Internal Robust Representations for Domain Generalization

Mohammad Rostami

Information Sciences Institute, University of Southern California
rostamim@usc.edu

Abstract

Model generalization under distributional changes remains a significant challenge for machine learning. We present consolidating the internal representation of the training data in a model as a strategy of improving model generalization.

Introduction

The classic premise of machine learning is that the training and the testing data are drawn from the same distribution. This assumption, however, is not valid when the input distribution changes during testing time, e.g., due to temporal drifts (Rostami 2021b). The distributional gap between the training and testing data will lead to poor model generalization. We argue that the model generalization can be improved if we consolidate the internally learned distribution of data and demonstrate that this general strategy can be used to challenges of domain adaptation and continual learning.

Consider a training dataset $\mathcal{D}_I = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ for a classification task which is drawn from an initial training distribution $P_I(\boldsymbol{x}, y)$. In a parametric formulation, we select a function $f_\theta(\cdot) : \mathcal{X} \rightarrow \mathcal{Y}$ and solve for the optimal parameter θ^* using empirical risk minimization over $\mathcal{D}_I$. In most cases, we can decompose the model into two sub-models. i.e., $f_\theta(\cdot) = c_{\boldsymbol{v}}(\cdot) \circ e_{\boldsymbol{w}}(\cdot)$, where $e_{\boldsymbol{w}}(\cdot)$ denotes an encoder and $c_{\boldsymbol{v}}(\cdot)$ denotes a classifier. The classifier can perform well if the internal marginal distribution of data, i.e., $e_{\boldsymbol{w}}(P_I(\boldsymbol{x})$, represents the input data as separable clusters. The reason behind poor generalization of the model in the presence of distribution gaps is the mismatch between the internally learned distribution $e_{\boldsymbol{w}}(P_I(\boldsymbol{x}))$ and the internal distribution of the testing data $e_{\boldsymbol{w}}(P_T(\boldsymbol{x}))$. We can improve the model generalization $P_T(\boldsymbol{x})$ by consolidating the internal distribution.

Unsupervised Domain Adaptation

In an unsupervised domain adaptation setting, the testing domain usually is a target domain at which we have access only to unlabeled data and the goal is to update the model using only unlabeled data. We improve the model generalization by matching the target domain distribution with the internal distribution at the output space of the encoder. To this end, we can select a probability metric $D(\cdot, \cdot)$ and then adapt the encoder by minimizing the empirical distribution between P_I and P_T, i.e., $D(e_{\boldsymbol{w}}(\boldsymbol{x}), e_{\boldsymbol{w}}(\boldsymbol{x}))$ We have used this strategy to address several challenges of unsupervised domain adaptation. (Stan and Rostami 2021; Rostami 2022; Stan and Rostami 2022; Rostami 2022; Rostami and Galstyan 2023).

Continual Learning

In a continual learning setting, the goal is to learn a sequence of tasks using a single model. The specific challenge is to maintain the model generalization on past learned tasks while the model is continually updated, i.e., catastrophic forgetting. Since the tasks are observed in sequence, the training data for a past task is not accessible after learning that task. To tackle forgetting, we can model the internal distribution $e_{\boldsymbol{w}}(P_I(\boldsymbol{x}))$ using a parametric distribution and always learn new tasks such that the internal distribution remains unchanged and is shared across all tasks. We have used this strategy to mitigate catastrophic forgetting with both annotated and unannotated subsequent tasks (Rostami et al. 2020; Rostami 2021a).

Conclusions

We conclude that a successful approach to improve model generalization is consolidating the internally distribution.

References

Rostami, M. 2021a. Lifelong domain adaptation via consolidated internal distribution. In *Advances in Neural Information Processing Systems*, volume 34, 11172–11183.

Rostami, M. 2021b. *Transfer Learning Through Embedding Spaces*. CRC Press.

Rostami, M. 2022. Increasing model generalizability for unsupervised domain adaptation. In *CoLLAs*.

Rostami, M.; and Galstyan, A. 2023. *TOvercoming Concept Shift in Domain-Aware Settings through Consolidated Internal Distributions*. The 2023 AAAI Conference on Artificial Intelligence.

Rostami, M.; Kolouri, S.; Pilly, P.; and McClelland, J. 2020. Generative continual concept learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 5545–5552.

Stan, S.; and Rostami, M. 2021. Unsupervised model adaptation for continual semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2593–2601.

Stan, S.; and Rostami, M. 2022. Unsupervised Model Adaptation for Source-free Segmentation of Medical Images. In *BMVC*.