Metacognitive Symbolic Distillation Framework for Multi-choice Machine Reading Comprehension

Anonymous ACL submission

Abstract

Previous research often utilizes symbolic distillation to transfer the reasoning abilities of large teacher models to smaller student models. However, when it comes to multi-choice machine reading comprehension (MMRC), solely learning from the rationales generated by the teacher model for correct options overlooks educational significance of understanding the reasons behind incorrect options. In education, metacognition requires individuals to actively identify errors when reading to deepen their understanding. To this end, we propose a novel framework for achieving metacognitive symbolic distillation. Initially, we prompt the teacher large language model (LLM) to generate rationales for each option in the MMRC dataset. Subsequently, the student model could be fine-tuned based on the MMRC data equipped with these rationales. Our experiments on two MMRC datasets demonstrate that our approach effectively enhances the performance of the small model compared with standard fine-tuned models and symbolic distilled models. Moreover, when the student model is large enough, upgrading the teacher model can lead to further improvements. We will make our code and data publicly available.

1 Introduction

Large language models (LLMs) have demonstrated promising performance on various tasks by employing chain-of-thought (CoT) prompting (Wang et al., 2023a). Generating justifications to elucidate their responses can enhance their accuracy. Nevertheless, deploying LLMs in compute-starved scenarios is challenging due to their large-scale parameters and high inference latency (Xu and McAuley, 2023). As a result, it is common to fine-tune a smaller model with human-labeled data to cater to specific requirements (Wang et al., 2023b).

To enhance the performance and interpretability of small models, existing studies commonly employ symbolic distillation to transfer the reasoning capabilities of large LLMs into smaller ones (West et al., 2022). In particular, a large teacher model is utilized to generate rationales for expected responses, which are employed to fine-tune a smaller student model (Li et al., 2023). However, within the realm of pedagogy, metacognition necessitates individuals to actively identify errors while reading, which is considered a crucial element in attaining a profound comprehension of texts (Dori et al., 2018). Especially in the context of Multi-choice Machine Reading Comprehension (MMRC) tasks, current symbolic distillation approaches solely focus on having the teacher LLM generate rationales for the correct answer options, while neglecting the identification of causes for the incorrect options.

In this study, we present a novel framework for achieving metacognitive symbolic distillation. Our framework employs CoT prompts to guide the teacher model to generate rationales for each option. To accomplish this, we adopt two strategies, namely **elimination strategy** and **sequential strategy**. In the first one, we instruct the teacher model to first generate rationales for incorrect options and conclude by generating a rationale for the correct option. In the second one, we prompt the teacher model to generate rationales in order of original options. Subsequently, our framework combines the original dataset along with the generated rationales to fine-tune the student model. This enables the student model to analyze the reasons behind incorrect options, considering the input context, question, and options, and finally provide the correct answer and its corresponding justifications.

We conduct extensive experiments on two MMRC benchmarks: MCTest (Richardson et al., 2013) and RACE (Lai et al., 2017). Our findings indicate that our framework can effectively enhance the performance of the student model compared with standard fine-tuned models and symbolic distillation models. Additionally, the performance of the student model can be further improved by up-

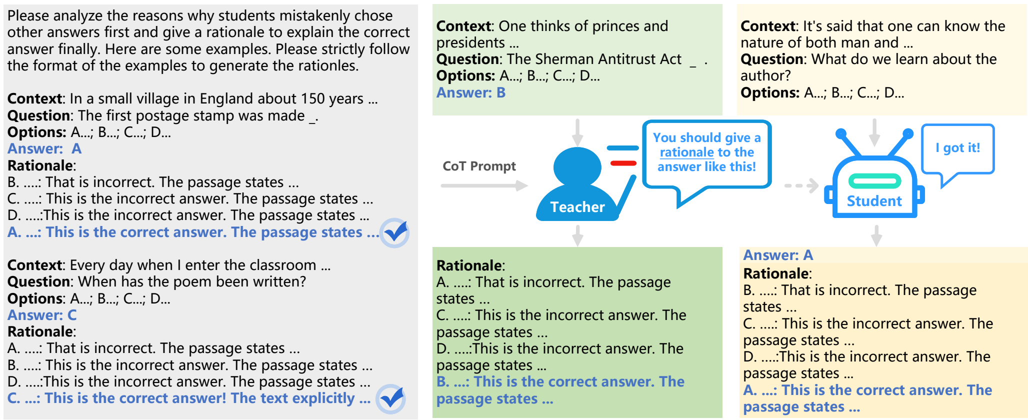

Figure 1: The overall architecture of metacognitive symbolic distillation. Initially, the teacher model will generate rationales for both the incorrect and correct options pertaining to the given input context, question, options, and answer. This process relies on a manually tailored CoT prompt within the dataset. Subsequently, the context, question, options, answer, and generated rationales will be employed to fine-tune the student model.

grading the teacher model. Notably, this improvement in performance is primarily observed when the student model itself is of a significant size.

Our contributions are summarized as follows:

(i) To the best of our knowledge, our work is the first to incorporate the metacognitive perspective into symbolic distillation.

(ii) We propose a framework for metacognitive symbolic distillation on MMRC task. Our approach involves formulating exemplars to prompt the teacher LLM to generate rationales for each option. Subsequently, we employ the dataset and the rationales to fine-tune the small student model.

(iii) Experiments on two MMRC datasets demonstrate that the proposed framework significantly improves the performance of the student model compared with standard fine-tuned models and symbolic distillation models. When the student model reaches a significant size, upgrading to a larger teacher model can lead to further improvements in the performance of the student model.

2 Metacognitive Symbolic Distillation

Task Description Prior to introducing our framework, we shall expound upon the problem formulation and notations. The MMRC dataset consists of multiple samples denoted as $\{(c, q, O, a)\}_{i=1}^{N}$, where c represents the context, q represents the question, $O = \{o_1, \cdots, o_n\}$ represents the set of candidate answer options, and a represents the only one right answer. In the MMRC task, the objective is to answer the question q by selecting the most appropriate answer a in the set of candidate answer options O based on the given context c. Our proposed approach leverages a teacher LLM $\mathcal{T}$ to generate rationales for each option within every sample of the dataset, thereby forming N new samples $\{(c, q, O, R, a)\}_{i=1}^{N}$, where $R = \{r_1, \cdots, r_4\}$ is the set of generated rationales. We utilize these samples to fine-tune a smaller student model $\mathcal{S}$.

Metacognitive Teacher The initial step is to curate a set of labeled CoTs based on the dataset to serve as prompts for $\mathcal{T}$. For each sample in MMRC dataset, we select K (*e.g.*, $K = 2$) samples from the dataset and manually customized CoTs denoted as $Z = \{z_1, \cdots, z_4\}$ for the four options (as shown in left part of Fig. 1) in every sample (c, q, O, a) to compose the prompt set $\mathcal{P} = \{(c, q, O, a, Z)\}_{k=1}^{K}$. This approach can provide LLM with exemplars, prompting LLM to generate rationales for each option in the MMRC dataset.

Moreover, we implement two answering strategies to regulate the sequence of CoTs in Z: In the case of the **elimination strategy**, the prioritization of CoTs is focused on incorrect options, while the CoT for the correct option is reserved until the final stage. Conversely, in the **sequential strategy**, the CoTs are arranged in the order of the options. Afterwards, we input each sample (c, q, O, a) in the dataset together with $\mathcal{P}$ into teacher LLM to obtain a training dataset $\mathcal{D}$ for the small student model. $\mathcal{D}$ consists of N samples $\{c, q, O, R, a\}_{i=1}^{N}$, where $R = \{r_1, \cdots, r_4\}$ are the rationales for every option, arranged in the order of incorrect options first and correct options last (elimination strategy) or the order of options (sequential strategy).

Metacognitive Student Once we have acquired the dataset $\mathcal{D}$, we employ the same answering strategy to fine-tune the student model $\mathcal{S}$. Given a context c, a question q, and options O, the student model $\mathcal{S}$ is supervised to generate a sequence of answer tokens concatenated with the rationale tokens. Upon adopting the elimination strategy, the rationales section of the student model's output will prioritize generating explanations for the incorrect answer options before providing the reasoning behind the correct answer option. When implementing the sequential strategy, the rationales section of the student model's output will generate explanations in order of options. To accomplish this, we fine-tune the student model using the standard language model loss function (Raffel et al., 2020).

3 Experiments

3.1 Settings

Datasets We evaluate our framework on two English MMRC benchmarks: MCTest (Richardson et al., 2013) and RACE (Lai et al., 2017). MCTest consists of 660 fictional stories, where each story is accompanied by four questions and four candidate answers. RACE is collected from middle and high school English exams in China. Compared with other MMRC datasets, RACE requires more reasoning to answer questions.

Baselines We conduct performance comparison with the following baselines. **Standard Fine-tuning:** We directly fine-tune a student model using MCTest and RACE, bypassing the need for the teacher model. Given a context, question, and options, the small model is trained to produce the accurate option as its output. The output of the small model does not include rationales. **Standard Symbolic Distillation:** By utilizing CoT prompting, the teacher model is employed to generate rationales for the correct options. Afterwards, the student model is tasked with generating a sequence of tokens that concatenates rationale and the corresponding correct option when provided with a context, question, and available options. In contrast to metacongnitive symbolic distillation, standard symbolic distillation solely generates justifications for the correct options.

Implementation Details For the teacher model, we select Llama-2-13b and GPT-3.5 turbo. For the student model, we opt for different sizes of T5. Apart from GPT-3.5 turbo, which is accessed via an API, the remaining models are locally deployed using Pytorch and Huggingface. The inference of the teacher models and the training and inference of the student models are conducted on an NVIDIA A800 80GB GPU, respectively. Exactly matching metric is employed for evaluation. Furthermore, for the experiments conducted on the RACE, we specifically utilize 3,000 articles from the high school training set of the RACE to form a distinct training set. Regarding the MCTest, we select 500 articles and divide them into training, validation, and testing sets.

Teacher	Student	RACE	MCTest
-	T5-ft	0.5921	0.8383
Llama-13b	T5-rt	0.6735	<u>0.8733</u>
	T5-or†	0.6881	0.8617
	T5-et†	<u>0.6941</u>	0.8633
GPT-3.5	T5-rt	0.7093	0.8683
	T5-or†	**0.7316**	0.8783
	T5-et†	0.7210	**0.8800**

Table 1: The performance of the student model (T5-3b) using different settings. T5-ft and T5-rt signify the T5-3b subsequent to the standard fine-tuning and symbolic distillation, respectively. T5-or and T5-et represent the implementation of sequential and elimination strategies of the proposed metacognitive symbolic distillation, respectively. The best results obtained using GPT-3.5 and Llama-13b as teacher models are represented by **boldface** and <u>underline</u>, respectively.

3.2 Results and Analysis

We first evaluate the performance of metacognitive symbolic distillation and all baselines. The comparison results are illustrated in Table 1.

Symbolic distillation outperforms standard fine-tuned models in all cases. By employing symbolic distillation, the student model demonstrates superior performance compared with the traditional fine-tuning approach, as evidenced by the exact matching metric scores of 0.5921 and 0.8383 on both datasets. This finding suggests that symbolic distillation not only enhances the interpretability of the student model in the MMRC task but also effectively enhances its inference capabilities.

In the majority of cases, metacognitive symbolic distillation demonstrates the ability to enhance the performance of student models compared with standard symbolic distillation. For RACE, regardless of whether the teacher model is Llama-13b or GPT-3.5, and whether the policy adopts the elimination (T5-et) or sequential strategy (T5-or), our framework consistently outperforms standard symbolic distillation (T5-rt). Similarly, for MCTest, our framework surpasses standard symbolic distillation when the teacher is GPT-3.5. While our framework may exhibit slightly lower performance than standard symbolic distillation on MCTest when the teacher model is Llama-13b, it is important to consider that this discrepancy could be attributed to the influence of the training size.

Metacognitive symbolic distillation exhibits improved performance with increasing teacher model size. When applying our proposed framework, whether utilizing the rationale strategy, it is observed that the performance of the student model improves as the size of the teacher model increases. Referring to Table 1, as for the RACE dataset, we can observe an enhancement in the performance of elimination strategy when transitioning from the Llama teacher (0.6941) to the GPT 3.5 teacher (0.7210). The similar phenomenon is observed in the MCTest as well. Contrarily, when employing standard symbolic distillation, a larger teacher model does not always result in improved performance of the student model. In the MCTest dataset, the utilization of the GPT 3.5 teacher results in a performance drop from 0.8733 to 0.8683 compared with using the Llama teacher.

We assess the performance of our proposed framework across various sizes of student models. We select three T5 models with different sizes and conduct metacognitive symbolic distillation using two teacher models: Llama-13b and GPT-3.5. These experiments are conducted on the RACE dataset. The results are presented in Figure 2. Our findings indicate that there is a performance improvement after distillation when the student model is larger. Furthermore, we observe that a larger teacher model can enhance the performance of the student model, especially when the student model itself is larger. When comparing the performance of student models distilled by two teachers of different sizes, the performance remains relatively similar regardless of whether the student is small (60m) or large (770m). However, when utilizing the 3b model, the student distilled by GPT-3.5 outperforms the student model distilled by Llama-13b.

We further divide the RACE training set into two smaller subsets, comprising 20% and 50% of the original data, respectively. Subsequently, we reevaluate the proposed framework on these subsets. As depicted in Figure 3, it is observed that the performance of the student model exhibited enhancement with the augmentation of training set.

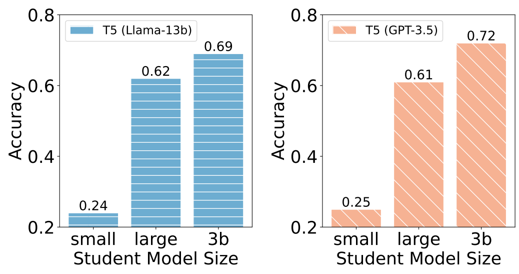

Figure 2: Accuracy of the T5-et with different sizes of student model (T5) on RACE. The subfigures on the left and right correspond to the teacher models Llama-13b and GPT-3.5, respectively.

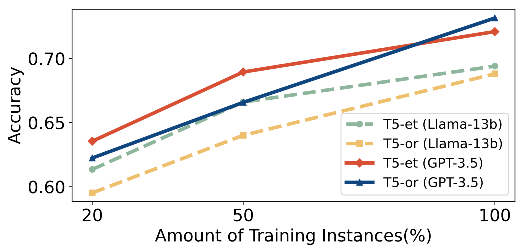

Figure 3: Accuracy of symbolic distillation strategies with different amount of training instances on RACE.

4 Conclusion

This paper presents a novel framework for metacognitive symbolic distillation within the domain of MMRC. While previous research focused on transferring the reasoning abilities of LLMs to smaller student models, our framework goes a step further by emphasizing the educational significance of comprehending the reasons behind both correct and incorrect options. Experimental results on two MMRC datasets demonstrate the effectiveness of our proposed framework, outperforming standard fine-tuned models and symbolic distillation models. Additionally, it is observed that upgrading the teacher model can lead to further improvements when the student model is large enough.

Limitations

Firstly, we focus lies solely on conducting experiments with T5 as the student. Secondly, our study primarily investigates the performance of T5 on the MMRC task, and we have not thoroughly explored its effectiveness in multiple-choice tasks involving open-ended questions. Lastly, while we use the rationales generated by the LLMs to improve the performance of the student model, not all rationales are accurate due to the hallucination of LLMs.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Yehudit Judy Dori, Zemira R Mevarech, and Dale R Baker. 2018. Cognition, metacognition, and culture in stem education. *Innovations in Science Education and Technology*, 24:386.

Olga Golovneva, Moya Chen, Spencer Poff, Martin Corredor, Luke Zettlemoyer, Maryam Fazel-Zarandi, and Asli Celikyilmaz. 2022. Roscoe: A suite of metrics for scoring step-by-step reasoning. *arXiv preprint arXiv:2212.07919*.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *stat*, 1050:9.

Jiaxin Huang, Shixiang Shane Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2022. Large language models can self-improve. *arXiv preprint arXiv:2210.11610*.

Yufan Jiang, Shuangzhi Wu, Jing Gong, Yahui Cheng, Peng Meng, Weiliang Lin, Zhibo Chen, et al. 2020. Improving machine reading comprehension with single-choice decision and transfer learning. *arXiv preprint arXiv:2011.03292*.

Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. 2017. Race: Large-scale reading comprehension dataset from examinations. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 785–794.

Liunian Harold Li, Jack Hessel, Youngjae Yu, Xiang Ren, Kai-Wei Chang, and Yejin Choi. 2023. Symbolic chain-of-thought distillation: Small models can also" think" step-by-step. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 2665–2679.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Qiu Ran, Peng Li, Weiwei Hu, and Jie Zhou. 2019. Option comparison network for multiple-choice reading comprehension. *arXiv preprint arXiv:1903.03033*.

Matthew Richardson, Christopher JC Burges, and Erin Renshaw. 2013. Mctest: A challenge dataset for the open-domain machine comprehension of text. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 193–203.

Peifeng Wang, Zhengyang Wang, Zheng Li, Yifan Gao, Bing Yin, and Xiang Ren. 2023a. Scott: Self-consistent chain-of-thought distillation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 5546–5558.

Ruida Wang, Wangchunshu Zhou, and Mrinmaya Sachan. 2023b. Let's synthesize step by step: Iterative dataset synthesis with large language models by extrapolating errors from small models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 11817–11831.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

Peter West, Chandra Bhagavatula, Jack Hessel, Jena Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck, and Yejin Choi. 2022. Symbolic knowledge distillation: from general language models to commonsense models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4602–4625.

Canwen Xu and Julian McAuley. 2023. A survey on model compression and acceleration for pretrained language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10566–10575.

Shuailiang Zhang, Hai Zhao, Yuwei Wu, Zhuosheng Zhang, Xi Zhou, and Xiang Zhou. 2020. Dcmn+: Dual co-matching network for multi-choice reading comprehension. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9563–9570.

A Appendix

A.1 Related Work

A.1.1 Chain-of-thought Prompting

Chain-of-thought (CoT) is a form of prompting that enhances the performance of LLMs by incorporating statements such as "let's think step by step" within the prompts (Wei et al., 2022). Current studies on CoTs primarily centers around improving the efficiency of CoT prompting and assessing the quality of CoTs. For instance, Wang et al. (2022) introduced a voting mechanism to determine the ultimate output from numerous CoTs generated by LLMs, giving preference to the one with the highest frequency of occurrences. Huang et al. (2022)

proposed a bootstrapping training approach that involves iteratively training on self-generated CoTs to enhance the performance of CoTs. Additionally, Golovneva et al. (2022) presented automatic metrics for assessing CoTs automatically. In this study, we aim to leverage CoTs containing incorrect options to enhance the performance and interpretability of small models.

A.1.2 Symbolic Distillation

Symbolic distillation (West et al., 2022), categorized as a form of knowledge distillation (Hinton et al., 2015), deviates from conventional techniques of distilling knowledge from soft representations such as logits. In symbolic distillation, LLMs are regarded as data generators for training smaller models. Particularly, Li et al. (2023) furnished an LLM with an unlabeled corpus, thereby prompting the LLM to discern labels and corresponding rationales for the unannotated data. The data that has been labeled by the LLM will subsequently be utilized as training data for the smaller models. Wang et al. (2023b) introduced the explain-then-generate strategy, leveraging an LLM to generate rationales for labels in the dataset. Subsequently, the LLM is utilized to generate training data for the small model based on the labels and the generated rationales. However, existing symbolic distillation methods fail to take into account the reasoning behind incorrect answers. In contrast, our work effectively incorporates both the rationales of correct and incorrect options through symbolic distillation, thereby enabling better supervision of the small model.

A.1.3 Multi-choice Machine Reading Comprehension

Multi-choice Machine Reading Comprehension (MMRC) aims to ascertain the accurate answer from a given context and question by selecting from a set of options. Initially, pre-trained language models like BERT (Devlin et al., 2018) were used to encode the context, questions, and options. Matching networks were then employed to score the options. Ran et al. (2019) introduced an option comparison network, which facilitated better reasoning by identifying word-level correlations among options. Zhang et al. (2020) proposed a dual co-matching network that effectively captured the bidirectional relationship among the document, question and options, enhancing their interactions modeling. Jiang et al. (2020) treated MMRC as multiple binary classification tasks, determining the final answer selection by independently calculating the confidence scores between each option and the context as well as the questions. However, although these methods improved the accuracy of answer selection based on the interaction between context, questions, and options, they lacked the ability to provide justifications for the chosen responses. Our work employs an LLM to generate justifications for the correct options (explaining why they are right) and corresponding rationales for the incorrect options (explaining why they are wrong) in the MMRC datasets. Subsequently, the datasets with generated rationales are employed to train a small student model, thereby bolstering the interpretability of the answers.

A.2 Implementation Details

We utilize Llama-2-13b[1] and GPT-3.5 turbo[2] as our teacher models, while T5[3] serves as the student model. The original datasets we work with are RACE[4] and MCTest[5]. Generating corresponding rationales by the teacher model typically requires approximately 24 hours. The training process of each student model takes around 16 hours. Due to time constraints, we were only able to conduct a single run for all experiments.

[1] https://huggingface.co/meta-llama/Llama-2-13b-chat-hf
[2] https://platform.openai.com/docs/models/gpt-3-5-turbo
[3] https://huggingface.co/google-t5
[4] https://www.cs.cmu.edu/ glai1/data/race/
[5] https://huggingface.co/datasets/sagnikrayc/mctest