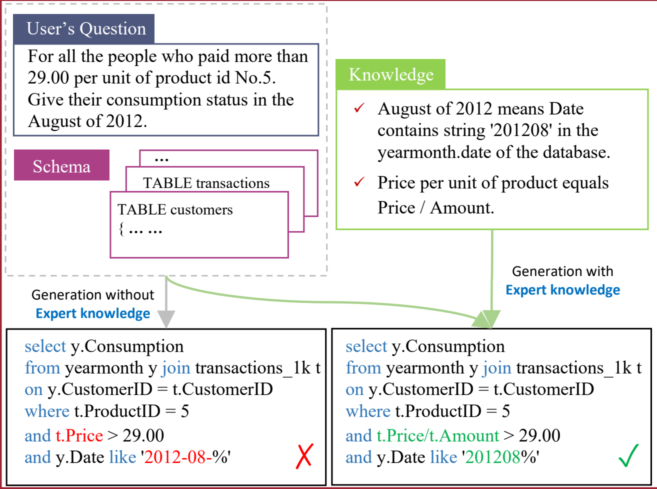
ACL 2024 Bangkok, Thailand

Knowledge-to-SQL: Enhancing SQL Generation with Data Expert LLM

Zijin Hong, Zheng Yuan, Hao Chen, Qinggang Zhang, Feiran Huang, Xiao Huang
Jinan University, The Hong Kong Polytechnic University

暨南大學 JINAN UNIVERSITY
THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Introduction



- The text-to-SQL task involves converting natural language question into executable SQL query.
- Given challenging questions and complex database structures, LLMs often make mistakes without sufficient specialized knowledge.
- Incorporating "Expert Knowledge" linking domain-specific content with the database and question can help.
- We propose a novel framework that can generate the required knowledge to assist in accurate SQL generation.

Training Details of Data Expert LLM (DELLM)

- **Supervised Fine-tuning of DELLM:** Firstly, the model is SFT based on the gold knowledge annotated by human experts, which can generate helpful knowledge preliminarily.
  - To enhance the model understanding, the input of SFT training incorporates the relevant table T for the user question Q in the corresponding database S.
  - Given human-annotated knowledge K^{gold} as the training label, the objective function of SFT is:

$$\mathcal{L}_{SFT} = -\log \Pr(K^{gold} \mid Q, S, T) = -\sum K^{gold} \log(K^{gen})$$

- **Preference Learning via Database Feedback:** Secondly, the model is further refined by PL based the feedback from the database's execution and the knowledge contribution to the ground-truth SQL.
  - The gold and generated knowledge are denoted by K^{gold} and K^{gen}, respectively. The execution result of the ground-truth SQL Y and the SQL generated by incorporating K^{gen} are denoted by V^{gold} and V^{gen}. We obtain the preference knowledge pairs K_w, K_l through indicator χ_{db} that evaluates the execution results and χ_{sql} that measures the contribution of the knowledge:

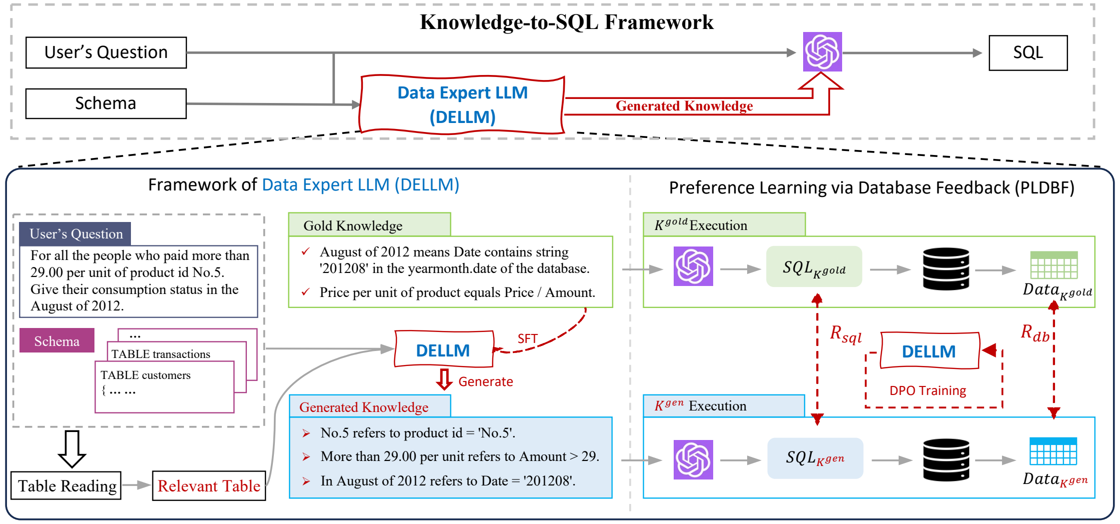
$$\mathcal{P}^{db}_{\{K_w,K_l\}} = \{K^{gold}, K^{gen} \mid \chi_{db}(V^{gold}, V^{gen}) = 0\}$$

$$\mathcal{P}^{sql}_{\{K_w,K_l\}} = \{K^{gold}, K^{gen} \mid \chi_{sql}(K^{gold}, Y) = 1, \chi_{sql}(K^{gen}, Y) = 0\}$$

  - Then, the DPO training is conducted as further preference learning refinement:

$$\mathcal{L}_{\mathrm{PL}}(\pi^{DPO}; \pi^{SFT}) = -\mathbb{E}_{\pi}[\log \sigma(\beta R(K_w) - \beta R(K_l))]$$

Proposed Method



- **SFT**: Supervised Fine-tunes the DELLM based on enhanced input and human-annotated knowledge.
- **PLDBF**: Optimizes knowledge generation through preference learning with database feedback.
- **Knowledge-to-SQL**: Integrates generated knowledge to assist an off-the-shelf text-to-SQL model, significantly improving the accuracy.

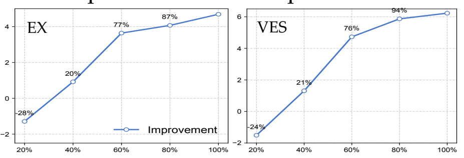
Experiments

	Models	EX		VES	
		w/o knowledge	w/ DELLM	w/o knowledge	w/ DELLM
BIRD	T5-3B	10.37	16.68 (+6.31)	13.62	20.84 (+7.22)
BIRD	GPT-3.5-Turbo	27.64	33.31 (+5.67)	28.64	36.12 (+7.48)
BIRD	GPT-4	33.25	37.94 (+4.69)	35.92	42.15 (+6.23)
BIRD	Claude-2	30.05	35.53 (+5.48)	32.97	39.71 (+6.74)
BIRD	GPT-3.5-Turbo + CoT	27.25	32.79 (+5.54)	29.16	35.51 (+6.35)
BIRD	DAIL-SQL + GPT-4	40.89	45.81 (+4.92)	45.13	51.59 (+6.46)
BIRD	MAC-SQL + GPT-4	43.65	48.92 (+5.27)	48.07	54.78 (+6.71)
Spider	GPT-3.5-Turbo	67.89	69.60 (+1.71)	68.33	70.16 (+1.83)
Spider	GPT-4	70.02	71.68 (+1.66)	71.03	72.82 (+1.79)

- The improvement brought by DELLM to different methods on Spider and BIRD

Model	Simp.	Mod.	Chall.	All
GPT-3.5-Turbo	35.58	14.60	17.61	27.64
GPT-3.5-Turbo + D	43.09	18.30	17.61	33.31
GPT-3.5-Turbo + E	50.27	31.81	20.42	41.98
GPT-4	41.05	21.13	21.13	33.25
GPT-4 + D	47.16	24.18	21.83	37.94
GPT-4 + E	54.01	36.38	31.69	46.67

- Compared to human expert annotations



- Performance on partial training data

- Contact: hongzijin@stu2020.jnu.edu.cn, xiaohuang@comp.polyu.edu.hk, huangfr@jnu.edu.cn (corresponding author)
- Code: https://github.com/Rcrossmeister/Knowledge-to-SQL Paper: https://arxiv.org/pdf/2402.11517