Not All Options Are Created Equal: Textual Option Weighting for Token-Efficient LLM-Based Knowledge Tracing

Jongwoo Kim[1]*, Seongyeub Chu[1]*, Bryan Wong[1], Mun Yong Yi[1]†
[1] KAIST, Republic of Korea
*co-first author, † corresponding author

Introduction

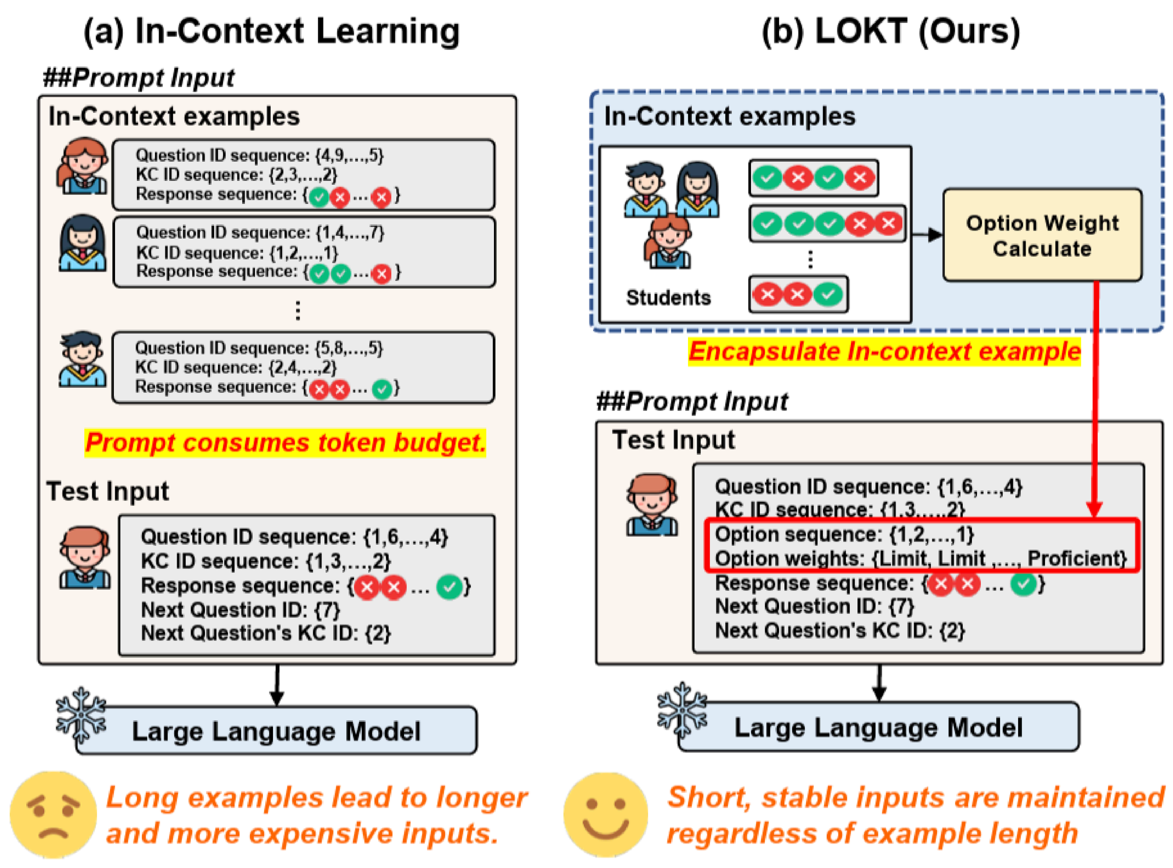
Knowledge Tracing (KT)

- Core method in learning analytics to model learners' knowledge state changes.
- **Challenge:** limited performance in cold-start settings.

LLM-based Knowledge Tracing

- Prior knowledge, Reasoning ability, Effectiveness in cold-start settings.
- In-context learning (ICL) provides flexibility and practicality without model parameter updating process.

Existing LLM-based approaches suffer from several challenges in terms of token usage.



Challenge 1. API cost

ICL incurs higher API cost as the number of few-shot examples increases, while improving knowledge tracing performance.

Challenge 2. Token usage limit

The token usage limit imposed by LLMs' attention span restricts the number of few-shot examples available for ICL.

Problem Definition

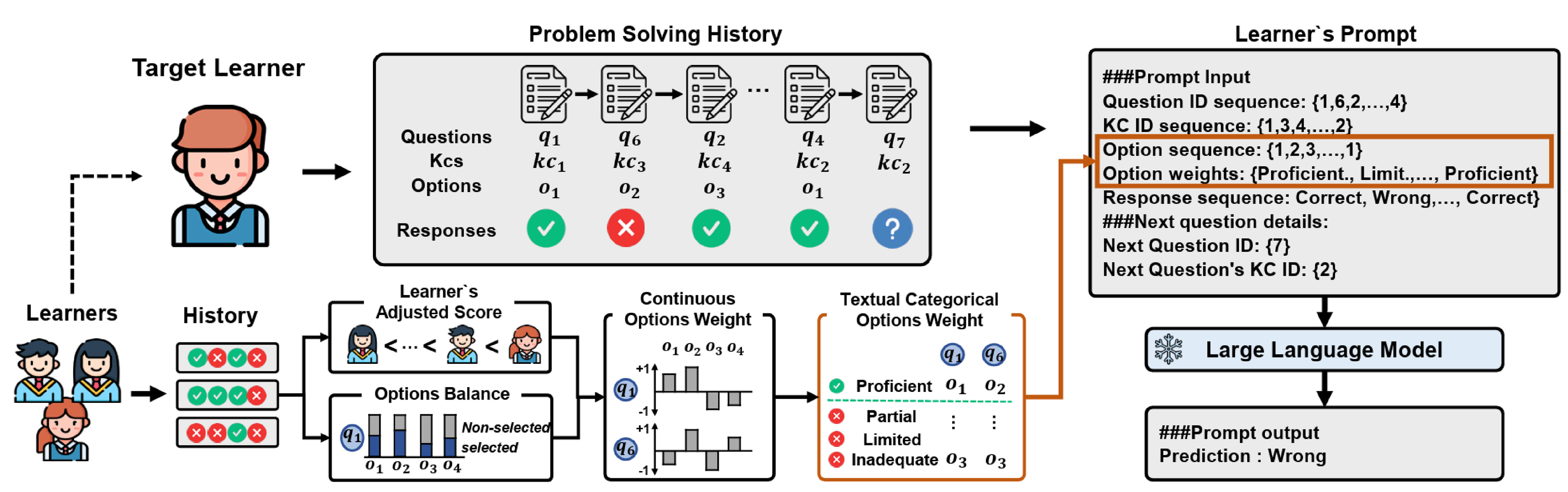
Knowledge Tracing

Models learners' evolving mastery of knowledge components by analyzing their past question-response interaction.

Few-shot Cold-start

Traces learners' knowledge states when only a small proportion of learners are available.

Methodology



Method 1. Learning from Option Information

Calculates option weights considering chosen/unchosen options and question difficulty.

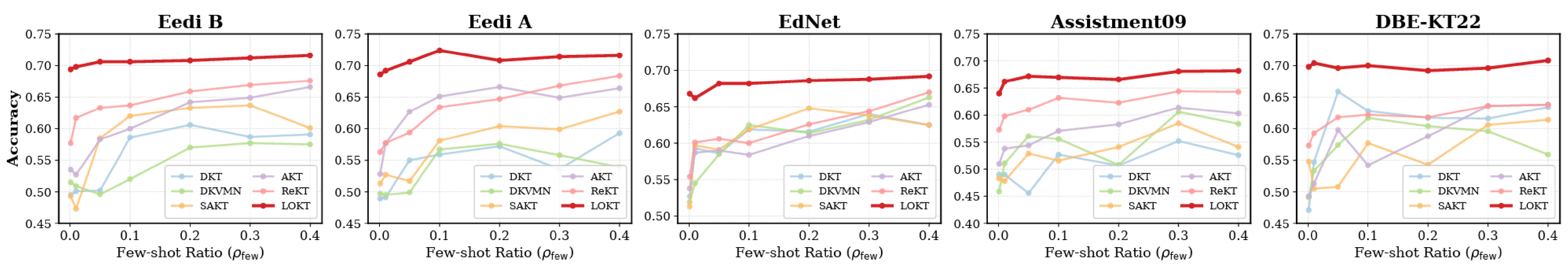
Method 2. Textual Categorical Option Weight

Converts the option weights into categorical text to enable LLMs to more effectively comprehend learners' proficiencies.

Results

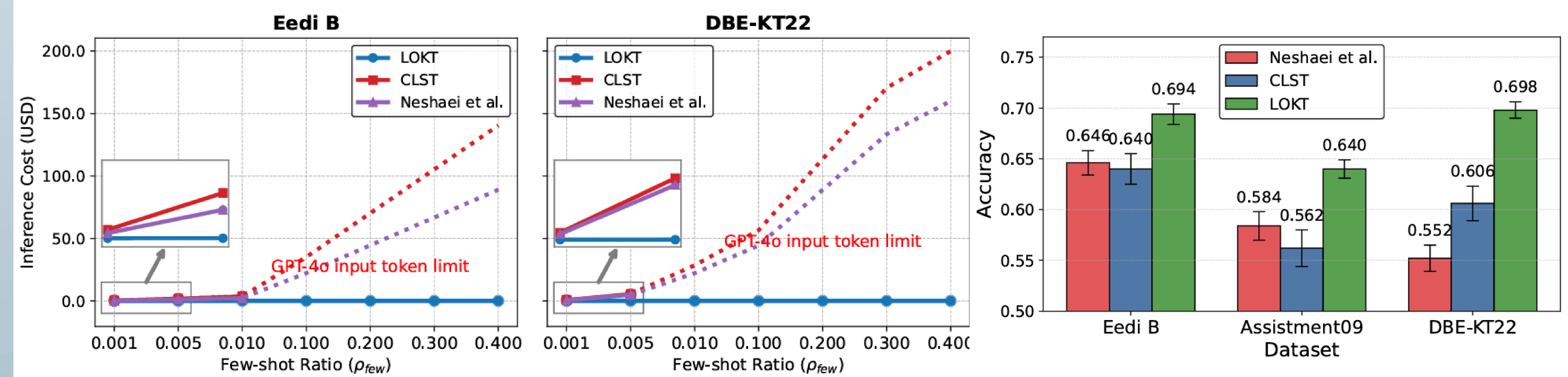
Performance in Cold-Start Settings

Model	Eedi B		Eedi A		Ednet		Assistment09		DBE-KT22	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
Non-LLM Based										
DKT	$0.495_{\pm0.044}$	$0.496_{\pm0.041}$	$0.489_{\pm0.061}$	$0.497_{\pm0.062}$	$0.527_{\pm0.050}$	$0.513_{\pm0.077}$	$0.490_{\pm0.067}$	$0.495_{\pm0.076}$	$0.471_{\pm0.048}$	$0.481_{\pm0.048}$
DKVMN	$0.515_{\pm0.048}$	$0.526_{\pm0.034}$	$0.503_{\pm0.023}$	$0.503_{\pm0.034}$	$0.539_{\pm0.021}$	$0.521_{\pm0.060}$	$0.459_{\pm0.054}$	$0.459_{\pm0.071}$	$0.490_{\pm0.034}$	$0.520_{\pm0.035}$
SAKT	$0.493_{\pm0.085}$	$0.494_{\pm0.092}$	$0.513_{\pm0.091}$	$0.497_{\pm0.093}$	$0.483_{\pm0.034}$	$0.540_{\pm0.092}$	$0.483_{\pm0.092}$	$0.540_{\pm0.099}$	$0.548_{\pm0.081}$	$0.623_{\pm0.083}$
AKT	$0.535_{\pm0.065}$	$0.476_{\pm0.024}$	$0.523_{\pm0.045}$	$0.538_{\pm0.048}$	$0.538_{\pm0.023}$	$0.645_{\pm0.048}$	$0.507_{\pm0.084}$	$0.574_{\pm0.071}$	$0.493_{\pm0.028}$	$0.659_{\pm0.059}$
ExtraKT	$0.573_{\pm0.018}$	$0.518_{\pm0.022}$	$0.561_{\pm0.018}$	$0.534_{\pm0.020}$	$0.552_{\pm0.016}$	$0.522_{\pm0.021}$	$0.568_{\pm0.018}$	$0.529_{\pm0.023}$	$0.571_{\pm0.022}$	$0.533_{\pm0.023}$
ReKT	$0.577_{\pm0.019}$	$0.522_{\pm0.023}$	$0.563_{\pm0.018}$	$0.538_{\pm0.021}$	$0.554_{\pm0.017}$	$0.527_{\pm0.019}$	$0.571_{\pm0.019}$	$0.532_{\pm0.024}$	$0.573_{\pm0.021}$	$0.537_{\pm0.024}$
LLM-Based (GPT-4o)										
(Neshaei et al., 2024)	$0.660_{\pm0.015}$	$0.679_{\pm0.014}$	$0.624_{\pm0.019}$	$0.615_{\pm0.018}$	$0.581_{\pm0.020}$	$\mathbf{0.656}_{\pm0.016}$	$0.632_{\pm0.015}$	$\mathbf{0.681}_{\pm0.009}$	$0.608_{\pm0.017}$	$0.688_{\pm0.014}$
CLST	$0.674_{\pm0.007}$	$0.678_{\pm0.011}$	$0.638_{\pm0.015}$	$0.660_{\pm0.016}$	$0.577_{\pm0.020}$	$0.616_{\pm0.018}$	$0.620_{\pm0.012}$	$0.651_{\pm0.017}$	$0.657_{\pm0.019}$	$0.693_{\pm0.018}$
LOKT	$\mathbf{0.694}_{\pm0.017}$	$\mathbf{0.683}_{\pm0.019}$	$\mathbf{0.686}_{\pm0.015}$	$\mathbf{0.698}_{\pm0.017}$	$\mathbf{0.668}_{\pm0.012}$	$0.646_{\pm0.011}$	$\mathbf{0.640}_{\pm0.016}$	$0.644_{\pm0.016}$	$\mathbf{0.698}_{\pm0.011}$	$\mathbf{0.737}_{\pm0.019}$



- Presents accuracy and F1 scores under an extreme cold-start setting(ρ_{few}=0.001) across five public datasets,.
- Demonstrates LOKT's effectiveness in knowledge tracing performance over other off-the-shelf traditional baselines and current SOTA LLM-based baselines.

Scalability and Efficiency



- LOKT compresses information through TCOW, maintaining a nearly constant prompt length in spite of increasing number of few-shot examples.
- LOKT consistently achieves the highest accuracy under a fixed token limit of 2000 tokens.

Effect of TCOW on KT Performance of LLMs

Method	Eedi B		Eedi A		EdNet		Assistment09		DBE-KT22	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
Continuous	0.656	0.625	0.672	0.694	0.622	0.505	0.620	0.614	0.658	0.707
Ordinal	0.668	0.617	0.676	0.696	0.630	0.525	0.630	0.627	0.670	0.714
TCOW	**0.694**	**0.683**	**0.686**	**0.698**	**0.668**	**0.640**	**0.640**	**0.644**	**0.698**	**0.737**

- TCOW consistently outperforms both continuous and ordinal representation of option weight, indicating that semantic structure is key to enabling LLMs' effective knowledge tracing.

Conclusion

- LOKT compresses learner interactions into textual categorical option weights within prompts, enabling large language models to efficiently understand learners' problem-solving histories.
- The effective performance of LOKT highlights that leveraging efficient compression techniques in in-context learning is key to achieving scalable and practical knowledge tracing with LLMs.

Paper Jongwoo Kim Seongyeub Chu