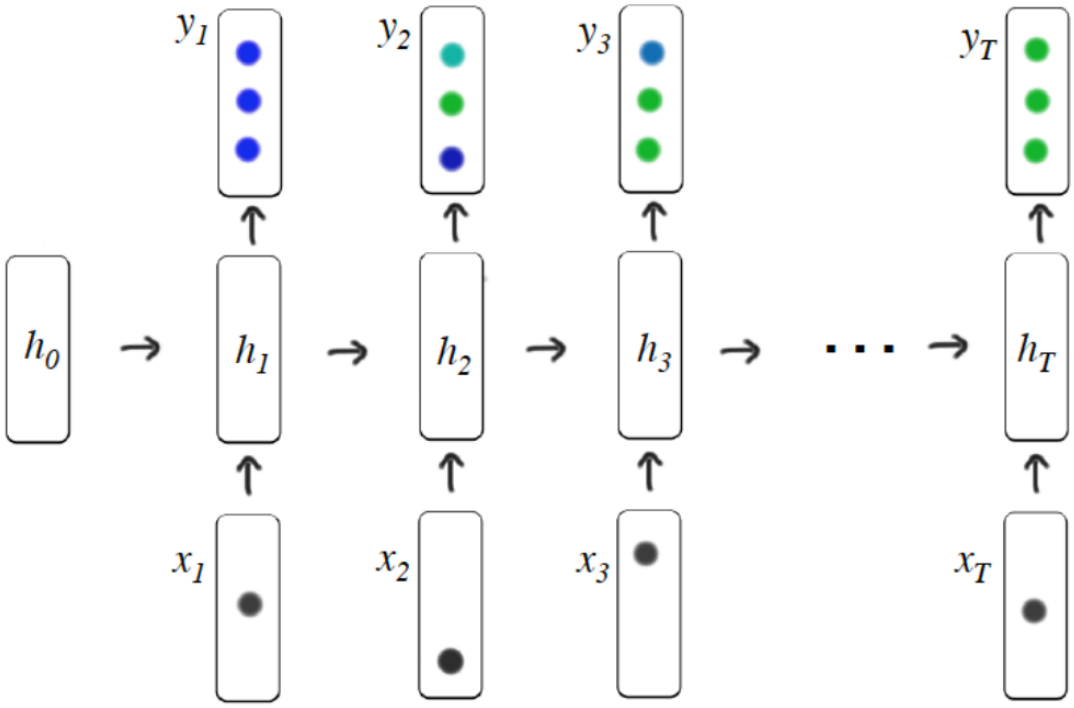
Models

Since the deep learning based KT models can be categorized into deep sequential models, memory augmented models, adversarial based models, graph based models and attention based models in our work, we mainly develop the DLKT models by these four categories in pyKT.

DKT

DKT is the first model that uses Recurrent Neural Networks (RNNs) to solve Knowledge Tracing.



Piech, Chris, et al. “Deep knowledge tracing.” Advances in neural information processing systems 28 (2015).

DKT+

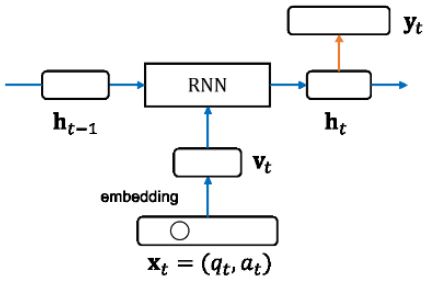
DKT+ introduces regularization terms that correspond to reconstruction and waviness to the loss function of the original DKT model to enhance the consistency in KT predi

Yeung, Chun-Kit, and Dit-Yan Yeung. “Addressing two problems in deep knowledge tracing via prediction-consistent regularization.” Proceedings of the Fifth Annual ACM Conference on Learning at Scale. 2018.

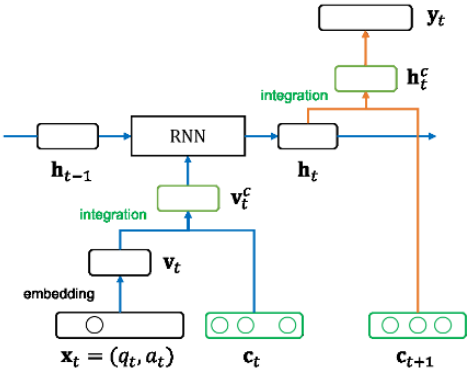
DKT-Forget

DKT-Forget explores the deep knowledge tracing model by considering the

forgetting behavior via incorporate multiple forgetting information.



(a) Architecture for deep knowledge tracing.

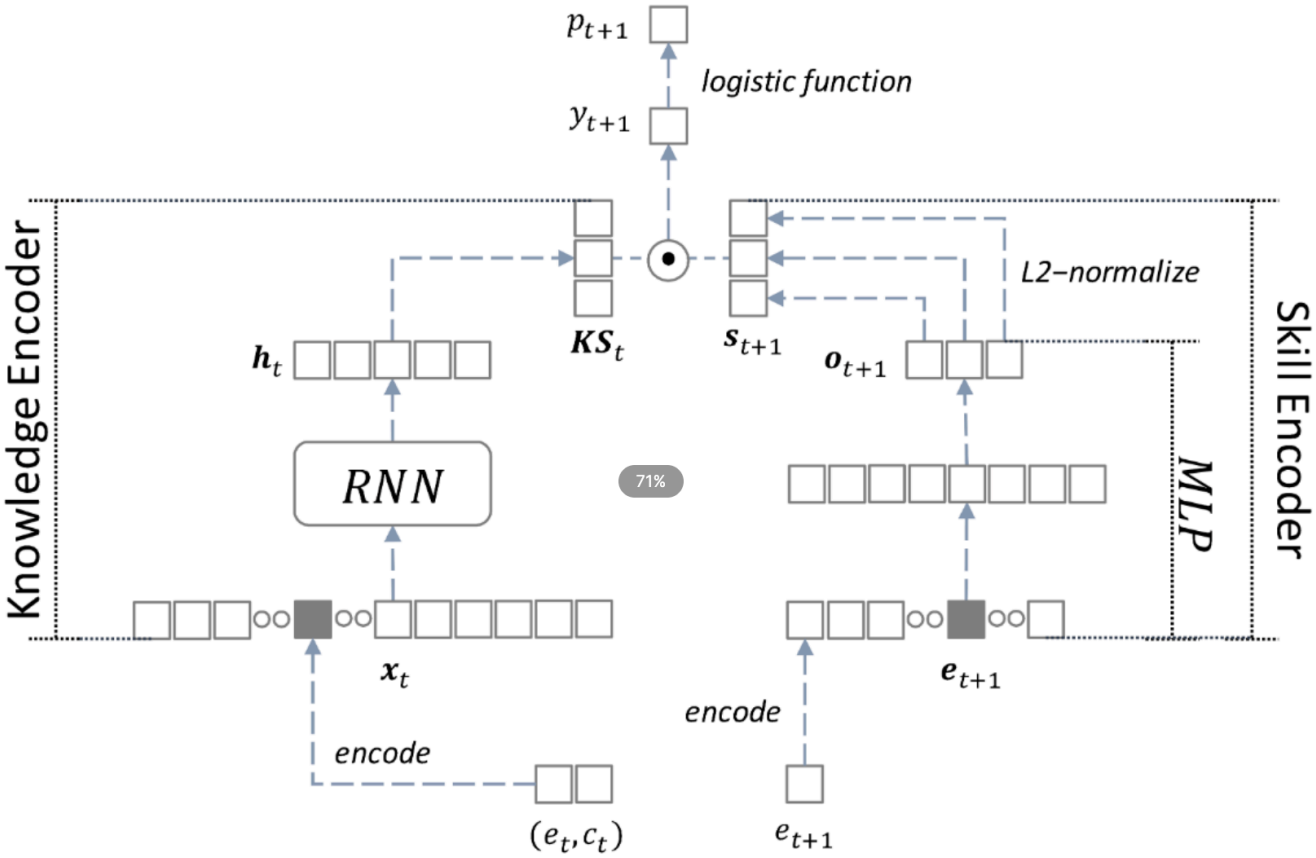


(b) Architecture for proposed model.

Nagatani, Koki, et al. “Augmenting knowledge tracing by considering forgetting behavior.” The world wide web conference. 2019.

KQN

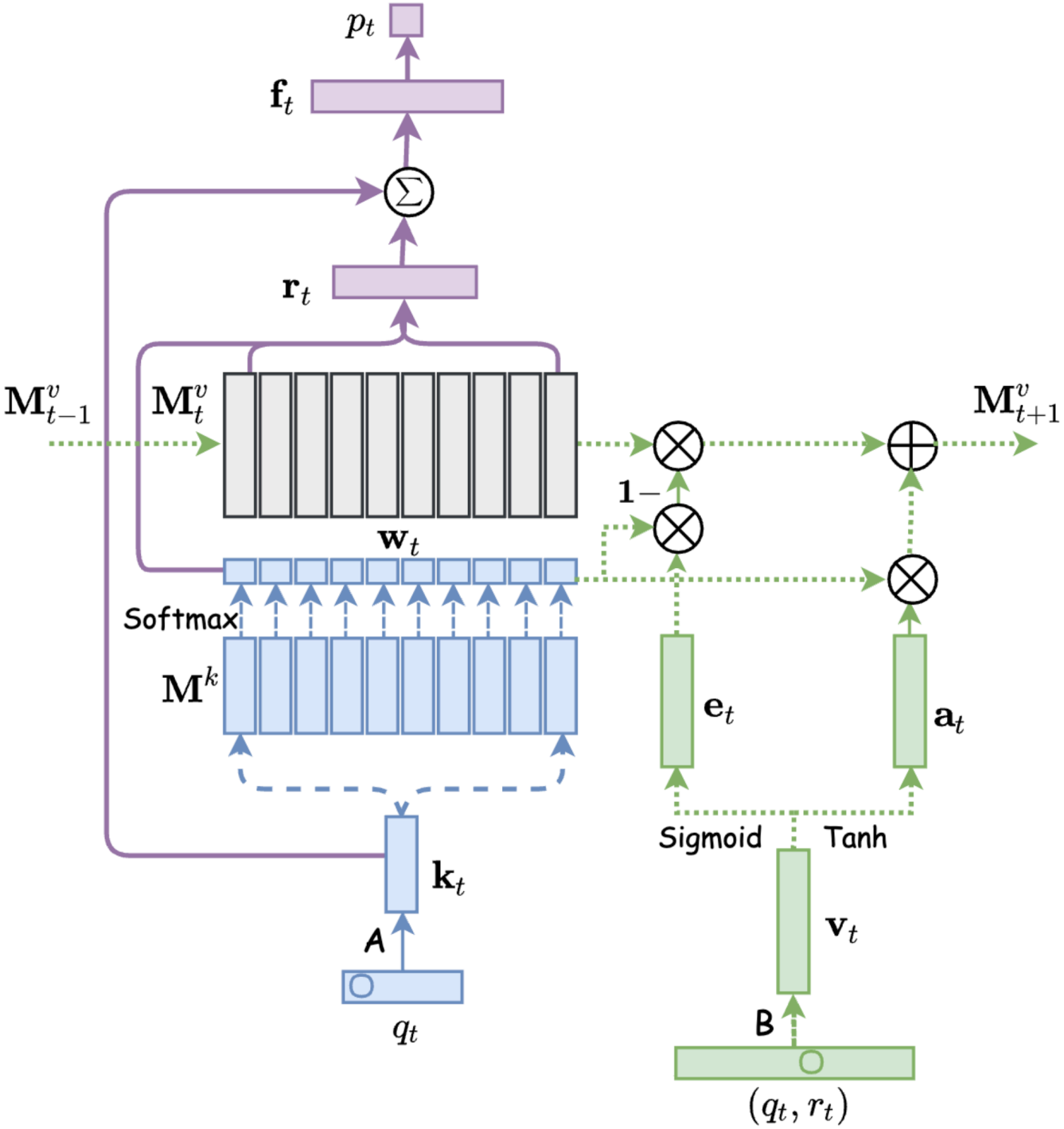
KQN uses neural networks to encode student learning activities into knowledge state and skill vectors, and calculate the relations between the interactions via dot product.



Lee, Jinseok, and Dit-Yan Yeung. “Knowledge query network for knowledge tracing: How knowledge interacts with skills.” Proceedings of the 9th international conference on learning analytics & Knowledge. 2019.

DKVMN

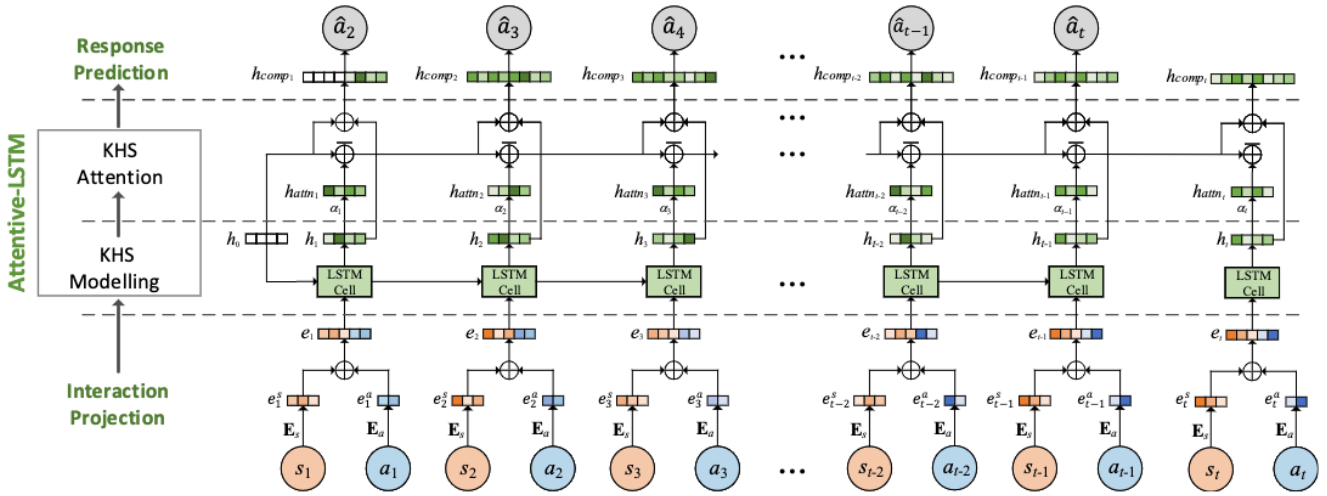
Dynamic key-value memory networks (DKVMN) exploit the relationships between latent KCs which are stored in a static memory matrix *key* and predict the knowledge mastery level of a student directly based on a dynamic memory matrix *value*.



Zhang, Jiani, et al. “Dynamic key-value memory networks for knowledge tracing.” Proceedings of the 26th international conference on World Wide Web. 2017.

ATKT

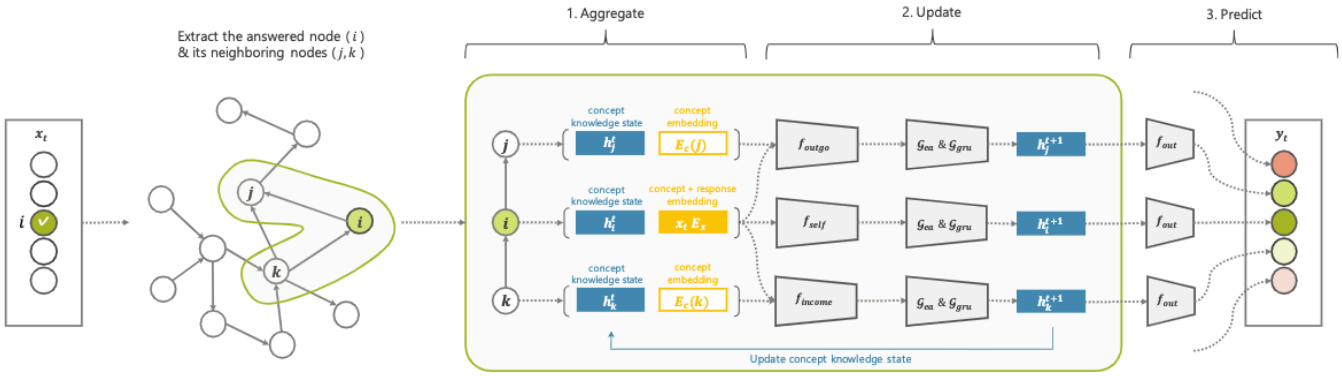
Adversarial training (AT) based KT method (ATKT) is an attention based LSTM model which apply the adversarial perturbations into the original student interaction sequence to reduce the the risk of DLKT overfitting and limited generalization problem.



Guo, Xiaopeng, et al. “Enhancing Knowledge Tracing via Adversarial Training.” Proceedings of the 29th ACM International Conference on Multimedia. 2021.

GKT

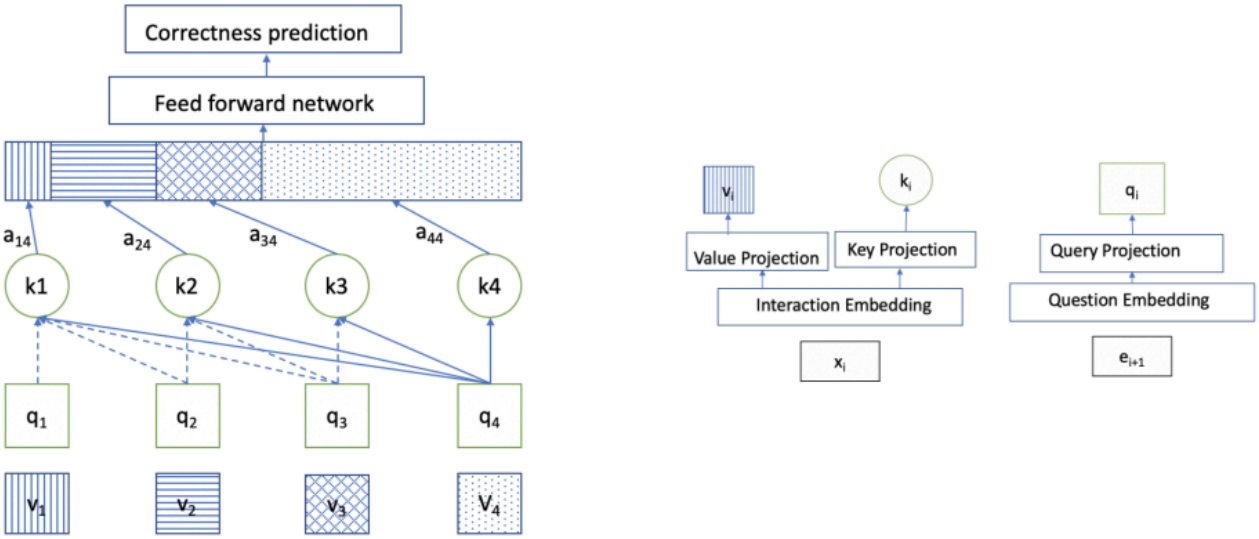
Graph-based Knowledge Tracing (GKT) is a GNN-based knowledge tracing method that use a graph to model the relations between knowledge concepts to reformulate the KT task as a time-series node-level classification problem.



Nakagawa, Hiromi, Yusuke Iwasawa, and Yutaka Matsuo. “Graph-based knowledge tracing: modeling student proficiency using graph neural network.” 2019 IEEE/WIC/ACM International Conference On Web Intelligence (WI). IEEE, 2019.

SAKT

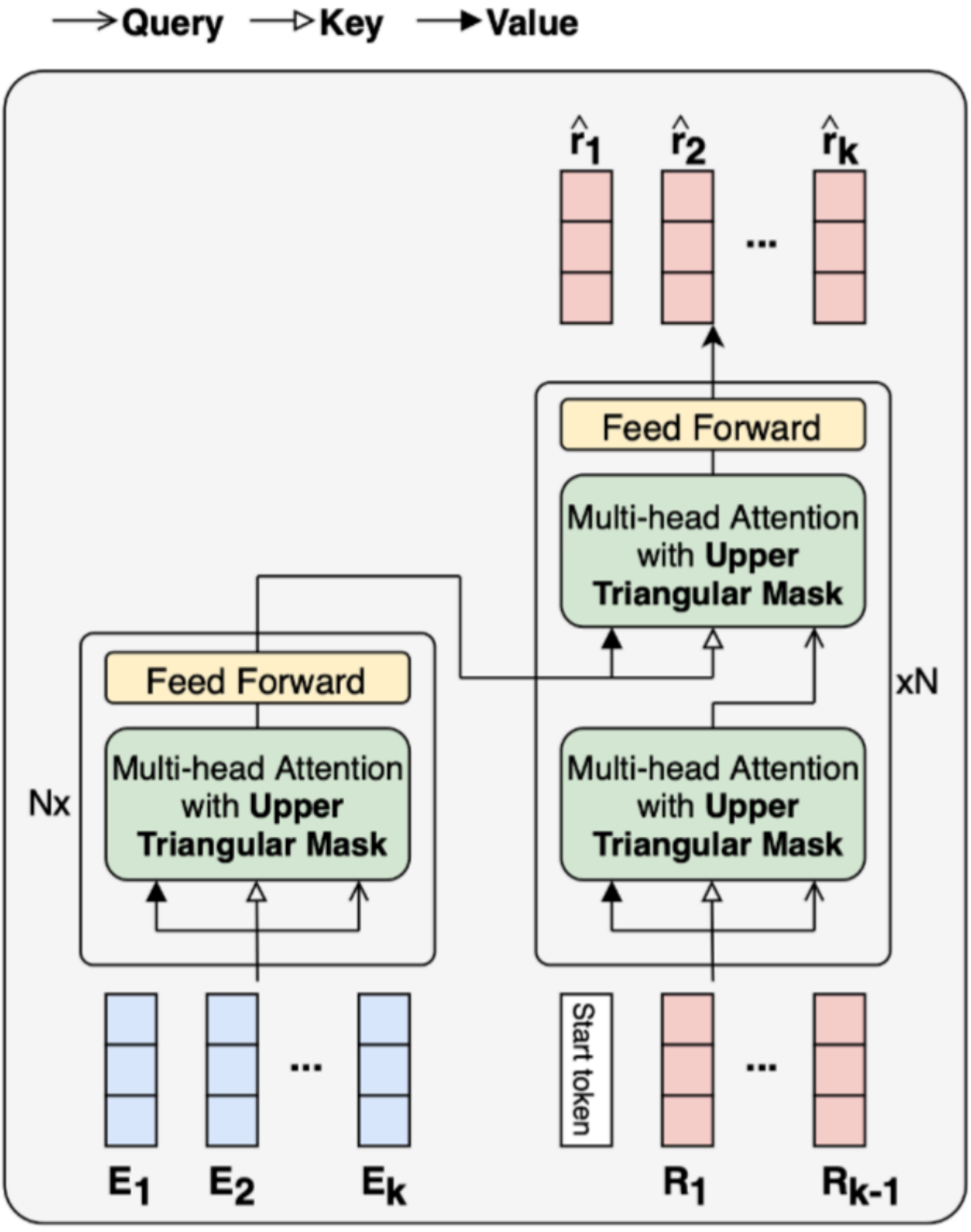
Self Attentive Knowledge Tracing (SAKT) use self-attention network to capture the relevance between the KCs and the students’ historical interactions.



Pandey, Shalini, and George Karypis. “A self-attentive model for knowledge tracing.” arXiv preprint arXiv:1907.06837 (2019).

SAINT

Separated Self-AttentIve Neural Knowledge Tracing(SAINT) is a typical Transformer based structure which embeds the exercises in encoder and predict the responses in decoder.

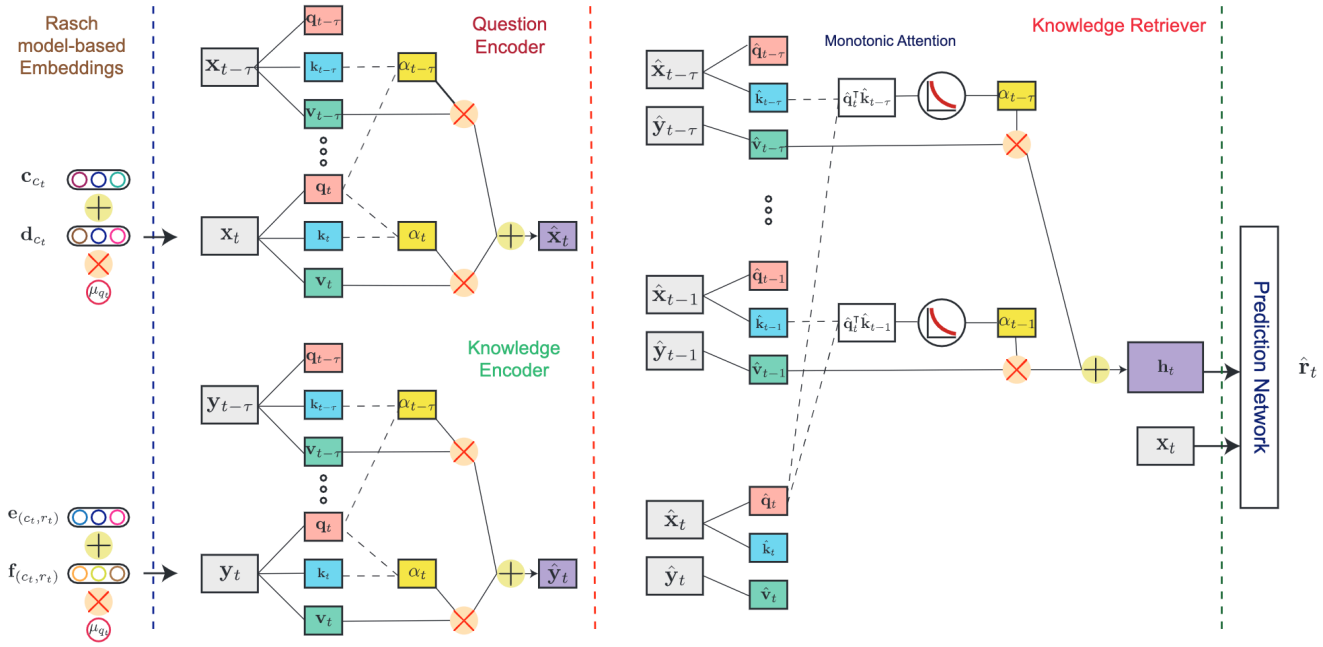


Choi, Youngduck, et al. “Towards an appropriate query, key, and value computation for knowledge tracing.” Proceedings of the Seventh ACM Conference on Learning@ Scale. 2020.

AKT

Attentive knowledge tracing (AKT) introduce a rasch model to

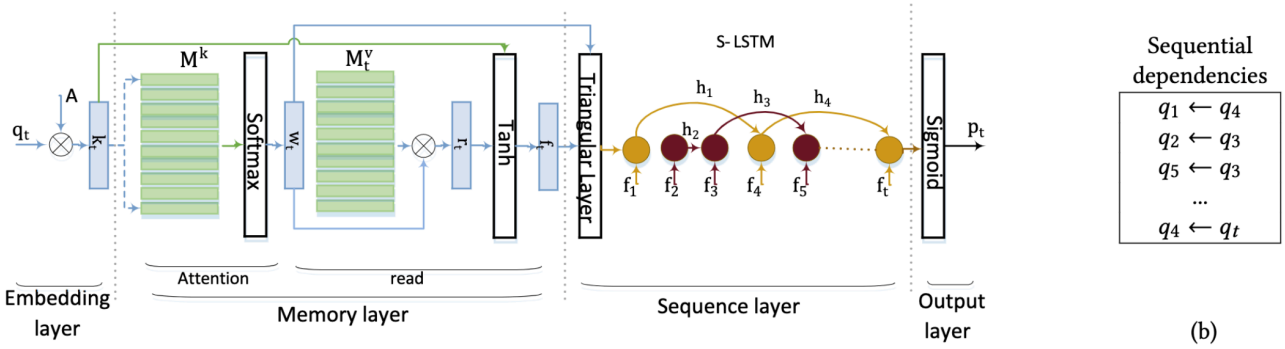
regularize the KC and question embeddings to discriminate the questions on the same KC, and modeling the exercise representations and the students' historical interactdion embeddings via three self-attention based modules.



Ghosh, Aritra, Neil Heffernan, and Andrew S. Lan. "Context-aware attentive knowledge tracing." Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020.

SKVMN

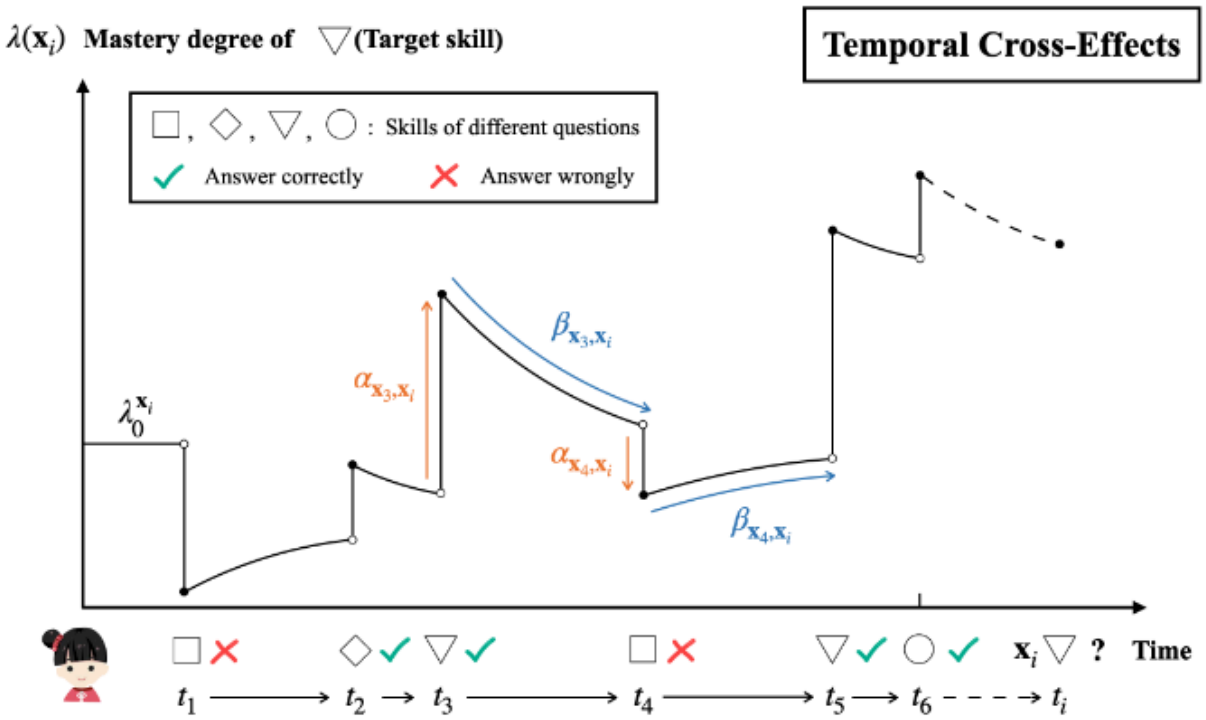
This model unifies the strengths of recurrent modeling capacity and the capability of memory networks to model the students' learning precocesses.



Abdelrahman, Ghodai, and Qing Wang. "Knowledge tracing with sequential key-value memory networks." Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019.

HawkesKT

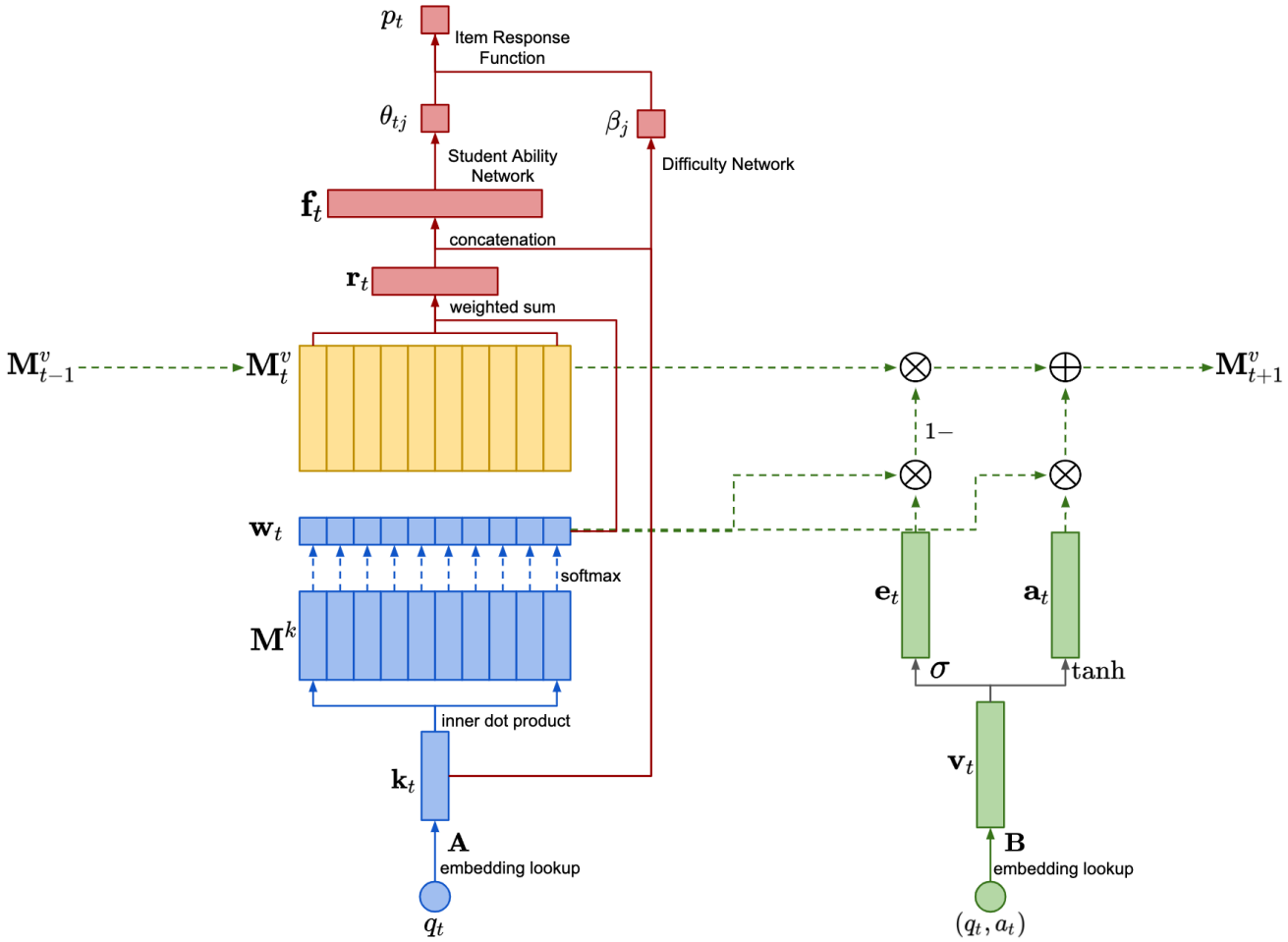
HawkesKT is the first to introduce Hawkes process to model temporal cross effects in KT.



Wang, Chenyang, et al. "Temporal cross-effects in knowledge tracing." Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021.

Deep-IRT

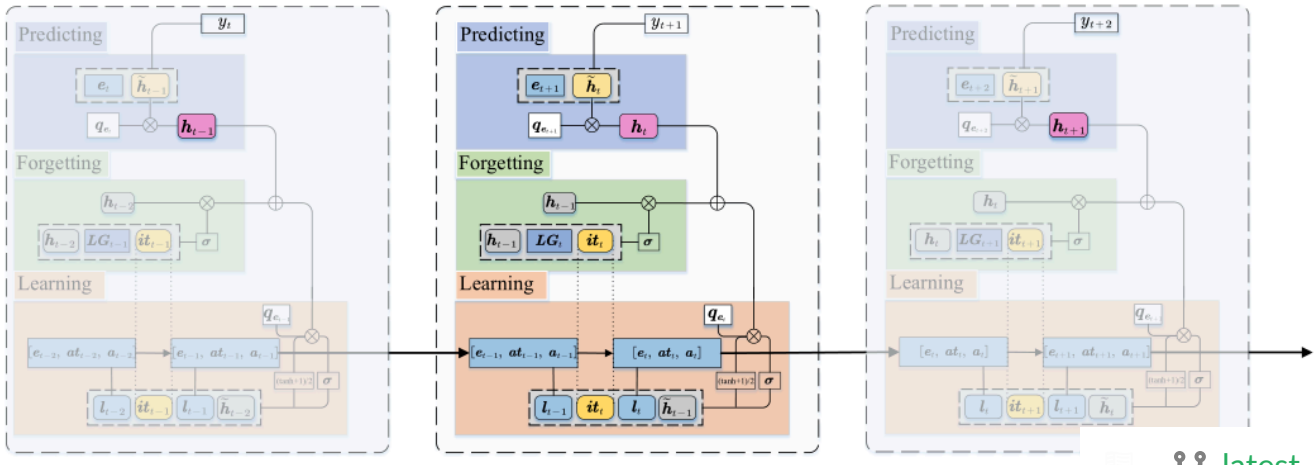
Deep-IRT is a synthesis of the item response theory (IRT) model and a knowledge tracing model that is based on the deep neural network architecture called dynamic key-value memory network (DKVMN) to make deep learning based knowledge tracing explainable.



Yeung, Chun-Kit. “Deep-IRT: Make deep learning based knowledge tracing explainable using item response theory.” arXiv preprint arXiv:1904.11738 (2019).

LPKT

Learning Processconsistent Knowledge Tracing(LPKT) monitors students’ knowledge state by directly modeling their learning process.

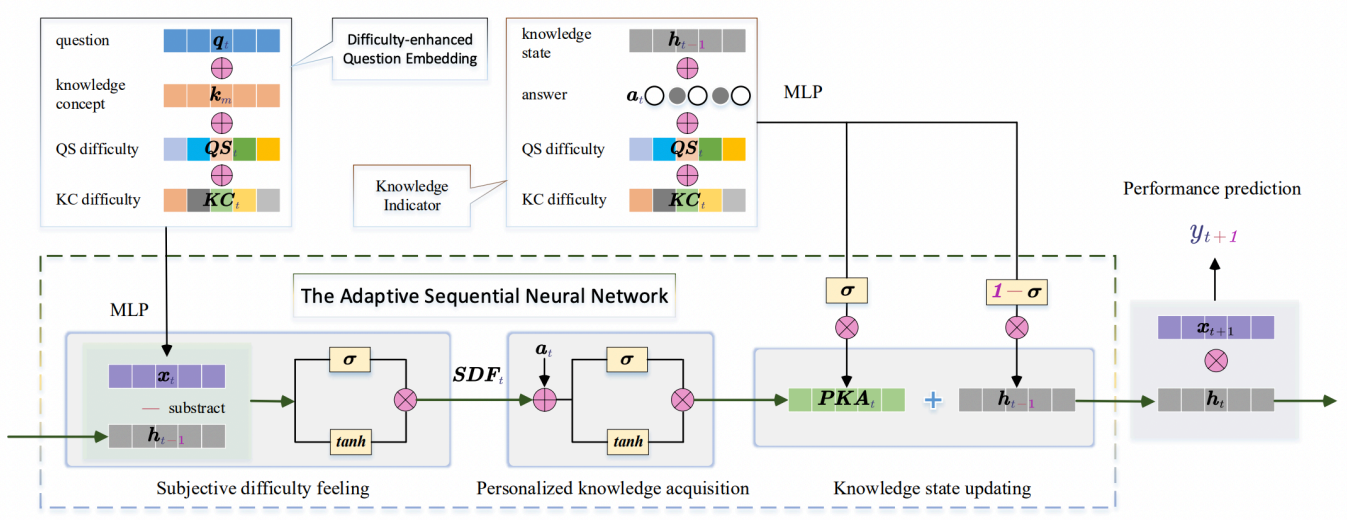


latest

Shen, Shuanghong, et al. “Learning process-consistent knowledge tracing.” Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021.

DIMKT

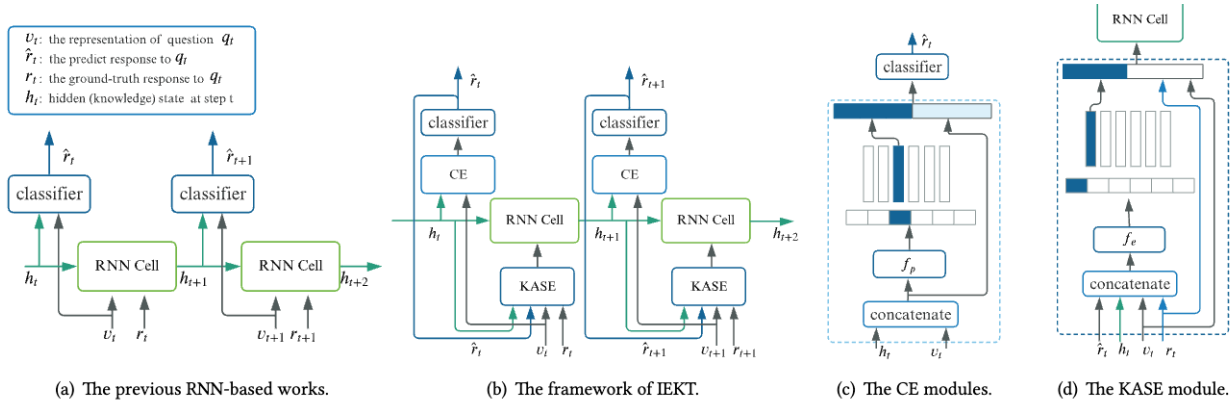
The DIfficulty Matching Knowledge Tracing (DIMKT) model explicitly incorporate the difficulty level into the question representation and establish the relation between students' knowledge state and the question difficulty level during the practice process.



Shen, Shuanghong, et al. "Assessing Student's Dynamic Knowledge State by Exploring the Question Difficulty Effect." Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022.

IEKT

Individual Estimation Knowledge Tracing (IEKT) estimates the students' cognition of the question before response prediction and assesses their knowledge acquisition sensitivity on the questions before updating the knowledge state.



(a) The previous RNN-based works. (b) The framework of IEKT. (c) The CE modules. (d) The KASE module.

Long, Ting, et al. "Tracing knowledge state with individual cognition and acquisition estimation." Proceedings of the 44th International ACM SIGIR Conference on Research and Information Retrieval. 2021.

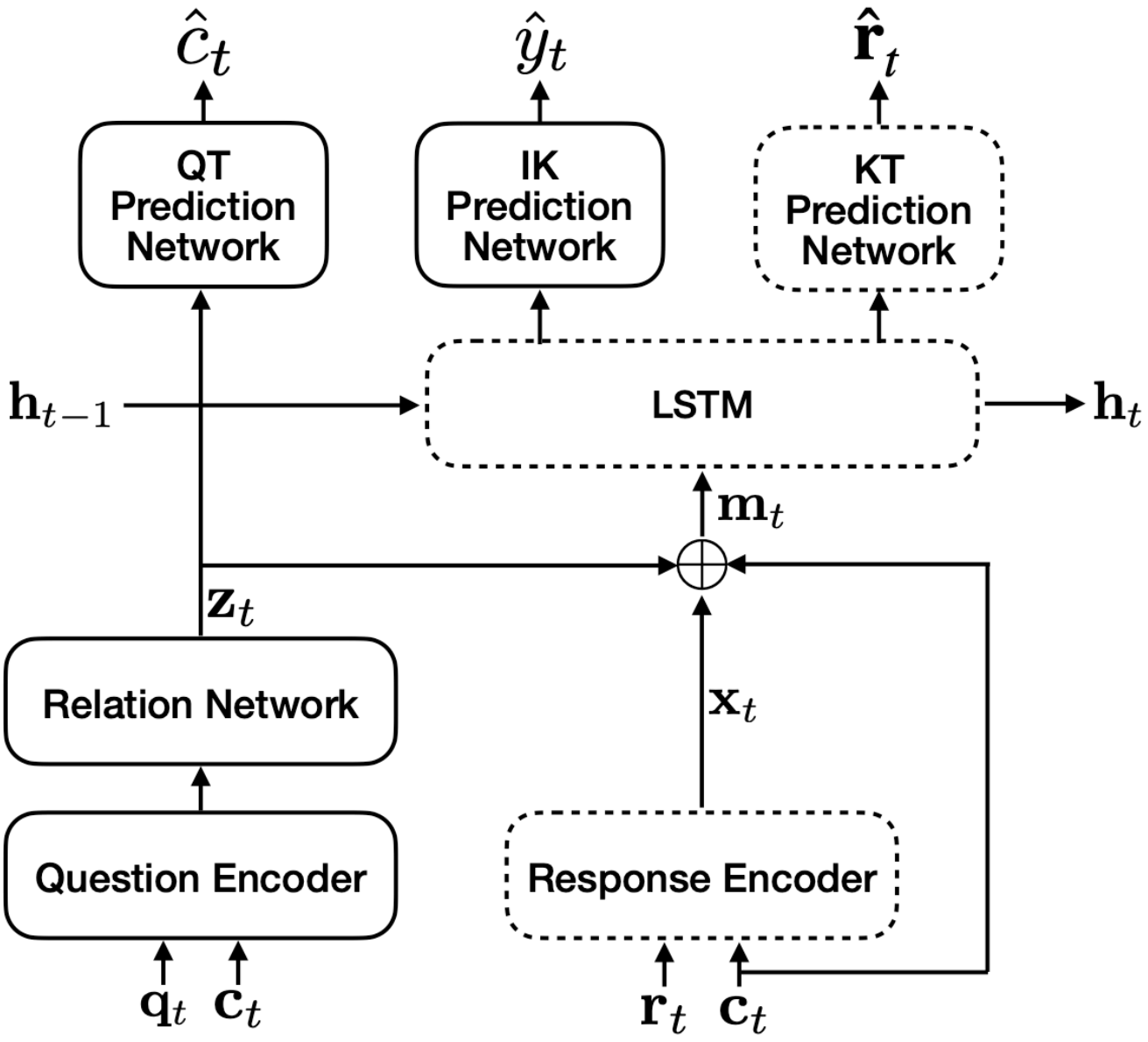
qDKT

qDKT(base) is a model same as DKT, but use the question ID as the input.

Sonkar, Shashank, et al. “qdkt: Question-centric deep knowledge tracing.” arXiv preprint arXiv:2005.12442 (2020).

AT-DKT

AT-DKT improve the prediction performance of the original deep knowledge tracing model with two auxiliary learning tasks including question tagging prediction task and individualized prior knowledge prediction task.



latest

Liu, Zitao, et al. “Enhancing deep knowledge tracing with auxiliary tasks.” Proceedings of the ACM Web Conference. 2023.

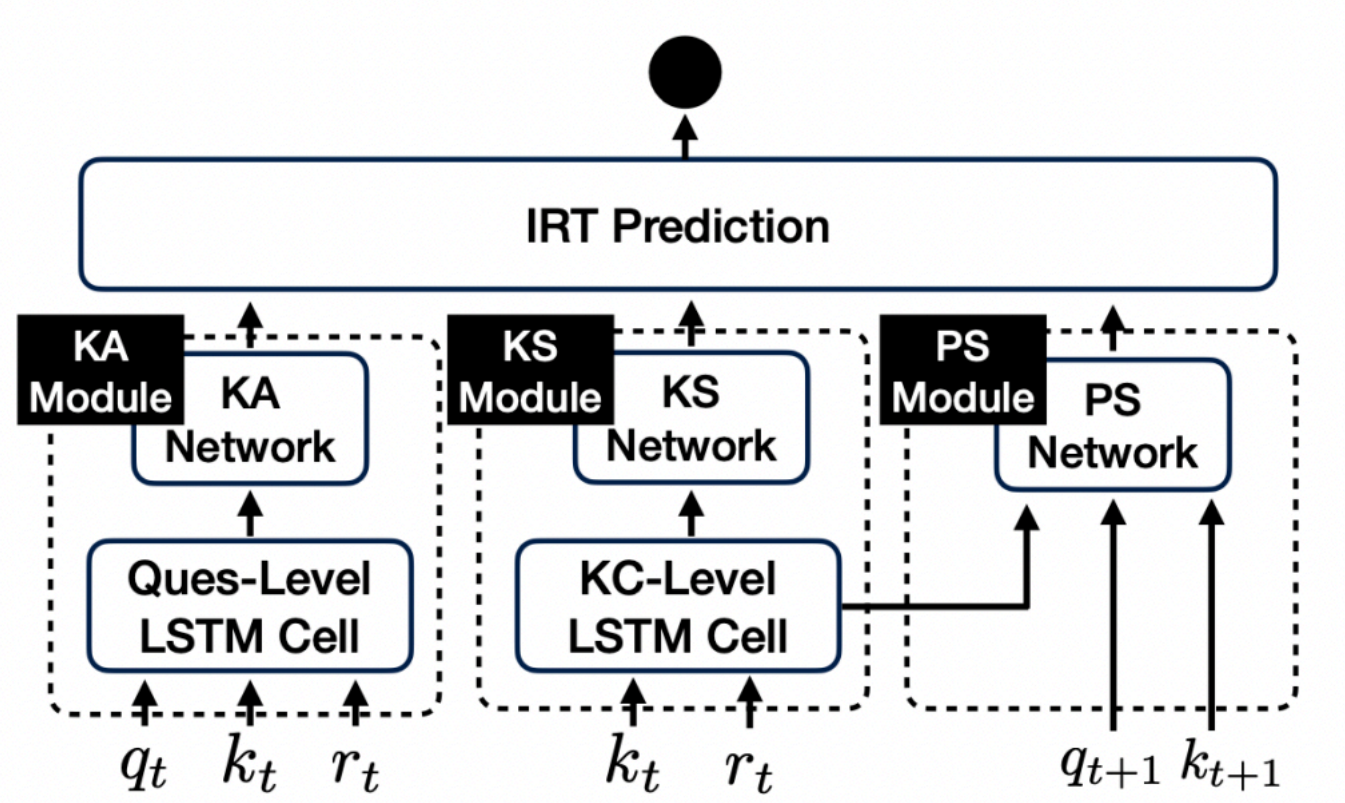
simpleKT

simpleKT is a strong but simple baseline method to deal with the KT task by modeling question-specific variations based on Rasch model and use the ordinary dot-product attention function to extract the time-aware information embedded in the student learning interactions.

Liu, Zitao, et al. “simpleKT: A Simple But Tough-to-Beat Baseline for Knowledge Tracing.” The Eleventh International Conference on Learning Representations. 2022.

QIKT

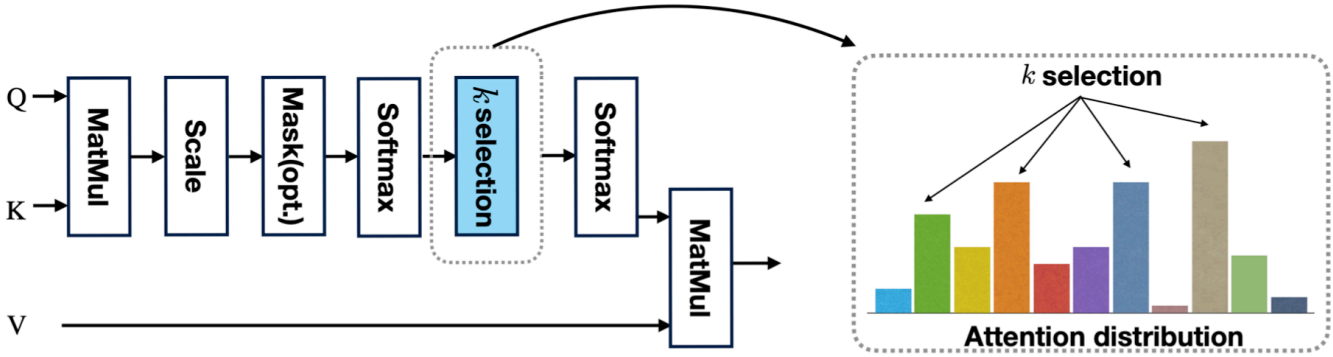
QIKT is a question-centric interpretable KT model that estimates students’ knowledge state variations at a fine-grained level with question-sensitive cognitive representations that are jointly learned from a question-centric knowledge acquisition module and a question-centric problem solving module.



Chen, Jiahao, et al. “Improving interpretability of deep sequential knowledge tracing models with question-centric cognitive representations.” The 37th AAAI Conference on Artificial Intelligence. 2023.

sparseKT-soft/topK

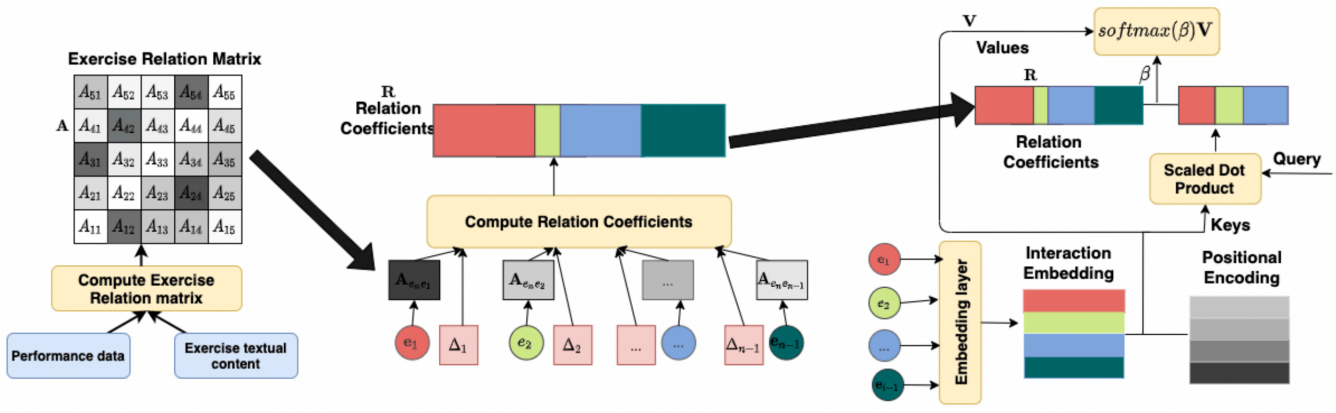
sparseKT incorporate a k-selection module to only pick items with the highest attention scores including two sparsification heuristics: (1) soft-thresholding sparse attention (sparseKT-soft) and (2) top-K sparse attention (sparseKT-topK).



Shuyan Huang, et al. "Towards Robust Knowledge Tracing Models via k-Sparse Attention." Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023.

RKT

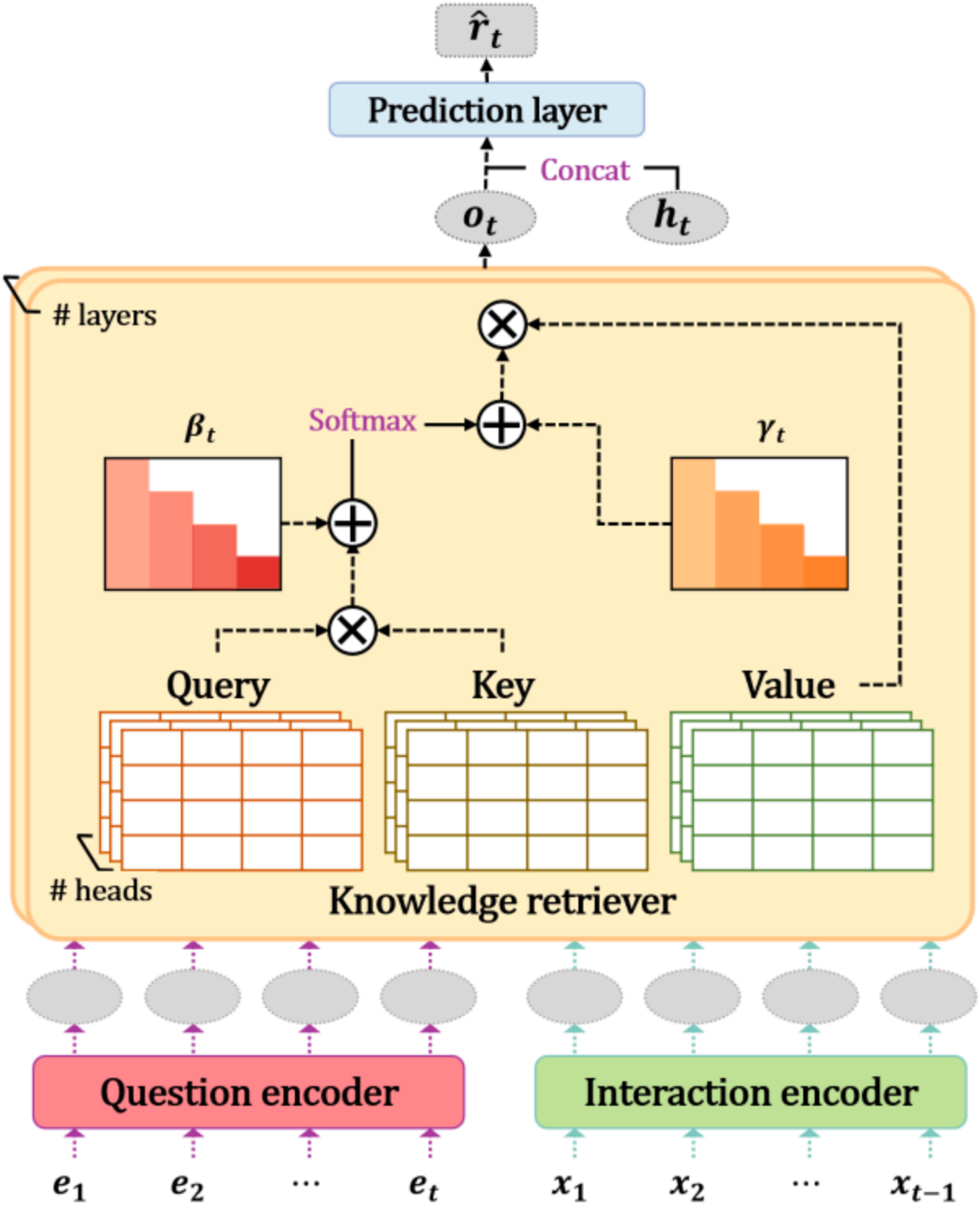
RKT contains a relation-aware self-attention layer that incorporates the contextual information including both the exercise relation information through their textual content as well as student performance data and the forget behavior information through modeling an exponentially decaying kernel function.



Pandey, Shalini, and Jaideep Srivastava. "RKT: relation-aware self-attention for knowledge tracing." Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.

FoLiBiKT

FoLiBi (Forgetting-aware Linear Bias) is a simple yet effective solution that introduces a linear bias term to explicitly model learners' forgetting behavior, compensating for the neglect of forgetting effects in existing attention-based Knowledge Tracing models。 We reproduced FoLiBi with AKT, namely FoLiBiKT.

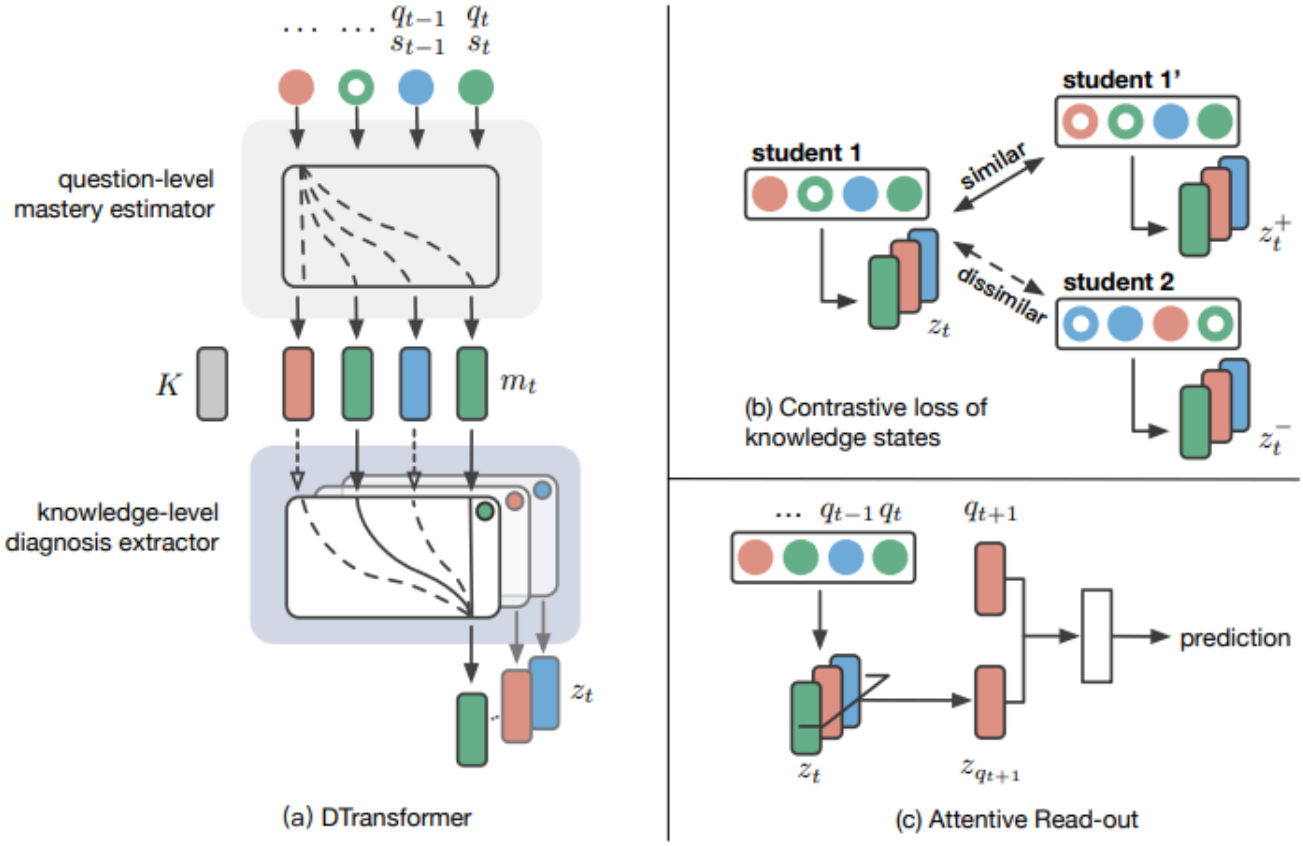


latest

Im, Yoonjin, et al. "Forgetting-aware Linear Bias for Attentive Knowledge Tracing." Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023.

Dtransformer

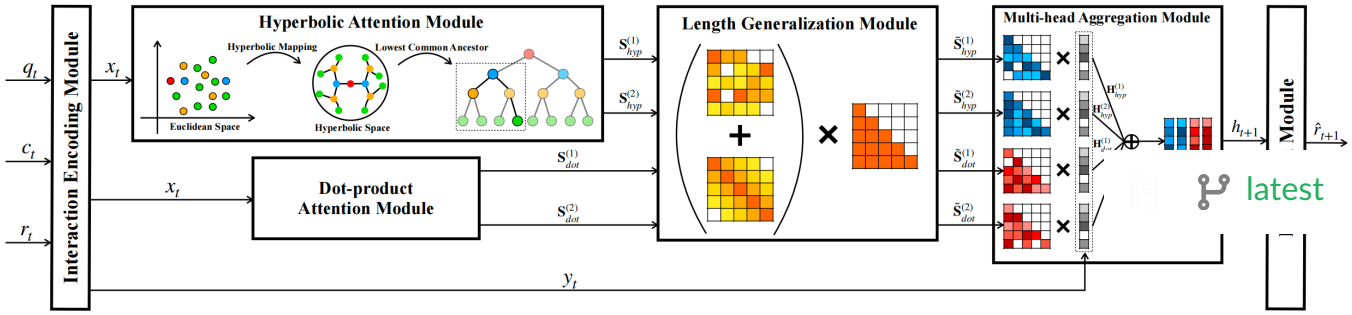
The Diagnostic Transformer (DTransformer) integrates question-level mastery with knowledge-level diagnosis using Temporal and Cumulative Attention (TCA) and multi-head attention for dynamic knowledge tracing. Moreover, a contrastive learning-based training algorithm enhances the stability of knowledge state diagnosis.



Yin, Yu, et al. “Tracing Knowledge Instead of Patterns: Stable Knowledge Tracing with Diagnostic Transformer.” Proceedings of the ACM Web Conference. 2023.

stableKT

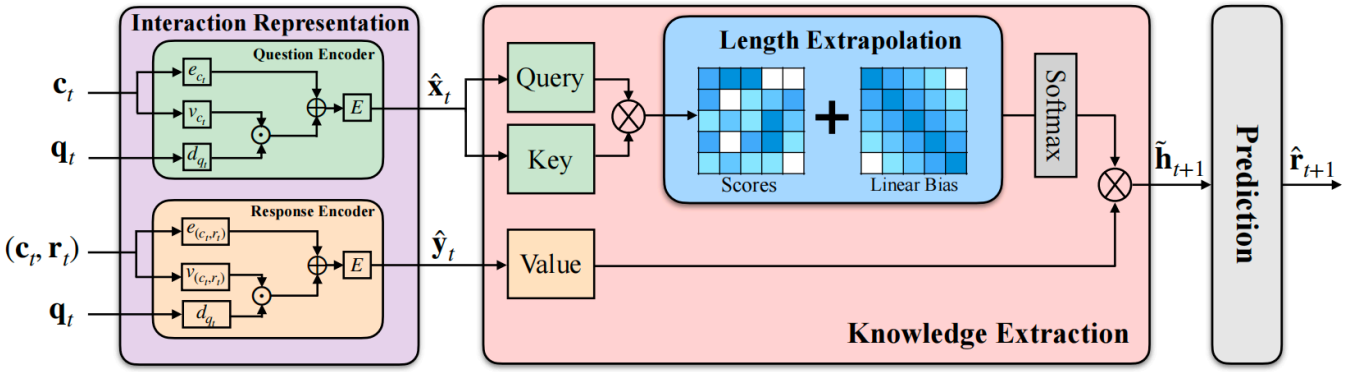
StableKT is able to learn from short sequences, maintain stable and consistent performance when generalizing on long sequences, and capture hierarchical relationships between questions and their associated KCs.



Li X, Bai Y, Guo T, et al. “Enhancing Length Generalization for Attention Based Knowledge Tracing Models with Linear Biases.” Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. 2024.

extraKT

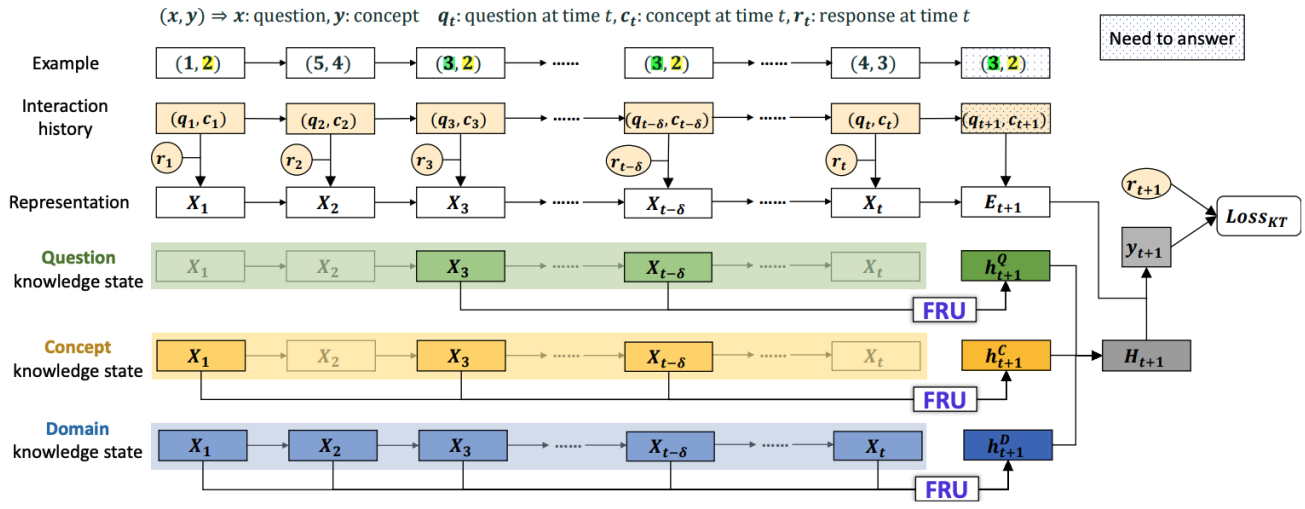
ExtraKT represents student interactions using a question encoder and a response encoder, and utilizes an efficient position embedding method to facilitate better extrapolation.



Li X, Bai Y, Guo T, et al. “Extending Context Window of Attention Based Knowledge Tracing Models via Length Extrapolation.” Proceedings of the 26th European Conference on Artificial Intelligence. 2024.

reKT

ReKT models student knowledge states from questions, concepts, and domains, leveraging a lightweight Forget-Response-Update (FRU) framework inspired by human cognition to achieve superior KT performance with minimal computational resources.

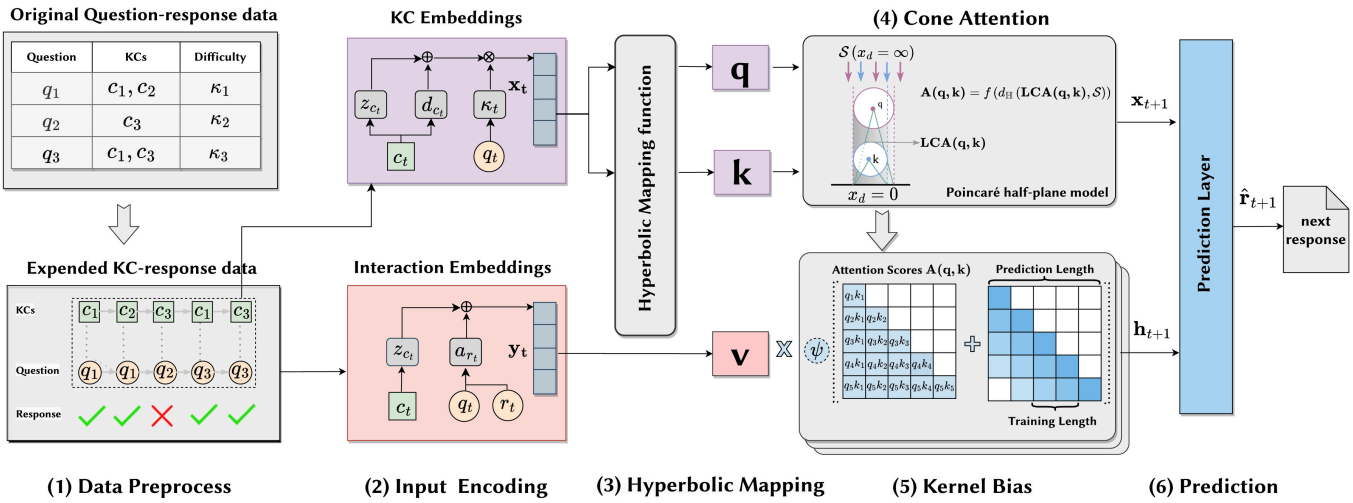


latest

Shen, Xiaoxuan, et al. “Revisiting Knowledge Tracing: A Simple and Powerful Model.” Proceedings of the 32nd ACM International Conference on Multimedia. 2024.

csKT

csKT is specifically designed to address the cold-start problem in knowledge tracing by introducing kernel bias and cone attention mechanisms to handle short student interaction sequences while maintaining prediction stability for longer sequences.



Bai, Youheng, et al. “csKT: Addressing cold-start problem in knowledge tracing via kernel bias and cone attention.” Expert Systems with Applications. 2025.

FlucKT

FlucKT enhances knowledge tracing by explicitly modeling both long-term cognitive trends and short-term cognitive fluctuations through a decomposition-based attention mechanism with causal convolution and a kernelized bias attention score penalty.

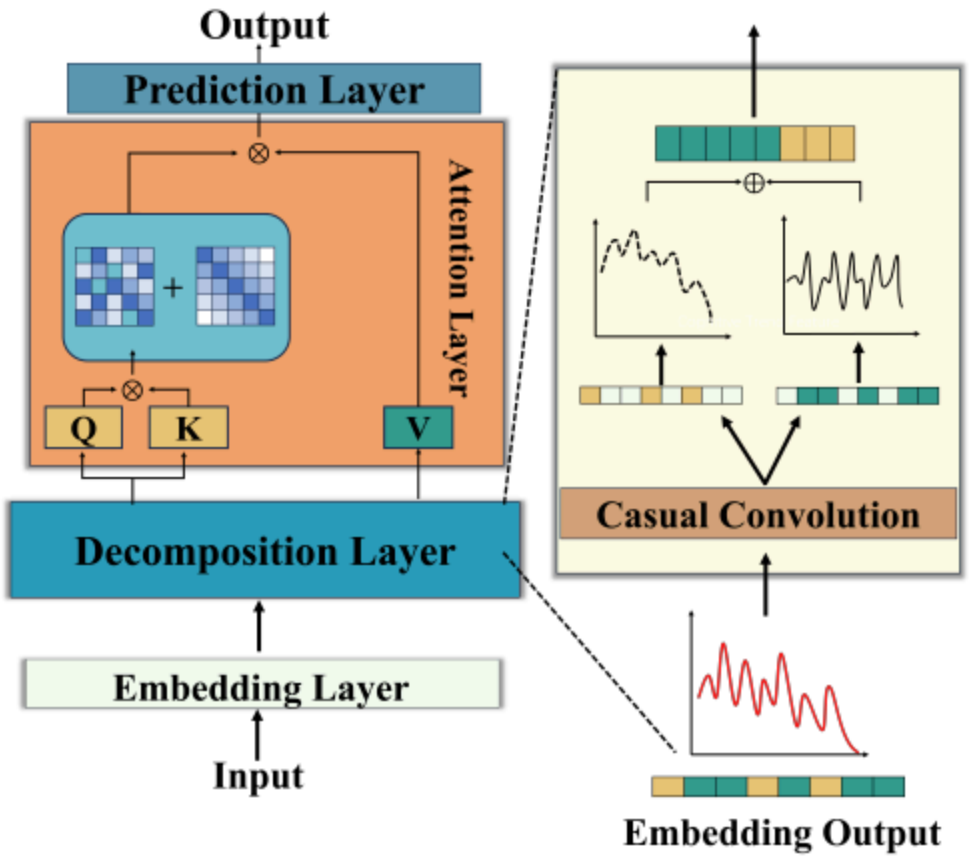
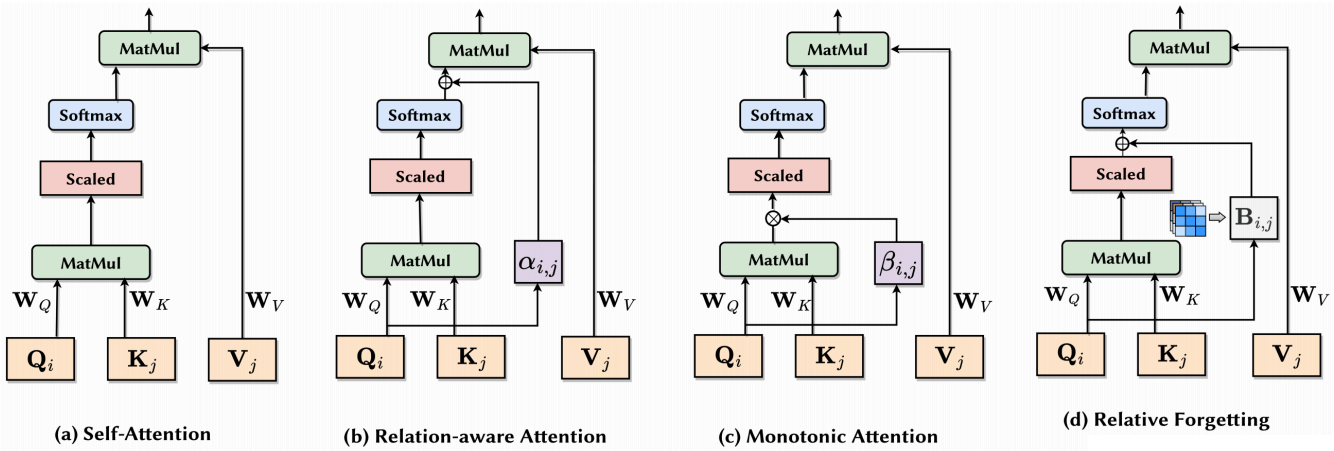


Figure 3: The overview of the proposed FlucKT framework.

Hou, Mingliang, et al. “Cognitive Fluctuations Enhanced Attention Network for Knowledge Tracing.” Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence. 2025.

lefoKT

LefoKT decouples forgetting patterns from problem relevance via relative forgetting attention to better model diverse forgetting behaviors in ever-growing interaction sequences.



latest

Bai, Youheng, et al. “Rethinking and Improving Student Learning and Forgetting Processes for Attention Based Knowledge Tracing Models.” Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence. 2025.

UKT

UKT represents student knowledge as probability distributions, tracks learning transitions with Wasserstein attention, and uses uncertainty-aware contrastive learning to improve prediction accuracy.

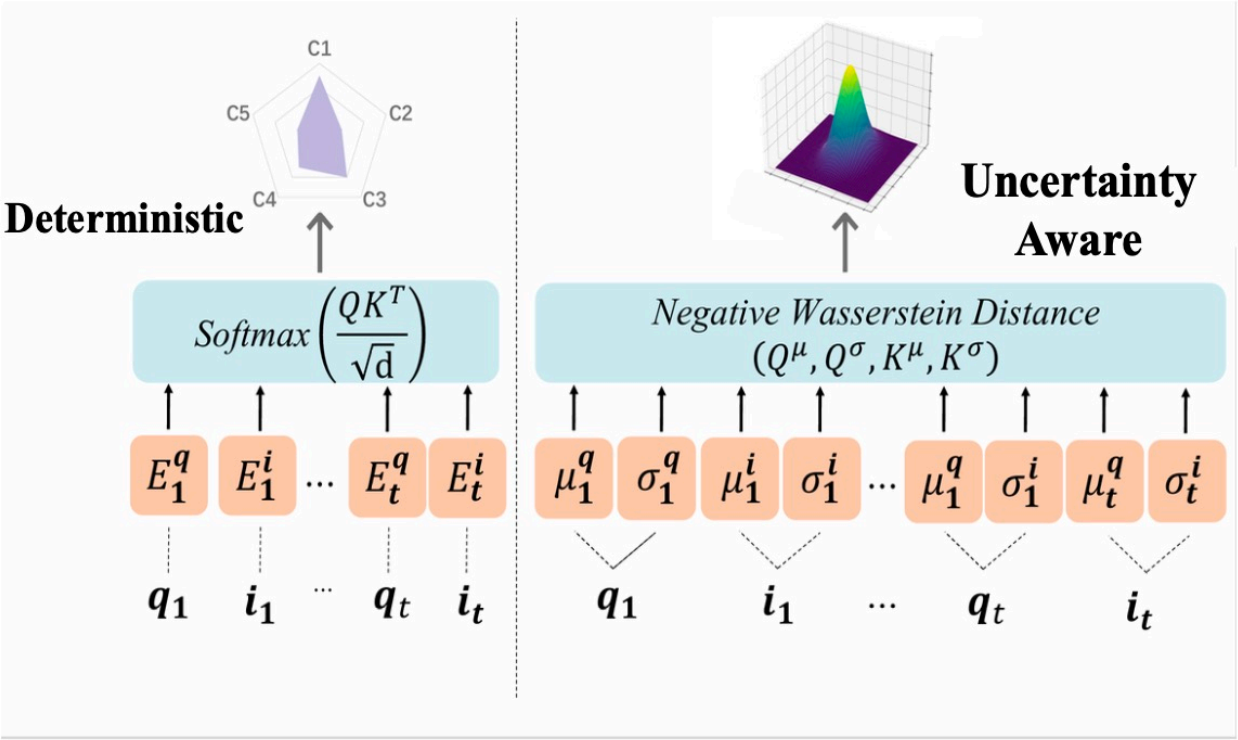
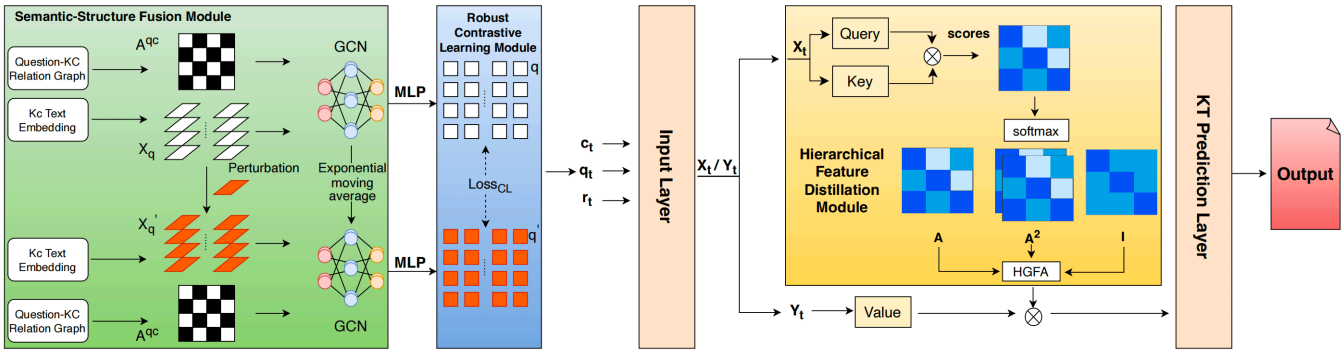


Figure 3: The differences between deterministic KT and UKT architectures.

Cheng, Weihua, et al. “Uncertainty-aware Knowledge Tracing.” Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence. 2025.

HCGKT

HCGKT integrates hierarchical graph filtering attention with contrastive learning and graph convolutions to model educational data relationships and predict student performance more accurately.



Huang, Z., Liu, Z. (2025, July). HCGKT: Hierarchical Contrastive Graph Knowledge Tracing with Multi-level Feature Learning. In International Conference on Artificial Intelligence in Education (pp. xxx). Cham: Springer Nature Switzerland.