Uncertainty-Aware Knowledge Tracing: Towards the Use of Subjective Logic

Rania Ait Chabane[1,2,*], Armelle Brun[1,2] and Azim Roussanaly[1,2]

[1]Université de Lorraine
[2]LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications)

Abstract
With the growing adoption of online educational applications, there is an increasing need for knowledge tracing (KT) models that estimate learners' mastery of concepts. In education, accurate models are not sufficient: they should not only provide explanations that are understandable to teachers and learners, but also indicate how reliable these predictions are, which leads to increased explainability. In this work, we introduce *Uncertainty-aware Knowledge Tracing using Subjective Logic (KT-USL)*, the first explainable and uncertainty-aware probabilistic KT model. KT-USL quantifies uncertainty through a *Subjective Logic opinion* that decomposes mastery into belief, disbelief, and epistemic uncertainty, together with a base-rate prior, and estimates a mastery value. This representation makes it possible to distinguish, for example, between low mastery due to consistent evidence and low mastery due to a lack of data, thereby supporting uncertainty-aware, explainable educational decisions.

The literature has shown that quantifying uncertainty leads to improved performance of deep models. We ask to what extent this conclusion stands for probabilistic models. To this aim, we evaluate KT-USL on three large-scale mathematics KT datasets, and compare it against interpretable probabilistic baselines and a state-of-the-art deep KT model under a unified experimental protocol. Experiments show that KT-USL outperforms classical probabilistic models. While deep models still achieve the highest predictive performance, they remain black boxes, thus KT-USL provides a favourable trade-off between uncertainty-aware explainability and predictive performance.

Keywords
Knowledge Tracing, Subjective Logic, Uncertainty, Explainable AI

1. Introduction

Knowledge Tracing (KT), that is a core component of intelligent tutoring systems, aims to maintain an estimate of what a learner knows as she interacts with instructional activities. Specifically, KT aims to (1) infer how a learner's mastery of underlying concepts evolves over time from her interactions [1], and (2) use this evolving mastery state to predict future performance on activities [2].

While next–question correctness is a common evaluation target, KT supports a broader range of downstream educational applications. It enables informed activity recommendation, provides richer learning analytics tools such as dashboards for teachers and learners [3], and supports, among other benefits, competency-based learning frameworks [4, 5]. Accurate estimation of a learner's mastery of concepts is essential for these uses, which rely on reliable and interpretable representations of the learning progress [6].

Issues related to model uncertainty arise in many machine learning settings, and education is no exception: they are further amplified by sparse and irregular learner traces, heterogeneous practice patterns, etc. In practice, these data characteristics primarily give rise to epistemic uncertainty—stemming from insufficient or incomplete evidence—which in turn leads to unstable predictions and undermines the usefulness of KT models [7]. In what follows, we use the term uncertainty to refer to epistemic uncertainty unless stated otherwise.

Yet, quantifying uncertainty is a key ingredient of model reliability and explainability, as quantifying model uncertainty provides valuable information about how confident a model is in its predictions

XAI-Ed 2026: Demystifying AI in Education and Learning Analytics through Explainability, Agency, and Transparency Workshop (XAI-Ed@LAK26), 27-28 April, 2026, Bergen, Norway
rania.ait-chabane@loria.fr (R. Ait Chabane); armelle.brun@loria.fr (A. Brun); azim.roussanaly@loria.fr (A. Roussanaly)
0009-0002-7630-1195 (R. Ait Chabane); 0000-0002-9876-6906 (A. Brun); 0000-0002-3311-3613 (A. Roussanaly)

[8][9]. For instance, in online education, such a model may predict a high probability of correctness, while being highly uncertain due to limited evidence. Let us consider a teacher-facing recommendation scenario in which the system suggests practice activities for a learner working on fractions. Suppose the learner has completed only two exercises on fraction multiplication, both answered correctly. A classical KT model, that does not represent uncertainty, may interpret these two correct responses as sufficient evidence of mastery and therefore confidently recommend advancing to fraction addition. In contrast, an uncertainty-aware KT model would express low confidence in the learner's mastery due to the limited amount of evidence. As a result, it may present and argue for two options: (1) remaining at the present level and recommending additional fraction multiplication exercises, and (2) advancing to fraction addition exercises. The final decision will remain with the teacher, who can rely on their expertise together with the model's uncertainty estimates to select the most appropriate activity.

In educational systems, explainable models can provide support to educators and learners. However, only a subset of KT models are explainable. Deep KT models approximate complex learning patterns, resulting in strong predictive accuracy [10, 11]. However, they are not explainable as they do not manage an interpretable mastery state [7, 12].

Probabilistic KT models provide interpretable parameters for learning, guessing, and slipping [1], and yield explicit mastery states. However, these explainable models are less performing than deep models.

None of these models quantify uncertainty, so their output is under the form of a single point estimate that conflates belief with evidence. In real-world KT deployments—where interaction histories are uneven and concept mappings imperfect [13]—this absence of uncertainty leads to overconfident or unreliable mastery estimates [7]. The UKT model [14] is one of the few models that quantifies predictive uncertainty. This model has shown to result in an increase in the prediction accuracy. Nonetheless, explainability remains limited due to the deep architecture and an output still made up of a single point.

To sum up, no KT model does jointly address explainability and uncertainty. On the one hand, recent deep KT models attempt to incorporate uncertainty to mitigate black-box opacity, but as these architectures are not inherently interpretable, the uncertainty they produce is not explicit so is difficult to exploit in practice. On the other hand, probabilistic KT models are explainable as they offer interpretable mastery representations. Yet, their performance remains limited and they do not quantify uncertainty.

Given the improvement in accuracy reached by uncertainty-aware deep KT models, we propose to introduce a probabilistic KT model that quantifies uncertainty, with the objective to increase accuracy.

Such a model will thus jointly address explainability and uncertainty. This model is the *Uncertainty-aware Knowledge Tracing using subjective logic (KT-USL)* model. KT-USL uses Subjective Logic to represent mastery. This representation provides (1) an explicit, interpretable estimate of concept mastery and (2) a principled, explicit modelling of uncertainty. By unifying explicit mastery and uncertainty estimation, KT-USL offers mastery states that can be directly leveraged in explainable educational applications, and its explicit modelling of uncertainty also improves robustness in sparse and uneven data settings by preventing overconfident updates when evidence is limited.

2. Related Work

This section covers prior work on the evolution of knowledge tracing models from classical models to deep architectures, on approaches for uncertainty modeling in machine learning and in KT, and on explainability both broadly and within the KT literature.

2.1. Knowledge Tracing

Over the past decades, KT models have evolved in both the techniques they employ and in the contextual information they leverage, ranging from probabilistic models [1, 15, 16] to deep neural architectures [2, 11, 17]. These models differ not only in their predictive accuracy, but also in the assumptions they rely on and the extent to which they provide an explicit and interpretable representation of concept mastery.

In practice, KT is often used as an intermediate layer in broader educational applications, such as adaptive

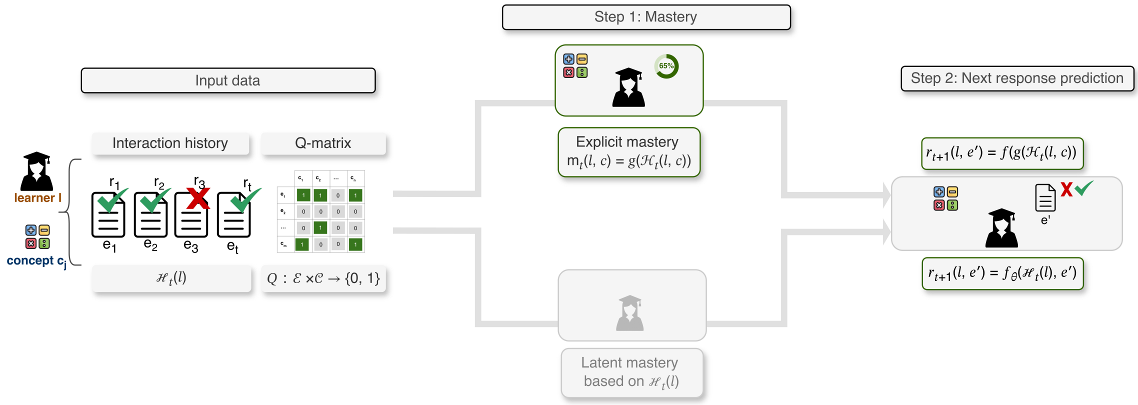

Figure 1: General framework for Knowledge Tracing.

learning systems, learning activity recommendation, early-warning systems, or teacher and learner dashboards. In such settings, two evaluation dimensions are particularly important: (1) predictive performance on learner responses, and (2) ability to estimate knowledge mastery in an interpretable and actionable way.

General KT Framework. Figure 1 illustrates the general framework of KT models.

Let L be the set of learners, E the set of exercises (questions, items, quizzes), and C the set of concepts. The Q-matrix encodes which concepts $c \in C$ are involved in each exercise e, with $Q(e, c) = 1$ if exercise e involves concept c, and 0 otherwise:

$$Q : E \times C \to \{0, 1\}.$$

For an exercise $e \in E$, we denote by

$$C(e) = \{c \in C : Q(e, c) = 1\} \tag{1}$$

the set of concepts involved in e.

When learner $l \in L$ attempts exercise $e \in E$, we observe a success score

$$r(l, e) \in [0, 1], \tag{2}$$

Depending on the context, $r(l, e)$ can be binary $r(l, e) \in \{0, 1\}$, or a normalized score.

At time t, the interaction history of learner l is

$$H_t(l) = \big\{(e, r(l, e)) \bigm| e \in E_t(l)\big\}, \tag{3}$$

where $E_t(l) \subseteq E$ denotes the set of exercises attempted by learner l prior to time t.

For a fixed concept c, we can restrict the learner's history to exercises involving c:

$$H_t(l, c) = \{(e_i, r(l, e_i)) \in H_t(l) : Q(e_i, c) = 1\}, \tag{4}$$

In probabilistic models (upper part of Figure 1), the knowledge state of learner l on concept c is explicitly defined as

$$m_t(l, c) = g\big(H_t(l, c)\big), \tag{5}$$

where g is a concept-specific aggregation function mapping the learner's interaction history on concept c to a mastery estimate $m_t(l, c) \in [0, 1]$.

Given a target exercise e' at time $t + 1$ for learner l, probabilistic KT models use the learner's mastery over the concepts in e' to predict the probability of answering e' correctly:

$$\hat{r}_{t+1}(l, e') = f\Big(m_t(l, c) \, : \, c \in C(e')\Big), \tag{6}$$

where f is a prediction function.

In deep KT models (lower part of Figure 1), the intermediate knowledge state $m_t(l, c)$ is not explicitly evaluated (this step appears grayed out). Instead, the model directly maps the history to a prediction:

$$\hat{r}_{t+1}(l, e') = f_\theta\big(H_t(l), e'\big), \tag{7}$$

where θ denotes the set of trainable parameters of the deep model.

In what follows, we briefly review two major paradigms of KT models and emphasize their explainability properties and their capacity (or lack thereof) to expose an explicit mastery state.

Probabilistic Models. The earliest and most influential probabilistic KT model is *Bayesian Knowledge Tracing (BKT)* [1]. BKT assumes, for each concept and each learner, a two-state Hidden Markov Model with a binary variable indicating whether the concept is *non-mastered* or *mastered.* Learning is modelled as a Markovian transition from non-mastery to mastery, parameterised by an initial mastery probability, a learning rate, a *guess* parameter (probability of a correct answer without mastery), and a *slip* parameter (probability of an incorrect answer despite mastery). Some variants also introduce a forgetting transition to allow the mastery state to revert over time [18]. BKT provides an explicit mastery estimate $m_t(l, c)$, updates it according to a probabilistic state-transition mechanism, and predicts future performance using a concept-specific version of equation ((6)).

These parameters are psychologically meaningful and make BKT attractive for educational applications, as the model directly yields a posterior probability of mastery for each learner–concept pair at each time step.

However, the original BKT formulation assumes independent concepts and concept-level parameters shared across all learners, which limits its ability to capture individual differences in learning. A large body of work therefore extends BKT along several dimensions. Individualised and hierarchical BKT variants introduce learner-specific or group-specific parameters to account for heterogeneity in learning rates and error tendencies [19]. Time-aware extensions incorporate forgetting and time gaps between interactions, allowing the mastery state to decay when learners do not practice [20]. In order to capture intermediate proficiency levels more precisely, Zhang and Yao [21] propose a three-state BKT with *non-mastered*, *currently mastering*, and *mastered* states instead of a two-state (non-mastered/mastered) model[1] [1].

A complementary line of work seeks finer-grained representations of mastery through fuzzy modelling. *Fuzzy Bayesian Knowledge Tracing (FBKT)* [22] extends classical BKT by replacing binary mastery states with fuzzy sets that assign graded membership to categories such as "low", "medium", or "high" mastery. This approach enables smoother transitions between states and can better accommodate partially correct or open-ended responses. Despite these advantages, fuzzy formulations produce membership degrees rather than calibrated probabilities and therefore do not provide a principled way to quantify uncertainty arising from limited evidence.

The literature adopts different taxonomies for probabilistic KT models. Some works distinguish probabilistic models that maintain an explicit mastery state (e.g., BKT) from logistic models that directly estimate the probability of a correct response, while others include logistic formulations within probabilistic approaches. We follow the latter convention and treat logistic KT models as a subclass of probabilistic models.

Logistic models.

Another family of knowledge tracing approaches formulates prediction as a logistic regression problem rather than an inference over latent Markov states. The most prominent representatives are *Learning Factors Analysis (LFA)* [15] and *Performance Factors Analysis (PFA)* [16], which model the

[1] Some notations have been adapted from the original papers to ensure consistency across the formalism used in this work.

probability of a correct response as a function of accumulated practice on the corresponding concept. LFA incorporates a concept intercept and the number of prior opportunities, whereas PFA distinguishes the contributions of successful and unsuccessful attempts. Extensions such as AFM [23], temporal decay variants [24], and logistic models with learner- or item-specific effects [25] further refine this line of work. Due to their transparency, robustness in sparse-data regimes, and low computational cost, logistic KT models remain accurate and reliable baselines in many practical settings. Within the general framework, these models instantiate the prediction function directly, without maintaining an explicit latent mastery state $m_t(l, c)$.

More expressive formulations have been developed by generalising logistic models to higher-dimensional feature interactions. *KT Machines (KTM)*, introduced by Vie and Kashima, leverage *factorization machines (FMs)* to incorporate rich side information about learners, items, and knowledge components [26].

Despite their flexibility and performance, logistic models—including LFA/PFA, AFM, temporal decay variants, and KTM—estimate only the probability of correctness and do not maintain an explicit mastery state which limits their explainability.

Deep Learning Models. The introduction of *Deep Knowledge Tracing (DKT)* [2] marked a major shift in KT research by replacing parametric probabilistic models with recurrent neural networks (RNNs/LSTMs). DKT learns temporal representations of learner interactions without relying on strong modelling assumptions. DKT is a dominant baseline in KT. In terms of the general framework, DKT implements the update function using a recurrent architecture and obtains predictions via equation ((7)), while the latent state $h_t(l)$ does not correspond to an explicit mastery estimate $m_t(l, c)$.

Subsequent work has expanded deep KT along several architectural directions. *Memory-augmented models*, notably *Dynamic Key-Value Memory Networks (DKVMN)* [11], introduce an external memory indexed by concepts to store and retrieve latent knowledge representations. *Attention-based models* such as *Self-Attentive KT (SAKT)* [17] leverage Transformer-style mechanisms to capture dependencies across past interactions more effectively than recurrent approaches. Another important direction is *graph-based KT*, which incorporates prerequisite or relational structure among knowledge components and performs message passing over concept graphs. More recently, *uncertainty-aware KT models* such as *UKT* [14] introduce stochastic or distributional components to quantify predictive uncertainty in deep architectures.

Recalling the two objectives of Knowledge Tracing defined in the introduction, deep learning KT models primarily address the second objective, namely predicting future performance, but do not explicitly infer how a learner's mastery of underlying concepts evolves over time from her interactions.

As a result, despite their strong predictive performance on large-scale datasets, deep KT models operate as black boxes whose latent representations cannot be mapped to explicit mastery states, as noted in recent surveys [6]. This significantly limits their explainability.

2.2. Uncertainty in Knowledge Tracing

Uncertainty is an important element to account for in machine learning systems, where predictions fundamentally depend on the quality, quantity, and representativeness of the available data [8]. In many domains, accounting for uncertainty is crucial because model outputs inform decision-making processes [27]. Educational technologies are no exception: their predictions guide decisions that directly affect learners.

Following the general definition of Walker et al. [28], uncertainty can be understood as any deviation from the unattainable ideal of possessing complete, deterministic knowledge of a system. A widely accepted distinction separates *epistemic uncertainty*, arising from incomplete knowledge, and *aleatory uncertainty*, arising from inherent randomness in the data itself [29].

Epistemic uncertainty arises when the model lacks information. In educational settings, it may be caused by sparse interaction histories, ambiguous or unobserved learner traits. Epistemic uncertainty is *reducible*: it can be decreased by collecting more targeted evidence or improving the representational quality of the data [29].

Aleatory uncertainty, by contrast, reflects intrinsic randomness in behaviour or observations. In education, this includes slips, lucky guesses, momentary attention lapses, or day-to-day performance variability. This uncertainty is *irreducible*: it can be *estimated* but not reduced by gathering additional data [29].

Motivated by these considerations, several studies have recently investigated ways to incorporate uncertainty into deep knowledge tracing. Christie et al. [30] introduce *Dynamic LENS*, a framework that combines variational autoencoders with Bayesian state-space models to represent learner knowledge as a Gaussian distribution and propagate epistemic uncertainty across time. Building on distributional representations, Cheng et al. [7] propose a KT model that encodes uncertainty through stochastic distribution embeddings, using a Wasserstein-based self-attention mechanism to track shifts in latent knowledge states and an aleatory-aware contrastive loss to enhance robustness to noisy responses. In parallel, Mitton et al. [12] explore predictive uncertainty through Bayesian neural networks with Monte Carlo dropout, enabling uncertainty estimates within a deep architecture. A complementary line of work by Ding and Larson [31] approaches uncertainty from the standpoint of loss-function regularisation, proposing explicit modifications to the cross-entropy objective to incorporate more principled uncertainty representations into learner response modelling.

Beyond deep learning, Fuzzy Bayesian Knowledge Tracing [22] extends classical BKT by replacing strictly binary mastery states with fuzzy sets. This allows responses to be partially correct. It also produces smoother transitions between learning states. However, the fuzziness models degrees of correctness, not epistemic uncertainty arising from missing or insufficient evidence.

Overall, uncertainty-aware KT is an emerging research direction. Existing approaches provide meaningful uncertainty estimates but are often embedded in complex deep architectures whose internal states and uncertainty signals remain difficult to interpret. To the best of our knowledge, no work has explored a probabilistic, uncertainty-aware approach.

2.3. Explainability

Explainability in machine learning reflects how easily humans can understand a model's internal computation and its decision [32]. A common distinction separates *white-box* and *black-box* models. White-box models have simple, transparent components whose internal operations can be interpreted directly by users. In contrast, black-box models exhibit complex, nonlinear mappings from inputs to outputs, which makes their operating mechanisms difficult to understand [33].

Following the taxonomy proposed in a recent explainable KT survey [9], explanation techniques fall into two groups: *ante-hoc* interpretability and *post-hoc* explanations. Ante-hoc approaches aim for interpretability by design and include *transparent models* and *models with intrinsic interpretability*. Post-hoc approaches produce explanations after training and include *model-agnostic* and *model-specific* methods. Original KT models are white-box and intrinsically interpretable.

Probabilistic and logistic KT models (e.g., BKT, LFA/PFA/AFM, KT machines models) express mastery through parameters with clear educational meaning such as initial knowledge, learning rate, item difficulty, or the effects of prior practice [1, 15, 16, 23, 18, 24]. These models provide explanations that are intuitive for educators and learners. [6]. On the other hand, deep KT models learn latent representations of learner behavior [2, 11, 17]. These representations do not map to interpretable mastery variables. Instead, predictions emerge from opaque neural computations. This opacity motivates research on both ante-hoc and post-hoc explainability [34, 9].

Several works redesign deep KT models to expose interpretable internal structures. Attention-based models highlight influential past interactions [17, 9]. Memory networks indexed by concepts or items attempt to visualise concept-level dynamics [11], although such interpretations are often assumptions rather than verified mechanisms.

Hybrid models embed cognitive principles, such as IRT components or forgetting curves, into deep architectures to gain some interpretability [6, 35]. Recent works also integrate Markov Blanket [36] or task-aware constraints using LLMs [37] to make deep KT more transparent. Post-hoc approaches explain a trained model without modifying its structure. Model-agnostic techniques such as LIME/SHAP-style feature attributions, perturbation analyses, or surrogate models have been adapted to KT [34, 9]. Model-specific approaches use gradients, relevance propagation, or genetic search to identify causal features [38]. These methods offer local explanations of which exercises or concepts influenced a prediction. However, they may lack faithfulness to the model's true computation [9, 6]. While several deep learning works aim to make their models more explainable, explainability is not achieved due to the complexity of their internal mechanisms. In contrast, probabilistic and logistic models remain inherently interpretable, which often makes them more suitable for trustworthy decision-making in online education.

3. The Proposed Model

Guided by prior work on KT and by the need for explainability in this task, we adopt a probabilistic approach. To improve accuracy and explainability, we extend this line of work by incorporating an explicit uncertainty component to account for epistemic uncertainty arising from limited data.

Before introducing the model we propose, that is built on the formalism of Subjective Logic, we briefly recall the key concepts of Subjective Logic that underpin this model.

3.1. Subjective Logic

Subjective Logic (SL) extends probabilistic reasoning by attaching an explicit *uncertainty mass* to probability estimates. This makes it suitable for domains where evidence is incomplete or inconsistent. It is widely used for modelling and analysing trust networks [39] and for extensions of Bayesian networks [40].

An agent's belief about a binary proposition is represented by an *opinion* ω defined as follows:

$$\omega = (b, d, u, a), \tag{8}$$

where b is belief (supporting evidence), d is disbelief (refuting evidence), u is epistemic uncertainty (uncommitted mass), and a is a base-rate prior that encodes the assumed probability in the absence of any evidence. The masses satisfy the constraint

$$b + d + u = 1, \tag{9}$$

SL projects an opinion to a single probability through the expectation $E(\omega)$ defined as follows:

$$E(\omega) = b + a\,u,$$

which corresponds to the evidence-supported belief b plus the portion of uncertainty u weighted by the base rate a.

A common grounding of opinions relies on a Beta distribution over a binary proposition, where r denotes the amount of supporting evidence (accumulated positive observations) and s the amount of refuting evidence (accumulated negative observations), and where $W > 0$ is the prior weight assigned to uncertainty:

$$b = \frac{r}{r + s + W}, \tag{10}$$

$$d = \frac{s}{r + s + W}, \tag{11}$$

$$u = \frac{W}{r + s + W}, \tag{12}$$

$$E(\omega) = \frac{r + aW}{r + s + W}. \tag{13}$$

As evidence accumulates, u decreases and the expectation moves smoothly from the prior a (no data) toward the empirical rate (many data). Importantly, the same expected mastery E can correspond to very different combinations of belief b and uncertainty u. This reveals whether E is supported by strong evidence or mainly by a prior applied to a large uncertainty mass (see Figure 2) [41, 39].

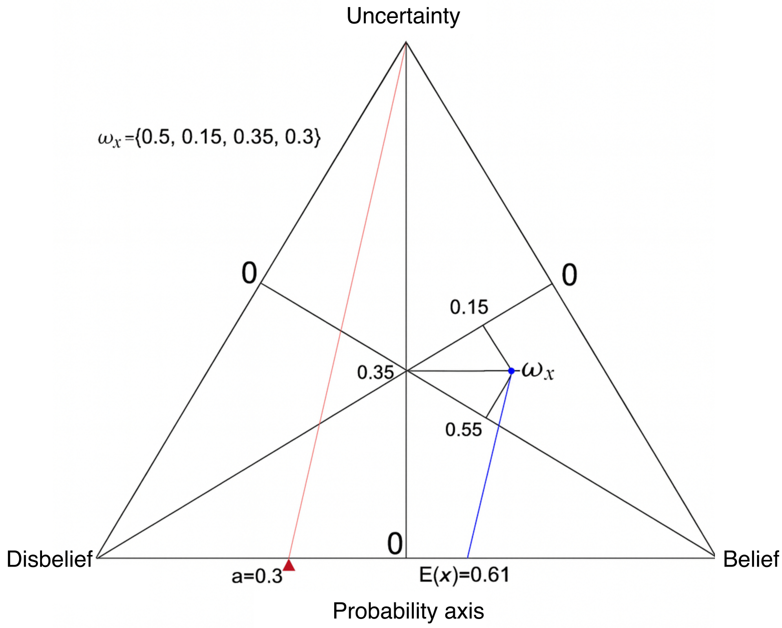

Figure 2: Subjective Logic simplex. An opinion $\omega = (b, d, u, a)$ lies inside the belief–disbelief–uncertainty triangle (b+d+u=1). Projecting the point toward the base-rate a on the probability axis yields $E(\omega) = b + a\,u$. Example shown: $\omega = (0.50, 0.15, 0.35, 0.30)$ gives $E = 0.50 + 0.30 \times 0.35 = 0.605 \approx 0.61$.

3.2. KT-USL: Uncertainty-aware Knowledge Tracing using Subjective Logic

Recall that our goal is to design a probabilistic KT model that explicitly quantifies uncertainty.

SL provides a principled decomposition of evidence into belief, disbelief, and uncertainty, clearly separating what is supported by observed data from what remains unknown. Such a decomposition is especially useful for accounting for epistemic uncertainty, which is frequent in educational settings. This in mind, representing a learner's mastery of concepts as an opinion in the sense of SL is particularly appealing for KT.

As for probabilistic KT models, the intended model will share the advantage of explicitly modeling learner mastery at the concept level, (see Figure 1 Step 1). It will differ by accounting for uncertainty to distinguish between low mastery supported by substantial evidence and low mastery that merely reflects insufficient information (and symmetrically for high mastery).

This model will also be comparable to uncertainty-aware deep KT models of the literature. Yet these models typically define learner mastery at the *exercise* level, rather than at the *concept* level. Moreover, while these models do integrate uncertainty internally, they do not provide an explicit uncertainty representation that can be extracted or reused for downstream tasks.

We thus propose KT-USL, a probabilistic uncertainty-aware knowledge tracing model that relies on subjective logic to both model concept-level mastery and quantify uncertainty.

3.2.1. Concept Mastery Viewed as an Opinion in SL

KT-USL follows the general KT framework introduced in Section 2.1. Like classical probabilistic KT models, it takes as input the learner's interaction history $H_t(l)$ defined in equation ((3)), updates the learner–concept mastery through an aggregation mechanism (equation ((5))), and ultimately predicts the correctness of the next response using the prediction function (equation ((6))).

The key difference lies in how the learner's mastery of concepts is represented. In the general framework, the mastery $m_t(l, c)$ is a single scalar in $[0, 1]$, implicitly treated as a fully reliable estimate. In contrast, KT-USL outputs, at each time step t, a mastery value for each learner–concept pair that incorporates uncertainty into its computation $m_t(l, c)$, together with an explicit quantification of how confident the model is in that mastery estimate $u_t(l, c)$.

KT-USL maintains a *subjective logic opinion* for each learner–concept pair, computed from the evidence accumulated up to time $t - 1$:

$$\omega_t(l, c) = \big(b_{t-1}(l, c),\ d_{t-1}(l, c),\ u_{t-1}(l, c),\ a\big), \tag{14}$$

where $b_{t-1}(l, c)$ denotes belief in mastery, $d_{t-1}(l, c)$ denotes disbelief, and $u_{t-1}(l, c)$ captures epistemic uncertainty.

From this temporally evolving opinion, the model derives both a mastery estimate and the associated uncertainty. These quantities are updated according to the subjective logic formulation introduced in Section 3.1 and the uncertainty definition in Equation 12. The mastery estimate is given by:

$$m_t(l, c) = g_{\mathrm{SL}}\big(H_t(l, c)\big), \tag{15}$$

where g_{SL} is the subjective-logic aggregation mechanism applied to the interaction history of learner l on concept c.

Concretely, this aggregation yields a subjective logic opinion $\omega_t(l, c)$, from which the mastery estimate is obtained via its expected value:

$$m_t(l, c) = E(\omega_t(l, c)) = b_{t-1}(l, c) + a\, u_{t-1}(l, c), \tag{16}$$

where $b_t(l, c)$, $d_t(l, c)$, and $u_t(l, c)$ are the belief, disbelief, and epistemic uncertainty at time t.

The uncertainty at time $t - 1$, computed from the accumulated evidence $r_{t-1}(l, c)$ and $s_{t-1}(l, c)$, is:

$$u_{t-1}(l, c) = \frac{W}{r_{t-1}(l, c) + s_{t-1}(l, c) + W}. \tag{17}$$

This modification is designed to provide two key benefits. First, it will enable more robust predictions by integrating uncertainty into the mastery computation, preventing overly confident updates when evidence is scarce. Second, the explicit representation of uncertainty will yield an intrinsically interpretable view of learner knowledge, thereby enhancing the explainability of the KT model.

3.2.2. Example

We set $a = 0.5$ and $W = 2$, where a represents the prior mastery assumed in the absence of evidence, and W is the prior weight assigned to uncertainty. A large W means that uncertainty decreases more slowly as evidence accumulates. Table 1 illustrates how the opinion for a learner–concept pair (l, c) evolves across successive interactions.

Table 1
Example for KT-USL on a single concept (values rounded).

Step	**Wins** $r_{l,c}$	**Fails** $s_{l,c}$	**Opinion** $(b_{l,c}, d_{l,c}, u_{l,c}, E_{l,c})$
0	0	0	(0.00, 0.00, 1.00, 0.50)
1	1	0	(0.33, 0.00, 0.67, 0.67)
2	3	1	(0.50, 0.17, 0.33, 0.67)
3	5	0	(0.71, 0.00, 0.29, 0.86)

- **Step 0 (no interactions).** With $r_{l,c} = 0$ and $s_{l,c} = 0$, the model has no evidence: $b_{l,c} = 0$, $d_{l,c} = 0$, and $u_{l,c} = 1$. The expected mastery is $E_{l,c} = 0.50$. The learner is treated as neither clearly mastering nor clearly failing the concept.
- **Step 1 (one correct answer).** After a single success ($r_{l,c} = 1, s_{l,c} = 0$), belief increases to $b_{l,c} \approx 0.33$, disbelief remains $d_{l,c} = 0$, and uncertainty drops to $u_{l,c} \approx 0.67$. The expected mastery rises to $E_{l,c} \approx 0.67$, but the high uncertainty indicates that the estimate relies on very limited evidence.
- **Step 2 (three successes, one failure).** With more practice ($r_{l,c} = 3, s_{l,c} = 1$), belief grows to $b_{l,c} = 0.50$, disbelief increases to $d_{l,c} \approx 0.17$, and uncertainty decreases to $u_{l,c} \approx 0.33$. The expected mastery remains around 0.67, but now this estimate is supported by a stronger evidence base and lower uncertainty.
- **Step 3 (five successes, no failures).** After additional correct answers ($r_{l,c} = 5, s_{l,c} = 0$), belief dominates with $b_{l,c} \approx 0.71$, disbelief returns to $d_{l,c} = 0$, and uncertainty drops further to $u_{l,c} \approx 0.29$. The expected mastery increases to $E_{l,c} \approx 0.86$. The learner is now estimated to strongly master the concept with relatively high confidence.

4. Experimental Setup

4.1. Datasets

We evaluate KT-USL on three widely used mathematics knowledge tracing datasets: Junyi, ASSISTments 2015, and Eedi (Task 1–2). Using datasets from the same subject domain ensures that cross-dataset evaluations remain comparable, while the three datasets still differ in content structure, annotation schemes, and interaction density.

Junyi The Junyi dataset is provided by the Junyi Academy learning platform and covers mathematics exercises across several grade levels.[2] Interactions include learner identifiers, exercise identifiers, timestamps, and a correctness binary label. We map each exercise to a unique knowledge component using the mapping file provided and discard interactions whose exercises cannot be mapped.

ASSISTments 2015 ASSISTments 2015 consists of interaction logs collected from the ASSISTments intelligent tutoring system, widely used in middle-school mathematics education. [3] The dataset records learner–problem interactions, including problem identifiers, associated concepts, and binary correctness labels indicating whether a learner answered a problem correctly.

Eedi (Task 1–2) Eedi is an online mathematics learning platform that provides multi-step metadata describing questions, learners, and answers. [4] We use the Task 1-2 version of the dataset, which provides primary logs, question metadata, answer timestamps, and hierarchical curriculum tags. We retain only Level-3 curriculum tags, which represent fine-grained knowledge components.

4.2. Preprocessing

All datasets are processed using a unified pipeline. Invalid interactions (missing identifiers, concepts, correctness values, or timestamps) are first removed. Questions are then mapped to concepts using dataset-specific metadata, timestamps are normalized, and interactions are sorted chronologically per learner. Interactions with unmatched concepts are discarded, and each learner's history is finally represented as a time-ordered sequence of tuples (`questionId`, `subjectId`, `isCorrect`, `timestamp`), yielding comparable traces across models.

The final statistics related to the three datasets are presented in Table 2. We report the global success rate, defined as the proportion of correct responses over all interactions in the dataset. This rate

[2]https://www.kaggle.com/datasets/junyiacademy/learning-activity-public-dataset-by-junyi-academy
[3]https://sites.google.com/site/assistmentsdata/datasets/2015-assistments-skill-builder-data
[4]https://www.kaggle.com/datasets/alejopaullier/eedi-external-dataset

indicates that the datasets are moderately imbalanced. We also report the mean number of interactions per learner–concept pair. This corresponds to the average number of observations available to estimate a learner's mastery of a given concept, which is particularly relevant for probabilistic knowledge tracing models that rely on learner–concept statistics.

Dataset	#Learners	#Interactions	#Concepts	Success rate	Mean interactions/(l,c)
EEDI	118,971	15,867,850	290	0.643	2.826
Junyi	72,758	16,217,311	171	0.699	7.025
ASSISTments 2015	19,917	708,631	100	0.706	5.672

Table 2
Summary statistics of the three processed datasets.

4.3. Baseline Models

We evaluate KT-USL against representative baselines from the three main families of knowledge tracing (KT): probabilistic models and one deep learning model. For consistency and reproducibility, the deep KT baseline (UKT) is implemented using the publicly available PyKT framework of Liu et al. [42]. It is important to note that, while this deep models provides strong predictive accuracy, it does not offer explicit mechanisms for explainability or uncertainty interpretation—features that KT-USL is specifically designed to provide.

This selection ensures (1) a fair comparison with classical interpretable approaches and (2) a state-of-the-art neural architecture.

The baselines considered are BKT [1], PFA [16], and UKT [14].

4.4. Metrics

We evaluate predictive performance using complementary metrics. *Accuracy* measures the proportion of correct predictions at a fixed decision threshold. *F1-score* balances precision and recall and is particularly informative under class imbalance. *AUC* assesses the ability of a model to rank correct responses above incorrect ones independently of the decision threshold. *ROC curves* visualize this ranking behaviour by showing the trade-off between true and false positive rates across thresholds.

4.5. Protocol

All experiments follow a unified evaluation protocol across datasets and models. For each learner, interaction sequences are ordered temporally and split into 80% for training and 20% for testing, ensuring that no future information leaks into the training phase. All attempts are retained. Models are evaluated in an online setting: during testing, they sequentially predict each interaction outcome and then update their internal state after observing the true response, which reflects real-world knowledge tracing usage.

5. Experiments and Results

In this section we evaluate the predictive performance of KT-USL on the three and ASSISTments 2015. Our primary goal is to assess whether KT-USL improves upon classical probabilistic models BKT and PFA, while preserving explicit, concept-level mastery representations. The UKT deep learning model is included as a state-of-the-art reference point that provides an upper bound on predictive performance rather than a direct competitor.

All models are evaluated using the unified protocol described in Section 4.5. Table 3 reports Accuracy, AUC, and F1-score for all models and datasets, while Figure 3 presents the corresponding ROC curves.

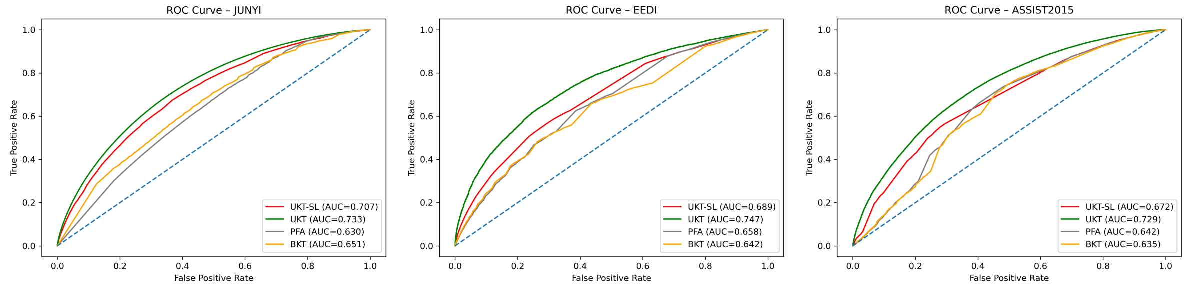

Figure 3: ROC curves on **Junyi**, **EEDI**, and **ASSISTments 2015**. Curves closer to the top-left corner indicate better ranking performance.

5.1. Predictive performance across datasets

		Model			
		Probabilistic			Deep
Dataset	**Metric**	**BKT**	**PFA**	**KT-USL**	**UKT**
Junyi	Accuracy	0.709	0.695	0.721	**0.732**
	AUC	0.651	0.630	**0.752**	0.749
	F1	0.814	0.797	0.889	**0.920**
EEDI	Accuracy	0.622	0.574	0.671	**0.719**
	AUC	0.642	0.658	0.691	**0.755**
	F1	0.683	0.575	**0.843**	0.806
ASSISTments 2015	Accuracy	0.665	0.506	0.687	**0.747**
	AUC	0.635	0.642	0.796	**0.807**
	F1	0.758	0.557	0.775	**0.865**

Table 3
Performance comparison across datasets in terms of Accuracy, AUC, and F1-score.

First of all, the reported values are consistent with those commonly observed in the literature [2, 14]. Considering the baseline BKT, its accuracy is slightly higher than the correct rate on Junyi, which means that its accuracy is not higher than a simple model that systematically predicts a mastery. On EEDI and ASSISTments 2015, the accuracy falls below the correct rate. This behavior can be attributed to the sparsity of both datasets. The effect is particularly pronounced on EEDI, which exhibits the lowest interaction density (2.83 interactions per learner–concept pair). As a result, BKT attains relatively modest Accuracy, especially on sparse datasets. Considering AUC and F1-score, conclusions are similar, with smaller values on EEDI than on denser datasets.

PFA underperforms BKT in terms of Accuracy and F1 across all datasets, with the largest drop observed on ASSISTments 2015 and EEDI. PFA relies on question difficulty as well as the cumulative number of past successes and failures. When the difficulty is poorly estimated or the success and failure counts are low, the resulting probability estimates become unstable, leading to degraded performance.

The proposed KT-USL model consistently outperforms BKT across all datasets and metrics. Gains are particularly strong on EEDI, the most sparse dataset, where KT-USL improves over BKT by 7.8% in Accuracy and 23.4% in F1-score. This improvement was expected and can be attributed to the explicit modeling of uncertainty through subjective logic, which is particularly beneficial in low-density interaction settings, where evidence accumulation is limited.

To assess whether the observed improvements are statistically meaningful, we conducted bootstrap confidence interval tests [43] comparing KT-USL against BKT. Results show that Accuracy and F1-score of KT-USL are significantly higher than BKT across all datasets. For instance, on **EEDI**, the improvement in F1-score ($\Delta = 0.0770$) is associated with a tight 95% confidence interval $[0.0766, 0.0775]$, indicating a robust and meaningful gain. Improvements are smaller on Junyi, but remain significant, where BKT already performs strongly due to the dataset's denser interaction structure.

In terms of the Area Under the Curve (AUC), which summarizes a model's overall ranking ability by measuring its capacity to assign higher scores to correct responses than to incorrect ones independently of any fixed decision threshold, KT-USL consistently outperforms probabilistic baselines across all datasets. Moreover, the relative improvements are more pronounced in AUC than in Accuracy, indicating that KT-USL achieves better discrimination between correct and incorrect responses across different decision thresholds.

As expected, the deep UKT model achieves the highest overall Accuracy and AUC across datasets, benefiting from its expressive sequential representation. However, an important observation is that KT-USL exceeds UKT in F1-score on Junyi, despite relying on a simpler probabilistic formulation.

The strong performance of UKT was expected given its training setting and internal structure. Unlike probabilistic models that update mastery independently for each learner–concept pair, UKT is trained on the full training set and processes each learner's entire interaction sequence across concepts using a self-attention mechanism. This enables the model to capture richer temporal and cross-concept dependencies, which in turn improves predictive performance.

In summary, the results demonstrate that KT-USL effectively bridges the gap between explainability and predictive performance. While deep models remain strong in absolute terms, the proposed probabilistic approach achieves competitive performance—particularly in terms of F1-score and AUC—while providing principled uncertainty quantification at the concept level, thereby improving explainability.

5.2. ROC curves analysis

Figure 3 presents the ROC curves obtained on Junyi, EEDI, and ASSISTments 2015. On Junyi, the proposed KT-USL model consistently dominates the probabilistic baselines across the entire range of false positive rates. Its ROC curve remains above those of BKT and PFA at all operating points. This behavior is also reflected in the higher AUC value and can be attributed to the relatively high interaction density of Junyi, which enables the uncertainty-aware belief updates of KT-USL to accumulate reliable evidence over time.

On EEDI, KT-USL clearly outperforms BKT and PFA in the low-to-mid false positive rate region, remaining superior up to approximately a false positive rate of 0.6. Beyond this region, the curves tend to converge, which suggests that when permissive thresholds are used, the advantage of uncertainty modeling decreases. This behavior can be explained by the characteristics of the EEDI dataset, which combines a relatively low correct rate (0.643) with a very sparse learner-concept interaction density (2.83 interactions per learner-concept pair). In this setting, many predictions rely on limited evidence. By explicitly representing uncertainty, KT-USL adopts a more conservative behavior when evidence is weak, which helps reduce false positives in low false positive rate regimes.

A similar pattern is observed on ASSISTments 2015, which exhibits relatively sparse learner-concept interactions (5.672 interactions per learner-concept pair). KT-USL outperforms BKT and PFA in the low to moderate false positive rate region (up to approximately 0.4) by avoiding overconfident predictions under limited evidence. As thresholds become more permissive, the ROC curves converge.

6. Conclusion

In this paper, we addressed the gap between explainability and predictive performance in Knowledge Tracing (KT) and investigate how uncertainty quantification can be leveraged to enhance both. Existing KT approaches tend to prioritize either predictive performance, such as deep neural architectures, or explainability, such as probabilistic models. However, none of these models jointly provide explicit mastery representations, to foster explainability, together with quantification of uncertainty.

To bridge this gap, we introduced KT-USL, an uncertainty-aware probabilistic KT model grounded in Subjective Logic. KT-USL represents each learner-concept pair as an opinion decomposing evidence into belief, disbelief, and uncertainty, complemented by a base-rate prior. The SL formalism yields both an interpretable mastery estimate and a principled measure of how strongly this estimate is supported by the available data. We also illustrated how the interaction between expected mastery and uncertainty can inform educational decisions, such as determining whether a learner is ready to advance or whether additional evidence should be collected through targeted practice.

Experiments conducted confirmed that uncertainty quantification contributes to increasing performance of probabilistic models, leading to the expected trade-off between explainability and performance.

Several directions remain for future work. On the modelling side, we plan to extend KT-USL to incorporate prerequisite and hierarchical relationships between concepts within a knowledge graph structure. This will allow mastery estimates to propagate through concept dependencies and better reflect the semantics of the domain. On the application side, we aim to integrate KT-USL into an adaptive learning platform that personalises learning paths based on learner mastery and the structure of the underlying knowledge graph.

Declaration on Generative AI

During the preparation of this work, the authors used a generative language model (ChatGPT, OpenAI) for limited language editing, specifically to suggest alternative English phrasings and to improve grammar and stylistic clarity in some paragraphs. No scientific content, modelling choices, experimental design, results, or interpretations were generated by the tool. All methodological decisions, analyses, and conclusions were conceived, verified, and approved by the authors, who take full responsibility for the content of this publication.

References

[1] A. T. Corbett, J. R. Anderson, Knowledge tracing: Modeling the acquisition of procedural knowledge, User Modeling and User-Adapted Interaction 4 (1994) 253–278.

[2] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, J. Sohl-Dickstein, Deep knowledge tracing, in: Advances in Neural Information Processing Systems 28 (NeurIPS 2015), Curran Associates, Inc., 2015, pp. 505–513. URL: https://papers.nips.cc/paper/5654-deep-knowledge-tracing.

[3] Y. Yilmaz, W. Al-Halabi, Y. Yılmaz, B. Musal, R. H. Ellaway, Developing a dashboard for faculty development in competency-based training programs: a design-based research project, Canadian Medical Education Journal 12 (2021) 48–64.

[4] J. Gervais, The operational definition of competency-based education, The Journal of Competency-Based Education 1 (2016) 98–106. doi:10.1002/cbe2.1011.

[5] C. Chappell, A. Gonczi, P. Hager, Competency-based education, in: Understanding Adult Education and Training, Routledge, London, UK, 2020, pp. 191–205.

[6] S. Shen, Z. Yu, X. Chen, L. Sun, M. Sun, X. Li, Y. Ma, A survey of knowledge tracing: Models, variants, and applications, IEEE Transactions on Learning Technologies 17 (2024) 1858–1879.

[7] W. Cheng, S. Lyu, Y. Yang, M. Minaei, L. Sun, Uncertainty-aware knowledge tracing, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, 2025.

[8] J. C. Helton, D. E. Burmaster, Aleatory or epistemic? does it matter?, Risk Analysis 26 (2005) 515–518.

[9] Y. Bai, Y. Zhang, X. Wang, et al., A survey of explainable knowledge tracing, arXiv preprint arXiv:2403.07279 (2024).

[10] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, J. Sohl-Dickstein, Deep knowledge tracing, in: Advances in Neural Information Processing Systems, volume 28, 2015.

[11] J. Zhang, X. Shi, I. King, D. Yeung, Dynamic key-value memory networks for knowledge tracing, in: Proceedings of the 26th International Conference on World Wide Web (WWW '17), ACM, 2017, pp. 765–774. doi:10.1145/3038912.3052580.

[12] J. Mitton, S. Noveck, A. Gottlieb, et al., Uncertainty-aware knowledge tracing models, arXiv preprint arXiv:2509.21514 (2025).

[13] X. Gao, et al., A hierarchical probabilistic framework for incremental knowledge tracing in classroom settings, arXiv preprint arXiv:2506.09393 (2025).
[14] W. Cheng, H. Du, C. Li, E. Ni, L. Tan, T. Xu, Y. Ni, Uncertainty-aware knowledge tracing, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, 2025, pp. 27905–27913. URL: https://ojs.aaai.org/index.php/AAAI/article/view/35007. doi:10.1609/aaai.v39i27.35007.
[15] H. Cen, K. Koedinger, B. Junker, Learning factors analysis – a general method for cognitive model evaluation and improvement, in: Proceedings of the 8th International Conference on Intelligent Tutoring Systems (ITS 2006), Springer, 2006, pp. 164–175.
[16] P. I. Pavlik, H. Cen, K. R. Koedinger, Performance factors analysis – a new alternative to knowledge tracing, Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED 2009) (2009) 531–538.
[17] S. Pandey, G. Karypis, A self-attentive model for knowledge tracing, arXiv preprint arXiv:1907.06837 (2019). URL: https://arxiv.org/abs/1907.06837.
[18] T. Käser, S. Klingler, A. G. Schwing, M. Gross, Dynamic Bayesian networks for student modeling, IEEE Transactions on Learning Technologies 10 (2017) 450–462.
[19] M. V. Yudelson, K. R. Koedinger, G. J. Gordon, Individualized Bayesian knowledge tracing models, in: International Conference on Artificial Intelligence in Education, Springer, Berlin, Heidelberg, 2013, pp. 171–180.
[20] Y. Qiu, Y. Qi, H. Lu, Z. A. Pardos, N. T. Heffernan, Does time matter? modeling the effect of time with bayesian knowledge tracing, in: Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011), 2011.
[21] K. Zhang, Y. Yao, A three learning states Bayesian knowledge tracing model, Knowledge-Based Systems 148 (2018) 189–201.
[22] F. Liu, S. Gong, W. Wang, Y. Liu, Fuzzy bayesian knowledge tracing, IEEE Transactions on Fuzzy Systems 30 (2021) 2412–2425.
[23] H. Cen, K. Koedinger, B. Junker, Learning factors analysis–a general method for cognitive model evaluation and improvement, in: Proceedings of the 8th International Conference on Intelligent Tutoring Systems, Springer, 2008, pp. 164–175.
[24] J. P. González-Brenes, Y.-T. Huang, P. Brusilovsky, Towards accurately interpreting student model parameters: A view from item response theory, in: Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), 2014, pp. 38–45.
[25] M. Khajah, R. V. Lindsey, M. C. Mozer, Integrating cognitive modeling and item response theory: A non-linear time-varying ability model for learning and assessment, in: Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), 2014, pp. 64–71.
[26] J.-J. Vie, H. Kashima, Knowledge tracing machines: Factorization machines for knowledge tracing, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 750–757. URL: https://doi.org/10.1609/aaai.v33i01.3301750. doi:10.1609/aaai.v33i01.3301750.
[27] L. Wimmer, E. Derner, S. Teso, A. Vergari, Quantifying aleatoric and epistemic uncertainty in machine learning: Are conditional entropy and mutual information appropriate measures?, in: Uncertainty in Artificial Intelligence, PMLR, 2023, pp. 2293–2303.
[28] W. E. Walker, P. Harremoës, J. Rotmans, J. P. Van Der Sluijs, M. B. Van Asselt, P. Janssen, M. P. Krayer von Krauss, Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support, Integrated Assessment 4 (2003) 5–17.
[29] E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Machine Learning 110 (2021) 457–506.
[30] S. T. Christie, C. Cook, A. N. Rafferty, Uncertainty-preserving deep knowledge tracing with state-space models, arXiv preprint arXiv:2407.17427 (2024).
[31] X. Ding, E. C. Larson, Incorporating uncertainties in student response modeling by loss function regularization, Neurocomputing 409 (2020) 74–82.
[32] F. Doshi-Velez, B. Kim, Considerations for evaluation and generalization in interpretable machine learning, in: W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, K.-R. Müller (Eds.), Explainable and Interpretable Models in Computer Vision and Machine Learning, Springer International Publishing,

Cham, 2018, pp. 3–17.
[33] C. Agarwal, A. Nguyen, M. W. Mahoney, K. R. Varshney, OpenXAI: Towards a transparent evaluation of model explanations, Advances in Neural Information Processing Systems 35 (2022) 15784–15799.
[34] Y. Lu, X. Chen, X. Li, W. Chen, X. He, Interpreting deep learning models for knowledge tracing, International Journal of Artificial Intelligence in Education 33 (2023) 519–542.
[35] C.-Q. Huang, et al., Xkt: Toward explainable knowledge tracing model with cognitive learning theories for questions of multiple knowledge concepts, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 7308–7325.
[36] B. Jiang, et al., Improving the performance and explainability of knowledge tracing via markov blanket, Information Processing & Management 61 (2024) 103620.
[37] H. Li, et al., Explainable few-shot knowledge tracing, Frontiers of Digital Education 2 (2025) 34.
[38] Q. Li, et al., A genetic causal explainer for deep knowledge tracing, IEEE Transactions on Evolutionary Computation 28 (2023) 861–875.
[39] C. Haydar, A. Roussanaly, A. Boyer, Local trust versus global trust networks in subjective logic, in: 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), volume 1, IEEE, 2013.
[40] A. Jøsang, L. Kaplan, Principles of subjective networks, in: Proceedings of the 19th International Conference on Information Fusion (FUSION), IEEE, 2016, pp. 1150–1157.
[41] A. Jøsang, Subjective Logic: A Formalism for Reasoning under Uncertainty, Springer, Cham, 2018.
[42] Z. Liu, S.-T. Yeung, S. Gao, X. Guo, M. Zhao, Y. Chen, B. Chen, X. Liu, Y. Yacoob, F. Rudzicz, D.-Y. Yeung, pykt: A python library to benchmark deep learning based knowledge tracing models, in: Advances in Neural Information Processing Systems, volume 35, 2022, pp. 18542–18555.
[43] K. Boyd, K. H. Eng, C. D. Page, Area under the precision–recall curve: Point estimates and confidence intervals, in: Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD), Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2013, pp. 451–466.

A. Online Resources

The implementation of **KT-USL**, along with scripts for data preprocessing and evaluation, is publicly available on GitHub at: https://github.com/anonymousreseach/KT_USL.git.