SCIENCE CHINA
Information Sciences

• LETTER •

June 2025, Vol. 68, Iss. 6, 160109:1–160109:2
https://doi.org/10.1007/s11432-024-4379-8

Special Topic: Enabling Techniques and Cutting-Edge Applications of Foundation Models

Dynamic prompt allocation and tuning for continual test-time adaptation

Chaoran CUI[1], Yongrui ZHEN[1], Shuai GONG[1], Chunyun ZHANG[1], Hui LIU[1] & Yilong YIN[2*]

[1]*School of Computing and Artificial Intelligence, Shandong University of Finance and Economics, Jinan 250014, China*
[2]*School of Software, Shandong University, Jinan 250101, China*

Received 26 November 2024/Revised 7 February 2025/Accepted 3 April 2025/Published online 19 May 2025

Citation Cui C R, Zhen Y R, Gong S, et al. Dynamic prompt allocation and tuning for continual test-time adaptation. Sci China Inf Sci, 2025, 68(6): 160109, https://doi.org/10.1007/s11432-024-4379-8

Continual test-time adaptation (CTTA) [1] is a recently emerged method, which is used for the adaptation of a pretrained source model to continuously evolving target distributions during test time. Its purpose is to accommodate the dynamic nature of real-world environments. CTTA is generally performed online; it processes unlabeled target data sequentially and adapts the model to each incoming data batch before the model starts making predictions.

A major problem in CTTA is catastrophic forgetting, where the model "forgets" the previously acquired knowledge as it continuously adapts to dynamic data distributions. To overcome this problem, previous studies have typically incorporated an additional regularization term to constrain the variation of model parameters. However, these studies cannot fundamentally resolve catastrophic forgetting because they rely on a single shared model, which should adapt to all target domains, inevitably leading to severe interdomain interference. When the shared model adapts to new target domains, its parameters are updated to optimize its performance on the new data. This process involves the risk of overwriting parameters that were crucial in previous domains, leading to catastrophic forgetting. Therefore, to fundamentally resolve catastrophic forgetting, it is essential to allow different domains to be learned in a separated way for CTTA.

In this study, we propose a novel dynamic prompt allocation and tuning (PAINT) method for CTTA. This method is based on recent advances in prompt tuning for deep neural networks. In prompt tuning, a few learnable prompt tokens are used as extra inputs to facilitate the rapid adaptation of a pretrained model to downstream tasks. Specifically, PAINT introduces domain-specific prompts for individual target domains, thereby partially disentangling the parameter space across different domains. During adaptation, samples from a target domain are expected to contribute solely to the optimization of their corresponding domain-specific prompts. Given the absence of domain identity for target samples, PAINT employs a query mechanism to dynamically determine whether the current samples come from a known or a new domain. For a known domain, the corresponding domain-specific prompt is directly selected from a memory buffer, whereas for a previously unknown domain, a new prompt is allocated in the buffer. Prompt tuning is subsequently performed using mutual information maximization along with structural regularization. The conceptual framework of PAINT is illustrated in Figure 1.

Problem definition. In CTTA, test samples obtained from continuously evolving target domains arrive sequentially in batches. Let $\mathcal{B} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_B\}$ be a batch of samples obtained from a certain target domain, where B is the batch size. A pretrained model adapts to $\mathcal{B}$ by updating its parameters with one gradient step; then, the model makes online predictions for each sample $\boldsymbol{x}_i \in \mathcal{B}$. Note that the domain identity of test samples and the total number of target domains are unknown. As a standard practice in CTTA, we consider the K-way image classification task and assume that all domains share the same category space.

Dynamic prompt allocation. In the proposed PAINT method, we employ a query mechanism to authorize test samples to decide by themselves which prompt to employ. A memory buffer $\mathcal{M} = \{(\boldsymbol{k}_1, \boldsymbol{P}_1), (\boldsymbol{k}_2, \boldsymbol{P}_2), \ldots\}$ is maintained to store domain-specific prompts, where each prompt $\boldsymbol{P}_j$ is paired with its corresponding key $\boldsymbol{k}_j$. Initially, $\mathcal{M}$ is empty and gradually expands as target samples arrive.

Upon receiving a batch of samples $\mathcal{B}$, the pretrained source model is used as a frozen feature extractor to obtain the visual features $\boldsymbol{q}_i$ for each sample $\boldsymbol{x}_i \in \mathcal{B}$. $\boldsymbol{x}_i$ uses $\boldsymbol{q}_i$ as a query and retrieves the most appropriate prompt by matching $\boldsymbol{q}_i$ with each prompt key $\boldsymbol{k}_j$ obtained from the memory buffer $\mathcal{M}$. We empirically used the cosine similarity as the matching function. All samples in $\mathcal{B}$ can make their own choices. Then, their choices are aggregated via majority voting to select the optimal prompt for each data batch; this is defined as $\boldsymbol{P}_s$ and is paired with key $\boldsymbol{k}_s$. In this study, we assume that each batch of target samples belongs to a single target domain; this is in line with the assumptions

* Corresponding author (email: ylyin@sdu.edu.cn)

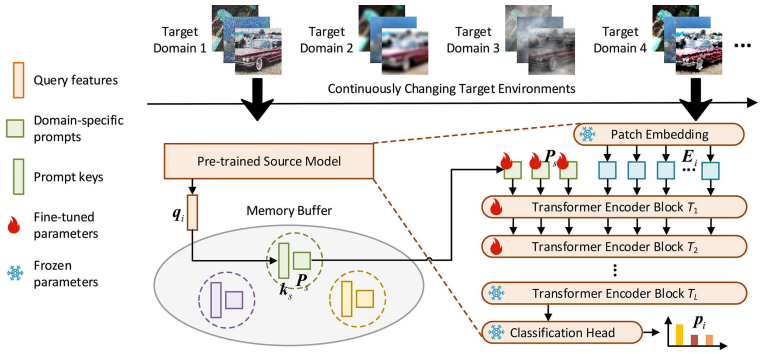

Figure 1 (Color online) Conceptual framework of the proposed PAINT method for CTTA.

made in most previous studies on CTTA [1, 2].

Note that data batch $\mathcal{B}$ probably comes from a target domain that has not been encountered before; as a result, none of the prompts in $\mathcal{M}$ corresponds to the domain of $\mathcal{B}$. To overcome this problem, we measure the reliability r_s of the selected prompt $\boldsymbol{P}_s$; r_s is calculated as the average matching score between the key $\boldsymbol{k}_s$ and all samples in $\mathcal{B}$. If r_s exceeds a predefined threshold η, $\boldsymbol{P}_s$ is considered a reliable prompt for the domain to which $\mathcal{B}$ belongs. Otherwise, $\boldsymbol{P}_s$ is discarded, and a new prompt is initialized in $\mathcal{M}$ to serve as the selection for $\mathcal{B}$. More details of the prompt allocation process can be found in Appendix B.2.

Prompt tuning. After determining the domain-specific prompt $\boldsymbol{P}_s$ that is suitable for the current data batch $\mathcal{B}$, $\boldsymbol{P}_s$ is optimized using $\mathcal{B}$. The objective is to minimize the mutual information loss $\mathcal{L}_{\text{mi}}$ [3] of the model's predictions. Minimizing $\mathcal{L}_{\text{mi}}$ serves a dual purpose: (1) it reduces the average entropy of the model's predictions, thereby preventing the model from producing ambiguous predictions for individual target samples; (2) it increases the entropy of the model's average output, thus enhancing prediction diversity within a batch and reducing the risk of the model's predictions collapsing into a limited number of categories.

Additionally, we incorporate an interpolation consistency regularization function $\mathcal{L}_{\text{ic}}$ [4] to ensure that the model's predictions remain consistent when applied to interpolated target samples. This is achieved by generating pseudo labels for samples based on the model's initial predictions, which are then mixed with other samples and their corresponding pseudo labels. The model is trained to minimize the cross-entropy loss between its predictions for these mixed samples and the mixed pseudo labels. This process enables the model to make accurate predictions not only for the original samples but also for the interpolated samples, thereby improving the model's ability to be generalized to new data in the target domain.

Finally, the optimization objective of PAINT integrates $\mathcal{L}_{\text{mi}}$ and $\mathcal{L}_{\text{ic}}$. In addition to optimizing $\boldsymbol{P}_s$, we also fine-tune the feature encoder of the pretrained model. To minimize the computational overhead, only the first three encoder blocks are updated during training. These shallow blocks are critical for capturing the common knowledge shared across all target domains. More details of the prompt tuning process can be found in Appendix B.3.

Experiments. We conducted experiments on four benchmark datasets, namely, CIFAR10-C, ImageNet-C, ImageNet-R, and CIFAR10-W. The proposed PAINT method was evaluated against several recently proposed CTTA methods based on ResNet and ViT backbones. The results showed that PAINT achieves state-of-the-art performance in CTTA. We verified the effectiveness of PAINT in handling gradually changing scenarios and its robust anti-forgetting capability regarding source knowledge. Moreover, we conducted ablation studies and hyperparameter analyses to validate the key components of PAINT. Visualization studies provide deep insights into PAINT. Details of the experimental settings and a comprehensive analysis of the results can be found in Appendix C.

Conclusion. Catastrophic forgetting, which is primarily caused by interdomain interference, is a major problem in CTTA. In this study, we introduced domain-specific prompts to guide model adaptation, thus facilitating the partial disentanglement of parameter spaces across different domains. As the domain identity for target samples is unknown, we employed a query mechanism to dynamically determine whether data come from a previously known or a new domain, followed by prompt tuning using mutual information maximization and structural regularization. The results obtained from the experiments conducted on four benchmark datasets demonstrate the effectiveness of the proposed PAINT method in CTTA. A limitation of this study is the assumption that each batch of the target samples corresponds to a single domain, which may not hold in practice. In future work, we will extend PAINT to mixed distribution shifts [5], where target samples in the same batch may span across multiple domains.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 62077033) and Taishan Scholar Program of Shandong Province (Grant Nos. tsqn202211199, tstp20221137).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Wang Q, Fink O, van Gool L, et al. Continual test-time domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022. 7201–7211
2 Döbler M, Marsden R A, Yang B. Robust mean teacher for continual and gradual test-time adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023. 7704–7714
3 Liang J, Hu D, Wang Y, et al. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 8602–8617
4 Zhang H Y, Cissé M, Dauphin Y N, et al. mixup: beyond empirical risk minimization. In: Proceedings of the 6th International Conference on Learning Representations, 2018
5 Niu S C, Wu J X, Zhang Y F, et al. Towards stable test-time adaptation in dynamic wild world. In: Proceedings of the 11th International Conference on Learning Representations, 2023