Improving Meta-learning for Low-resource Text Classification and Generation via Memory Imitation

Yingxiu Zhao[1], **Zhiliang Tian**[1,*], **Huaxiu Yao**[2], **Yinhe Zheng**[3], **Dongkyu Lee**[1]
Yiping Song[4], **Jian Sun**[3], **Nevin L. Zhang**[1]
[1]The Hong Kong University of Science and Technology, Hong Kong SAR, China
[2]Stanford University, [3]Alibaba Group
[4]Department of Computer Science, Peking University, Beijing, China
{yzhaocx,ztianac,dleear,lzhang}@cse.ust.hk,huaxiu@cs.stanford.edu
{zhengyinhe.zyh,jian.sun}@alibaba-inc.com,songyiping@pku.edu.cn,

Abstract

Building models of natural language processing (NLP) is challenging in low-resource scenarios where only limited data are available. Optimization-based meta-learning algorithms achieve promising results in low-resource scenarios by adapting a well-generalized model initialization to handle new tasks. Nonetheless, these approaches suffer from the *memorization overfitting* issue, where the model tends to memorize the meta-training tasks while ignoring support sets when adapting to new tasks. To address this issue, we propose a memory imitation meta-learning (MemIML) method that enhances the model's reliance on support sets for task adaptation. Specifically, we introduce a task-specific memory module to store support set information and construct an imitation module to force query sets to imitate the behaviors of some representative support-set samples stored in the memory. A theoretical analysis is provided to prove the effectiveness of our method, and empirical results also demonstrate that our method outperforms competitive baselines on both text classification and generation tasks.

1 Introduction

Building natural language processing (NLP) models in low-resource scenarios is of great importance in practical applications because labeled data are scarce. Meta-learning-based methods (Thrun and Pratt, 2012) have been commonly used in such scenarios owing to their fast adaptation ability. Notable successes have been achieved by meta-learning on low-resource NLP tasks, such as multi-domain sentiment classification (Yu et al., 2018; Geng et al., 2019) and personalized dialogue generation (Madotto et al., 2019; Song et al., 2020; Zheng et al., 2020).

Among different meta-learning approaches (Hospedales et al., 2021), optimization-based approaches have been widely used in various low-resource NLP scenarios (Madotto et al., 2019; Qian and Yu, 2019; Li et al., 2020; Mi et al., 2019) because they are model-agnostic and easily applicable. Concretely, optimization-based meta-learning algorithms aim to learn a well-generalized global model initialization θ that can quickly adapt to new tasks within a few steps of gradient updates. In the meta-training process, we first train θ on a *support set* (i.e., a few training samples of a new task i) to obtain task-specific parameters θ'_i. Then, we optimize θ based on the performance of θ'_i on a *query set* (i.e., another set of samples in task i).

Despite its effectiveness, optimization-based meta-learning algorithms usually suffer from the *memorization overfitting* issue [1] (Yin et al., 2020; Rajendran et al., 2020), where the learned model tends to solve all the meta-training tasks by memorization, rather than learning how to quickly adapt from one task to another via support sets. This is acceptable for training process, but results in poor generalization on the meta-testing sets, because the memorized model does not have knowledge of those tasks and does not know how to utilize the base learner to learn new tasks. Hence, this issue hinders the model from capturing task-specific characteristics from support sets and thus prevents the model from adapting to distinct new tasks (Rajendran et al., 2020). For instance, in personalized dialogue generation, this implies that the dialog model cannot adapt to individual users based on short conversation histories and hence fails to generate personalized responses.

Several works have been proposed to tackle the memorization overfitting issue for regression and image classification tasks. Some studies try to explicitly regularize the model parameters (Yin et al.,

*Corresponding author

[1]Memorization overfitting is different from the overfitting in conventional supervised learning (Hawkins, 2004). The latter means that the model overfits to the training tasks and fails to generalize to the testing tasks.

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Volume 1: Long Papers, pages 583 - 595