NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Efficient Cross-Task Prompt Tuning for Few-Shot Conversational Emotion Recognition

Yige Xu[1,2], Zhiwei Zeng[1,2], Zhiqi Shen[2]
[1]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly
[2]School of Computer Science and Engineering
Nanyang Technological University, Singapore
yige002@e.ntu.edu.sg, {zhiwei.zeng,zqshen}@ntu.edu.sg

Outlines

- Background and Motivation
- Methodology
- Main Results
- Conclusion

Background and Motivation

- Task Description for Emotional Recognition in Conversation
  - Given a conversation history $c = [(u_1, s_1), (u_2, s_2), \ldots, (u_N, s_N)]$ with N utterances, where u_i and s_i is the content and the speaker information for the i-th utterance.
  - The target is to predict the emotional category $e = [e_1, e_2, \ldots, e_N]$ for each utterances.

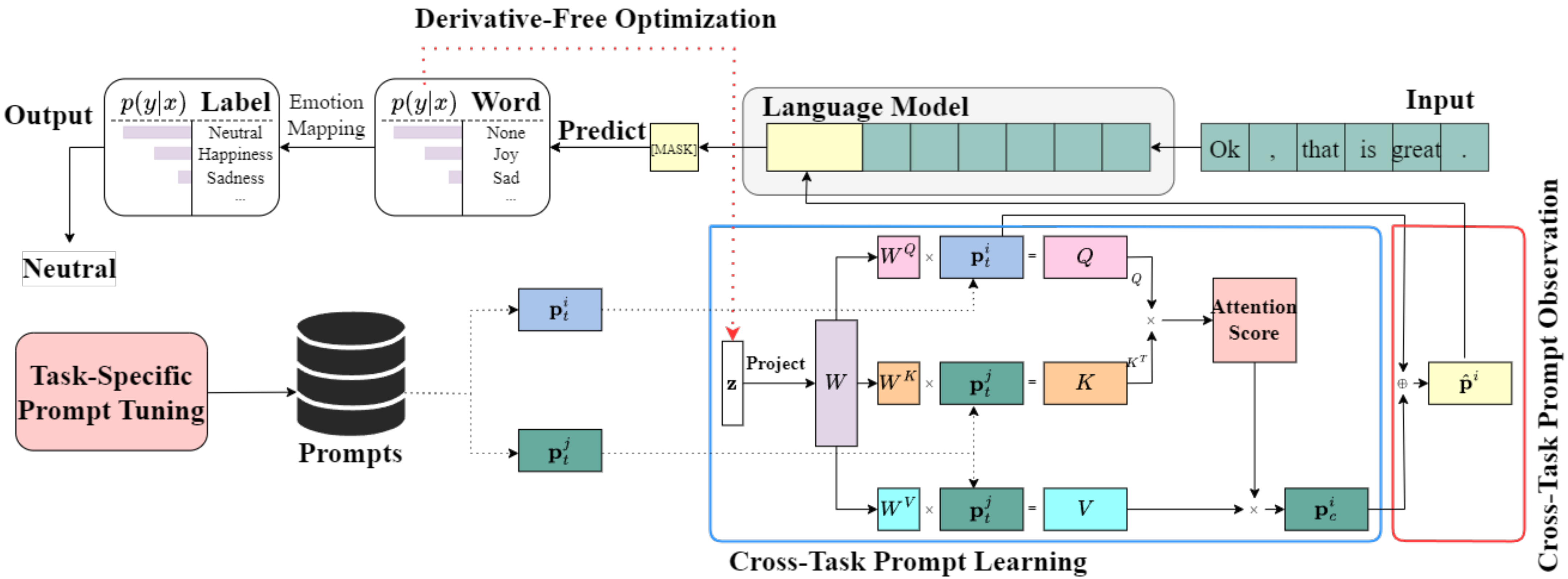
Background and Motivation

- Motivation
  - LMs are becoming larger and larger
  - Limitations in low-resource scenarios.

How to improve the training efficiency, include the **data-efficiency** and **parameter-efficiency**

Methodology

- Cross-Task Prompt Tuning



Methodology

- Cross-Task Prompt Tuning
  - Cross-task prompt learning
    - Learn from task-specific knowledge

$$P(x) = \mathrm{concat}[f(\mathbf{p}_c^i, \mathbf{p}_t^i); x]$$

$$\mathbf{h} = \sum_{j, j \neq i} \mathrm{MHA}(\mathbf{p}_t^i, \mathbf{p}_t^j),$$

$$\mathrm{MHA}(\mathbf{p}_t^i, \mathbf{p}_t^j) = \sum_{\mathrm{head}} \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V,$$

Here $\mathbf{p}_c^i$ is the learned cross-task knowledge and $\mathbf{p}_t^i$ is the task-specific knowledge.

Methodology

- Cross-Task Prompt Tuning
  - Cross-task prompt learning
    - Learn from emotional knowledge
      - Rewrite the verbalizer based on the assumption that some emotional labels refer to the same emotion though they have different textual representations.

$$
\begin{aligned}
\hat{\mathcal{V}} =&\{v \mid \forall v \in \mathcal{V}_i, i = 1, 2, \cdots, n\}, \\
h :& \hat{\mathcal{V}} \rightarrow \mathcal{V}, \\
v(\mathbf{h}) =& g\Big(p(\texttt{[MASK]} = v \mid \mathbf{h}) \mid v \in \mathcal{V}\Big),
\end{aligned}
$$

Methodology

- Cross-Task Prompt Tuning
  - Cross-task prompt observation
    - Some cross-task knowledge is useful
    - Some other cross-task knowledge is useless
    - Utilize a gate mechanism to observe the learned knowledge

$$\hat{\mathbf{p}}^i = f(\mathbf{p}_c^i, \mathbf{p}_t^i) = \mathbf{z}_i'' \otimes \mathbf{p}_t^i + (\mathbb{I} - \mathbf{z}_i'') \otimes \mathbf{p}_c^i,$$

Methodology

- Cross-Task Prompt Tuning
  - Derivate-free Optimization (DFO)
    - Covariance Matrix Adaptation Evolution Strategy (CMA-ES)
      - Simulates a population of potential solutions, treating each solution as an evolving population
      - Adapted by the covariant matrix
      - Optimize a vector z
    - Adapt to CTPT
      - Maps the vector z to the prompts and other learnable parameters

$$\mathbf{p}_t = \mathbf{A}\mathbf{z} + \mathbf{p}_0,$$

Main Results

- Cross-Task Prompt Tuning
  - Experiments

Dataset	Domain	# Emotions	# Conv.	# Utter.
EC (Chatterjee et al., 2019)	Tweet	4	30,160/2,755/5,509	90,480/8,265/16,527
DailyDialog (Li et al., 2017)	Daily Chat	7	11,118/1,000/1,000	87,170/8,069/7,740
MELD (Poria et al., 2019)	TV Show Scripts	7	1,038/114/280	9,989/1,109/2,610
EmoryNLP (Zahiri and Choi, 2018)	TV Show Scripts	7	659/89/79	7,551/954/984
IEMOCAP (Busso et al., 2008)	Daily Chat	6	100/20/31	4,758/1,000/1,622

Table 1: Statistics of five ERC datasets. $a/b/c$ indicates the number of examples in the training set, development set, and testing set, respectively.

Main Results

- Cross-Task Prompt Tuning
  - Experiments

	Model	EC	DailyDialog	MELD	EmoryNLP	IEMOCAP
Baselines	KET (Zhong et al., 2019)	0.1296	0.0909	0.0897	0.1312	0.1646
	TUCORE-GCN (Lee and Choi, 2021)	0.1918	0.2029	0.2596	0.1311	0.1527
	EmotionFlow (Song et al., 2022b)	0.4084	0.3749	0.2934	0.1465	0.1699
	SPCL (Song et al., 2022a)	0.4269	0.3699	0.2941	0.1499	0.1873
	TSPT	0.6274	0.4996	0.2521	0.1613	0.2877
Ours	TSPT + CTPL	0.6226	0.5193	0.2732	0.1724	0.2829
	CTPT (w/o. BP)	<u>0.6394</u>	<u>0.5571</u>	**0.3212**	<u>0.1902</u>	<u>0.3124</u>
	CTPT (w. BP)	**0.6405**	**0.5588**	<u>0.3128</u>	**0.2057**	**0.3182**

Table 2: Performance of different ERC datasets under the few-shot settings ($k = 16$). "TSPT" indicates task-specific prompt tuning, "CTPT" indicates cross-task prompt tuning. The result of EC and DailyDialog are micro-averaged F_1, and the result of other datasets are weighted macro-F_1. We **bolded** the best result and <u>underline</u> the second best.

Main Results

- Cross-Task Prompt Tuning
  - Experiments

Source Task \ Target Task	EC	DailyDialog	MELD	EmoryNLP	IEMOCAP
EC	\	**0.5119**	**0.2438**	0.0307	0.1684
DailyDialog	**0.5276**	\	0.2400	0.0308	0.2204
MELD	0.4579	0.4834	\	0.0245	0.2313
EmoryNLP	0.3642	0.1804	0.1315	\	**0.2658**
IEMOCAP	0.3870	0.2192	0.1104	**0.0599**	\

Table 4: Performance of zero-shot transfers. The task-specific prompt of the target task is excluded during the training stage. We **bolded** the best zero-shot transfer result for each target task.

Conclusion

- We strictly define the task of the few-shot setting for ERC.
- We propose a cross-task prompt tuning (CTPT) method to tackle this problem utilizing the cross-task knowledge.
- Experiments on ERC benchmarks show that CTPT can not only outperforms baseline models in the **few-shot setting** but also obtain a surprising result in the **zero-shot transfer**.
- CTPT is **data-efficient** and **parameter-efficient**.

Thank you