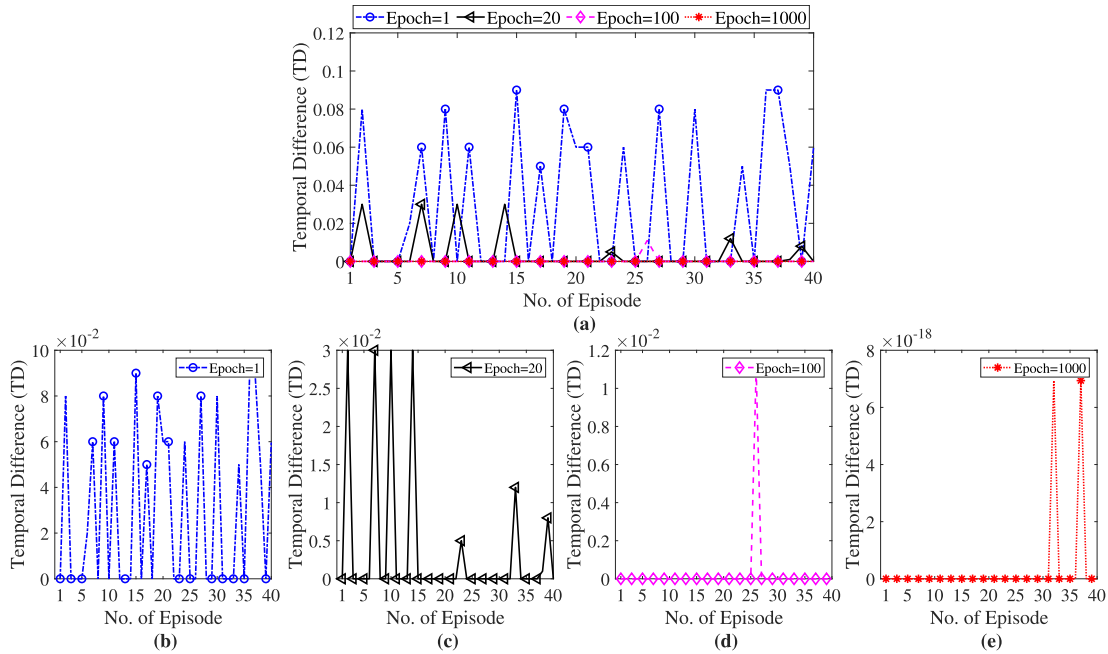


FIGURE 5. Comparing temporal difference error for parameter value $\alpha = 0.9$ and number of epochs; (a) Epoch = 1, 20, 100, 1000 (b) Epoch = 1 (c) Epoch = 20 (d) Epoch = 100 (e) Epoch = 1000.

To elaborate on the better convergence of the proposed framework when a higher learning rate value of $\alpha = 0.9$ is employed, the TD error is summarized in Fig. 5. Compared to Fig. 4, the results in Fig. 5 indicate a significant decrease in the TD error for higher training epoch values. Fig. 5(b)-(e) presents the TD error for the training epoch values of 1 to 1000, respectively. Contrary to the results presented in Fig. 4(b)-(e), the results in Fig. 5(b)-(e) indicate that the TD error reduces for every training epochs value, whereas the change is more dominant for a higher value of learning rate α in Fig. 5(b)-(e). The TD error reduces from 9×10^{-2} to 7×10^{-18} as the training epoch approach 1000. It is interesting to observe that, for every increment in training epochs, the TD error exhibits a significant decrease, which was not the case for the lower learning rate value in Fig. 4. This decrease in TD error over the episodic learning progress justifies the learning progress rate result in Fig. 3(d). The higher value of the training epoch demonstrates a better result than lower training epochs. The experimental observations indicate that, with a higher learning rate $\alpha = 0.9$ and 1000 training epoch, the PAKES becomes less erroneous than other parametric scenarios and makes qualitatively better personalized learning path recommendations to improve the overall learning outcomes.

In summary, the convergence analysis presented in Figs. 4 and 5 justifies our intention of employing a higher learning rate of 0.9 and 1000 training epochs for the proposed framework. With the optimized parameter values, the proposed framework exhibits better convergence within a few episodes than the conventional approaches. In the later section, we present the performance comparison of the proposed framework equipped with the optimized parameters to the traditional CDRS framework.

3) ADAPTABILITY EFFECT OF LEARNING WITH COMPARISON

Finally, we studied the effect of PAKES on the learning improvement of learners and compared the results with the CDRS. After selecting the optimal parametric configuration ($\alpha = 0.9, epoch = 1000$), the learning progress score of the learners is calculated using (3). The experimental results in Fig. 6 illustrate the relationship between the percentage of the learning progress score L_{ps} with a varying number of training epochs. The learners interacting with the proposed framework achieve an L_{ps} from 23.99 to 62.28 for 1000 training epochs, whereas the CDRS fails to achieve this learning score, which is from 18.89 to 41.39. Moreover, the proposed framework attains $L_{ps} = 53.23$ within 200 training epochs, whereas the higher L_{ps} of the learners interacting with the CDRS is 41.39% at 1000 epochs. The comparative numerical analysis of the proposed and baseline systems is presented in Table 6. This analysis illustrates that, compared to the CDRS, the proposed framework gains in L_{ps} from 5.1% to 20.89% with 1 to 1000 epochs. The maximum gain in L_{ps} is 22.89% for PAKES with $L_{ps} = 60.13\%$ compared to CDRS at 600 epochs. These results demonstrate that the proposed framework offers advantages for optimal learning path recommendations to the learners as per their personalized educational experiences, learning desires, and understanding compared to the CDRS. It adapted the learners with their different learning requirements to enhance the overall learning gain throughout the learning process. The most exciting fact these graphs show that exploits the reasonable diagnostic model is that it is more productive than using the RL approach to optimize learning instructions. This is the principal reason that the proposed framework with the GenMA model and Q-learning algorithm shows the potential results as claimed by the CDRS with 3PL and actor-critic algorithms. Moreover, the proposed

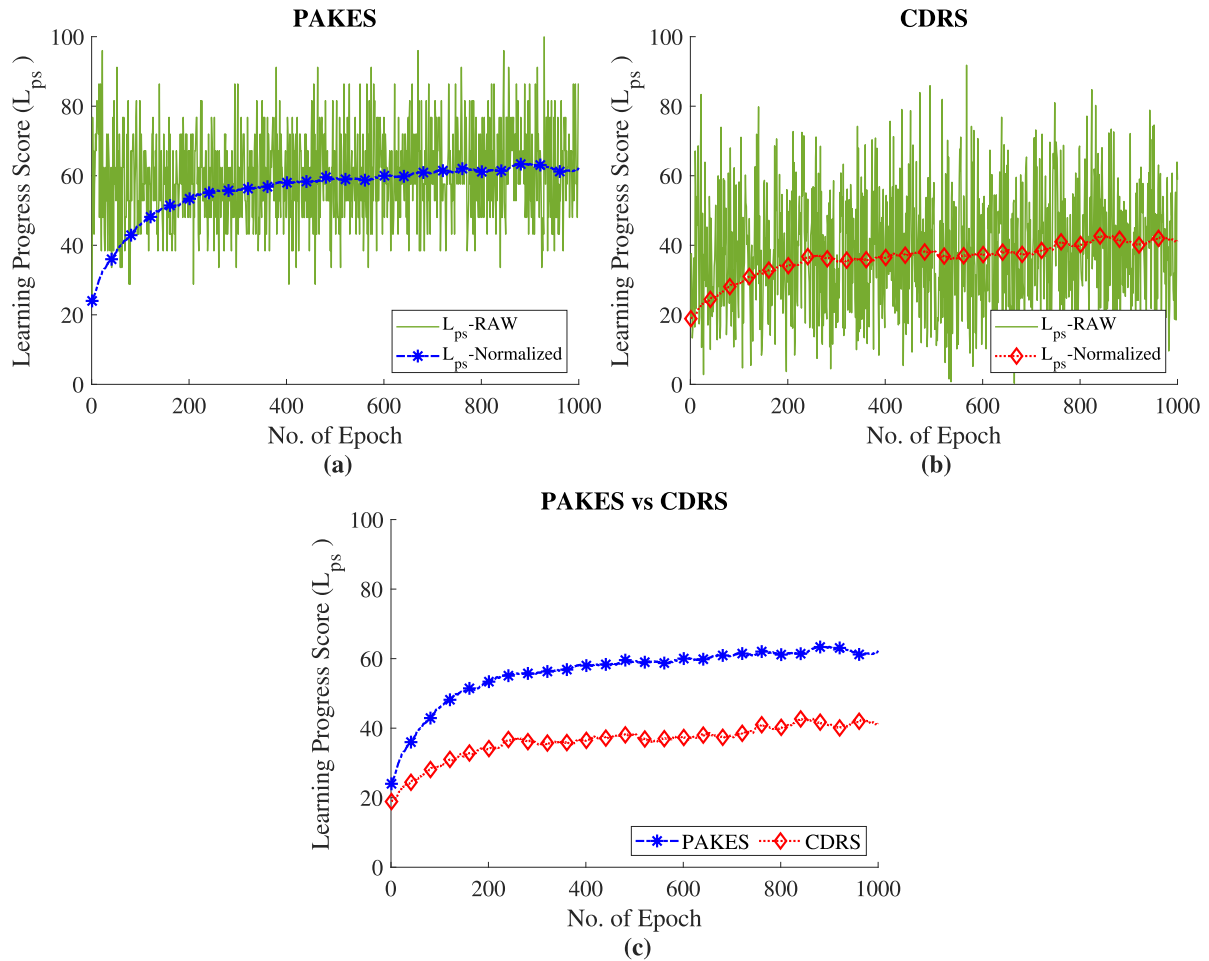


FIGURE 6. For each epoch, the curve shows the average learning progress score being made by (a) proposed PAKES and (b) baseline model (CDRS) recommendations, (c) illustrates the comparative analysis between PAKES and CDRS.

TABLE 6. Comparative analysis of the learning progress score between PAKES and CDRS model.

Epochs	L_{ps} out of 100		Gain%
	(PAKES)	(CDRS)	
1	23.99	18.89	5.1
200	53.23	33.96	19.27
400	58.00	36.38	21.62
600	60.13	37.24	22.89
800	61.23	40.25	21.03
1000	62.28	41.39	20.89

framework also employs the concept of affective states, which is a growing research area in educational data mining [56], [57]. By detecting the affective knowledge states of the learners to employ the GenMA model, the proposed framework significantly provides personalized instructional sequencing and allows learners to learn at their own pace. The experiment illustrates that the proposed framework not merely outperforms the CDRS but also delivers personalized adaptive learning instructions.

V. CONCLUSION AND FUTURE WORK

In this work, we investigated the effect of RL-aided ALSs on the learning performance of learners by combining data-driven and theory-driven approaches. A knowledge extraction strategy called PAKES was proposed for the ALSs to recommend personalized pedagogical sequences. The proposed framework ensures the optimal balance between the learner-control and system-control policies. Equipped with a GenMA model, PAKES can predict ongoing learning performance to model the latent knowledge state of learners. The proposed framework employed a Q-learning algorithm to recommend the best learning paths for learners. It can efficiently handle individual learners based on their learning curves by exploiting the adaptability methodology from learning analytics. The experimental results demonstrate that PAKES outperform the baseline strategy (CDRS), by improving the learning progress score L_{ps} from 18.89% to 23.99% (5.1% gain), and from 41.39% to 62.28% (20.89% gain) for 1 to 1000 epochs, respectively. Moreover, the performance gain achieved in current ALSs by employing the optimal learning paths recommended by PAKES implies the progress of the learners in the overall learning trajectory. Future

extensions of this work can explore the advancements for capturing the hidden information of the learner and adapting the RL-based policies to optimize learning outcomes with a real-time implementation in the educational system.

REFERENCES

- [1] T. P. Tran and D. Meacham, "Enhancing learners' experience through extending learning systems," *IEEE Trans. Learn. Technol.*, vol. 13, no. 3, pp. 540–551, Jul./Sep. 2020.
- [2] D. Miliband, "Personalised learning: Building a new relationship with schools," in *Proc. Speech Minister State School Standards North England Educ. Conf.*, Belfast, Northern Ireland, Jan. 2004, pp. 1–25.
- [3] 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," *Int. J. Artif. Intell. Educ.*, vol. 24, no. 4, pp. 470–497, Dec. 2014.
- [4] K. R. Koedinger, E. Brunskill, R. S. J. D. Baker, E. A. McLaughlin, and J. Stamper, "New potentials for data-driven intelligent tutoring system development and optimization," *AI Mag.*, vol. 34, no. 3, pp. 27–41, Sep. 2013.
- [5] N. M. Noh, A. Ahmad, S. A. Halim, and A. M. Ali, "Intelligent tutoring system using rule-based and case-based: A comparison," *Proc.-Social Behav. Sci.*, vol. 67, pp. 454–463, Dec. 2012.
- [6] Y. Su, Q. Liu, Q. Liu, Z. Huang, Y. Yin, E. Chen, C. Ding, S. Wei, and G. Hu, "Exercise-enhanced sequential modeling for Student performance prediction," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018.
- [7] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum learning for reinforcement learning domains: A framework and survey," 2020, *arXiv:2003.04960*.
- [8] Q. Liu, S. Tong, C. Liu, H. Zhao, E. Chen, H. Ma, and S. Wang, "Exploiting cognitive structure for adaptive learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 627–635.
- [9] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," 2015, *arXiv:1511.06939*.
- [10] S. Minn, Y. Yu, M. C. Desmarais, F. Zhu, and J.-J. Vie, "Deep knowledge tracing and dynamic Student classification for knowledge tracing," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 1182–1187.
- [11] A. Iglesias, P. Martínez, R. Aler, and F. Fernández, "Learning teaching strategies in an adaptive and intelligent educational system through reinforcement learning," *Int. J. Speech Technol.*, vol. 31, no. 1, pp. 89–106, Aug. 2009.
- [12] F. Wang, "Learning teaching in teaching: Online reinforcement learning for intelligent tutoring," in *Future Information Technology*. Berlin, Germany: Springer, 2014, pp. 191–196.
- [13] K. N. Martin and I. Arroyo, "Agentx: Using reinforcement learning to improve the effectiveness of intelligent tutoring systems," in *Proc. Int. Conf. Intell. Tutoring Syst. (ITS)*, Akagoas, Brazil, Aug./Sep. 2004, pp. 564–572.
- [14] M. Chi, K. VanLehn, D. Litman, and P. Jordan, "Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies," *User Model. User-Adapted Interact.*, vol. 21, nos. 1–2, pp. 137–180, Apr. 2011.
- [15] X. Li, H. Xu, J. Zhang, and H.-H. Chang, "Optimal hierarchical learning path design with reinforcement learning," 2018, *arXiv:1810.05347*.
- [16] R. Huang, J. M. Spector, and J. Yang, *Educational Technology: A Primer for the 21st Century*. Singapore: Springer, 2019, pp. 1–248.
- [17] G. Conole, T. Klobucar, C. Rensing, J. Konert, and E. Lavoué, "Design for teaching and learning in a networked world," in *Proc. 10th Eur. Conf. Technol. Enhanced Learn. (EC-TEL)*, Toledo, Spain, Sep. 2015, pp. 1–647.
- [18] H. Yago, J. Clemente, and D. Rodriguez, "Competence-based recommender systems: A systematic literature review," *Behav. Inf. Technol.*, vol. 37, nos. 10–11, pp. 958–977, Nov. 2018.
- [19] A. Cully and Y. Demiris, "Online knowledge level tracking with data-driven Student models and collaborative filtering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 10, pp. 2000–2013, Oct. 2020.
- [20] J. Knobbout and E. Van Der Stappen, "Where is the learning in learning analytics? A systematic literature review on the operationalization of learning-related constructs in the evaluation of learning analytics interventions," *IEEE Trans. Learn. Technol.*, vol. 13, no. 3, pp. 631–645, Jul. 2020.
- [21] S. Doroudi, V. Alevan, and E. Brunskill, "Where's the reward?" *Int. J. Artif. Intell. Educ.*, vol. 29, no. 4, pp. 568–620, 2019.
- [22] Y. Chen, X. Li, J. Liu, and Z. Ying, "Recommendation system for adaptive learning," *Appl. Psychol. Meas.*, vol. 42, no. 1, pp. 24–41, Jan. 2018.
- [23] A. S. Lan and R. G. Baraniuk, "A contextual bandits framework for personalized learning action selection," in *Proc. EDM*, 2016, pp. 424–429.
- [24] B. Clement, D. Roy, P.-Y. Oudeyer, and M. Lopes, "Multi-armed bandits for intelligent tutoring systems," 2013, *arXiv:1310.3174*.
- [25] B. Clement, P.-Y. Oudeyer, and M. Lopes, "A comparison of automatic teaching strategies for heterogeneous student populations," *Int. Educ. Data Mining Soc.*, Raleigh, NC, USA, Jun./Jul. 2016, pp. 330–335.
- [26] B. Clement, D. Roy, P.-Y. Oudeyer, and M. Lopes, "Online optimization of teaching sequences with multi-armed bandits," in *Proc. 7th Int. Conf. Educ. Data Mining (EDM)*, London, U.K., Jul. 2014, pp. 1–4.
- [27] M. Z. Islam, K. Mehmood, and H. S. Kim, "Reinforcement learning based interactive agent for personalized mathematical skill enhancement," in *Proc. Int. Conf. Math. Aspects Comput. Inf. Sci. (MAICS)*, Gebze, Turkey, Nov. 2019, pp. 395–407.
- [28] B. Clement, D. Roy, P.-Y. Oudeyer, and M. Lopes, "Developmental learning for intelligent tutoring systems," in *Proc. 4th Int. Conf. Develop. Learn. Epigenetic Robot.*, Oct. 2014, pp. 426–433.
- [29] X. Tang, Y. Chen, X. Li, J. Liu, and Z. Ying, "A reinforcement learning approach to personalized learning recommendation systems," *Brit. J. Math. Stat. Psychol.*, vol. 72, no. 1, pp. 108–135, Feb. 2019.
- [30] C. Tan, R. Han, R. Ye, and K. Chen, "Adaptive learning recommendation strategy based on deep Q-learning," *Appl. Psychol. Meas.*, vol. 44, no. 4, pp. 251–266, Jun. 2020.
- [31] R. Han, K. Chen, and C. Tan, "Curiosity-driven recommendation strategy for adaptive learning via deep reinforcement learning," *Brit. J. Math. Stat. Psychol.*, vol. 73, no. 3, pp. 522–540, 2020.
- [32] Z. Huang, Q. Liu, C. Zhai, Y. Yin, E. Chen, W. Gao, and G. Hu, "Exploring multi-objective exercise recommendations in online education systems," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1261–1270.
- [33] R. Wu, Q. Liu, Y. Liu, E. Chen, Y. Su, Z. Chen, and G. Hu, "Cognitive modelling for predicting examinee performance," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, Buenos Aires, Argentina, Jul. 2015, pp. 1017–1024.
- [34] J.-J. Vie, F. Popineau, E. Bruillard, and Y. Bourda, "A review of recent advances in adaptive assessment," in *Learning Analytics: Fundamentals, Applications, and Trends* (Studies in Systems, Decision and Control). Cham, Switzerland: Springer, 2017, pp. 113–142.
- [35] P. Dwivedi, V. Kant, and K. K. Bharadwaj, "Learning path recommendation based on modified variable length genetic algorithm," *Educ. Inf. Technol.*, vol. 23, no. 2, pp. 819–836, Mar. 2018.
- [36] S. Wang, H. Lin, H.-H. Chang, and J. Douglas, "Hybrid computerized adaptive testing: From group sequential design to fully sequential design," *J. Educ. Meas.*, vol. 53, no. 1, pp. 45–62, Mar. 2016.
- [37] M. Ventura, V. Shute, and Y. J. Kim, "Assessment and learning of qualitative physics in Newton's playground," in *Proc. Int. Conf. Artif. Intell. Educ. (AIED)*, Memphis, TN, USA, Jul. 2013, pp. 579–582.
- [38] J.-J. Vie, F. Popineau, Y. Bourda, and É. Bruillard, "Adaptive testing using a general diagnostic model," in *Proc. Educ. Conf. Technol. Enhanced Learn. (EC-TEL)*, Lyon, France, Sep. 2016, pp. 331–339.
- [39] R. Howard, *Dynamic Programming and Markov Processes*. New York, NY, USA: MIT Press, 1960.
- [40] S. Doroudi, P. S. Thomas, and E. Brunskill, "Importance sampling for fair policy selection," in *Proc. 33rd Conf. Uncertainty Artif. Intell. (UAI)*, Sydney, NSW, Australia, Aug. 2017, pp. 1–16.
- [41] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.
- [42] S. Bhatnagar and K. M. Babu, "New algorithms of the Q-learning type," *Automatica*, vol. 44, no. 4, pp. 1111–1119, Apr. 2008.
- [43] E. Even-Dar and Y. Mansour, "Learning rates for Q-learning," *J. Mach. Learn. Res.*, vol. 5, pp. 1–25, Dec. 2004.
- [44] J. Peng and R. J. Williams, "Incremental multi-step Q-learning," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1994, pp. 226–232.
- [45] J. N. Tsitsiklis, "Asynchronous stochastic approximation and Q-learning," *Mach. Learn.*, vol. 16, no. 16, pp. 185–202, Sep. 1994.
- [46] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.
- [47] B. Deonovic, P. Chopade, M. Yudelson, J. de la Torre, and A. A. von Davier, "Application of cognitive diagnostic models to learning and assessment systems," in *Handbook of Diagnostic Classification Models*. Cham, Switzerland: Springer, 2019, pp. 437–460.

- [48] R. P. Chalmers, "mirt: A multidimensional item response theory package for the R environment," *J. Stat. Softw.*, vol. 48, no. 6, pp. 1–29, 2012.
- [49] D. Andrich and I. Marais, "A course in Rasch measurement theory," in *Measuring in the Educational, Social and Health Sciences*. Singapore: Springer, 2019.
- [50] L. T. DeCarlo, "On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix," *Appl. Psychol. Meas.*, vol. 35, no. 1, pp. 8–26, Jan. 2011.
- [51] L. Cai, "Metropolis-hastings robbins-monro algorithm for confirmatory item factor analysis," *J. Educ. Behav. Statist.*, vol. 35, no. 3, pp. 307–335, Jun. 2010.
- [52] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 1008–1014.
- [53] H. Zhu, F. Tian, K. Wu, N. Shah, Y. Chen, Y. Ni, X. Zhang, K.-M. Chao, and Q. Zheng, "A multi-constraint learning path recommendation algorithm based on knowledge map," *Knowl.-Based Syst.*, vol. 143, pp. 102–114, Mar. 2018.
- [54] M. Salehi, I. N. Kamalabadi, and M. B. G. Ghouschi, "Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering," *Educ. Inf. Technol.*, vol. 19, no. 4, pp. 713–735, Dec. 2014.
- [55] L. E. Yelle, "The learning curve: Historical review and comprehensive survey," *Decis. Sci.*, vol. 10, no. 2, pp. 302–328, Apr. 1979.
- [56] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.
- [57] M. Anoopkumar and A. M. J. M. Z. Rahman, "A review on data mining techniques and factors used in educational data mining to predict Student amelioration," in *Proc. Int. Conf. Data Mining Adv. Comput. (SAPIENCE)*, Mar. 2016, pp. 122–133.



MUHAMMAD ZUBAIR ISLAM (Student Member, IEEE) received the B.S. degree in computer science from the University of Agriculture Faisalabad (UAF), Pakistan, in 2017, and the master's degree in information and communication engineering from Sejong University, Seoul, South Korea, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Intelligent Mechatronics Engineering. Along with his study, he is also working as a Research Assistant with the Mobile Intelligent Embedded System (MINES) Laboratory, Sejong University. His research interests include applications of artificial intelligence and predictive models in education data mining (EDM), next-generation communication technologies (tactile internet), and edge intelligence.



RASHID ALI (Member, IEEE) received the B.S. degree in information technology from Gomal University, Pakistan, in 2007, the master's degree in computer science (advanced network design) and the master's degree in informatics from the University West, Sweden, in 2010 and 2013, respectively, and the Ph.D. Diploma degree in information and communication engineering from the Department of Information and Communication Engineering, Yeungnam University, South Korea, in February 2019. From 2007 to 2009, he worked at Wateen Telecom Private Ltd., Pakistan, as a WiMAX Engineer with the Operations and Research Department. From July 2013 to June 2014, he worked at COMSATS University Islamabad (Vehari), Pakistan, as a Lecturer. He has also served as a Postdoctoral Research Fellow for the Department of Information and Communications Engineering, Yeungnam University. He is currently working as an Assistant Professor at the School of Intelligent Mechatronics Engineering, Sejong University, South Korea. His research interests include next-generation wireless local area networks (IEEE 802.11ax/ah), unlicensed wireless networks in 5G, the Internet of Things, performance evaluation of wireless networks, named-data/information-centric networking, reinforcement learning techniques for wireless networks, and federated reinforcement learning for next-generation WLANs.



AMIR HAIDER received the B.S. degree in electronics engineering from International Islamic University, Islamabad, Pakistan, in 2012, the M.S. degree in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2014, and the Ph.D. degree from the Division of Electronics and Electrical Engineering, Dongguk University, Seoul, South Korea, in 2019. Before that, he served as a Lecturer for the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Pakistan, from August 2014 to August 2015, and later from September 2019 to February 2020, after completing of his Ph.D. degree. He was as a Postdoctoral Researcher at the School of Intelligent Mechatronics Engineering, Sejong University, Seoul, from March 2020 to February 2021. He is currently working as an Assistant Professor at the School of Intelligent Mechatronics Engineering, Sejong University. His research interests include URLLC for beyond 5G mobile networks, artificial intelligence aided wireless communication systems, MIMO antenna design, optical communication systems, and wireless positioning technologies.



MD ZAHIDUL ISLAM is currently a Professor of computer science with the School of Computing, Mathematics, and Engineering, Charles Sturt University, Australia. He is serving as the Director for the Data Science Research Unit (DSRU) with the Faculty of Business, Justice, and Behavioural Sciences. His main research interests include data mining, classification, and clustering algorithms, missing value imputation, data cleaning and pre-processing, privacy preserving data mining, privacy issues due to data mining on social network users, and applications of data mining in real life.



HYUNG SEOK KIM (Member, IEEE) received the B.S. degree from the School of Electrical Engineering, Seoul National University, South Korea, in 1996, and the M.S. and Ph.D. degrees from the School of Electrical Engineering and Computer Engineering, Seoul National University, in 1998 and 2004, respectively. From 2003 to 2004, he held a Visiting Researcher position with the University of Virginia. From 2004 to 2006, he was with the Telecommunication Research and Development Center, Samsung Electronics. He is currently a Faculty Member with the Department of Information and Communication Engineering, Sejong University. His research interests include networking with special emphasis on wireless communication and visual light communication.