TransPrompt: Towards an Automatic Transferable Prompting Framework for Few-shot Text Classification

Chengyu Wang[1], Jianing Wang[2], Minghui Qiu[1], Jun Huang[1], Ming Gao[2]

[1] Alibaba Group, [2] School of Data Science and Engineering, East China Normal University

Key Contributions

- Based on continuous prompt embeddings, we propose *TransPrompt*, a transferable prompting framework for few-shot learning across similar tasks.
- In *TransPrompt*, we employ a multi-task meta-knowledge acquisition procedure to train a meta-learner that captures cross-task transferable knowledge. Two de-biasing techniques are further designed to make it more task-agnostic and unbiased towards any tasks. After that, the meta-learner can be adapted to target tasks with high accuracy.
- Extensive experiments show that *TransPrompt* outperforms single-task and cross-task strong baselines over multiple NLP tasks and datasets. *TransPrompt* also outperforms strong fine-tuning baselines when learning with full training sets.

Introduction

Background. Prompt-based approaches fine-tune BERT-style PLMs in a few-shot learning setting, which adapt PLMs into producing specific tokens corresponding to each class, instead of learning the prediction head. Despite the remarkable success, we notice that current prompt-based approaches may have a few limitations. For few-shot learning, the performance of downstream tasks is still constrained by the number of training instances. It would be highly desirable if the model can acquire the transferable knowledge from similar NLP tasks before it is adapted to specific tasks with few samples. A natural question arises: *how can we design a prompting framework for BERT-style models that captures transferable knowledge across similar NLP tasks to improve the performance of few-shot learning?*

Our Work. We present *TransPrompt*, a prompting framework that allows PLMs to capture *cross-task transferable knowledge* for few-shot text classification. *TransPrompt* firstly employs a *Multi-task Meta-knowledge Acquisition* (MMA) procedure to learn the transferable representations of prompt encoders and PLMs jointly across similar NLP tasks. To reduce *over-fitting* and make the underlying PLM *more task-agnostic* and *less unbiased* towards any specific tasks, we propose two *de-biasing* techniques, namely *prototype-based de-biasing* and *entropy-based de-biasing*. After MMA, *TransPrompt* takes the *Task-aware Model Specification* (TMS) step to be adapted to specific tasks.

TransPrompt: The Proposed Framework

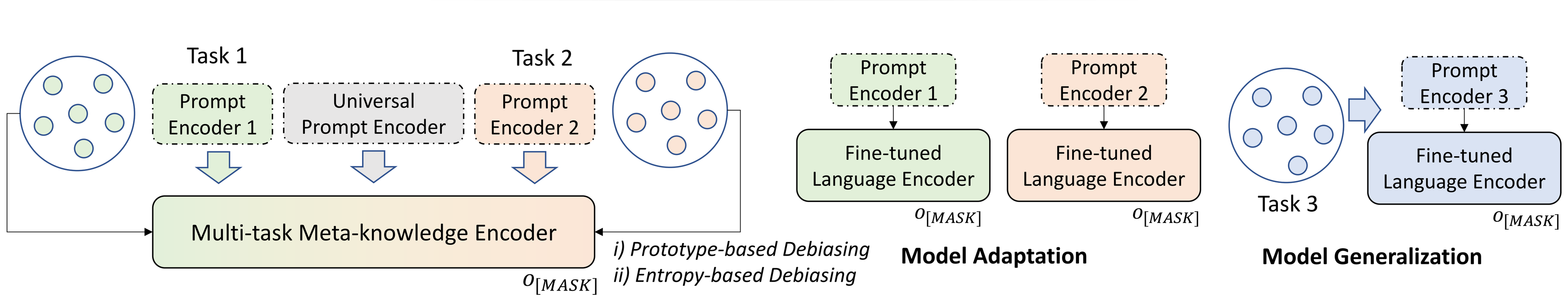

Figure 1:The high-level architecture of the *TransPrompt* framework. In the toy example, Task 1 and Task 2 are existing tasks, while Task 3 is a new task for the meta-learner to generalize.

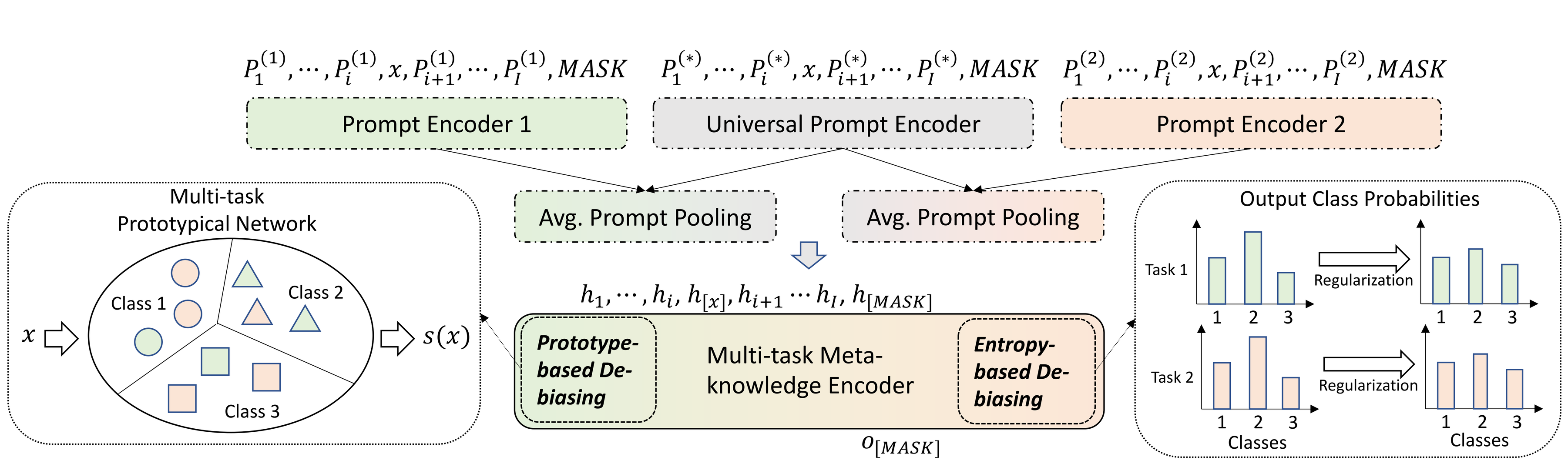

Figure 2:The model architecture of the meta-learner training process during MMA. For simplicity, we assume there are two tasks and three classes for few-shot text classification.

TransPrompt: The Proposed Framework

We assume that there are K training samples associated with each class $y \in \mathcal{Y}$ in each task $\mathcal{T}_m$. Hence, we have a training set $\mathcal{D}_m$ for each task $\mathcal{T}_m$, each containing $N \times K$ samples. The total number of training instances of M tasks is $N \times K \times M$. In *TransPrompt*, we train a *meta-learner* $\mathcal{F}_{meta}$ with parameters initialized from any PLMs, based on the M few-shot training sets $\mathcal{D}_1, \cdots, \mathcal{D}_M$. After that, $\mathcal{F}_{meta}$ is adapted to each task $\mathcal{T}_m$ based on its own training set $\mathcal{D}_m$. The task-specific model is denoted as $\mathcal{F}_m$. As $\mathcal{F}_{meta}$ is designed to digest the *transferable knowledge* across tasks, rather than simple multi-task learning, $\mathcal{F}_{meta}$ can also be adapted to previously unseen tasks. Due to the data privacy or computation efficiency issues, when the few-shot training set $\tilde{\mathcal{D}}$ of a similar task $\tilde{\mathcal{T}}$ is not available during the training process of $\mathcal{F}_{meta}$, we explore how *TransPrompt* can be used to generate an accurate model $\tilde{\mathcal{F}}$ based on $\mathcal{F}_{meta}$ and $\tilde{\mathcal{D}}$. In this case, $\mathcal{F}_{meta}$ does not have any knowledge of the new task $\tilde{\mathcal{T}}$ when it is trained during MMA.

Experiments

Few-shot Results. For few-shot learning, We assume that each class only has 16 training instances. The underlying PLM is the RoBERTa large model.

Method	Task: Sentiment Analysis			Task: NLI		Task: Paraphrase		Avg.
	SST-2	MR	CR	MNLI	SNLI	MRPC	QQP	
Single-task Baselines								
Fine-tuning	81.42	76.15	84.50	54.17	44.45	73.28	59.64	67.66
LM-BFF (man)	90.75	86.60	90.50	63.62	70.77	74.05	60.27	76.65
LM-BFF (auto)	91.62	87.25	91.80	64.25	71.21	74.23	60.59	77.28
P-tuning	91.85	86.60	91.75	62.41	70.28	66.42	60.57	75.70
Cross-task Baselines								
Fine-tuning (mtl)	83.37	79.30	84.75	41.32	48.14	53.12	59.31	64.19
Meta Fine-tuing	86.32	83.85	88.42	48.52	58.20	71.56	67.12	72.00
LM-BFF (mtl)*	91.97	87.45	90.70	69.09	75.90	50.00	67.40	76.07
P-tuning (mtl)*	93.12	87.75	91.35	68.83	74.24	70.83	69.99	79.44
TransPrompt	93.58	88.80	92.00	71.90	76.99	75.98	75.80	82.15

Table 1:The few-shot testing results of *TransPrompt* and baselines in terms of accuracy (%).

Full Training Set Results. We also evaluate our framework with full training sets available. The underlying PLM is the RoBERTa base model.

Method	Task: Sentiment Analysis			Task: NLI		Task: Paraphrase		Avg.
	SST-2	MR	CR	MNLI	SNLI	MRPC	QQP	
Single-task Baselines								
Fine-tuning	93.00	90.15	90.90	82.87	87.87	72.28	89.53	86.65
LM-BFF (man)	93.65	88.50	90.98	87.23	91.10	88.75	85.12	89.33
LM-BFF (auto)	93.81	88.75	91.25	87.01	91.51	88.97	83.12	89.20
P-tuning	93.69	90.10	90.25	87.17	91.67	88.97	90.87	90.38
Cross-task Baselines								
Fine-tuning (mtl)	94.72	90.65	91.05	87.10	91.80	69.85	90.20	87.91
Meta Fine-tuing	95.70	91.25	91.42	83.67	89.48	78.92	89.72	88.59
LM-BFF (mtl)*	95.41	90.45	91.50	86.76	88.25	69.36	90.32	87.43
P-tuning (mtl)*	95.30	90.40	90.08	86.97	91.48	68.87	90.59	87.67
TransPrompt	96.05	91.78	91.59	88.70	91.88	86.87	91.27	91.16

Table 2:Results of *TransPrompt* and baselines with full training sets in terms of accuracy (%).

Conclusion

In this work, we present the *TransPrompt* framework for few-shot learning across similar NLP tasks based on continuous prompt embeddings. Experimental results show that *TransPrompt* consistently outperforms strong baselines in both few-shot learning and standard fine-tuning settings. Additionally, we find that the meta-learner trained by *TransPrompt* can be adapted to previously unseen tasks easily. In the future, we will explore how *TransPrompt* is applied to other PLMs apart from BERT-style models and other NLP tasks.