Knowledge Tracing: Brief Theory, Selected Algorithms, and Proposed Experiments

Nghia Duong-Trung

20.03.2025

1 Introduction

Deep Knowledge Tracing (**DKT**) [1, 4, 2, 5, 1] is a powerful machine learning approach that leverages sequence modeling to track and predict student performance on a variety of skills. In traditional educational settings, instructors often rely on one-time assessments to estimate learners' mastery levels. However, such single-point tests may not capture the complex and evolving nature of student knowledge over time. By contrast, a DKT system continuously collects data on a student's interactions (e.g., questions answered, correctness of responses) and updates its beliefs about the student's skill mastery accordingly.

With this individualized modeling technique, it becomes possible to offer targeted interventions at precisely the moment a learner most needs them. For example, if the model infers from recent incorrect responses that a student has not fully grasped a particular subtopic, it can recommend additional exercises or explanations on that subtopic. Conversely, if the learner shows consistent success on certain skills, the system can suggest more advanced tasks or potentially reduce the time spent reviewing already-mastered material.

A key advantage of knowledge tracing models is the ability to operate at scale and in real-time. Educators can thus glean insights from large cohorts of students, identifying patterns of common misconceptions or points of confusion across the population. Meanwhile, individual learners benefit from a dynamic, data-driven experience in which each step of the learning process is tailored to their current needs and skill gaps.

In our proposed solutions, we have developed a pipeline that integrates data preparation, model training, and real-time inference into a cohesive framework. The system begins by converting raw interaction logs (e.g., CSV files of student records) into consistent sequences of `(skill, correctness)` pairs, applying appropriate truncation or padding as necessary. Next, it feeds these sequences into a deep learning model, typically an LSTM-based DKT, which continuously adapts its internal state to reflect each learner's evolving mastery profile. Finally, the trained model can be deployed to provide immediate feedback, skill-level interventions, and long-term analytics for course administrators, enabling a cycle of ongoing improvement in teaching practices.

Taken together, this solution aims to address the challenges of personalized, evidence-based education by capturing the nuance of learning trajectories. Through careful design of the data flow and model architecture, instructors and instructional designers can unlock new possibilities for adaptive learning, boosting engagement and improving student outcomes across diverse learning contexts.

2 Personalized Knowledge Tracing: Conceptual Model

In modern educational contexts, *personalized knowledge tracing (PKT)* [6, 3] describes a family of methods that leverage machine learning (ML) to estimate, in real-time, the evolving knowledge state of individual learners. This section provides deeper insights into how PKT addresses the requirements outlined in your document (i.e., didactic levels and methods of internal differentiation), and highlights relevant ML considerations for successful implementation. At its core, PKT deals with two main tasks:

1. **Knowledge State Estimation:** Determine the learner's current understanding of a set of skills or concepts.
2. **Content Recommendation:** Predict the most beneficial next activity or resource that would improve the learner's mastery.

Formally, we can define:

- A *learner* $i \in \{1, 2, \ldots, N\}$.
- A *skill set* $\{s_1, s_2, \ldots, s_K\}$ relevant to the curriculum or domain.
- A *sequence of learning interactions* per learner: $\{(q_1, r_1), (q_2, r_2), \ldots\}$, where q_t corresponds to a specific item or task (typically associated with one or more underlying skills), and r_t is the learner's response (e.g., correct/incorrect, partial).
- A *knowledge state* vector $\boldsymbol{\theta}_i^t \in \mathbb{R}^K$ that represents learner i's proficiency across the K skills after t interactions.

The aim of the PKT framework is to learn a function f such that:

$$\boldsymbol{\theta}_i^{t+1} = f\big(\boldsymbol{\theta}_i^t,\ q_{t+1},\ r_{t+1}\big), \tag{1}$$

and simultaneously estimate $P\big(r_{t+1} = 1 \mid \boldsymbol{\theta}_i^t, q_{t+1}\big)$, i.e., the probability that the learner will respond correctly to the next item. Methods such as Bayesian Knowledge Tracing (BKT), Deep Knowledge Tracing (DKT), and other neural approaches can be adapted here.

3 Challenges & Opportunities

1. **Fine-Grained Skill Representation:** The didactic framework (Nano to Macro levels) implies complex hierarchical structures of knowledge. ML models often need to embed these hierarchical relationships (e.g., sub-skills within macro-skills) to produce accurate predictions.
2. **Sparsity of Data:** In real classrooms, learners may only answer a small number of items per day. Overcoming limited data within short timespans is a key issue.

3. **Explainability:** Educators often prefer interpretable models (like BKT) to black-box neural networks for transparency in decision-making at the *micro* and *meso* levels.

4 Alignment with Didactic Levels

Nano Level (Individual Activities):

- **Real-Time Feedback:** PKT automatically updates learner knowledge states after each micro-task (e.g., a single question) and can supply instant feedback (hints, corrections).
- **Task Adaptation:** By continually inferring mastery, the system can recommend the *next best question or exercise* to bridge knowledge gaps quickly. This scaffolding approach addresses requirements such as individual tasks and immediate support measures.

Micro Level (Individual Lessons):

- **Adaptive Lesson Flow:** For a given lesson, PKT helps track each student's skill mastery in real time, allowing the educator (or the system) to dynamically allocate more class time to concepts where the group as a whole shows lower understanding.
- **Formative Assessment Integration:** Micro-level PKT insights allow teachers to identify frequent misconceptions during the lesson. Teachers can then run targeted mini-interventions (short explanations, group problem-solving).

Meso Level (Sequences of Lessons/Modules):

- **Informed Sequencing and Curriculum Chaining:** PKT data aggregated over several lessons can inform how future modules are introduced. For instance, if many learners consistently struggle with foundational algebraic manipulation, the teacher might build in extra practice sessions or reorder the topics.
- **Spiral Review:** Since PKT captures historical data over multiple lessons, it can propose periodic *spiral reviews* of previously taught content at strategic intervals. This aligns with meso-level planning (e.g., ensuring continuous reinforcement).

Macro Level (Long-Term Curriculum):

- **Curriculum Mapping and Global Analytics:** Over months or a full academic year, PKT's longitudinal data highlights persistent knowledge gaps or advanced competencies. This helps educational leaders make curriculum-level adjustments.
- **Longitudinal Learner Profiling:** Each student's knowledge trajectory over an entire program can be visualized, enabling personalized pathways. Learners who exhibit accelerated growth in certain areas may be offered advanced opportunities.

5 Supporting Methods of Internal Differentiation

Individual & Differentiated Assignments: By quantifying each student's mastery probabilities, PKT naturally facilitates *tiered* or *individualized* tasks. If a learner's mastery probability for "concept A" is high, the system presents more advanced tasks; if it is low, the system proposes remediation.

Flexible Learning Paths: A PKT-driven platform can define unique sequences of learning materials for each student. Thus, at any time, a learner who demonstrates readiness can branch into more challenging content, while another may revisit essentials.

Various Social Forms (Individual, Partner, Group Work): While PKT typically tracks individuals, it can also inform group formation. For instance, the teacher may pair a *high-mastery* student with a *low-mastery* student for peer tutoring, or group students with complementary skills to foster cooperative learning.

Diverse Media & Materials: PKT can track the *effectiveness* of different media on each student's learning progress. If analytics show that a learner progresses faster with videos than with purely text-based content, the system can recommend more video resources.

Adapting Learning Times: Because PKT data is individualized, the system flags which students need extra practice or extended deadlines. Conversely, advanced learners can be "fast-tracked" without feeling held back.

Individual Objectives: Teachers can establish *learning milestones* in the system. Each student's PKT profile indicates whether a milestone is met, and the teacher can refine these goals in consultation with the learner (e.g., in a weekly review).

Promote Independence & Responsibility: PKT dashboards typically visualize a learner's progress in real time. Such transparency motivates learners to set personal targets, track improvement, and reflect on study strategies, fostering self-directed learning behaviors.

Stimulating Learning Environment: Although PKT focuses on data-driven logic, it also facilitates *meaningful tasks* in a dynamic classroom. A teacher aware of the system's recommendations might create a "learning station" with tasks specifically addressing the most common knowledge gaps, thereby creating a sense of purpose in the physical environment.

Feedback & Reflection: PKT's continuous assessment loop ensures learners can receive feedback that goes beyond single-letter grades. Teachers might schedule short *reflection sessions* where students review their mastery profiles, discuss challenges, and set new goals.

Diagnostics (Assessment & Monitoring): Fundamentally, PKT is a **diagnostic mechanism** that provides fine-grained details on each individual. The system highlights when a concept was first introduced, how many attempts were made, and where the learner stands at present. This data can also feed into more formal assessments (e.g., shaping midterm or final exam reviews).

6 Additional Machine Learning Perspectives

Model Selection & Architecture: Various PKT architectures exist:

- **Bayesian Knowledge Tracing (BKT):** A simpler, more interpretable approach where each skill is modeled by a Hidden Markov Model. It assumes a relatively small set of parameters and can yield easily explainable "mastery probabilities."
- **Deep Knowledge Tracing (DKT):** A recurrent neural network model (often LSTM or GRU) that learns latent representations of student knowledge states. More flexible but less interpretable.
- **Hierarchical or Graph-Based Models:** When skills are hierarchical or networked (which is often the case in multi-level didactic frameworks), graph-based or hierarchical Bayesian approaches can capture prerequisite relationships.

Evaluation Metrics: Standard metrics include:

- **AUC (Area Under the ROC Curve):** Measures the model's ability to discriminate between correct and incorrect responses.
- **RMSE or Log-Loss:** Typical metrics for probability predictions.
- **Explainability / Interpretability:** A more qualitative aspect, but crucial for teacher trust.

Data Pipeline Considerations:

- **Data Integration:** Collecting interactions from multiple sources (e.g., LMS logs, in-class quizzes, homework platforms).
- **Privacy and Ethics:** PKT systems handle sensitive learner data. Ensuring data protection and ethical usage is paramount.
- **Real-Time Adaptation:** Implementing PKT in real classrooms often requires cloud or local servers with low-latency response times, so recommendations are immediate and actionable.

Scaling and Deployment:

- **Integration with Existing LMS:** Many schools already use Learning Management Systems (LMS). PKT modules should integrate smoothly to avoid disruption.

- **Teacher Training:** Teachers require training to interpret PKT outputs (dashboards, mastery charts) and translate them into pedagogical decisions.

- **Fairness and Equity:** PKT-based systems must ensure they do not systematically disadvantage any group. For example, if the question bank heavily targets certain cultural contexts, certain students might receive less-accurate feedback.

7 Remarks

Personalized knowledge tracing (PKT) aligns with, and directly supports, the didactic framework from *nano* to *macro* levels. From the learner's perspective, PKT ensures that each new task or resource is optimally matched to their current proficiency. From the instructor's perspective, PKT offers powerful, data-driven insights into the class as a whole, aiding the integration of internal differentiation strategies (e.g., varied assignments, multiple learning paths, targeted feedback). On the machine learning side, implementing PKT involves considerations of model architecture, data integration, interpretability, and deployment in real-world environments. By thoroughly accounting for these details, PKT can not only track knowledge but also help reshape the educational process around personalization, equity, and efficiency.

8 Prototype Implementation

8.1 Sample Dataset

A sample user_logs.csv with 120 data points. Simply copy the lines into a file named user_logs.csv, place it in the same folder as your scripts, and it will serve as input for data_preparation.py. The file has 6 different students (IDs 101 through 106), each with 20 interactions, and each interaction has:

- A student ID
- A timestamp in ascending chronological order for that student
- A skill ID between 0 and 9
- A correctness value (0 or 1)

8.2 Data Preparation

Run the Data Preparation script to generate NumPy arrays: python data_preparation.py –iput_csv="user_logs.csv" –output_dir="./prepared_data" –max_seq_len=20

This reads the CSV, groups interactions by student ID, sorts them by timestamp, and then produces:

- skill_sequences.npy
- correctness_sequences.npy
- skill2idx_map.npy

8.3 Run the DKT Training script

python dkt_training.py –data_dir=”./prepared_data” –batch_size=4 –hidden_size=16 –num_epochs=3 –learning_rate=1e-3 –device=”CPU”

When you see a prediction tensor with shape torch.Size([6, 50, 10]), it means:

- 6: The batch size – i.e., there are 6 student sequences in the current mini-batch.
- 50: The sequence length for each student. In a knowledge-tracing context, this corresponds to 50 time steps (e.g., each time step is a question the student answered).
- 10: The number of skills (or skill "slots") that the model can predict at each time step.

Put differently, for each of the 6 students (dimension 0), at each of the 50 time steps (dimension 1), the model is outputting 10 probabilities (dimension 2). Each of these 10 probabilities corresponds to how likely the student is to answer a particular skill correctly on the *next* question.

References

[1] Abdelrahman, G., Wang, Q., Nunes, B.: Knowledge tracing: A survey. ACM Computing Surveys **55**(11), 1–37 (2023)

[2] Casalino, G., Grilli, L., Limone, P., Santoro, D., Schicchi, D., et al.: Deep learning for knowledge tracing in learning analytics: an overview. TeleXbe (2021)

[3] Liu, X., Zhu, J., Pan, C., Huang, C., Song, Y., Cao, X.: Personalized learning made simple: A deep knowledge tracing model for individual cognitive development. In: CCF Conference on Computer Supported Cooperative Work and Social Computing. pp. 462–472. Springer (2023)

[4] Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L.J., Sohl-Dickstein, J.: Deep knowledge tracing. Advances in neural information processing systems **28** (2015)

[5] Šarić-Grgić, I., Grubišić, A., Gašpar, A.: Twenty-five years of bayesian knowledge tracing: a systematic review. User Modeling and User-Adapted Interaction **34**(4), 1127–1173 (2024)

[6] Wang, C., Sahebi, S.: Continuous personalized knowledge tracing: Modeling long-term learning in online environments. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. pp. 2616–2625 (2023)