Multi-Stage LLM Fine-Tuning with a Continual Learning Setting

Changhao Guan[1], Chao Huang[1], Hongliang Li[1]

, You Li[1], Ning Cheng[1], Zihe Liu[1]

, Jinan Xu[1], Yufeng Chen[*1], Jian Liu[2]

[1]Beijing Jiaotong University, Beijing, China

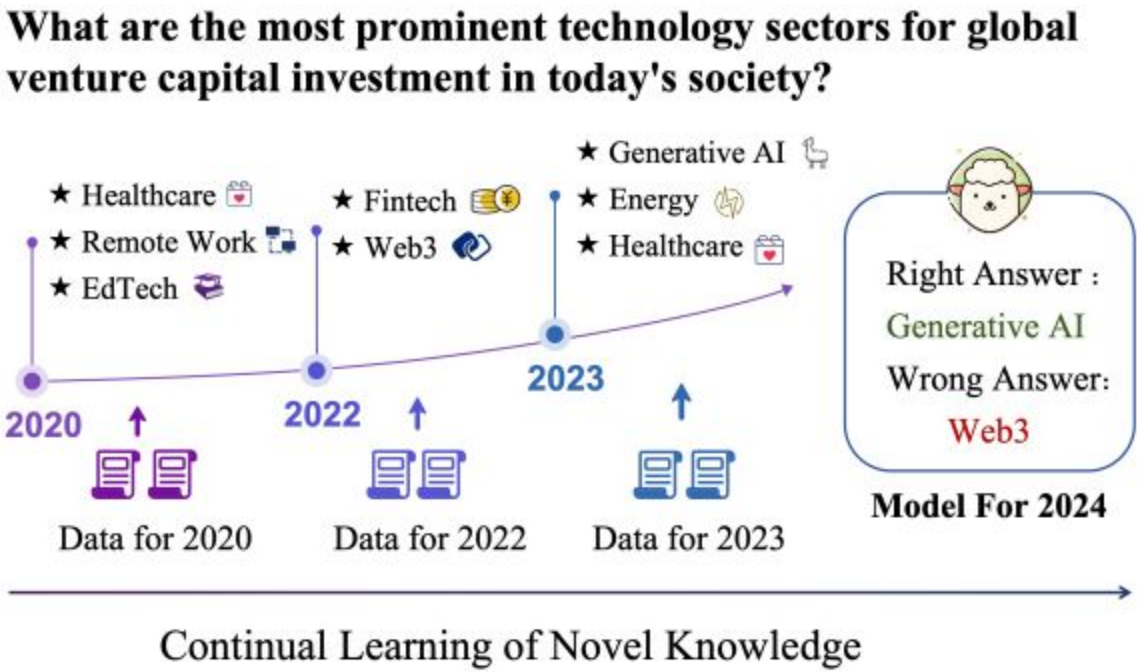
[2]University of Science and Technology Beijing, Beijing, China

{guanchanghao, huangchao, hongliangli, youlee}@bjtu.edu.cn

{ningcheng,23120386, jaxu, chenyf}@bjtu.edu.cn, jian.liu@ustb.edu.cn

Year of Publication :- 2025

Introduction



Introduction

- Large Language Models (LLMs) are considered complex knowledge repositories as they have ability to represent diverse general information [1,2].
- However, it is necessary to fine-tune them on customized datasets when applying them to specific domains [3,4].
- In addition, there is a continual learning requirement, especially when domain knowledge rapidly changes[5,6].
- Pilot experiments show that employing the standard fine-tuning methods for LLMs significantly degrades their performance. However, this problem has not received much research attention.

[1] Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., ... & Agarwal, S. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165, 1(3), 3.
[2] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35, 27730-27744.
[3] Xu, R., Luo, F., Zhang, Z., Tan, C., Chang, B., Huang, S., & Huang, F. (2021). Raise a child in large language model: Towards effective and generalizable fine-tuning. arXiv preprint arXiv:2109.05687.
[4] Xie, T., Wan, Y., Huang, W., Yin, Z., Liu, Y., Wang, S., ... & Hoex, B. (2023). Darwin series: Domain specific large language models for natural science. arXiv preprint arXiv:2308.13565.
[5] McCann, B., Keskar, N. S., Xiong, C., & Socher, R. (2018). The natural language decathlon: Multitask learning as question answering. arXiv preprint arXiv:1806.08730.
[6] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. arXiv preprint arXiv:2004.10964.

Introduction - Problems in Continual Learning

- Potential knowledge conflict [7,8]. When a domain undergoes rapid changes, potential conflicts between new and old knowledge may arise, potentially leading to “hallucinations” in LLMs.
- Incomparable amount of fine-tuning data compared to pre-training data [9,10]. Compared to the extensive data leveraged during pre-training, the domain specific data available for fine-tuning is typically scarce, making it challenging to adapt the model’s parameters to fit for fine-tuning.
- Proposed a new approach for fine-tuning LLMs in the multi-stage continual learning settings using a preference-based forgetting strategy and self-distillation based data augmentation.

[7] Longpre, S., Perisetla, K., Chen, A., Ramesh, N., DuBois, C., & Singh, S. (2021). Entity-based knowledge conflicts in question answering. arXiv preprint arXiv:2109.05052.
[8] Liu, Y., Yao, Z., Lv, X., Fan, Y., Cao, S., Yu, J., ... & Li, J. (2024). Untangle the knot: Interweaving conflicting knowledge and reasoning skills in large language models. arXiv preprint arXiv:2404.03577.
[9] Jiang, G., Jiang, C., Xue, S., Zhang, J. Y., Zhou, J., Lian, D., & Wei, Y. (2023). Towards anytime fine-tuning: Continually pre-trained language models with hypernetwork prompt. arXiv preprint arXiv:2310.13024.
[10] Dong, G., Yuan, H., Lu, K., Li, C., Xue, M., Liu, D., ... & Zhou, J. (2023). How abilities in large language models are affected by supervised fine-tuning data composition. arXiv preprint arXiv:2310.05492.

Related Works - Fine tuning LLMs

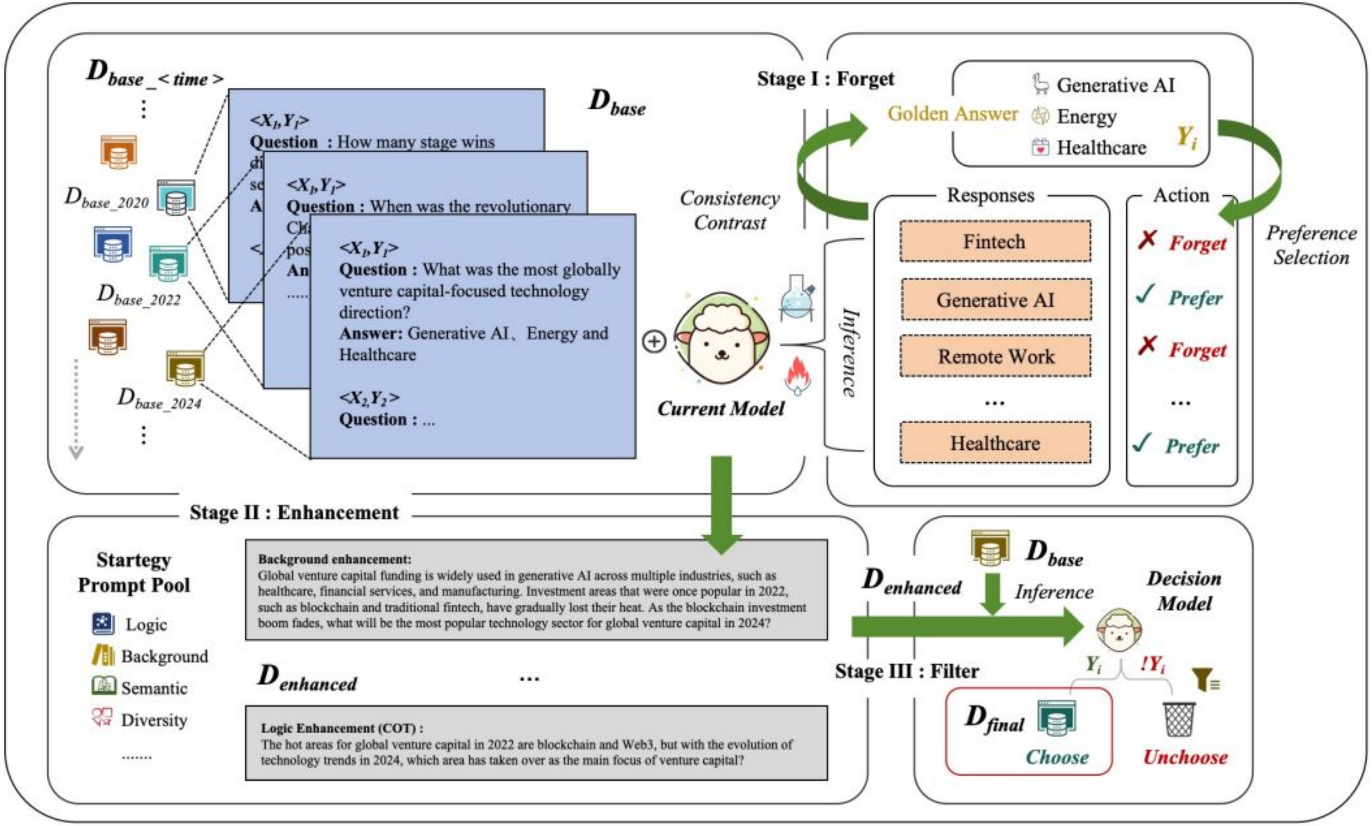
- Fine-tuning is a widely adopted approach to adopt LLMs to new domains and tasks using domain specific data [11,12].
    - Fine-tuning for complex instructions.
    - Fine-tuning for specific domains.
- Solely lying on fine-tuning often struggles to acquire new knowledge when facing significant domain shifts[13].
- Existing works have 2 stage approach [14].
    - First Stage - Fine-tuning to acquire domain knowledge.
    - Second Stage - Fine-tuning to enhance task specific capabilities.

[11] Ding, R., Han, X., & Wang, L. (2022). A unified knowledge graph augmentation service for boosting domain-specific NLP tasks. arXiv preprint arXiv:2212.05251.
[12] Zheng, J., Hong, H., Liu, F., Wang, X., Su, J., Liang, Y., & Wu, S. (2024). Fine-tuning large language models for domain-specific machine translation. arXiv preprint arXiv:2402.15061.
[13] Emelin, D., Bonadiman, D., Alqahtani, S., Zhang, Y., & Mansour, S. (2022). Injecting domain knowledge in language models for task-oriented dialogue systems. arXiv preprint arXiv:2212.08120.
[14] Han, R., Ren, X., & Peng, N. (2020). ECONET: Effective continual pretraining of language models for event temporal reasoning. arXiv preprint arXiv:2012.15283.

Related Works - Continual Learning with LLMs

- Traditional CL methods [15].
  - Regularization based.
  - Replay Based.
  - Architecture based strategies.
- Continuous fine tuning strategy [16] .
- Modular continual learning [17].
- Forget-before-learn [18].

Approach - Overview



Preference Based Learning Bias

- Let (x, y) be a training example.
- x is utilized as an input and apply the model K times to get a prediction set Y for each input x.
- Measure compatibility between each element of the set Y and the desired y.
- Divide Y into subsets Y_{align} and $Y_{conflict}$.
- Main motivation is to bias the model to generate responses similar to those in Y_{align} and avoid those in Y_{conf}.

Preference Losses

- Positive preference loss.

$$\mathcal{L}_{\text{PP}} = - \sum_{y' \in Y_{\text{align}}} \log \left(\frac{\pi_\theta(y' \mid x)}{\pi_{\text{ref}}(y' \mid x)} \right)$$

- Negative preference loss.

$$\mathcal{L}_{\text{NP}} = \sum_{y' \in Y_{\text{conf}}} \log \left(\frac{\pi_\theta(y' \mid x)}{\pi_{\text{ref}}(y' \mid x)} \right)$$

- Total loss

$$\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{PP}} + \beta \cdot \mathcal{L}_{\text{NP}}$$

Data Augmentation with Self Distillation

- Using the LLMs themselves for augmentation.
- Augmentation strategies.
  - Background knowledge integration - LLM is asked to provide more background knowledge in order for the input to contain more context related information.
  - Logic-Compatible Expansion - LLM is asked to incorporate the logic-related information to expand the semantic complexity of the input.
  - Paraphrase augmentation - This method involves rewriting and rephrasing the original example to write more similar examples with various structures and expressions.
- Based on the above strategies, a new pair (x’, y) can be created for any training example (x,y).

Dynamic Data Selection Strategy

- To evaluate the effectiveness and validity if the augmented data for model training.
- Based on heuristic criteria.
  - Mutual Information.

$$MI(x'; \hat{x}') = \sum_{w, \hat{w}'} N(w, \hat{w}') \log \left(\frac{N(w, \hat{w}')}{N(w) N(\hat{w}')} \right)$$

  - Indication from LLMs.

    By this criterion, it is measured whether x' produces same result as x.

- The filtered set of high-quality samples will be used as input for subsequent augmentation and fine-tuning processes.

Experimental Setup - Datasets

- Considered rapidly evolving domains: Natural sciences, medicine, technology, transportation, tourism, finance and social sciences.
- Total 21,000 question- answer pairs with 3000 question-answer pairs for each domain.
- In this, there are 1000 examples shared by any 2 domains to evaluate cross domain conflict situations.
- 6000 additional samples as general purpose samples to indicate model's performance in domain agnostic setting.

Experimental Setup - Evaluation Settings

- Domain independent continual learning.
  - Use the domain independent dataset (6000 samples) for fine-tuning in the initial stage.
  - Manually edit the answers and use the revised dataset for fine-tuning.
  - Generate the same number of examples compatible with the fine-tuning data as evaluation set.
- Cross domain scenarios.
  - Conducting continual learning using cross domain data by gradually adding domains one by one.
  - Manual verification to ensure that a minimum 1000 examples are shared between 2 domains.
  - Before fine-tuning, answers were edited to be different from previous domain to mimic domain dispute.
  - Each fine-tuning stage contains, 2000 domain independent examples and 1000 cross domain conflict examples.

Experimental Setup - Evaluation Metric

- Knowledge Gain Ratio (KGR) - Assessing model's improvement in learning dynamically evolving knowledge.
- Post-injection Accuracy - To measure overall accuracy improvement in a given test set.

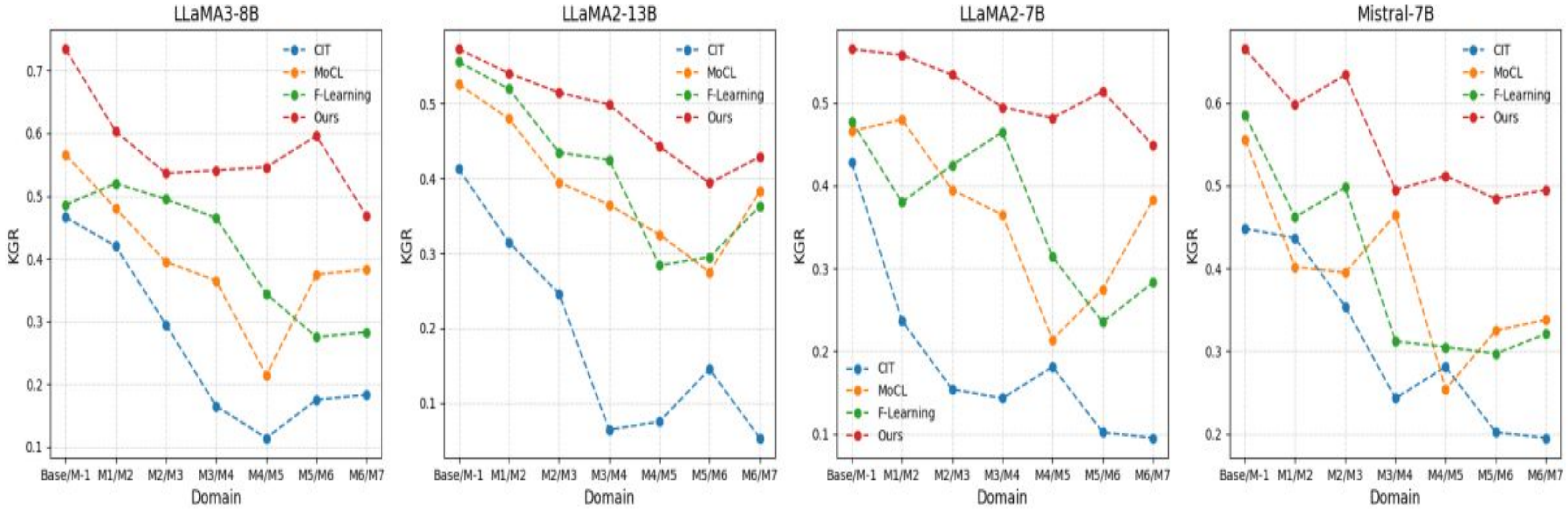
Experimental Setup - Baselines and backbone LLMs

- Baselines
  - Continual instruction fine-tuning (CIF).
  - Modular continual learning (MoCL).
  - Forgetting before learning (F-Learning).
- Backbone LLMs.
  - Llama2.
  - Llama3.
  - Misral7B.

Results - Domain independent continual learning

Eval	Stage1 (Initial)		Stage2		Stage3		Stage4		Stage5	
	KGR	ACC	KGR	ACC	KGR	ACC	KGR	ACC	KGR	ACC
Llama2-7B										
CIT (2023b)	50.50	66.55	$44.30_{\downarrow 6.20}$	$46.10_{\downarrow 20.45}$	$12.50_{\downarrow 38.00}$	$25.39_{\downarrow 41.16}$	$12.20_{\downarrow 38.30}$	$24.23_{\downarrow 42.32}$	$11.70_{\downarrow 38.80}$	$17.60_{\downarrow 48.95}$
MoCL (2024a)	–	–	$48.20_{\uparrow 3.90}$	$49.60_{\uparrow 3.50}$	$45.50_{\uparrow 33.00}$	$47.70_{\uparrow 22.31}$	$46.80_{\uparrow 34.60}$	$47.10_{\uparrow 22.87}$	$26.20_{\uparrow 14.50}$	$27.35_{\uparrow 9.75}$
F-Learning (2023)	–	–	$54.40_{\uparrow 10.10}$	$57.50_{\uparrow 11.40}$	$49.40_{\uparrow 36.90}$	$52.25_{\uparrow 26.86}$	$49.20_{\uparrow 37.00}$	$49.75_{\uparrow 25.52}$	$21.80_{\uparrow 10.10}$	$20.90_{\uparrow 3.30}$
Ours	–	–	$60.20_{\uparrow 15.90}$	$62.55_{\uparrow 16.45}$	$69.40_{\uparrow 56.90}$	$69.65_{\uparrow 44.26}$	$68.60_{\uparrow 56.40}$	$67.53_{\uparrow 43.30}$	$75.80_{\uparrow 64.10}$	$75.49_{\uparrow 57.89}$
Llama2-13B										
CIT (2023b)	68.90	81.95	$32.20_{\downarrow 36.70}$	$34.20_{\downarrow 47.75}$	$26.80_{\downarrow 42.10}$	$33.55_{\downarrow 48.40}$	$24.60_{\downarrow 44.30}$	$32.85_{\downarrow 49.10}$	$11.20_{\downarrow 57.70}$	$17.85_{\downarrow 64.10}$
MoCL (2024a)	–	–	$41.60_{\uparrow 9.40}$	$41.95_{\uparrow 7.75}$	$50.40_{\uparrow 23.60}$	$51.25_{\uparrow 17.70}$	$48.70_{\uparrow 24.10}$	$50.00_{\uparrow 17.15}$	$25.80_{\uparrow 14.60}$	$26.50_{\uparrow 8.65}$
F-Learning (2023)	–	–	$43.30_{\uparrow 11.10}$	$43.80_{\uparrow 9.60}$	$59.30_{\uparrow 32.50}$	$60.70_{\uparrow 27.15}$	$53.90_{\uparrow 29.30}$	$54.90_{\uparrow 22.05}$	$33.60_{\uparrow 22.40}$	$33.90_{\uparrow 16.05}$
Ours	–	–	$66.30_{\uparrow 34.10}$	$67.25_{\uparrow 33.05}$	$76.50_{\uparrow 49.70}$	$77.40_{\uparrow 43.85}$	$66.20_{\uparrow 41.60}$	$65.45_{\uparrow 32.60}$	$76.50_{\uparrow 65.30}$	$77.65_{\uparrow 59.80}$
Llama3-8B										
CIT (2023b)	65.90	81.55	$48.80_{\downarrow 17.10}$	$49.90_{\downarrow 31.65}$	$31.90_{\downarrow 34.00}$	$34.05_{\downarrow 47.50}$	$30.30_{\downarrow 35.60}$	$35.30_{\downarrow 46.25}$	$23.40_{\downarrow 42.50}$	$27.70_{\downarrow 53.85}$
MoCL (2024a)	–	–	$70.40_{\uparrow 21.6}$	$70.65_{\uparrow 20.75}$	$52.50_{\uparrow 20.60}$	$52.65_{\uparrow 18.60}$	$63.30_{\uparrow 33.00}$	$63.65_{\uparrow 28.35}$	$31.30_{\uparrow 7.90}$	$30.95_{\uparrow 3.25}$
F-Learning (2023)	–	–	$67.00_{\uparrow 18.20}$	$67.45_{\uparrow 17.55}$	$58.40_{\uparrow 26.50}$	$57.95_{\uparrow 23.90}$	$61.40_{\uparrow 31.10}$	$61.20_{\uparrow 25.90}$	$57.70_{\uparrow 34.30}$	$57.90_{\uparrow 30.20}$
Ours	–	–	$83.60_{\uparrow 34.80}$	$83.90_{\uparrow 34.00}$	$69.40_{\uparrow 37.50}$	$69.55_{\uparrow 35.50}$	$69.20_{\uparrow 33.90}$	$69.20_{\uparrow 39.03}$	$74.80_{\uparrow 51.40}$	$74.60_{\uparrow 46.90}$
Mistral-7B										
CIT (2023b))	61.00	74.20	$41.20_{\downarrow 19.80}$	$44.65_{\downarrow 29.55}$	$38.50_{\downarrow 22.50}$	$38.90_{\downarrow 35.30}$	$32.80_{\downarrow 28.20}$	$36.40_{\downarrow 37.80}$	$21.60_{\downarrow 39.40}$	$23.50_{\downarrow 50.70}$
MoCL (2024a)	–	–	$54.40_{\uparrow 13.20}$	$51.25_{\uparrow 6.60}$	$62.50_{\uparrow 24.00}$	$59.10_{\uparrow 20.20}$	$58.79_{\uparrow 25.99}$	$57.25_{\uparrow 20.85}$	$47.10_{\uparrow 25.50}$	$46.00_{\uparrow 22.50}$
F-Learning (2023)	–	–	$48.30_{\uparrow 7.10}$	$51.50_{\uparrow 6.85}$	$58.10_{\uparrow 19.60}$	$57.65_{\uparrow 18.75}$	$54.20_{\uparrow 21.40}$	$55.30_{\uparrow 18.90}$	$33.80_{\uparrow 12.20}$	$34.90_{\uparrow 11.40}$
Ours	–	–	$64.60_{\uparrow 23.40}$	$65.72_{\uparrow 21.07}$	$72.60_{\uparrow 34.10}$	$72.80_{\uparrow 33.90}$	$71.20_{\uparrow 38.40}$	$71.58_{\uparrow 35.18}$	$77.50_{\uparrow 55.90}$	$77.34_{\uparrow 53.84}$

Results - Cross Domain Setting



Discussion - Ablation study

- Impact of Preference based Learning Bias.

Method	ACC(%)	KGR(%)
CIT	48.80	49.90
PBL (Ours)	**59.00**	**59.30**

- Impact of Data Augmentation with Self-Distillation.

Method	ACC (%)	KGR (%)
No Argument	49.90	48.80
+ BKI	58.85	58.60
+ LCE	52.80	52.50
+ PA	56.40	64.10
CA (Ours)	**69.90**	**69.30**

Discussion - Ablation Study

- Impact of Dynamic Data Selection Strategy.

Method	**ACC (%)**	**KGR (%)**
No Argument	49.90	48.80
Data Argument	69.90	69.30
+ RS (50%)	79.85	79.60
+ RS (25%)	69.85	63.50
+ RS (12.5%)	70.40	66.70
DS (Ours)	**81.20**	**82.50**

Discussion - Analysis of Knowledge Retention

- 3000 data points were added in the first round of the experiment.
- Model’s knowledge retention and forgetting of original domain data were monitored in every round of the experiment.
- Knowledge Retention Rate (KRR) is introduced as an additional evaluation metric.
- The traditional CIT method results in a significant decline in both ACC and KRR after multiple training rounds, indicating a severe catastrophic forgetting phenomenon and a sharp deterioration in the model’s ability to recall original information.

Conclusion

- This work introduced a novel approach that incorporates conflict-based learning to address knowledge conflicts.
- This work introduces a self-distillation based data augmentation approach to enhance training data.
- Extensive experiments were done through which the method demonstrated significant improvements in both knowledge acquisition efficiency and long term retention of previously learned information.

References

1. Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., ... & Agarwal, S. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165, 1(3), 3.
2. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35, 27730-27744.
3. Xu, R., Luo, F., Zhang, Z., Tan, C., Chang, B., Huang, S., & Huang, F. (2021). Raise a child in large language model: Towards effective and generalizable fine-tuning. arXiv preprint arXiv:2109.05687.
4. Xie, T., Wan, Y., Huang, W., Yin, Z., Liu, Y., Wang, S., ... & Hoex, B. (2023). Darwin series: Domain specific large language models for natural science. arXiv preprint arXiv:2308.13565.
5. McCann, B., Keskar, N. S., Xiong, C., & Socher, R. (2018). The natural language decathlon: Multitask learning as question answering. arXiv preprint arXiv:1806.08730.
6. Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. arXiv preprint arXiv:2004.10964.
7. Longpre, S., Perisetla, K., Chen, A., Ramesh, N., DuBois, C., & Singh, S. (2021). Entity-based knowledge conflicts in question answering. arXiv preprint arXiv:2109.05052.
8. Liu, Y., Yao, Z., Lv, X., Fan, Y., Cao, S., Yu, J., ... & Li, J. (2024). Untangle the knot: Interweaving conflicting knowledge and reasoning skills in large language models. arXiv preprint arXiv:2404.03577.
9. Jiang, G., Jiang, C., Xue, S., Zhang, J. Y., Zhou, J., Lian, D., & Wei, Y. (2023). Towards anytime fine-tuning: Continually pre-trained language models with hypernetwork prompt. arXiv preprint arXiv:2310.13024.
10. Dong, G., Yuan, H., Lu, K., Li, C., Xue, M., Liu, D., ... & Zhou, J. (2023). How abilities in large language models are affected by supervised fine-tuning data composition. arXiv preprint arXiv:2310.05492.
11. Ding, R., Han, X., & Wang, L. (2022). A unified knowledge graph augmentation service for boosting domain-specific NLP tasks. arXiv preprint arXiv:2212.05251.
12. Zheng, J., Hong, H., Liu, F., Wang, X., Su, J., Liang, Y., & Wu, S. (2024). Fine-tuning large language models for domain-specific machine translation. arXiv preprint arXiv:2402.15061.
13. Emelin, D., Bonadiman, D., Alqahtani, S., Zhang, Y., & Mansour, S. (2022). Injecting domain knowledge in language models for task-oriented dialogue systems. arXiv preprint arXiv:2212.08120.
14. Han, R., Ren, X., & Peng, N. (2020). ECONET: Effective continual pretraining of language models for event temporal reasoning. arXiv preprint arXiv:2012.15283.
15. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences, 114(13), 3521-3526.
16. Xin, C., Lu, Y., Lin, H., Zhou, S., Zhu, H., Wang, W., ... & Sun, L. (2024, May). Beyond full fine-tuning: Harnessing the power of LoRA for multi-task instruction tuning. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 2307-2317).