Continual Text-to-SQL

Yongrui Chen

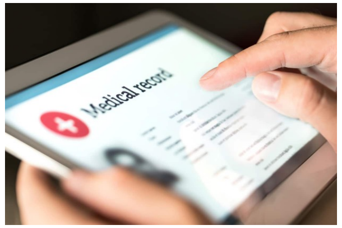
School of Computer Science and Engineering, Southeast University

Background

- Relational databases store a vast amount of today's information and provide the foundation of applications.

customer relations management

financial markets



medical records

- Structured Query Language (SQL)
  - Accessing relational databases
  - Machine understandable, Quick and efficient
  - Not user-friendly, requires deep understanding of the database and SQL syntax

Background

- Text-to-SQL
  - (Semantic Parsing) Transform Natural Language Question (NLQ) to SQL
  - Input: NLQ + database schema
  - Output: an SQL query

Database Schema

Cars_data

id	mpg	cylinders	edispl	horsepower	weight	accelerate	year

Car_names

make_id	model	make

Model_list

model_id	maker	model

Car_makers

id	maker	full_name	country

NLQ

For the cars with 4 cylinders, which model has the largest horsepower?

Desired SQL

```
SELECT T1.model
FROM car_names AS T1 JOIN cars_data AS T2
ON T1.make_id = T2.id
WHERE T2.cylinders = 4
ORDER BY T2.horsepower DESC LIMIT 1
```

Background

Different scenarios

- Single-Table Text-to-SQL (Zhong et al., 2017)
  - Only one table
  - No complex SQL syntax
  - **Solution:** Encoder + multi-task
- Multi-Table Text-to-SQL (Yu et al., 2018)
  - Multiple tables joined by foreign keys
  - Complex SQL syntax, e.g., GROUP BY, nested query, …
  - **Solution:** Encoder-Decoder
- Conversational Text-to-SQL (Yu et al., 2019)
  - Multiple tables joined by foreign keys
  - Complex SQL syntax, e.g., GROUP BY, nested query, …
  - Multiple NLQs
  - In context
  - **Solution:** Encoder-Decoder + Context information

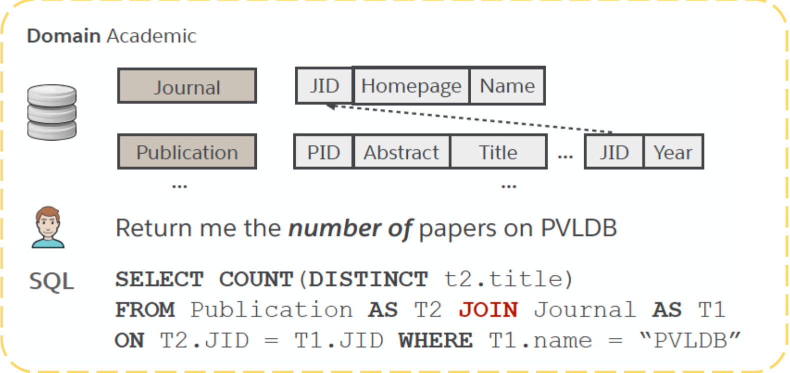
Question:

How many CFL teams are from York College?

SQL:

```
SELECT COUNT CFL Team FROM
CFLDraft WHERE College = "York"
```

WikiSQL



Spider

D_2 : Database about shipping company containing 13 tables

C_2 : **Find the names of the first 5 customers.**

Q_1 : What is the customer id of the most recent customer?

S_1 : `SELECT customer_id FROM customers ORDER BY date_became_customer DESC LIMIT 1`

Q_2 : What is their name?

S_2 : `SELECT customer_name FROM customers ORDER BY date_became_customer DESC LIMIT 1`

Q_3 : How about for the first 5 customers?

S_3 : `SELECT customer_name FROM customers ORDER BY date_became_customer LIMIT 5`

SParC

Background

Different scenarios

- Fixed data set (WikiSQL, Spider, Sparc,…)
    - Unchanging data distribution
    - Inability to adapt or expand the behavior over time
    - Uncommon in real-world scenarios

- Data stream (What we care about)
    - Changing data distribution
    - Require adapting or expanding the behavior over time
    - Common in real-world scenarios
    - Ideal Intelligence

In-domain

customer relations management

- New product categories
- Target customer change

⋮

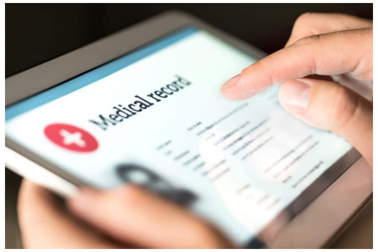
In-domain

financial markets

- Business expansion
- New Stock Issuance
- Policy Adjustment

⋮

In-domain



medical records

- New Diseases (Covid 19)
- Section upgrade

⋮

Cross-domain

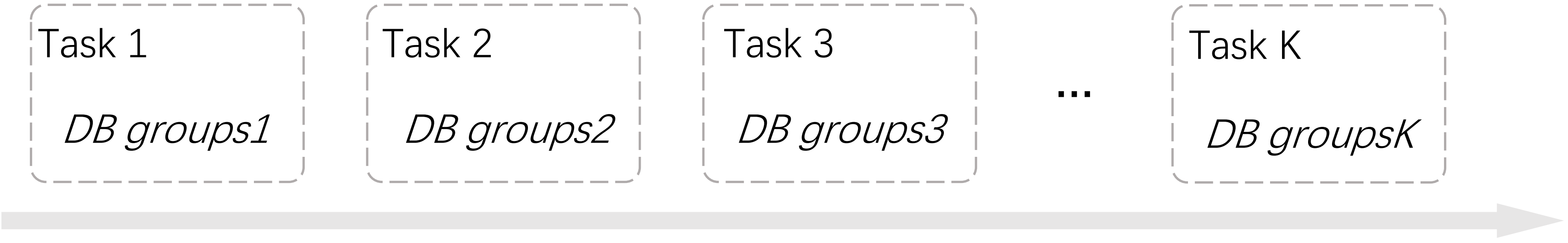
smart speakers

- New actions, skills, …
- Import knowledge

⋮

Challenge

A task stream is regarded as a sequence of *tasks*, each of which is a text-to-SQL task, but for a database not previously seen.



- Limited supervised data.
  - High cost of SQL annotation → A small amount of supervised data
- Costly full volume retraining.
  - Training models from scratch on all seen tasks → Too costly
- Catastrophic forgetting.
  - Continually fine-tune models for each task → Performance drop on the previous task

Our Work 1

Learn from Yesterday: A Semi-Supervised Continual Learning Method for Supervision-Limited Text-to-SQL Task Streams

Yongrui Chen[1], Xinnan Guo[1], Tongtong Wu[1,2], Guilin Qi[1], Yang Li[3], Yang Dong[4]

1 School of Computer Science and Engineering, Southeast University
2 Department of Data Science & AI, Monash University
3 Alibaba Group
4 Ant Group

AAAI 2023

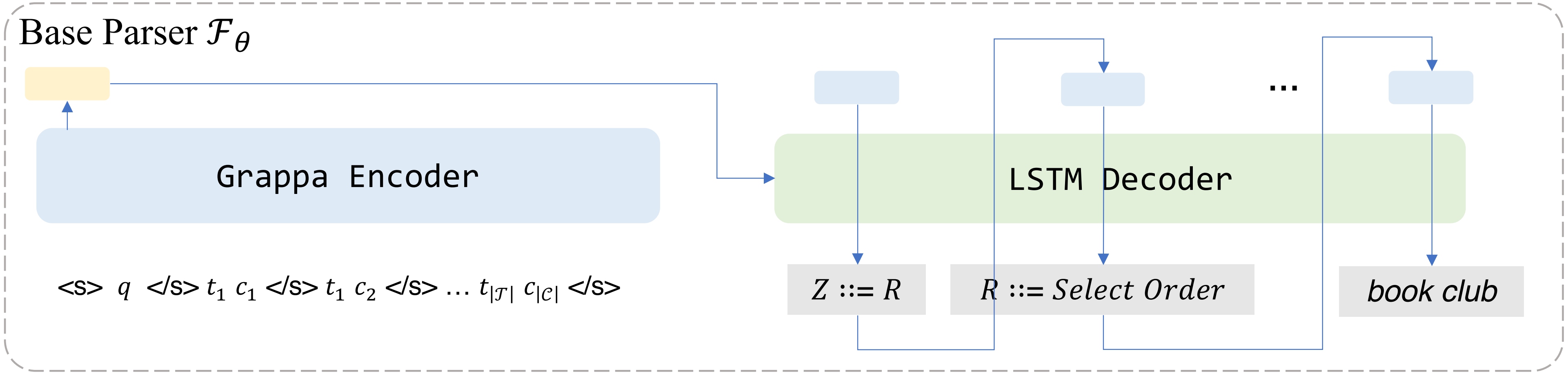
A base text-to-SQL parser

Following SOTA (Wang et al., 2019), (Lin et al., 2020), (Cao et al., 2021), we construct a strong encoder-decoder model as the base parser $\mathcal{F}_\theta$.

- A strong table semantic parsing PLM, Grappa (Yu et al., 2021), as the encoder;
- An LSTM as the decoder;
- Grammar-based decoding using SemQL (Guo et al., 2019);

SemQL Grammar

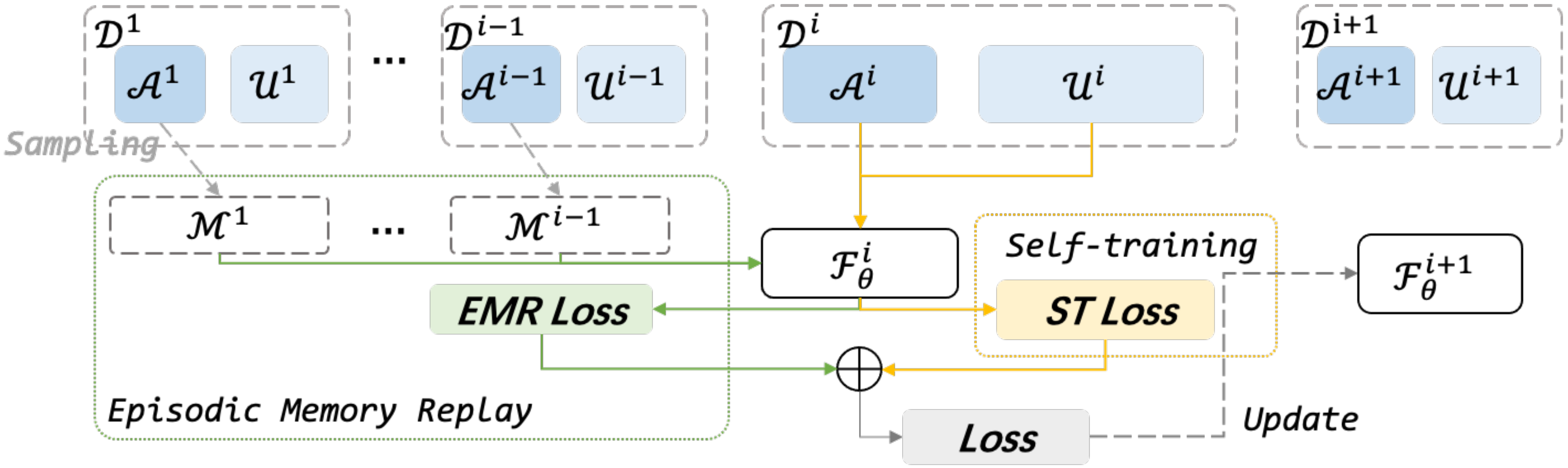
$$
\begin{aligned}
Z &::= intersect\ R\ R \mid union\ R\ R \mid except\ R\ R \mid R \\
R &::= Select \mid Select\ Filter \mid Select\ Order \\
&\quad \mid Select\ Superlative \mid Select\ Order\ Filter \\
&\quad \mid Select\ Superlative\ Filter \\
Select &::= A \mid A\ A \mid A\ A\ A \mid A\ A\ A\ A \mid A\ A \cdots A \\
Order &::= asc\ A \mid desc\ A \\
Suerlative &::= most\ A \mid least\ A \\
Filter &::= and\ Filter\ Filter \mid or\ Filter\ Filter \\
&\quad \mid > A \mid > A\ R \mid < A \mid < A\ R \\
&\quad \mid \geq A \mid \geq A\ R \mid = A \mid = A\ R \\
&\quad \mid \neq A \mid \neq A\ R \mid between\ A \\
&\quad \mid like\ A \mid not\ like\ A \mid in\ A\ R \mid not\ in\ A\ R \\
A &::= \max C\ T \mid \min C\ T \mid count\ C\ T \\
&\quad \mid sum\ C\ T \mid avg\ C\ T \mid none\ C\ T \\
C &::= column \\
T &::= table
\end{aligned}
$$



A Vanilla Solution

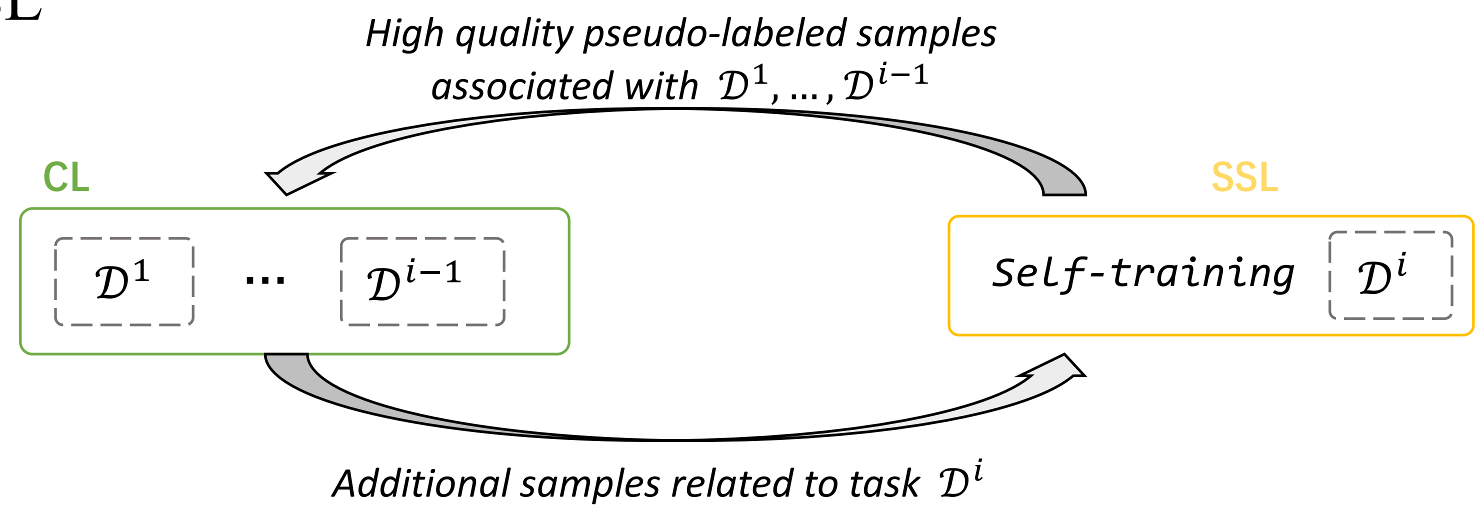
- Semi-Supervised Learning (SSL) eases over-fitting and improves model generalizability.
- Continual Learning (CL) provides an alternative cost-effective training paradigm.

- Rigidly combines CL and SSL
- SSL method: *Self-training (ST)*
  - Most used in semantic parsing
  - Good performance on few-shot text-to-SQL, (Guo et al., 2022)
- CL method: *Episodic Memory Replay (EMR)*
  - Simple process, suitable for complex structured text-to-SQL models
  - Best performance in CL methods for semantic parsing, (Li et al., 2021)



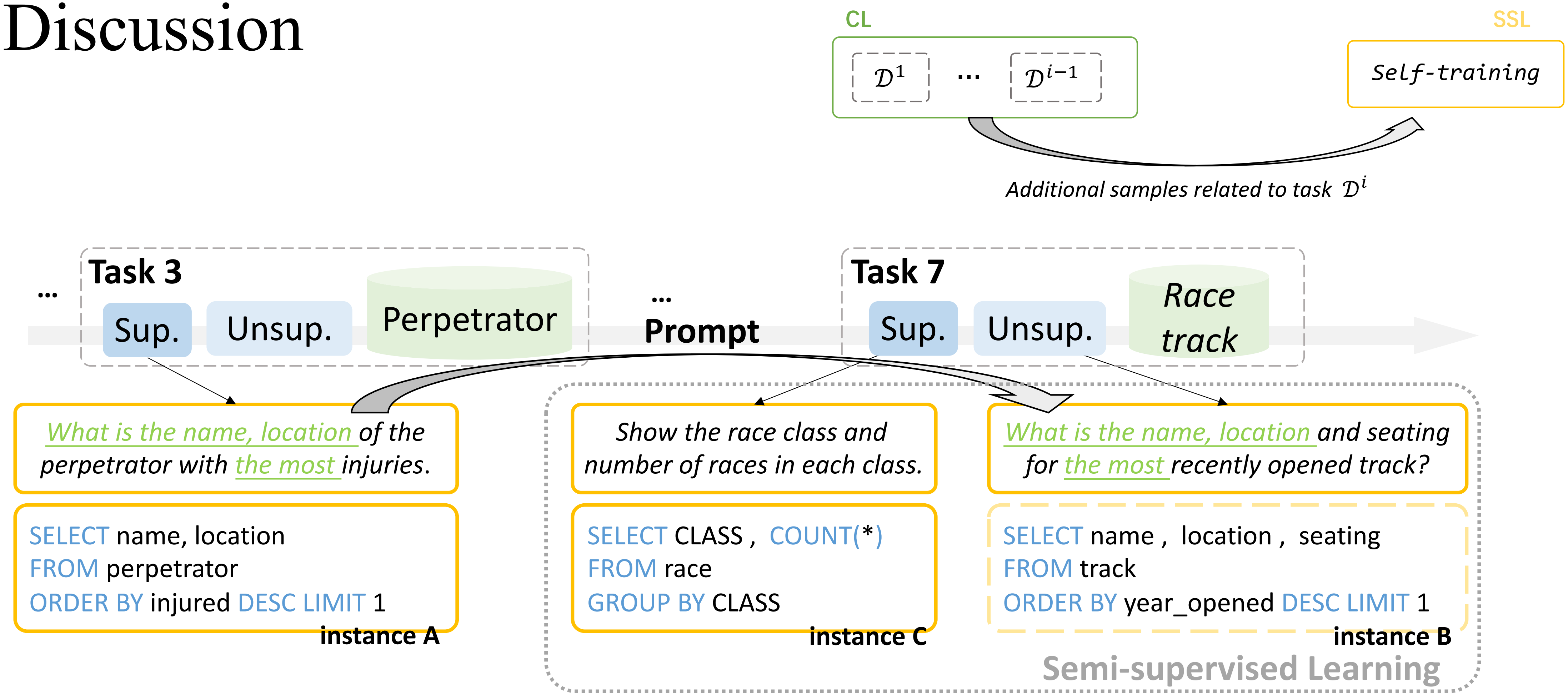
Discussion

- Differences in learning goals
  - SSL focuses on the optimal solution on a single task (current task).
  - CL is more concerned with maintaining performance on previous tasks.
- Promotion of CL and SSL



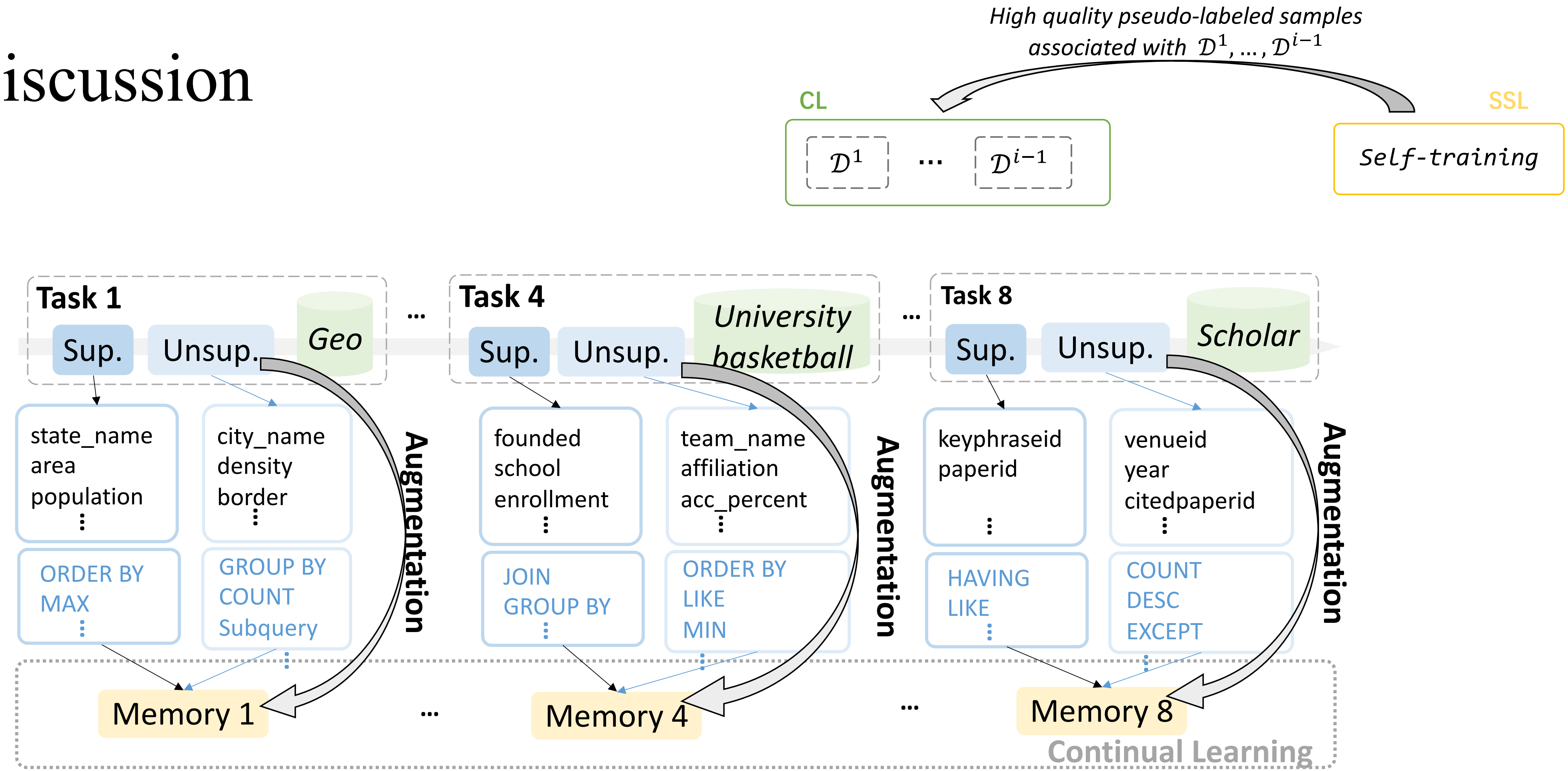
- Features of Text-to-SQL tasks
  - SQL structures & schema instances
  - Combinatorial generalization

Discussion



- Although instances A and B are associated with different databases, the target SQL of A is similar to that of B.
- The parser might learn from A on how to predict the pseudo-label of B.

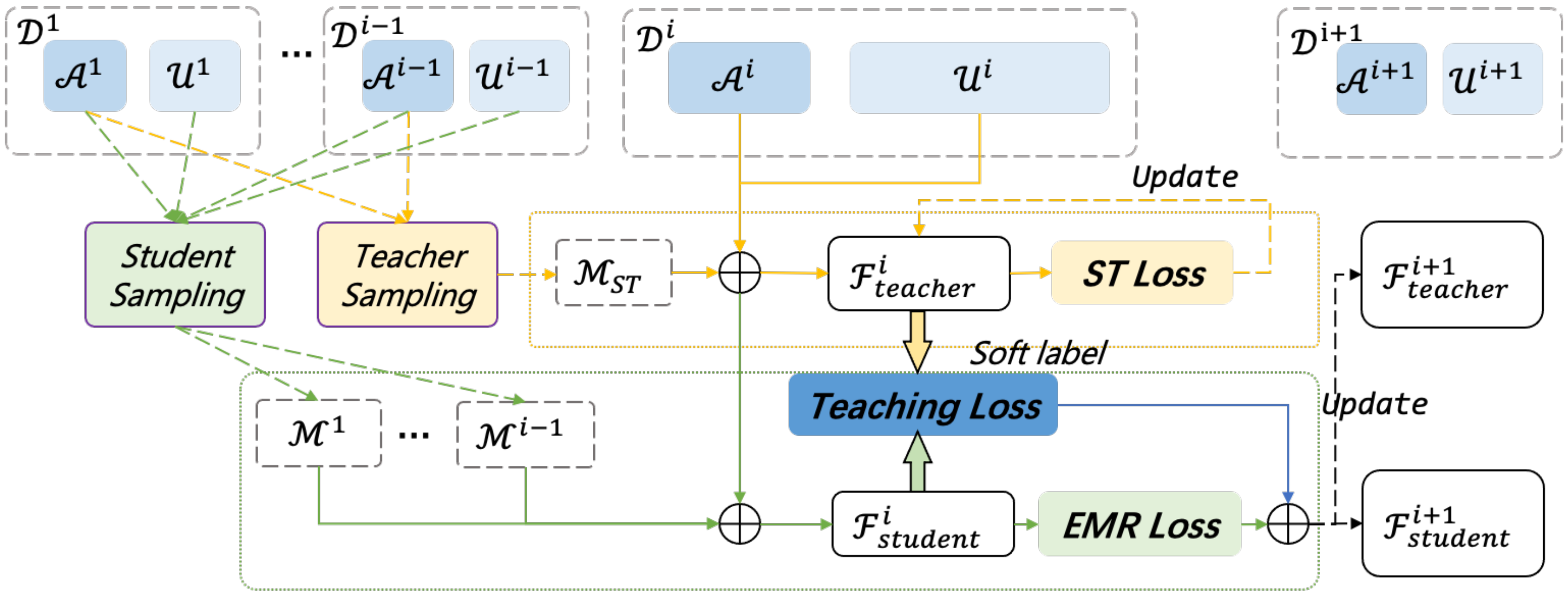
Discussion



- High-quality pseudo-labeled instances can also enrich the memory of past tasks.

Soft-fusion Network

- *Teacher-Student architecture* for the differences in learning goals of CL and SSL.
  - TEACHER → SSL;   STUDENT → CL;
- *Dual sampling* for the complementarity of CL and SSL.
  - Teacher sampling and student sampling
  - provides effective information supplement for TEACHER and STUDENT



Results

- Overall Results

Table 1: Experimental results for comparison with baselines.

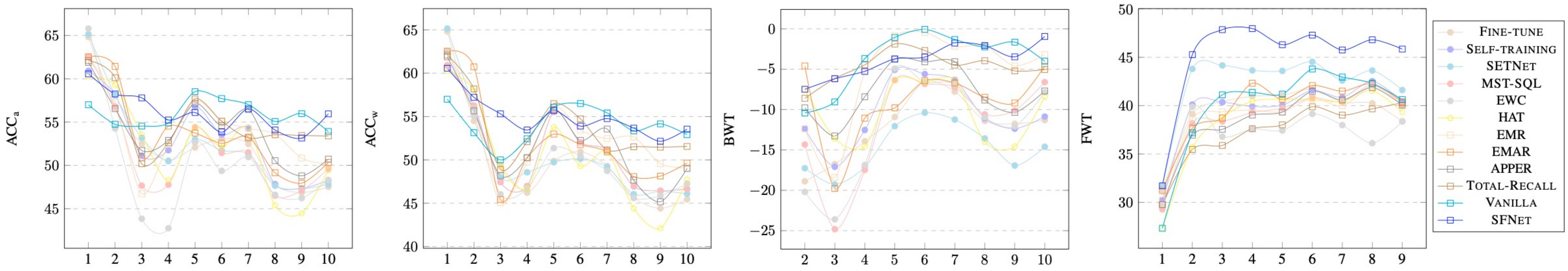
	Method	Spider				WikiSQL			
		ACC_a	ACC_w	BWT	FWT	ACC_a	ACC_w	BWT	FWT
	FINE-TUNE	47.5	45.5	-11.3	38.4	69.8	69.2	-1.5	63.0
SSL-only methods	SELF-TRAINING (Goldwasser et al. 2011)	48.3	46.6	-10.9	40.4	70.4	69.9	-3.3	63.5
	SETNET (Wang et al. 2020b)	47.8	46.1	-14.6	41.6	70.7	70.2	-2.1	61.8
	MST-SQL (Guo et al. 2022)	49.6	47.3	-6.6	40.7	70.7	70.1	-1.7	61.7
CL-only methods	EWC (Kirkpatrick et al. 2016)	48.3	47.2	-7.9	38.4	70.0	69.6	-2.0	61.4
	HAT (Serrà et al. 2018)	49.4	47.7	-8.4	39.3	70.0	69.6	**-1.4**	61.8
	EMR (Wang et al. 2019)	50.1	49.1	-3.2	40.3	71.1	70.7	-2.2	63.1
	EMAR (Han et al. 2020)	50.3	49.6	-4.6	40.4	70.8	70.5	-1.5	62.7
	APPER (Mi et al. 2020)	50.7	49	-7.7	40.0	70.2	69.9	-3.0	62.7
	TOTAL-RECALL (Li, Qu, and Haffari 2021)	53.4	51.6	-5.1	40.3	71.5	71.1	-2.1	62.7
Our Solutions	VANILLA	53.9	52.9	-4.0	40.6	72.2	71.9	-2.0	64.0
	SFNET	**56.0**	**53.6**	**-1.0**	**45.9**	**73.6**	**73.3**	-2.3	**65.6**
	ORACLE (all tasks w/o Unsup.)	62.9	63.4	5.2	48.7	73.1	72.7	2.6	64.2

- VANILLA outperforms all the baselines in terms of ACC_a and ACC_w;
- SFNET further improves its ACCa by 1.4% (WikiSQL) and 2.1% (Spider) and achieves SOTA performance in almost all metrics on two datasets.

Results

- Results till the Seen Tasks

ACCa, ACCw, BWT, FWT till the seen tasks on Spider after learning on each task sequentially.



- SFNET (blue) is always more stable than the other baselines in all metrics and this stability becomes more pronounced as the number of tasks grows.
- Almost all methods in BWT improves slightly as the number of tasks increases.

Our Work 2

Parameterizing Context: Unleashing the Power of Parameter-Efficient Fine-Tuning and In-Context Tuning for Continual Table Semantic Parsing

Yongrui Chen, Shenyu Zhang, Guilin Qi, Xinnan Guo

School of Computer Science and Engineering, Southeast University

NeurIPS 2023

A base text-to-SQL parser

we utilize T5 (Raffel et al., 2020) as the base parser $\mathcal{F}_\theta$.

In particular, each input pair X = (Q, S) is flattened into a plain text,

$$\mathcal{X}' = t_1 : c_1^{t_1}, c_2^{t_1}, \ldots, c_{m_1}^{t_1}; t_2 : c_1^{t_2}, c_2^{t_2}, \ldots, c_{m_2}^{t_2}; \ldots | \mathcal{Q},$$

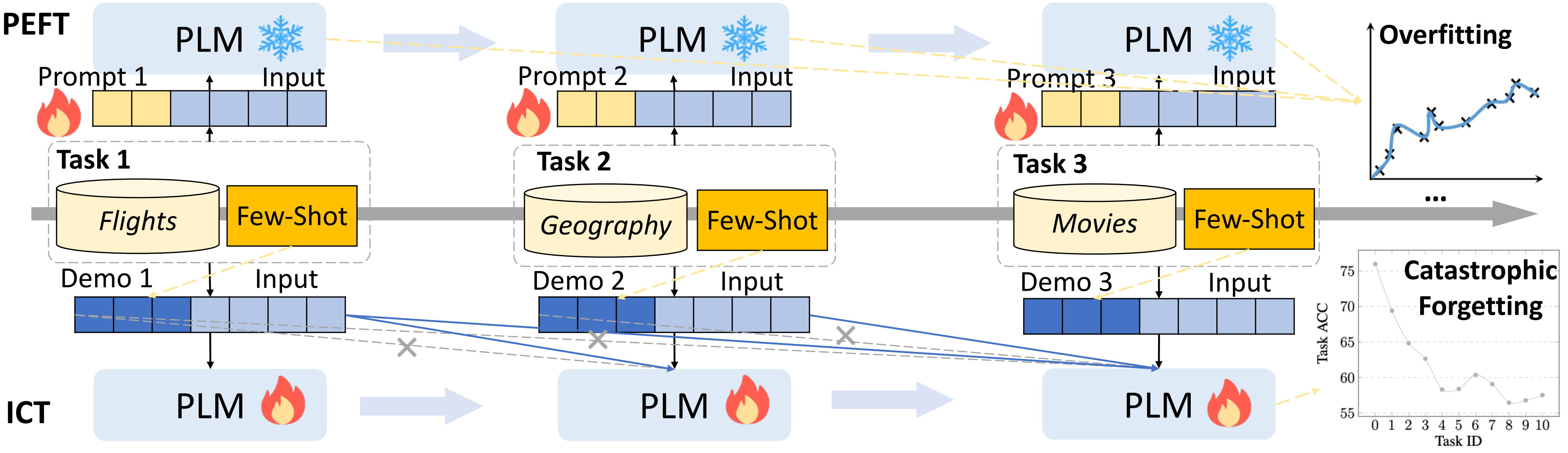
where $c_j^{t_i}$ denotes the j-th column name of the i-th table, and ":", ",", and "|" are predefined separators.

$$P(\tilde{\mathcal{Y}}|\mathcal{X}', \theta) = \prod_{j=1}^{|\tilde{\mathcal{Y}}|} P(\tilde{y}_j|\mathbf{X}', \tilde{y}_{<j}, \theta)$$

A Vanilla Solution

- In-context tuning (ICT) for limited supervision;
- Parameter-efficient fine-tuning (PEFT) for catastrophic forgetting by saving soft prompts;

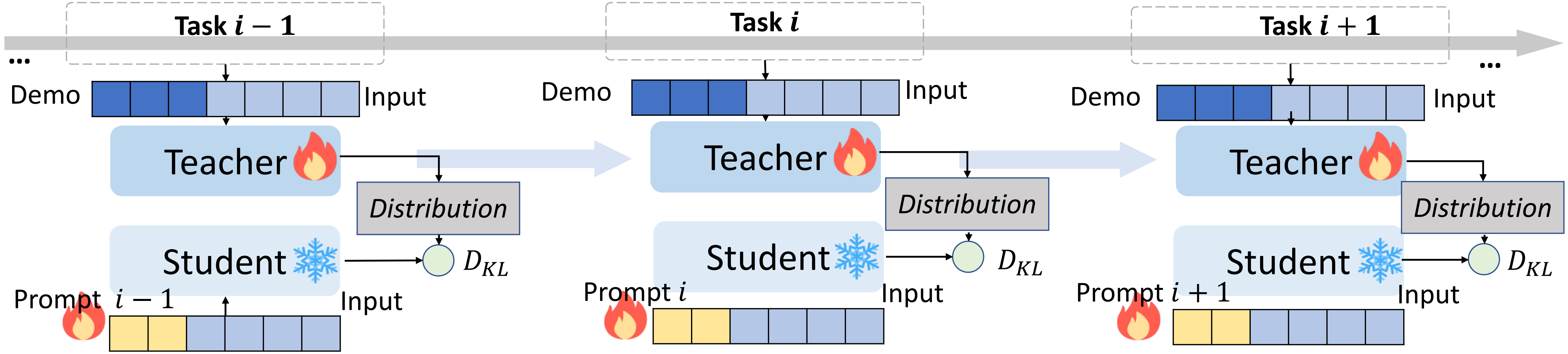
- Rigidly combines ICT and PEFT
- ICT method: *In-context Tuning*
  - Convenient in semantic parsing
  - Good performance on few-shot learning, (Min, et al., 2022)
- PEFT method: *Prompt-Tuning*
  - Simple process, suitable for complex structured text-to-SQL models
  - Best performance in CL methods for semantic parsing, (Zhu et al., 2022)



Context-Compressed Continual (C3) Parser

Teacher-student framework：

- TEACHER addresses the few-shot problem using ICT, which procures contextual information by demonstrating a few training examples.
- STUDENT leverages the proposed PEFT framework to learn from the teacher's output distribution, and subsequently compresses and saves the contextual information to the prompts, without storing any training examples.



Results

- Overall Results

Table 1: Experimental results for comparison with baselines in 3 random task orders. Means and standard variations are reported. The absence of standard deviation for PEFT and C3 is due to the fact that their performance is order-independent. ♠ indicates using the replayed memory of size 15 and ♣ indicates using additional unsupervised data.

Backbone	Method	Spider-Stream			Combined-Stream		
		TA (%)	EA (%)	MD (%)	TA (%)	EA (%)	MD (%)
GRAPPA-LARGE (340M)	FINE-TUNING	$56.9_{1.0}$	$54.6_{1.0}$	$-18.8_{1.5}$	$37.6_{1.8}$	$43.9_{0.9}$	$-39.1_{2.2}$
	MAML [8]	$52.2_{1.3}$	$49.1_{1.5}$	$-19.5_{2.2}$	$31.3_{1.3}$	$37.2_{1.4}$	$-43.8_{1.5}$
	ICT [13]	$57.0_{1.4}$	$54.3_{2.2}$	$-17.1_{2.1}$	$37.9_{4.0}$	$43.9_{1.8}$	$-37.4_{4.4}$
	EWC [28]	$57.5_{3.3}$	$55.1_{2.4}$	$-17.7_{3.9}$	$37.0_{1.9}$	$44.1_{0.9}$	$-38.4_{2.2}$
	HAT [29]	$57.8_{2.9}$	$54.8_{3.4}$	$-17.0_{3.3}$	$38.5_{4.1}$	$45.0_{2.0}$	$-37.6_{5.5}$
	EMR♠ [27]	$65.2_{0.2}$	$62.9_{0.6}$	$-9.4_{0.8}$	$60.9_{0.7}$	$58.6_{1.8}$	$-10.3_{1.8}$
	EMAR♠ [30]	$62.8_{1.2}$	$60.8_{1.2}$	$-10.5_{1.1}$	$63.1_{1.8}$	$60.8_{0.9}$	$-7.7_{2.6}$
	APPER [31]	$57.9_{1.6}$	$55.8_{1.7}$	$-17.2_{2.6}$	$37.1_{2.4}$	$44.0_{0.6}$	$-38.3_{2.9}$
	TR♠ [11]	$57.9_{1.2}$	$55.1_{1.4}$	$-15.8_{1.5}$	$59.7_{1.0}$	$56.3_{1.1}$	$-11.9_{0.7}$
	SFNET♣ [7]	-	-	-	$60.7_{0.9}$	$57.0_{1.9}$	$-6.0_{1.2}$
T5-BASE (220M)	EMR♠ [27]	$60.3_{0.9}$	$56.6_{0.1}$	$-13.4_{0.2}$	$62.6_{0.4}$	$60.0_{1.4}$	$-6.5_{0.6}$
	EMAR♠ [30]	$57.2_{0.4}$	$52.7_{1.0}$	$-16.7_{0.8}$	$62.1_{0.2}$	$58.1_{1.1}$	$-6.5_{0.3}$
	PEFT	65.7	64.5	0.0	63.8	66.2	0.0
	C3	$67.5_{0.3}$	$66.5_{0.2}$	$0.0_{0.0}$	$66.3_{0.2}$	$67.6_{0.2}$	$0.0_{0.0}$
T5-LARGE (770M)	MULTI-TASK	$76.3_{0.5}$	$76.2_{1.0}$	$3.2_{0.7}$	$70.0_{0.9}$	$71.1_{0.7}$	$1.7_{0.2}$
	PEFT	69.8	67.4	0.0	67.3	70.0	0.0
	C3	$\mathbf{71.1_{0.3}}$	$\mathbf{69.7_{0.5}}$	$\mathbf{0.0_{0.0}}$	$\mathbf{68.3_{0.5}}$	$\mathbf{70.6_{0.4}}$	$\mathbf{0.0_{0.0}}$

Results

- Results till the Seen Tasks

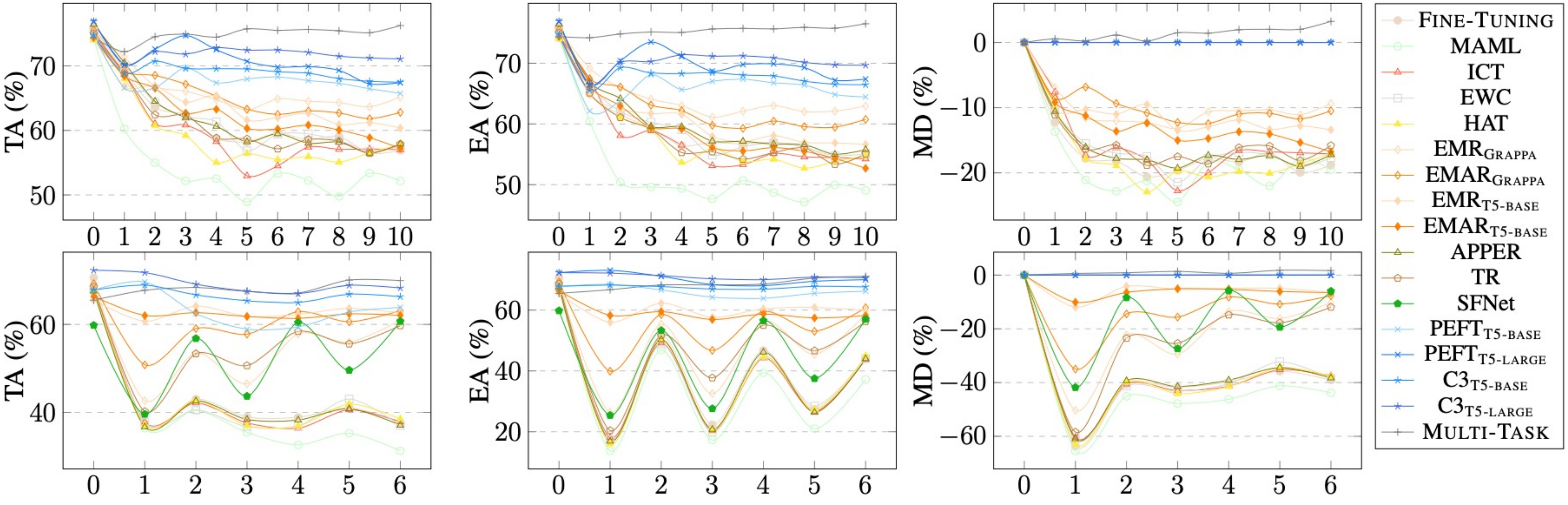


Figure 4: TA (%), EA (%), and MD (%) till the seen tasks of Spider-Stream (upper) and Combined-Stream (bottom) after learning on each task. Only the means are reported in 3 random task orders.

Results

- Using LLM as the teacher

Table 3: Performance of C3 using GPT as the TEACHER parser.

STUDENT	TEACHER	Spider-Stream		Combined-Stream	
		TA (%)	EA (%)	TA (%)	EA (%)
T5-BASE	text-davinci-003	66.3	64.8	65.5	66.8
	T5-LARGE	$\mathbf{67.5_{0.3}}$	$\mathbf{66.5_{0.2}}$	$\mathbf{66.3_{0.2}}$	$\mathbf{67.6_{0.2}}$
T5-LARGE	text-davinci-003	**71.3**	**69.6**	67.6	70.0
	T5-LARGE	$71.1_{0.3}$	$69.7_{0.5}$	$\mathbf{68.3_{0.5}}$	$\mathbf{70.6_{0.4}}$

Thank You !

References

- Wang, B., Shin, R., Liu, X., Polozov, O., & Richardson, M. (2019). Rat-sql: Relation-aware schema encoding and linking for text-to-sql parsers. arXiv preprint arXiv:1911.04942.
- Lin, X. V., Socher, R., & Xiong, C. (2020). Bridging textual and tabular data for cross-domain text-to-sql semantic parsing. arXiv preprint arXiv:2012.12627.
- Cao, R., Chen, L., Chen, Z., Zhao, Y., Zhu, S., & Yu, K. (2021). LGESQL: line graph enhanced text-to-SQL model with mixed local and non-local relations. arXiv preprint arXiv:2106.01093.
- Yu, T., Wu, C. S., Lin, X. V., Wang, B., Tan, Y. C., Yang, X., ... & Xiong, C. (2021, January). GraPPa: Grammar-Augmented Pre-Training for Table Semantic Parsing. In ICLR.
- Guo, J., Zhan, Z., Gao, Y., Xiao, Y., Lou, J. G., Liu, T., & Zhang, D. (2019). Towards complex text-to-sql in cross-domain database with intermediate representation. arXiv preprint arXiv:1905.08205.

References

- Wang, X., Sun, H., Qi, Q., & Wang, J. (2020). SETNet: A Novel Semi-Supervised Approach for Semantic Parsing. In ECAI 2020 (pp. 2236-2243). IOS Press.
- Qi, Q., Wang, X., Sun, H., Wang, J., Liang, X., & Liao, J. (2020). A novel multi-task learning framework for semi-supervised semantic parsing. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 2552-2560.
- Yin, P., Zhou, C., He, J., & Neubig, G. (2018). Structvae: Tree-structured latent variable models for semi-supervised semantic parsing. arXiv preprint arXiv:1806.07832.
- Cao, R., Zhu, S., Liu, C., Li, J., & Yu, K. (2019). Semantic parsing with dual learning. arXiv preprint arXiv:1907.05343.
- Ye, H., Li, W., & Wang, L. (2019). Jointly learning semantic parser and natural language generator via dual information maximization. arXiv preprint arXiv:1906.00575.
- Konstas, I., Iyer, S., Yatskar, M., Choi, Y., & Zettlemoyer, L. (2017). Neural amr: Sequence-to-sequence models for parsing and generation. arXiv preprint arXiv:1704.08381.
- Kate, R., & Mooney, R. (2007, April). Semi-supervised learning for semantic parsing using support vector machines. In Human Language Technologies 2007: the Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers (pp. 81-84).
- Corro, C., & Titov, I. (2018). Differentiable perturb-and-parse: Semi-supervised parsing with a structured variational autoencoder. arXiv preprint arXiv:1807.09875.
- Chen, Z., Chen, L., Li, H., Cao, R., Ma, D., Wu, M., & Yu, K. (2021). Decoupled dialogue modeling and semantic parsing for multi-turn text-to-SQL. arXiv preprint arXiv:2106.02282.

References

- Goldwasser, D., Reichart, R., Clarke, J., & Roth, D. (2011, June). Confidence driven unsupervised semantic parsing. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies(pp. 1486-1495).
- Lialin, V., Goel, R., Simanovsky, A., Rumshisky, A., & Shah, R. (2020). Update frequently, update fast: Retraining semantic parsing systems in a fraction of time. arXiv preprint arXiv:2010.07865.
- Li, Z., Qu, L., & Haffari, G. (2021). Total recall: a customized continual learning method for neural semantic parsers. arXiv preprint arXiv:2109.05186.
- Brahma, D., Verma, V. K., & Rai, P. (2021). Hypernetworks for Continual Semi-Supervised Learning. arXiv preprint arXiv:2110.01856.
- Luo, Y., Wong, Y., Kankanhalli, M., & Zhao, Q. (2022). Learning to Predict Gradients for Semi-Supervised Continual Learning. arXiv preprint arXiv:2201.09196.
- Smith, J., Balloch, J., Hsu, Y. C., & Kira, Z. (2021, July). Memory-efficient semi-supervised continual learning: The world is its own replay buffer. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
- Wang, L., Yang, K., Li, C., Hong, L., Li, Z., & Zhu, J. (2021). Ordisco: Effective and efficient usage of incremental unlabeled data for semi-supervised continual learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5383-5392).
- Baucum, M., Belotto, D., Jeannet, S., Savage, E., Mupparaju, P., & Morato, C. W. (2017, June). Semi-supervised deep continuous learning. In Proceedings of the 2017 International Conference on Deep Learning Technologies (pp. 11-18).

References

- Karim, N., Khalid, U., Esmaeili, A., & Rahnavard, N. (2022). CNLL: A Semi-supervised Approach For Continual Noisy Label Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3878-3888).
- Furlanello, T., Lipton, Z., Tschannen, M., Itti, L., & Anandkumar, A. (2018, July). Born again neural networks. In International Conference on Machine Learning (pp. 1607-1616). PMLR.
- Zhang, Y., Xiang, T., Hospedales, T. M., & Lu, H. (2018). Deep mutual learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4320-4328).