MetaCD: A Meta Learning Framework for Cognitive Diagnosis based on Continual Learning

Jin Wu[1,2], Chanjin Zheng [1]

[1]Shanghai Institute of Artificial Intelligence for Education, East China Normal University

[2]School of Computer Science and Technology, East China Normal University

Email: 52275901018@stu.ecnu.edu.cn, chjzheng@dep.ecnu.edu.cn

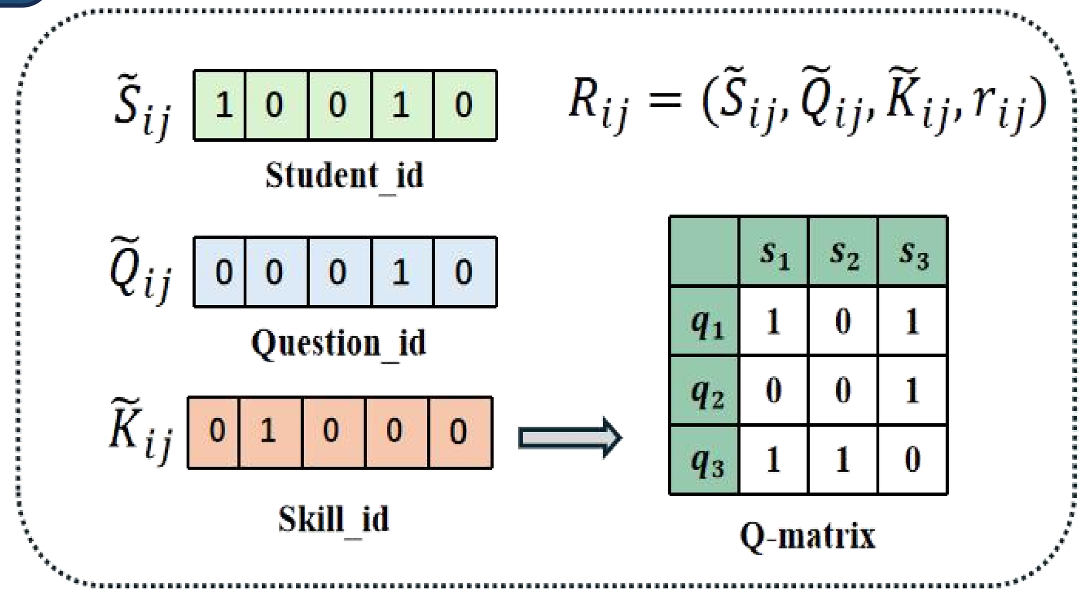
華東師範大學 EAST CHINA NORMAL UNIVERSITY

IAIE Shanghai 上海智能教育研究院 INSTITUTE of AI EDUCATION, SH

华东师范大学计算机科学与技术学院 School of Computer Science and Technology

NEURAL INFORMATION PROCESSING SYSTEMS

Background & Motivation



Definition: Cognitive Diagnosis models students' problem-solving processes to assess skill mastery and identify cognitive weaknesses.

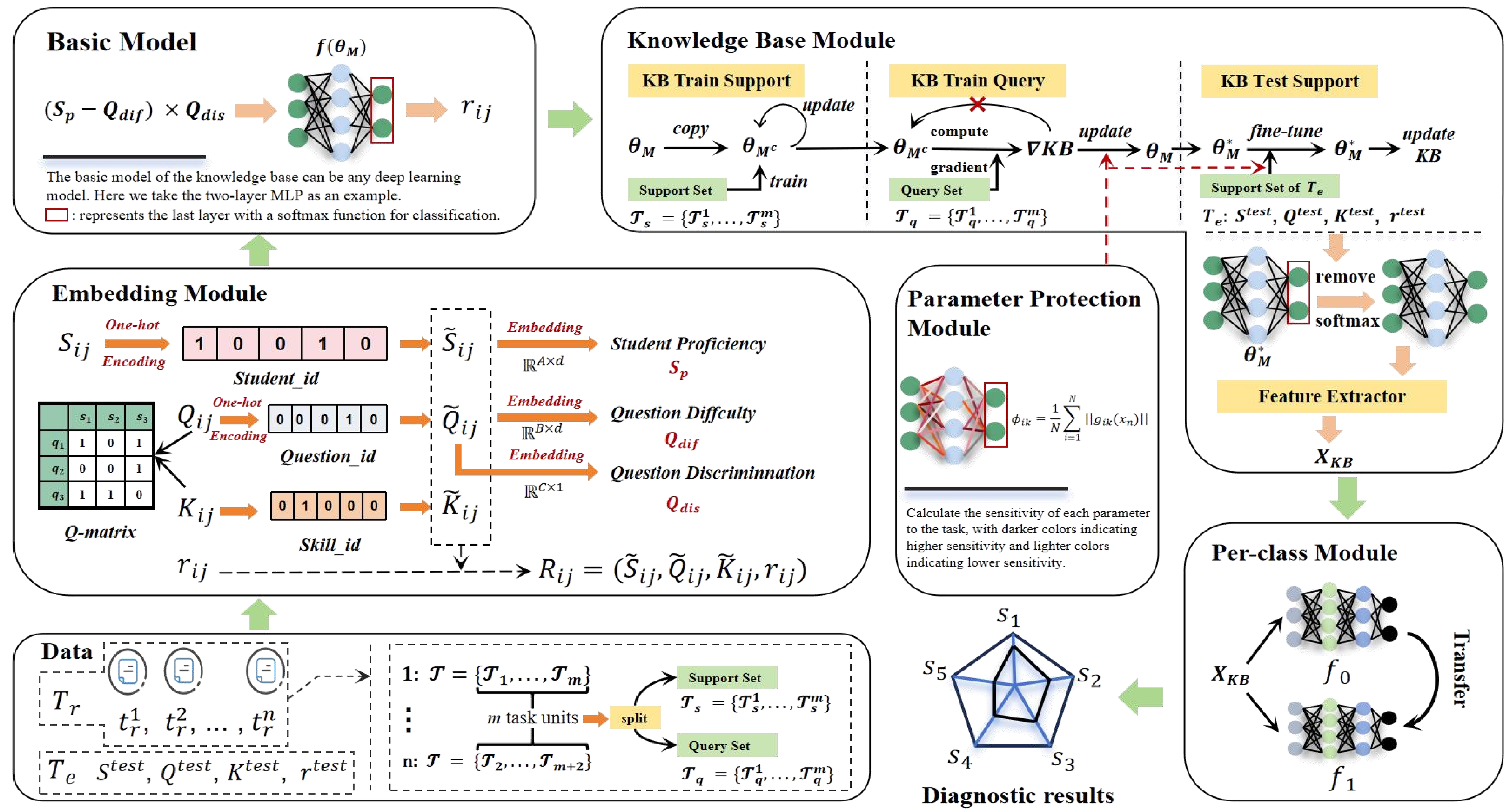
Core Assumptions

1. Static State: Cognitive status remains constant within a specific period.
2. State-Driven Behavior: Student actions are determined by latent cognitive states.

Challenges

- Long-Tail Problem: Data sparsity caused by uneven question response rates.
- System Dynamics:
  - ➢ Student Evolution: Proficiency changes as new skills are learned or old ones fade.
  - ➢ Task Updates: The continuous introduction of new learning tasks alters data patterns.

Methodology

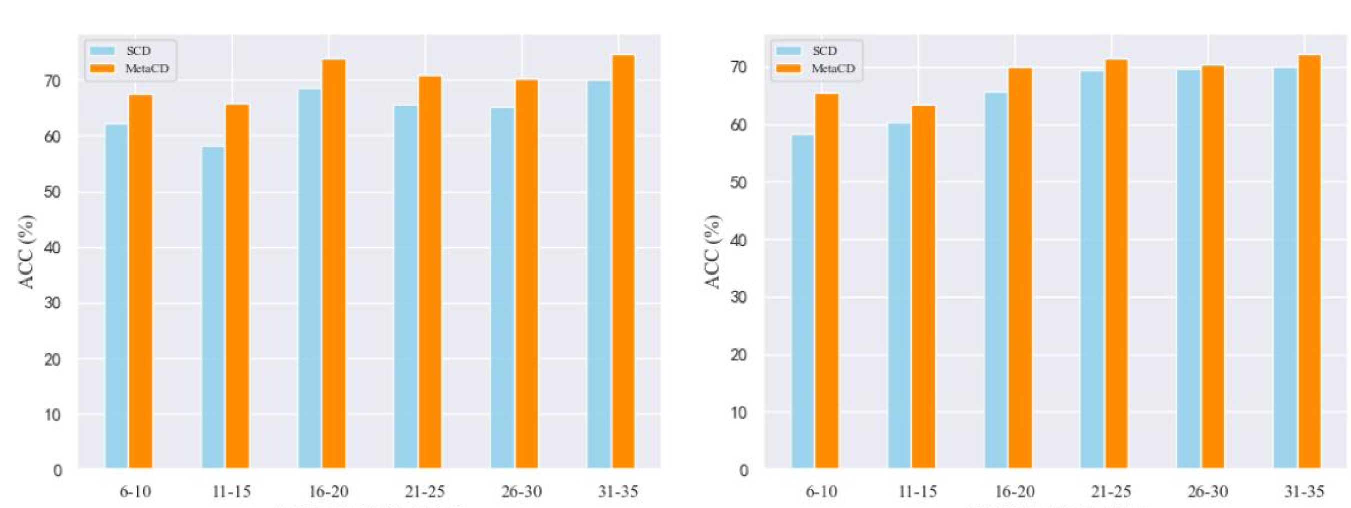


Experimental Results

Work	ASSIST2009_2010			ASSIST2017		
	ACC	RMSE	AUC	ACC	RMSE	AUC
MetaCD	**0.753**	0.425	**0.771**	**0.715**	**0.439**	**0.726**
IRT	0.654	0.472	0.681	0.658	0.464	0.668
MIRT	0.707	0.461	0.716	0.668	0.461	0.678
DINA	0.644	0.495	0.680	0.613	0.519	0.654
BCD	0.729	0.426	0.763	0.701	0.447	0.713
NCD	0.726	0.441	0.752	0.685	0.453	0.699
RCD	0.724	0.427	0.761	0.694	0.450	0.709
SCD	0.731	**0.423**	0.729	0.703	0.442	0.710

➢ **RQ1:** Can MetaCD achieve high accuracy in cognitive diagnosis compared to baseline?

MetaCD outperforms all baselines in both ACC and AUC on widely used datasets.



➢ **RQ2:** Can MetaCD maintain strong generalization under long-tailed data distributions?

Generalization on Long-tailed Data As shown in Figure above, **MetaCD** consistently outperforms SCD, demonstrating robust generalization even under sparse data distributions.

	$Task_1$	$Task_2$	$Task_3$	$Task_4$
$T_1^* \mid T_1$	0.771 / 0.771	/	/	/
$T_2^* \mid T_2$	0.719 / 0.598	0.703 / 0.721	/	/
$T_3^* \mid T_3$	0.706 / 0.533	0.689 / 0.557	0.700 / 0.715	/
$T_4^* \mid T_4$	0.693 / 0.506	0.675 / 0.531	0.686 / 0.519	0.697 / 0.701

➢ **RQ3:** Can MetaCD remain stable and adaptable in task-incremental learning?

We evaluate **MetaCD** in a task-incremental setting. Trained sequentially on four datasets after NIPS2020, MetaCD with parameter protection achieves a BWT of -0.04 vs. -0.217 without it, showing reduced forgetting and better stability.

Model	ASSIST2009_2010	ASSIST2012_2013	ASSIST2017	CDBD_a0910
MetaCD	0.753	0.725	0.715	0.712
w/o KB	0.719	0.685	0.687	0.668
w/o PPM	0.733	0.708	0.699	0.693
w/o Per-class	0.742	0.713	0.706	0.705

➢ **RQ4:** What is the impact of each module through ablation study?

Ablation results demonstrate that all components are essential. The KB module is pivotal, causing the largest accuracy drop (-3.65%) upon removal. Meanwhile, the PPM module (-1.8%) effectively filters redundant knowledge, and the Per-class module proves crucial for resolving fuzzy class boundaries.