TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 31/38 pp1819–1837
DOI: 10.26599/TST.2024.9010196
Volume 31, Number 3, June 2026

Research Article

Multi-Affection Prompt Learning for Sentiment, Emotion, and Sarcasm Joint Detection in Conversations

Yazhou Zhang, Yang Yu, Xiangkai Wang*, Xiang Li*, Hui Liang, and Prayag Tiwari

Abstract: As a new trend in Natural Language Processing (NLP), prompt tuning has been explored to provide a reliable answer without requiring massive labeled samples and training learning in sentiment and emotion detection tasks. However, how to effectively encode the commonness and uniquesness across difference affections into prompts sets a limit to the potential of multi-affection joint detection. To fill this gap, we propose a multi-affection prompt (MAP) learning framework that takes both the commonness of multiple affections and the uniqueness of specific affection into consideration. More specifically, two different prompt encoders are first proposed to elaborate the multi-task shared prompt and the task-specific prompt, respectively. Second, a multi-task prompt interaction learning layer is proposed to capture the correlation between the multi-task and task-specific prompts. MAP adopts separate multi-task and task-specific prompts to learn different vectors for different affection tasks, thus mitigating the affection discrepancy of the [MASK] token in the masked language modeling task. Extensive experiments on two benchmark datasets show that our proposed method can significantly improve the multi-task generalization capability of PLMs, and yield better results than other state-of-the-art (SOTA) baselines, by the margin of 2.7% and 3.4%.

Key words: sentiment analysis; affection detection; prompt learning; multi-task learning

1 Introduction

The enormous popularity of instant messaging platforms (e.g., WhatsAPP, Tiktok, WeChat, etc.) offers a variety of opportunities for humans to exchange their opinions and attitudes, leading to tight interpersonal communications. Such communications are often multi-modal with textual (e.g., natural language), visual (e.g., facial gestures) and audio (e.g., tones) channels, and also multi-affective in that different types of affections, such as sentiment, emotion and sarcasm, are often expressed in a mixed manner. The interactions of different modalities and inter-correlations between different affection types bring opportunities as well as challenges for multi-modal affection detection, especially in a conversational context[1–3].

• Yazhou Zhang is with College of Intelligence and Computing, Tianjin University, Tianjin 300350, China. E-mail: yazzhang@polyu.edu.hk.
• Yang Yu is with School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China. E-mail: yuyang19980818@outlook.com.
• Xiangkai Wang is with Shandong Zhengyun Information Technology Limited Company, Jinan 250353, China. E-mail: wangxiangkai@sdas.org.
• Xiang Li is with Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China. E-mail: xiangli@sdas.org.
• Hui Liang is with Zhengzhou University of Light Industry, Zhengzhou 450002, China. E-mail: huiliangi@zzuli.edu.cn.
• Prayag Tiwari is with School of Information Technology, Halmstad University, Halmstad 30118, Sweden. E-mail: prayag.tiwari@hh.se.
* To whom correspondence should be addressed.
Manuscript received: 2024-04-07; revised: 2024-08-25; accepted: 2024-10-11

Recently, with the emergence of a range of large-scale pre-trained language models (PLMs), e.g., bidirectional encoder representations from transformers (BERT)[4], a lite BERT (ALBERT)[5], robustly optimized BERT (RoBERTa)[6], etc., many fine-tuning PLMs based multi-task learning approaches have been proposed. In view that such works often leverage the shared knowledge from correlative tasks to improve individual performance, and they have achieved remarkable success in multi-affection joint detection[7, 8]. However, as PLMs become larger and more complicated than ever, natural language processing research has steped into the era of large language models (LLMs), e.g., generative pre-trained transformer (GPT-3) (175 B parameters)[9], GPT-3.5 (1800 B parameters)*, text-to-text transfer transformer (T5) (540 B parameters) as compared to GPT-1 (0.11 B parameters) and BERT (0.15 B parameters). Such LLMs have been pre-trained with an huge number of training samples (over 200 B token), and thus possess outstanding prior knowledge.

Why prompting tuning instead of fine-tuning now? The continued training of LLMs, commonly known as fine-tuning, presents diminishing returns in terms of performance improvements relative to its high computational costs and resource demands. This has prompted researchers to explore more efficient methods to leverage the extensive pre-trained knowledge embedded within these models under constrained resources. Prompt tuning emerges as a compelling alternative, specifically designed to bridge the gap between the extensive pre-training phase of LLMs and their application to downstream tasks without the need for extensive parameter updates. Unlike fine-tuning, which often requires updating a significant portion of the model's parameters, prompt tuning involves crafting task-specific prompts that elicit the LLM to apply its pre-existing knowledge to new tasks effectively. This method not only significantly reduces the computational overhead but also mitigates the risk of model overfitting to smaller task-specific datasets.

Prompt learning has rapidly developed into a robust framework that supports diverse applications across various natural language processing (NLP) scenarios, such as sentiment classification and recommendation systems, achieving state-of-the-art (SOTA) results with minimal changes to the model architecture[10–12]. By integrating succinct, contextually relevant prompts, researchers have successfully demonstrated that LLMs can achieve accurate classification results by re-contextualizing their responses according to the designed prompts. The advantages of prompt tuning extend beyond resource efficiency; it also offers greater flexibility in model deployment and faster adaptability to new domains or tasks, making it an increasingly popular choice among researchers and practitioners aiming to maximize the utility of pre-trained models under operational constraints.

In addition, prompt based approaches often build a predefined template (e.g., "The movie involves many famous actors. Overall, it is [MASK]."), which could not model different distributions of the [MASK] token in different tasks. Hence, multi-task prompt learning comes into view, and prefers to construct a shared template to learn the correlation across different tasks[13]. Compared with the traditional approaches, multi-task prompt learning involves less modeling and parameter updating, as shown in Fig. 1. For instance, Asai et al.[14] treated source prompts as encodings of large-scale source tasks into a small number of parameters and trained an attention module to interpolate the source prompts and a target prompt for

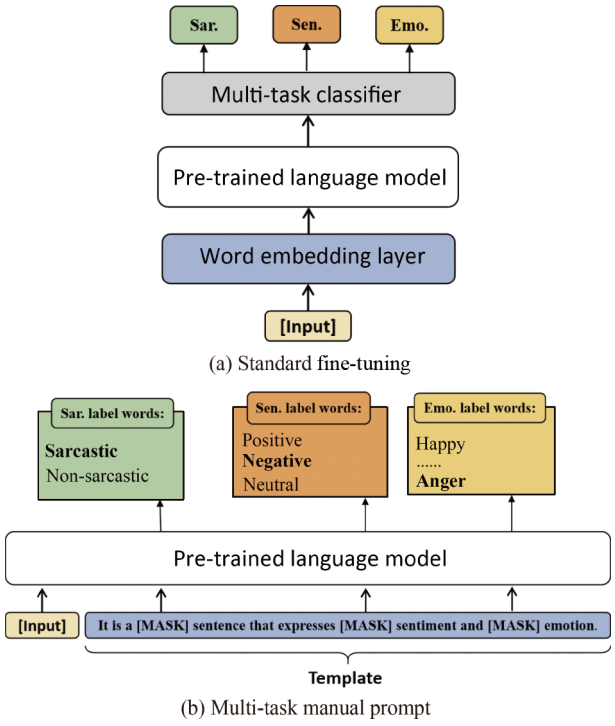

Fig. 1 Comparison between (a) fine-tuning-based method, and (b) prompt learning-based method.

*https://chat.openai.com/

every instance in the target task. Liu et al.[15] presented hierarchical prompt (HiPro) learning, a simple and effective method for jointly adapting a pre-trained vision language model (VLM) to multiple downstream tasks. Although they achieved success in multiple downstream tasks compared to many state-of-the-art parameter-efficient tuning approaches, the shared template (a.k.a. the shared knowledge) could only model the commonalities across tasks, while the uniqueness of a specific task (i.e., the unique knowledge implied by an individual task) had been neglected.

To fill this gap, in this paper, we propose a multi-affection prompt (MAP) learning framework for multi-modal sentiment, emotion and sarcasm joint detection in conversations. More specifically, we adopt a specialization-generalization training strategy. In the specialization stage, the task-specific prompts and a multi-task shared prompt are proposed. Each task-specific prompt is specifically trained on the data corresponding to its respective task, while the multi-task shared prompt indicates the global prompt tokens which are trained by minimizing the total loss returned by multiple masked labels corresponding to multiple tasks.

In the generalization stage, a multi-task prompt interaction learning layer is proposed to capture the correlation between the multi-task and task-specific prompts. MAP adopts separate multi-task and task-specific prompts to learn different vectors for different affection tasks, thus mitigating the affection discrepancy of the [MASK] token in the masked language modeling task.

The main contributions of this paper are summarized as follows:

- We make the first attempt to incorporate the commonness and uniqueness across correlative tasks into a multi-task prompt learning model.
- We adopt separated soft prompts to learn the task-specific prompts with different kinds of affection knowledge.
- A multi-task prompt interaction learning framework is proposed to capture the interaction between the multi-task and task-specific prompts.
- Extensive experiments on two benchmark datasets show the superiority of the proposed method over other SOTA baselines.

The rest of this paper is organized as follows. Section 2 briefly outlines the related work. In Section 3, we describe the proposed multi-task learning approach in detail. In Section 4, we report the empirical experiments and analyze the results. Section 5 concludes the paper and points out future research directions.

2 Related Work

In this section, we focus on investigating related models covering sentiment, emotion and sarcasm.

2.1 Sentiment, emotion, and sarcasm detection

Sentiment often refers to the expression of attitude or opinion towards objects, which is classified into three polarities: positive, negative, and neutral in natural language processing. Emotion reflects the individual psychological state, such as joy, happiness, fear, anger, disgust, surprise, etc. Sarcasm is a subtle form of figurative language where its literal sentiment is contrary to the true sentiment expressed in the utterance. Sentiment, emotion, and sarcasm are considered fundamental tasks in the domain of affective computing, which are tightly linked with each other. For example, emotions like sadness and anger are intrinsically associated with negative sentiments, and sarcasm often exists in the expression of intense sentiment. Given the strong correlation between these three tasks, many scholars have engaged in these fields and achieved remarkable works. We briefly introduce previous studies on multi-affection analysis.

Sentiment and emotion. Numerous sentiment and emotion analysis models have been proposed and evaluated on the above-mentioned datasets. For instance, Zhang et al.[16, 17] presented a density matrix based textual and visual representation model for multi-modal sentiment analysis. Majumder et al.[18] presented the use of different recurrent neural networks (RNNs) to keep track of the emotional states, termed DialogueRNN, and achieved state-of-the-art performance. Zhang and Li[19] used the variational autoencoder (VAE) to extract the latent emotional features for financial emotion recognition. Xiao et al.[20] proposed a multi-channel attentive framework for their new task, sentiment fusion and analysis. Now, there have been few models that focus on simultaneously detecting sentiment and emotion. Khare et al.[21] used a cross-modal Transformer based multi-task framework to learn multi-modal embeddings. Chen et al.[22] presented the EPE model: Emoji pre-trained feature enhanced sentiment analysis, and

collected 8 million tweets and selected 5 million tweets with pre-trained emoji with context using the BERT model.

Sarcasm. Utilizing multi-modal information for sarcasm detection has been explored recently. Cai et al.[23] designed a multi-modal sarcasm detection model in a hierarchical fusion manner. Liang et al.[24] constructed a stacked graphical structure and updated multi-modal representation using graph convolutional networks (GCNs) for sarcasm detection. Liu et al.[25] proposed a quantum inspired multi-task learning model, which used quantum measurement to model the relationship between sentiment and sarcasm. Majumder et al.[26] proposed to use tensor product to model the interaction between sentiment and sarcasm. Zhang et al.[27] used a retrospective reader in the detection process, which included two parallel modules: sketchy reading and intensive reading. Wen et al.[28] presented a dual incongruity perceiving (DIP) network consisting of two branches to mine the sarcastic information from factual and affective levels. For the factual aspect, they introduced a channel-wise reweighting strategy to obtain semantically discriminative embeddings, and leveraged gaussian distribution to model the uncertain correlation caused by the incongruity.

Different with the above-mentioned approaches, our model focuses on modeling the shared knowledge and unique knowledge from relative tasks.

2.2 Multi-task learning

Traditional multi-task learning. Multi-task learning aims to optimize several learning tasks simultaneously and leverage their shared information to help improve the generalization for each task. Recently, it is intuitively clear that multi-task learning has been applied to various NLP tasks and obtained outstanding results. For instance, Akhtar et al.[29] leveraged multi-modal and contextual information for identifying sentiment and emotion simultaneously. Zhang et al.[30] proposed a cross-modal cross-task attention-based multi-task learning model for sentiment and emotion joint analysis. Chanhan et al.[31] treated sentiment analysis and emotion recognition as auxiliary tasks to boost the performance of sarcasm detection by incorporating the knowledge of sentiment and emotion. Maity et al.[32] investigated the effect of sentiment, emotion, and sarcasm for cyberbullying detection in multi-modal code-mixed memes setting. Liu et al.[33] designed a quantum probability-driven multi-task network for sentiment and sarcasm classification. The above literature proves that the multi-task learning paradigm has shown awesome performance on various downstream tasks. However, it is expensive and time-consuming to fine-tune these PLMs with task-special data, where this approach limits the potential of PLMs.

Prompt tuning multi-task learning methods. Prompt tuning is an emerging paradigm for adapting pre-trained language models to downstream tasks, as a lightweight alternative to the full fine-tuning approach. A series of recent research works have explored methods based on prompt learning. The typical methods mainly adopt manually designed (discrete) prompts to adapt to various tasks. In view of it being cumbersome and laborious, numerous outstanding prompt-designed engineering works have been proposed and provide a platform for subsequent studies. Tang et al.[34] utilized contextualized continuous prompts to elicit knowledge from PLMs for natural language generation. Han et al.[35] proposed prompt-tuning based on logic rules to construct prompts with several sub-prompts for many-class text classification. Wu et al.[36] presented a novel adversarial training strategy and adopted soft prompts to learn the relevance of different domains for cross-domain sentiment analysis.

In the context of prompt learning, several approaches attempted to devise a shared prompt that is suitable for different tasks instead of fine-tuning multi-task models with specific data and parameters. Fu et al.[13] designed a polyglot prompt framework to model the unified semantic correlation among different languages and tasks. He et al.[37] presented a learnable hyper-prompts transformer model for supporting flexible information sharing among tasks. PromptonmyViT[38] explored multi-task prompt knowledge to capture information about relevant tasks and model the interaction with video tasks. Akyürek et al.[39] created a prompt-based multi-task learning framework for measuring social biases, which is presented in two formats: question-answer and premise-hypothesis.

Such methods have focused on modeling the common prompt template across different tasks and have achieved remarkable progress. However, it is evident that these prior works have overlooked the impact of task-specific prompts. In contrast, we initially attempt to model the commonalities across multi-task prompts as well as the uniqueness of task-specific prompts by employing a novel multi-task

prompt interactive learning framework.

2.3 Hard and soft prompting

(1) Pattern exploiting training

Pattern exploiting training (PET) is a relatively classic method in the domain of prompt learning. It involves modeling tasks as cloze-style problems, optimizing the selection of final output words to complete these tasks. While PET entails optimizing the parameters of the entire model, it requires significantly fewer data compared to traditional fine-tuning methods. This approach effectively harnesses the intrinsic capabilities of language models, making it especially advantageous in scenarios with limited available data. Now, this approach has been applied to numerous downstream NLP tasks. For example, Schick and Schütze[40] presented GENPET, a method based on PET, that enables finetuning of generative language models using both instructions and labeled examples. They proved that GENPET was a highly data-efficient method. Zeng et al.[41] proposed a multiple pattern exploiting training method (MPET) for solving the challenge of few-shot learning. Different from the original PET, MPET constructed multiple patterns to enhance the model's generalization capability. In addtion, they took the MPET as an auxiliary task, and jointly optimized classification and MPET.

(2) Soft prompt tuning

Soft prompt tuning (SPT) leverages learned prompts to enhance language models, allowing these prompts to be integrated into the model while also being storable separately. This flexibility enables each downstream task to require only the training of a small subset of parameters. Currently, there are three main forms of soft prompting: prompt tuning[42], prefix tuning[43], and P-tuning[44].

(a) Prompt tuning involves inserting trainable tokens, or soft prompts, directly into the input sequence before the task-specific data is processed by the model. The model then adjusts these tokens during training to better fit the task at hand. For example, Han et al.[35] proposed to encode the prior knowledge of a classification task into rules, then designed sub-prompts according to the rules, and finally combine the sub-prompts to handle text classification. (b) Prefix tuning adds a sequence of trainable vectors, or prefixes, to the beginning of the input sequence. These prefixes are optimized during training to guide the model's attention and processing mechanisms toward relevant features of the task. For example, Choi and Lee[45] presented a task-agnostic prompt tuning method for the program and language generation tasks, CodePrompt, that combines input-dependent prompt template and corpus-specific prefix tuning. (c) P-Tuning takes a slightly different approach by introducing trainable patterns that act as templates. These patterns are designed to capture and manipulate the model's latent knowledge to improve performance on specific tasks.

However, our proposed MAP approach is quite distinct from the above-mentioned prompt tuning approaches. While PET and soft prompt tuning methods integrate prompts at a shallow level, our MAP framework embeds a deeper, more nuanced interaction of prompts through its multi-task prompt interaction layer. This layer not only adapts prompts based on the task-specific demands but also dynamically adjusts these interactions according to the contextual information present in the input data.

3 Proposed Approach

In this section, we introduce our proposed multi-task prompt learning framework for multi-affection joint analysis in detail, which includes specialization and generalization stage.

3.1 Problem definition

Assume that there are N conversation samples in the dataset, the i-th conversation D_i contains R multi-modal utterances, where y_k means the sentiment (sa) and emotion (ec) and sarcasm (sd) labels of the k-th utterance, where $i \in [1,2,...,N]$, $k \in [1,2,...,R]$.

Now, we summarize our research problem as: Given one multi-speaker conversation including R multi-modal utterances, how to jointly detect their sentiments, emotions, and sarcasms without tuning great parameters? It could be written as

$$\zeta = \prod_{k} p\left(y_k^{\mathrm{sa}}, y_k^{\mathrm{ec}}, y_k^{\mathrm{sd}} | P_k^{\mathrm{sa}}, P_k^{\mathrm{ec}}, P_k^{\mathrm{sd}}, \Theta\right) \tag{1}$$

where y denotes the predicted label, P represents the prompt, and Θ denotes the parameter set.

3.2 Overview of framework

The architecture of our MAP framework is depicted in Fig. 2, encompassing two pivotal stages: the specialization stage and the generalization stage. During the Specialization Stage, distinct prompt encoders are engineered to develop both task-specific

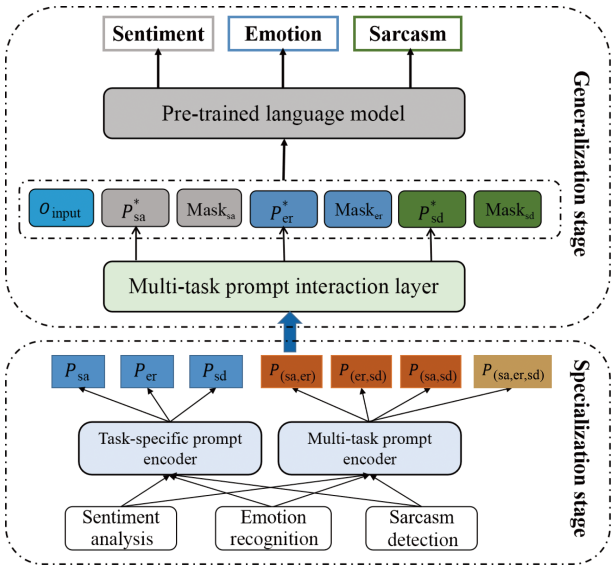

Fig. 2 Overall framework of our approach.

prompts (P_{sa} for sentiment analysis, P_{ec} for emotion recognition, and P_{sd} for sarcasm detection) and multi-task shared prompts ($P_{(\text{sa, ec})}$, $P_{(\text{ec, sd})}$, $P_{(\text{sa, sd})}$, and $P_{(\text{sa, ec, sd})}$). These encoders utilize advanced neural network architectures to refine prompt responsiveness to specific task demands while ensuring robustness through shared learning across multiple tasks.

In the Generalization Stage, we introduce the multi-task prompt interaction layer (MPIL), a sophisticated neural network layer designed to fuse and reconcile the diverse prompts derived from the earlier stage. The MPIL, leveraging state-of-the-art hierarchical attention mechanisms within its gating networks, dynamically adjusts the influence of individual task-specific and shared prompts based on the contextual relevance provided by the input data. This process allows for the nuanced integration of learned features, enhancing the model's ability to handle complex language processing tasks. Each gating network within the MPIL meticulously processes the associative patterns between different prompts, yielding three optimized prompts (P^*_{sa}, P^*_{ec}, and P^*_{sd}) that encapsulate refined task-specific insights.

The final output from the MPIL, consisting of these enhanced prompts, is then seamlessly integrated into a PLM or an LLM. This integration facilitates a sophisticated prompt-based interaction with the underlying language model, enabling precise and contextually aware predictions across the specified tasks of sentiment analysis, emotion recognition, and sarcasm detection.

3.3 Prompt generation

Task-specific prompt encoder. For critical NLP tasks such as sentiment analysis, emotion recognition, and sarcasm detection, we meticulously design task-specific prompt tokens to fine-tune the model's responsiveness to specific linguistic features. Illustrated in Fig. 3a, the process begins by converting the input text, containing [MASK] tokens representing potential target words or phrases, into word embedding vectors through a specialized embedding layer. This conversion is foundational, as it transforms raw text into a format that neural networks can process.

We create a complex template using several trainable prompt elements (labeled p_0 to p_m). These elements start with random settings but are designed to be highly flexible through training. They are important because they are directly trained to improve the model's understanding of specific task details. At the core of the task-specific prompt encoder is a bidirectional long short-term memory network (BiLSTM). This network optimizes the prompt elements in a continuous space, capturing both forward and backward connections in text sequences. This ensures that each element can adjust to the context of information flow, which is key for understanding complex language structures and meanings.

A two-layer neural network (MLP) works alongside the BiLSTM to encourage distinct learning. This network is crucial for refining the prompt elements so they not only contribute to the language model's smooth output but also improve its accuracy in specific tasks. By promoting distinctness, the MLP helps shape the prompt elements into clearer signals that can guide the language model more effectively. The settings of the prompt encoder—including those of the BiLSTM and MLP—are constantly updated through feedback based on how well the model performs in specific tasks. This ensures that the prompt elements improve in response to new information and challenges, helping to integrate these elements effectively into the language model's framework.

Let $\mathcal{M}$ refers to pre-trained language model and given an input x_{input} and the label y, with the help of the downstream loss function $\mathcal{L}$, we can optimize the prompt $h_j (0 \leqslant j \leqslant m)$ by the parameters θ:

$$\theta = \text{argmin}\mathcal{L}\left(\mathcal{M}\left(x_{\text{input}}, y\right)\right) \tag{2}$$

The encoder is optimized to obtain the continuous

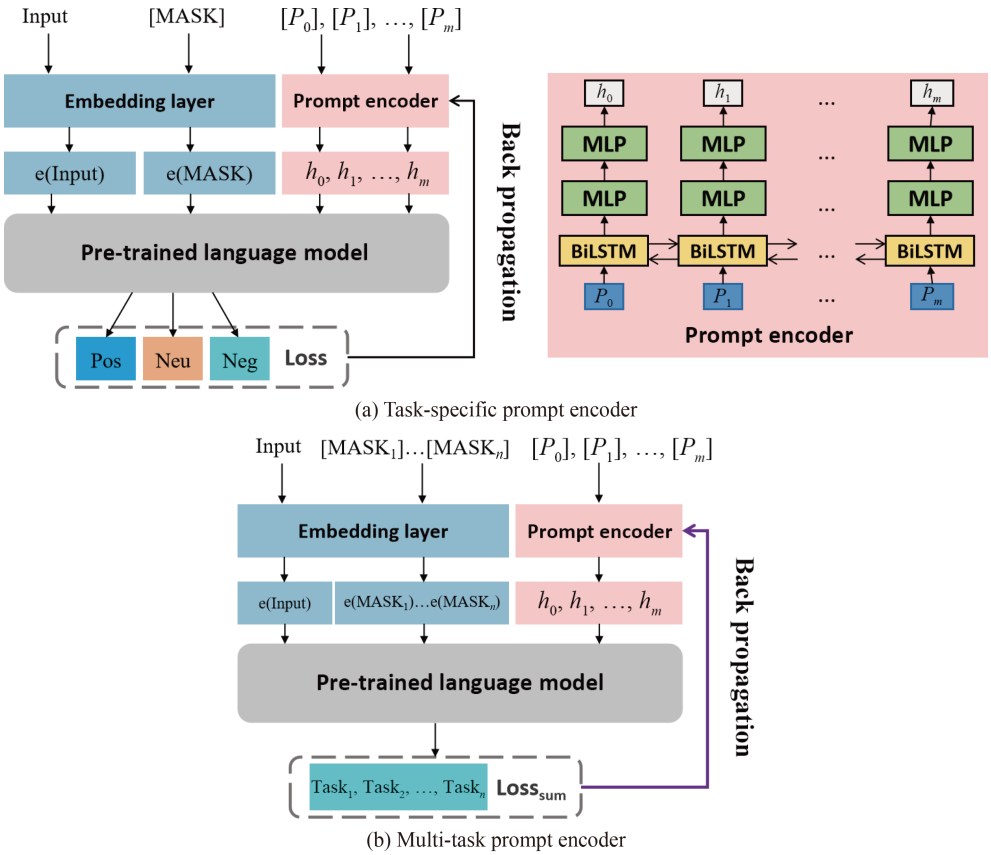

Fig. 3 Prompt generator framework. (a) illustrates the prompt tokens training strategy, taking the sentiment analysis task as an example. (b) shows the method of multi-task shared prompt generation by calculating the total loss of multiple masked labels.

prompt vectors:

$$P_i = \mathrm{MLP}_i(\mathrm{BiLSTM}_i(p_0, p_1, \cdots, p_m)) \tag{3}$$

where $i \in \{\mathrm{sa}, \mathrm{ec}, \mathrm{sd}\}$ and m is the length of prompt tokens.

Multi-task prompt encoder. As shown in Fig. 3b, for multi-task shared prompt training, we insert n [MASK] tokens corresponding to n tasks accompanied by input. The prompt tokens can be optimized through backpropagation and the $\mathcal{L}_{\mathrm{sum}}$ can be calculated as:

$$\mathcal{L}_{\mathrm{sum}} = \sum \mathcal{L}_i\left(\mathcal{M}\left(x_{\mathrm{input}}, y_i\right)\right) \tag{4}$$

where $i \in \{\mathrm{sa}, \mathrm{ec}, \mathrm{sd}\}$ and y_i indicates the label corresponding to i task. Finally, we can train the embedding for the multi-task shared prompt by minimizing the $\mathcal{L}_{\mathrm{sum}}$. By Eq. (3), we can train the multi-task prompts: $P_{(\mathrm{sa,ec})}$, $P_{(\mathrm{ec,\,sd})}$, $P_{(\mathrm{sa,\,sd})}$, and $P_{(\mathrm{sa,\,ec,\,sd})}$.

3.4 Multi-task prompt interaction layer

The MPIL framework, as shown in Fig. 4, plays a pivotal role in processing both task-specific and multi-task prompts, initially developed during the specialization stage. This framework is designed to meticulously balance the commonness and uniqueness of these prompts, optimizing the model's performance across varied tasks. It begins by applying a linear transformation to each input prompt, adjusting its dimensional attributes to prepare for more refined processing. This transformation is crucial as it standardizes the input for the next phase, which involves a self-attention mechanism. This mechanism, represented as $\mathrm{Gating} = \mathrm{LinearTrans}(\mathrm{Self-Attention}(*))$, enables the network to selectively concentrate on pertinent parts of the prompt data, depending on the

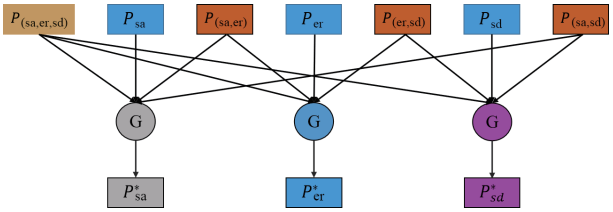

Fig. 4 Multi-task prompt interaction module.

task-specific requirements and current input context.

Following this initial processing, the MPIL generates weight scores for each task—α_{sa}, α_{er}, and α_{sd}—corresponding to sentiment analysis, emotion recognition, and sarcasm detection, respectively. These scores quantify the relevance of each prompt to the task, influencing the model's decision-making process, as detailed in Eq. (5):

$$\begin{aligned}
\alpha_{\text{sa}} &= \text{Gating}\left(W_{\text{sa}}\left(P_{\text{sa}},\ P_{(\text{sa, er})},\ P_{(\text{sa, sd})},\ P_{(\text{sa, er, sd})}\right)+b_{\text{sa}}\right),\\
\alpha_{\text{er}} &= \text{Gating}\left(W_{\text{er}}\left(P_{\text{er}},\ P_{(\text{sa, er})},\ P_{(\text{er, sd})},\ P_{(\text{sa, er, sd})}\right)+b_{\text{er}}\right),\\
\alpha_{\text{sd}} &= \text{Gating}\left(W_{\text{sd}}\left(P_{\text{sd}},\ P_{(\text{sa, sd})},\ P_{(\text{er, sd})},\ P_{(\text{sa, er, sd})}\right)+b_{\text{sd}}\right)
\end{aligned} \tag{5}$$

where W_{sa}, W_{er}, W_{sd} are weights and b_{sa}, b_{er}, b_{sd} are weighted biases.

To learn the uniqueness of task-specific prompts, the final prompts corresponding to tasks can be calculated by summing the weights of input prompts with source task-specific prompts, which are formulated as

$$\begin{aligned}
P^*_{\text{sa}} &= (1+\alpha^1_{\text{sa}})P_{\text{sa}} + \alpha^2_{\text{sa}}P_{(\text{sa, er})} + \alpha^3_{\text{sa}}P_{(\text{sa, sd})} + \alpha^4_{\text{sa}}P_{(\text{sa, er, sd})},\\
P^*_{\text{er}} &= (1+\alpha^1_{\text{er}})P_{\text{sa}} + \alpha^2_{\text{er}}P_{(\text{sa, er})} + \alpha^3_{\text{er}}P_{(\text{er, sd})} + \alpha^4_{\text{er}}P_{(\text{sa, er, sd})},\\
P^*_{\text{sd}} &= (1+\alpha^1_{\text{sd}})P_{\text{sa}} + \alpha^2_{\text{sd}}P_{(\text{sa, sd})} + \alpha^3_{\text{sd}}P_{(\text{er, sd})} + \alpha^4_{\text{sd}}P_{(\text{sa, er, sd})}
\end{aligned} \tag{6}$$

where $\alpha_{\text{sa}} = \left(\alpha^1_{\text{sa}}, \alpha^2_{\text{sa}}, \alpha^3_{\text{sa}}, \alpha^4_{\text{sa}}\right)^{\text{T}}$, $\alpha_{\text{er}} = \left(\alpha^1_{\text{er}}, \alpha^2_{\text{er}}, \alpha^3_{\text{er}}, \alpha^4_{\text{er}}\right)^{\text{T}}$, $\alpha_{\text{sd}} = \left(\alpha^1_{\text{sd}}, \alpha^2_{\text{sd}}, \alpha^3_{\text{sd}}, \alpha^4_{\text{sd}}\right)^{\text{T}}$. They are 4-dimensional vectors.

3.5 Composing prompts for tasks

Given an original input O_{input}, and three target tasks: T_{sa}, T_{er}, T_{sd} represent sentiment analysis, emotion recognition and sarcasm detection task, respectively. With the help of the prompt template P_T that consists of sub-prompts of multiple tasks, each task is redefined as predicting the word at the [MASK] position. The final template can be formulated as follows:

$$P_T = P^*_{\text{sa}}[\text{MASK}_{\text{sa}}],\ P^*_{\text{er}}[\text{MASK}_{\text{er}}],\ P^*_{\text{sd}}[\text{MASK}_{\text{sd}}] \tag{7}$$

where P_{sa}, P_{er}, P_{sd} represent the prompt of sentiment, emotion and sarcasm task respectively, and MASK_{sa}, MASK_{er}, MASK_{sd} represent the predicted label word of sentiment, emotion and sarcasm task respectively. Then P_T is fed into the pre-trained language model M, each [MASK] is predicted by maximizing the score of the language model, and the result is mapped to the label of the corresponding task. Considering the template contain multiple masked tokens, we adopt the following formula for making predictions:

$$p\left(y_i \mid O_{\text{input}},\ P_T\right) = p\left([\text{MASK}]_i = v_i \mid O_{\text{input}},\ P_T\right) \tag{8}$$

Where $i \in \{\text{sa, er, sd}\}$, y_i represents the predicted label corresponding to each task, and V_{y_i} is a subset of the whole vocabulary of i task.

4 Experiment

We conduct experiments to assess the efficacy of our method, aiming to address the following questions:

Q1: Does our model enhance the generalization capabilities of multi-task learning compared to traditional frameworks?

Q2: Is the multi-task shared prompt template (P_T) capable of effectively capturing the commonness and uniqueness among prompts from different tasks?

Q3: How effectively does our model assimilate knowledge from PLMs/LLMs?

4.1 Datasets and implementation details

Datasets. We construct our experiments on two benchmark datasets, including Memotion[46] and MUStARD[31]. The statistics of each dataset are listed in Table 1. We adopt the Micro-F1 (Mi-F1) as an evaluation metric in our experiments. We also introduce a balanced accuracy (Acc) metric for a few-shot study.

Implementation details. We use BERT[4] as the backbone pre-trained language model. In addition, we also use GPT-3, Flan-T5, and GPT-3.5 as the backbone large language models. We use the [unused] tokens as the trainable prompt tokens, and the length of the initialization prompt tokens is set to 10. We optimize our model with Adam using the learning rate of e^{-5} and fine-tune our model for 10 epochs. The batch size is 64, and the weight decay is set to e^{-2}.

4.2 Baselines

The state-of-the-art text sentiment analysis baselines are as follows:

Multi-head attention BiLSTM (MHA-BiLSTM): we apply a multi-head attention-based bidirectional long-short memory on the target utterance to extract the textual features respectively. Its advantages lie in capturing contextual information effectively through bidirectional long-short-term memory networks.

BERT&ALBERT: It uses BERT or ALBERT to produce the textual utterance vectors and feeds them into SVM for sarcasm, sentiment and emotion analysis. To explore the role of the context, we also represent the

Table 1 Dataset statistics.

Dataset	Task	Class	Size
Memotion	Sarcasm	Sarcastic	5448
		Non-sarcastic	1554
	Sentiment	Positive	631
		Negative	4160
		Neutral	2201
	Emotion (Humor)	Very funny	2188
		Not funny	1692
		Funny	2370
		Humor	742
MUStARD	Sarcasm	Sarcastic	345
		Non-sarcastic	345
	Sentiment	Positive	210
		Negative	391
		Neutral	89
	Emotion	Anger	97
		Excited	18
		Fear	14
		Sad	121
		Surprised	29
		Frustrated	57
		Happy	143
		Neutral	198
		Digust	39

context using BERT and merge them with the target utterance. Its advantages lie in possessing strong representational capabilities.

RCNN-RoBERTa[47]: It utilizes pretrained RoBERTa vectors to represent the utterance and uses an recurrent convolutional neural network (RCNN) to obtain its contextual representation. The final classification is performed by the softmax layer. Its advantages lie in significantly enhancing emotion and sarcasm detection accuracy by combining RNN's ability to learn contextual information with RoBERTa's deep pre-training advantages.

Multi-task graph attention network (MT-GAT): It builds three interactive undirected graphs over textual conversations and applies three standard GATs to update the utterance representations (which are pre-trained by BERT) interactively for sentiment, emotion and sarcasm classification. Its advantages lie in effectively differentiating and processing feature representations for different tasks using interactive graph attention networks.

RETripletG[48]: It learns emotion enriched embeddings by updating the vectors of emotion bearing words like fun, offense, angry, etc. It performs better than other pre-trained embeddings on the downstream tasks combining different classifiers, and has achieved SOTA results. Its advantages lie in enhancing word representation with rich emotional knowledge.

The state-of-the-art multi-task learning baselines are as follows:

UPB multi-task learning (UPB-MTL)[49]: It uses ALBERT to represent the textual utterance and uses visual geometry group net (VGG-16) to represent the accompanying image. Its advantages lie in demonstrating an effective use of shared knowledge from related tasks.

FAT-MTL[50]: It designs a genetic program tree, and applies a three layers neural network and fuzzy logic to predict the inter-task covariance matrix. We revise its code to make it be suitable for our three affection detection tasks. Its advantages lie in combining genetic programming and fuzzy logic to learn inter-task covariance.

A-MTL[31]: It designs two attention mechanisms, e.g., intrasegment and intersegment, to learn the relation between different segments and the relation within the same segment. Its advantages lie in effectively integrating intra-modal and inter-modal information.

Two state-of-the-art prompt learning approaches are:

PTR[35]: It proposes prompt tuning with rules (PTR) for many-class text classification and apply logic rules to construct prompts with several sub-prompts. It is a promising approach to take advantage of both human prior knowledge and PLMs. Its advantages lie in maintaining high performance with simplified prompt adjustments and reduced parameter updates.

ATTEMPT[51]: It is highly parameter-efficient (e.g., updates 2300 times fewer parameters than full fine-tuning) while achieving high task performance using knowledge from high-resource tasks. It has the similar advantages against PTR.

Two state-of-the-art LLMs are:

Flan-T5: It is an open-source, sequence-to-sequence, large language model that can reframe various tasks into a text-to-text format, such as translation, linguistic acceptability, sentence similarity, and document summarization. It can be used to generate text based on a prompt or input.

GPT-3/3.5: Developed by OpenAI, GPT-3.5 represents an intermediate iteration between GPT-3 and GPT-4, enhancing the capabilities of its

predecessor with improved training techniques and a larger dataset. This model excels in understanding and generating natural language and code, making it highly effective for a variety of applications including conversational agents, content creation, and programming assistance. GPT-3.5 is widely recognized as one of the most advanced large language models available, showing remarkable versatility and precision in task execution.

4.3 Main results and analysis

Table 2 illustrates the performance of our proposed MAP learning framework alongside various baselines on the Memotion and MUStARD datasets. We observe that MHA-BiLSTM, which employs a bidirectional LSTM combined with multi-head attention, yields the lowest scores among all models in sarcasm detection and sentiment analysis on the MUStARD dataset, with sentiment and sarcasm scores of 34.02% and 43.59%, respectively. Its primary limitation is its relatively simplistic approach to text representation, which lacks the deep human language understanding required for these complex tasks. The performance improvement of BERT over MHA-BiLSTM is substantial, with BERT scoring up to 10.08% higher in sentiment analysis on the MUStARD dataset, achieving a sentiment score of 44.10%. This enhancement is attributed to BERT's robust representation capabilities, derived from its transformer architecture, which is superior to traditional RNNs and GloVe embeddings.

ALBERT, a streamlined variant of BERT, demonstrates comparable performance across tasks, suggesting that its optimizations retain much of BERT's representational power while reducing resource consumption. However, its improvements are not statistically significant compared to BERT, with only a slight sentiment score increase on the MUStARD dataset from 44.10% to 46.59%. RCNN-RoBERTa shows significant advancements, especially in emotion recognition on the Memotion dataset, where it improves by approximately 1.55% over BERT, increasing the emotion recognition score from 45.28%

Table 2 Results of different baselines on two datasets. The best scores are marked in bold. "–" indicates not available in Memotion dataset. The data marked in red indicate the improvement of our model compared with other state-of-the-art baselines (Flan-T5 and GPT-3.5).

(%)

Paradigm	Baseline	Memotion			MUStARD		
		Sentiment	Emotion	Sarcasm	Sentiment	Emotion	Sarcasm
Pre-train+Finetune	**MHA-BiLSTM**	42.06	22.61	47.02	34.02	23.69	43.59
	BERT	34.59	39.90	49.15	44.10	30.67	64.68
	ALBERT	33.21	44.46	50.75	46.59	32.65	63.77
	RCNN-RoBERTa	36.67	47.45	51.72	46.48	33.14	65.16
	MT-GAT	34.77	45.28	51.38	47.22	33.20	66.13
	RETripletG+AttnNet	-	-	47.92	-	-	-
	RETripletG+SVM	-	-	56.67	-	-	-
	A-MTL	34.27	40.44	59.85	49.47	33.10	72.57
	UPB-MTL	34.49	45.18	51.83	47.70	33.60	65.41
	FAT-MTL	32.54	37.49	43.89	42.17	31.82	64.07
Prompt+Finetune	**PTR**	38.45	49.51	55.33	50.41	36.71	68.61
	ATTEMPT	37.41	48.44	53.41	51.23	35.78	68.12
	Flan-T5	45.11	55.01	65.15	58.81	43.29	74.54
	GPT-3.5	47.57	56.78	64.77	61.81	46.83	76.67
	Ours (+BERT)	40.56	51.24	57.43	53.36	37.74	69.41
	+context	–	–	–	54.61	38.15	70.54
	Ours (+Flan-T5)	**48.25** (+3.14)	**57.77** (+2.76)	**67.83** (+2.68)	**61.36** (+2.55)	**46.24** (+2.95)	**76.81** (+2.41)
	+context	–	–	–	60.14	46.09	76.35
	Ours (+GPT-3.5)	**50.16** (+2.59)	**58.22** (+1.44)	**67.67** (+2.90)	62.74	49.33	78.26
	+context	–	–	–	**64.73** (+2.92)	**52.18** (+2.24)	**80.05** (+3.37)

(MT-GAT as a close comparison) to 47.45%. The combination of RoBERTa's enhanced pre-training and RCNN's ability to model contextual relationships allows for superior long-range dependency modeling, crucial for capturing subtler emotional cues. Nonetheless, its design limits its applicability to text-only scenarios, neglecting multi-modal information.

MT-GAT, which integrates Graph Attention Networks with BERT embeddings, showcases superior performance over both BERT and ALBERT by learning interactive, task-specific representations. This model shows a notable improvement in sarcasm detection on the MUStARD dataset, scoring 66.13% compared to BERT's 64.68%, reflecting a 1.45% increase. The ability of MT-GAT to adaptively learn from the relational structure of data offers significant advantages over traditional models that process text in isolation, particularly in understanding the complex dynamics of conversational sarcasm. For RETripletG+ SVM, it is noted for its high performance in sarcasm detection on Memotion with a score of 56.67%. This model leverages enriched emotional knowledge embedded in word representations, making it highly effective for tasks requiring deep emotional insights. UPB-MTL, utilizing both textual and visual cues via ALBERT and VGG-16, demonstrates varied performance across tasks. It marginally outperforms RCNN-RoBERTa in sarcasm detection with a score of 65.41% compared to 65.16%. This slight improvement underscores UPB-MTL's effective leveraging of shared knowledge across tasks, though it shows limited advancement in sentiment analysis and emotion recognition where deeper contextual understanding might be required. FAT-MTL shows underwhelming performance, particularly due to its non-reliance on pre-trained language models and a design not tailored for the nuanced demands of affection classification tasks. It scores 64.07% in sarcasm on MUStARD, which is lower compared to more specialized models like A-MTL and MT-GAT. A-MTL excels in sarcasm detection, achieving the highest score among all models with 72.57% on MUStARD. This performance illustrates the strength of its attention mechanisms that effectively harness both intra-modal and inter-modal information, making it the best-performing model for detecting sarcasm.

Turning to prompt-based fine-tuning methods, the PTR methodology achieves notable results across the tasks on the Memotion and MUStARD datasets. On MUStARD, PTR achieves a sarcasm detection score of 68.61%, which is a significant improvement compared to some baseline models like BERT, which scores 64.68%. Similarly, in sentiment analysis, PTR scores 50.41%, which outperforms the baseline model MHA-BiLSTM that scores 34.02%. PTR's strength lies in its ability to integrate human-like reasoning within the tuning process by incorporating rules that guide the prompt construction. Despite its strengths, PTR's reliance on predefined rules can limit its flexibility. In contrast, ATTEMPT shows a strong performance on the MUStARD dataset, scoring 68.12% in sarcasm detection. This score is closely competitive to PTR's 68.61%, demonstrating its efficacy in handling nuanced tasks like sarcasm detection. In sentiment analysis, ATTEMPT records a score of 51.23%, which again closely matches PTR and shows a significant improvement over some pre-trained models like ALBERT. ATTEMPT's main advantage is its highly parameter-efficient design, which updates far fewer parameters compared to full model fine-tuning, making it both efficient and effective. However, it relies heavily on the ability of the attention mechanism to capture relevant features, which might not always align with the specific nuances of every task.

The culmination of our experimental evaluations robustly underscores the efficacy of the MAP learning framework. This advanced framework, especially when augmented with LLMs such as Flan-T5 and GPT-3.5, exhibits superior performance across diverse linguistic metrics. These LLMs, with their extensive model parameters and profound capabilities in human language comprehension, markedly surpass traditional approaches and our previous implementations that utilized BERT. Specifically, integrating our MAP framework with Flan-T5 not only harmonizes with but also significantly enhances the intrinsic capabilities of the model. This synergy results in a notable performance uplift, with Flan-T5 boosting sarcasm detection scores on the MUStARD dataset by over 3%. This improvement is not merely quantitative but also qualitative, markedly enhancing model responsiveness and accuracy in detecting nuanced linguistic elements.

The strategic inclusion of multi-task prompts within our MAP framework has proven pivotal in its success. These context-aware prompts enable the model to effectively interpret nuanced language subtleties, which are often overlooked by less advanced systems. This capability is particularly crucial in complex tasks such as sarcasm detection, where understanding

contextual cues is paramount. The enhanced contextual awareness afforded by our framework significantly aids in discerning the underlying tone and intent of expressions, thereby substantially reducing false positives and negatives in both sarcasm detection and sentiment analysis. For detailed experimental setups and further explanations, please refer to Section 4.7.

4.4 Secondary results on CMMA dataset

Given the small size of the MUStARD dataset, we aim to add a new experimental dataset to further validate our approach. It is crucial to select benchmark datasets that include annotations for sentiment, emotion, and sarcasm. The only dataset that meets these criteria is the Chinese multi-modal multi-affection conversation (CMMA) dataset[52]†. Additionally, the CMMA dataset is large-scale, containing 3000 multi-party conversations and 21 795 multi-modal utterances collected from various TV series. Therefore, we have selected CMMA as the new experimental benchmark. More statistics can be found in Table 4.

Based on the experimental results presented in Table 3, it is evident that our proposed MAP method significantly outperforms existing state-of-the-art baselines across multiple tasks, particularly when integrated with Flan-T5 and GPT-3.5. Specifically, our model achieves a notable improvement in sentiment analysis, emotion recognition, and sarcasm detection, as highlighted by the red-marked data. For instance, when combined with Flan-T5, our model improves sentiment analysis by 4.23%, emotion recognition by 2.82%, and sarcasm detection by 2.91%. The results are further enhanced with the inclusion of context, which drives performance to 65.11% in sentiment analysis, 63.56% in emotion recognition, and 65.47% in sarcasm detection, marking the best scores across all tasks. These substantial gains underscore the effectiveness of our framework, particularly in capturing the nuanced interrelations between tasks and efficiently adapting the underlying LLMs for complex multi-modal sentiment, emotion, and sarcasm detection tasks. This finding is consistent with the analysis of our main experimental results.

4.5 STP vs. MTP

In this section, we detail the outcomes of our empirical study comparing single-task prompt-tuning (STP) and multi-task prompt-tuning (MTP) across three datasets. The principle of multi-task learning, which is predicated on the efficacy of shared representations, forms the backbone of our methodology. For this purpose, we specifically designed and utilized prompts

†https://github.com/annoymity2022/Chinese-Dataset

Table 3 Results of different baselines on CMMA dataset. The best scores are marked in bold. The data marked in red indicate the improvement of our model compared with other state-of-the-art baselines (Flan-T5 and GPT-3.5).

(%)

Paradigm	Baseline	CMMA		
		Sentiment	Emotion	Sarcasm
Pre-train+ Finetune	**MHA-BiLSTM**	51.45	45.23	52.60
	BERT	54.89	56.18	53.61
	ALBERT	54.74	56.44	53.65
	RCNN-RoBERTa	55.47	56.89	54.28
	MT-GAT	55.12	56.67	54.44
	RETripletG+AttnNet	54.38	56.44	55.16
	RETripletG+SVM	55.67	56.80	56.23
	A-MTL	57.72	58.16	57.33
	UPB-MTL	55.46	56.29	53.86
	FAT-MTL	54.88	56.74	54.26
Prompt+ Finetune	**PTR**	57.04	57.89	57.17
	ATTEMPT	56.74	57.67	57.03
	Flan-T5	59.20	60.73	61.30
	GPT-3.5	60.14	59.55	60.77
	Ours (+BERT)	57.35	57.58	57.24
	+context	57.76	57.63	58.02
	Ours (+Flan-T5)	64.37 (+4.23)	**63.55** (+2.82)	**64.21** (+2.91)
	+context	**64.67**	63.28	64.33
	Ours (+GPT-3.5)	64.84 (+4.70)	63.30 (+2.57)	64.00 (+2.00)
	+context	**65.11**	**63.56**	**65.47**

Table 4 Dataset statistics of CMMA.

		CMMA		
		Train	Validation	Test
Sentiment	Positive	1889	559	923
	Neutral	9324	2694	2249
	Negative	2575	793	789
Emotion	Anger	1610	375	382
	Sadness	620	168	99
	Joy	1950	406	182
	Surprise	534	514	371
	Disgust	821	118	19
	Fear	477	92	43
	Neutral	7776	2373	2865
Sarcasm	Sarcastic	2172	525	446
	Non-sarcastic	11616	3521	3515

such as P^*_{sa} for sentiment analysis, P^*_{er} for emotion recognition, and P^*_{sd} for sarcasm detection. These prompts, integral to our single-task training approach, were systematically employed to enhance the model's task-specific performance.

Upon examining the results in Table 5, it is evident that MTP consistently outperforms STP across the array of tasks tested. MTP not only improves performance metrics across sentiment analysis, emotion recognition, and sarcasm detection but also offers significant efficiencies in training duration. This comparative superiority can be attributed to the multi-prompt template's capability to integrate and leverage commonalities across diverse tasks, thereby optimizing the learning process. The multi-prompt template in MTP effectively harnesses the potential of shared learning, allowing the model to generalize across tasks while retaining the ability to focus on task-specific nuances. This dual benefit is critical in tasks that require a nuanced understanding of context and subtleties in language, which are often pivotal in areas like sarcasm detection and emotion recognition.

4.6 Few-shot learning

Given the effectiveness of prompts in facilitating knowledge acquisition from pretrained language models, we embarked on few-shot experiments, training the model with 5, 10, 15, 20, 25 samples. Figure 5 delineates the few-shot learning outcomes across tasks on three distinct datasets. The experimental findings from the Memotion dataset underscore an enhancement in task performance proportional to the increment in training samples. This improvement is primarily attributed to the nature of the Memotion dataset, which, unlike conversational datasets, is bereft of contextual richness, making the augmentation of training samples a critical factor for achieving desirable results. Conversely, in the

Table 5 Results of STP vs. MTP.

(%)

Dataset	Method	Task		
		Sentiment	Emotion	Sarcasm
Memotion	STP	38.11	48.46	54.15
	MTP	**40.56**	**51.24**	**57.43**
MUStARD	STP	51.22	35.48	67.62
	MTP	**53.36**	**37.74**	**69.41**
CMMA	STP	53.37	54.21	53.43
	MTP	**57.35**	**57.58**	**57.24**

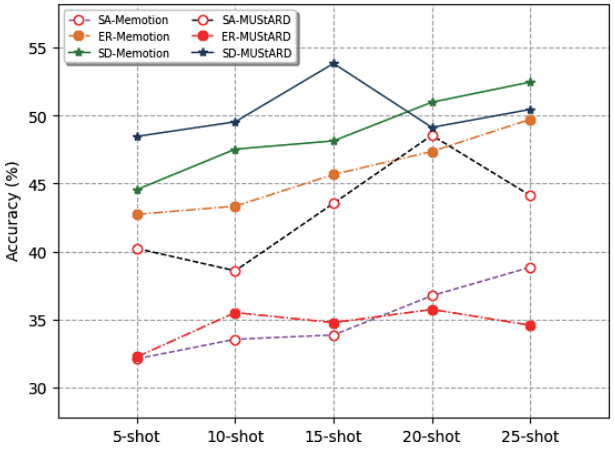

Fig. 5 Few-shot learning ability of different settings. Abbreviations: SA: Sentiment analysis, ER: Emotion recognition, SD: Sarcasm detection.

MUStARD dataset, optimal performance in sentiment analysis, emotion recognition, and sarcasm detection was observed with 20, 5, and 15 samples respectively, showcasing our method's prowess in few-shot learning scenarios. These results not only affirm the efficacy of our approach in contexts with limited data but also highlight the nuanced differences in dataset characteristics and their implications on task-specific performance. It is evident that our proposed method not only adapts to but also capitalizes on the unique challenges presented by few-shot learning, demonstrating its versatility and effectiveness across varying task demands and dataset peculiarities.

4.7 Ablation study

We conducted an ablation study focusing on the construction of the prompt template and the contribution of the gating network, with the results detailed in Table 6.

Prompt template for multi-task learning. To ascertain the impact of individual sub-prompt components, we devised templates by omitting (a) no P^*_{sa} (sentiment prompt), (b) no P^*_{er} (emotion prompt), and (c) no P^*_{sd} (sarcasm prompt), respectively. The findings, as illustrated in Table 6, indicate a marked decline in task performance upon the removal of any task-specific prompt. For instance, in the Memotion dataset, GPT-3.5 achieved the highest scores with the complete prompt set, showing 50.16% in sentiment, 58.22% in emotion, and 67.67% in sarcasm detection. However, when any prompt component was removed, there was a noticeable drop in these performance metrics. This trend is consistent across other datasets as

Table 6 Results of ablation experiments.

(%)

Model	Memotion			MUStARD			CMMA		
	Sentiment	Emotion	Sarcasm	Sentiment	Emotion	Sarcasm	Sentiment	Emotion	Sarcasm
No P^*_{sa}	31.51	48.61	54.23	42.78	32.35	65.41	48.36	54.84	53.59
No P^*_{er}	35.69	42.34	52.65	43.65	31.66	64.62	54.44	50.63	51.77
No P^*_{sd}	37.82	47.88	50.42	45.62	33.59	61.37	55.13	55.42	50.17
No gating network	35.65	46.68	55.64	49.44	34.51	62.65	55.23	54.49	51.55
Complete framework	**40.56**	**51.24**	**57.43**	**53.36**	**37.74**	**69.41**	**57.35**	**57.58**	**57.24**

well, such as MUStARD and CMMA, where GPT-3.5's performance significantly outpaced other models when the full prompt set was employed.

The removal of P^*_{sa}, P^*_{er}, or P^*_{sd} not only led to a decline in the performance of the associated tasks but also negatively impacted other related tasks. This suggests that while each task-specific prompt is finely tuned for its respective task, the presence of prompts related to other tasks contributes meaningfully to the overall performance. For instance, in the CMMA dataset, GPT-3.5 shows a strong performance across all tasks when the full prompt set is used (64.84% in sentiment, 63.30% in emotion, and 64.00% in sarcasm detection). The impact of prompt removal is especially evident when comparing these results with those of BERT, which shows significantly lower scores across all tasks, emphasizing the importance of the specialized prompts and the interconnectedness between them.

The contribution of the gating network. To evaluate the gating networks' effectiveness, we eliminated these networks, thereby treating all sub-prompts as a singular, extended prompt, as documented in Table 6.

Overall, the outcomes from our ablation experiments demonstrate suboptimal performance across various methodologies, affirming the integral role of our approach in bolstering multi-task prompt learning. The decline in performance when omitting either specific components or the gating mechanism itself emphasizes the sophistication of our model in navigating the intricacies of multi-task learning. Each component within our framework—be it the tailored prompts or the dynamic gating networks—plays a pivotal role in achieving nuanced understanding and performance across tasks. This synergy is crucial for harnessing the full potential of multi-task learning, as evidenced by the significant improvements achieved through their integration.

4.8 Impact of large language model

In this study, we designate BERT as the default foundational language model. Although BERT demonstrates significant strengths in modeling long-range dependencies, our interest extends to the capabilities of other PLMs. Due to constraints in resources and computation, the selection of a PLM was considered a fixed configuration rather than a variable hyperparameter within our research. Consequently, we examined three of the most recent and renowned LLMs to evaluate their performance, with the experimental outcomes presented in Table 7.

The experimental data indicates that GPT-3.5 consistently outperforms other models across sentiment, emotion, and sarcasm detection tasks. In the memotion dataset, GPT-3.5 demonstrates substantial improvements over GPT-3, with sentiment detection increasing by 15.8%, emotion detection by 8.6%, and sarcasm detection by 12.2%. Flan-T5, although generally behind GPT-3.5, shows competitive performance, particularly in sarcasm detection, where

Table 7 Effect of different LLMs.

(%)

Dataset	**LLMs**	**Task**		
		Sentiment	Emotion	Sarcasm
Memotion	GPT-3	43.32	53.59	60.31
	GPT-3.5	**50.16**	**58.22**	**67.67**
	Flan-T5	48.25	57.77	67.83
	BERT	40.56	51.24	57.43
MUStARD	GPT-3	57.77	41.29	73.86
	GPT-3.5	**62.74**	**49.33**	**78.26**
	Flan-T5	61.36	46.24	76.81
	BERT	54.61	38.15	70.54
CMMA	GPT-3	60.33	59.82	60.53
	GPT-3.5	**64.84**	63.30	64.00
	Flan-T5	64.37	**63.55**	**64.21**
	BERT	57.35	57.58	57.24

it slightly surpasses GPT-3.5 in the Memotion dataset. In the MUStARD dataset, GPT-3.5 again leads, with improvements of 8.6% in sentiment, 19.5% in emotion, and 6.0% in sarcasm over GPT-3. Flan-T5 follows closely, indicating its robust capability across tasks. In the CMMA dataset, GPT-3.5 outperforms GPT-3 with a 7.5% improvement in sentiment detection and maintains a strong performance in emotion and sarcasm detection, though Flan-T5 edges out slightly in emotion and sarcasm tasks. BERT, while still performing adequately, lags behind the other models, especially in complex, multi-modal tasks, showing the advantages of more advanced models like GPT-3.5 and Flan-T5 when integrated with the MAP framework. This underscores the effectiveness of GPT-3.5, particularly in handling nuanced tasks within multi-modal datasets, while also highlighting the competitive performance of Flan-T5 in specific tasks. These results suggest that the evolution of large language models will continue to enhance the classification accuracy achievable through prompt learning methodologies.

4.9 Effect of context range

This subsection aims to explore the impact of context length on the performance of MAP. We select and evaluate the context length from the pool {0, 1, 2, 3, All}, where All represents using all the contexts. The experimental results are shown in Fig. 6.

In general, longer context length will allow the model to access more information, thus making accurate prediction. We can notice that there is a slight improvement when the context length ranges from 0 to 2, and the performance slightly decreases when the context length ranges from 3 to All. This shows that too short can not provide supplementary knowledge where too long will introduce excessive amounts of noise. Hence, taking the previous two utterances before the target utterance into consideration may be a optimal choice. Additionally, processing long contextual information demands increased computational resources, thereby constraining the model's utility.

4.10 Error analysis

In this section, we present a few misclassification cases to explore the potential limitations of the proposed model, as shown in Tables 8 and 9.

In the sentiment analysis task presented in Table 8, the true label of the first case is negative, which signifies a negative sentiment. Nevertheless, the model erroneously predicted the label as neutral. The expression "I'M TRIGGERED" is frequently utilized

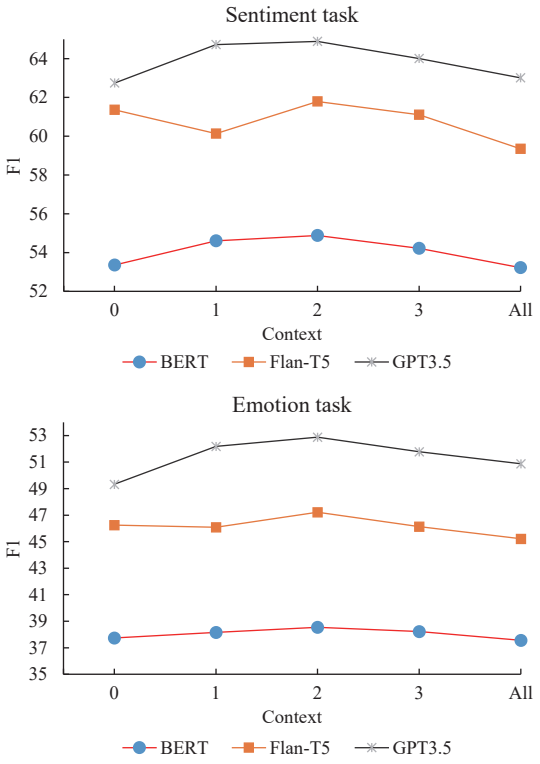

Fig. 6 Experiments of the varying context lengths.

Table 8 Few error cases for sentiment analysis and emotion recognition.

No.	Sentence	Sentiment& Emotion	
		True	Prediction
1	I'M TRIGGERED quickmeme.com.	Neg.	Neu.
2	WHAT I ACT LIKE WHEN I SPEAK SPANISH.	Pos.	Neu.
3	Rowan Atkinson's Family How we think it is. How it actually is.	Neg.	Neu.
4	No don't I beg of you!	Fear	Anger
5	What's wrong with you?	Disgust	Anger

Table 9 Few error cases for sarcasm detection.

No.	Sentence	Sarcasm	
		True	Prediction
1	NOSE BLEED??? JUDGE! SLOW PLAY! quickmeme.com.	Not Sar.	Sar.
2	YOU BRAKE OUR LEASE I'LL BREAK YOUR FACE memecenter.com MemeCenter.	Sar.	Not Sar.
3	I don't think I'll be able to stop thinking about it!	Sar.	Not Sar.
4	I've got some feelers out. In the meantime, listen to this.	Not Sar.	Sar.

in online discourse to convey strong negative emotions or a sense of offense. The model's inability to grasp the intensity and negative nuance of this phrase likely contributed to the misclassification. This suggests that the model may benefit from additional training data or fine-tuning to more accurately recognize and categorize such emotionally laden expressions. In the second case, the true label is positive, denoting a positive sentiment. Yet, the model inaccurately classified it as neutral. Without further context, pinpointing the precise cause of this misclassification is challenging. Regarding the third case, the provided text lacks adequate context to accurately assess the sentiment. Absent further details, it remains difficult to ascertain the exact reasons behind the misclassification.

In the emotion recognition task detailed in Table 8, the first case's true label is fear, denoting a fearful sentiment. Contrarily, our model inaccurately classified it as anger. The provided text exudes a sense of desperation and appeal, traits typically aligned with fear rather than anger. This discrepancy implies the model's inadequacy in detecting the nuanced fear embedded within the statement, opting instead to categorize it as an expression of anger. Such misclassification underscores the necessity for additional model training or refinements to adeptly discern between fear and anger.

The phrase "What's wrong with you?" frequently serves to convey dissatisfaction or negative judgment regarding someone's actions, potentially embodying both anger and disgust. In this instance, the model may have erroneously prioritized anger as the predominant emotion, despite the true sentiment being more closely related to disgust. This error indicates a requirement for the model to enhance its comprehension of the subtle distinctions and contextual nuances between anger and disgust, thereby improving its classificatory precision.

Across these scenarios, the model exhibited challenges in accurately categorizing the text into the correct emotional states, particularly distinguishing anger from other negative emotions like fear and disgust. This pattern suggests a possible deficiency in the model's grasp of specific linguistic indicators, contextual understanding, or the fine distinctions among these emotions.

For the task of sarcasm detection in Table 9, the error analysis reveals that the proposed prompt learning model faces challenges in accurately classifying sarcasm. It frequently misclassifies statements as sarcasm when they are not intended to be sarcastic and vice versa. The model seems to struggle with capturing the subtle linguistic cues, context, and tone that indicate sarcasm accurately.

To enhance the model's performance, it may benefit from additional training data that specifically focuses on sarcasm detection. The training data should include diverse examples of sarcastic expressions, encompassing various linguistic patterns, context, and tones. Fine-tuning the model using labeled examples that explicitly highlight the differences between sarcastic and non-sarcastic statements could also help improve its classification accuracy for sarcasm.

Overall, the error analysis highlights the need for further refinement and improvement of the proposed prompt learning model to better understand and classify sarcasm accurately. Additional training data, fine-tuning, and prompt engineering techniques could contribute to enhancing the model's performance in detecting sarcasm.

5 Conclusion

In this paper, we propose a multi-task prompt framework for multi-affection joint analysis, which models the commonness and uniqueness of task-specific prompts and multi-task prompts. Experiments on two benchmark datasets show our method outperforms these SOTA baselines that show the effectiveness of our proposed method (**Q1**). We also explore multi-task prompt-tuning that can better adapt the pre-trained language model to downstream tasks by performing experiments on STP vs MTP and ablation study (**Q2**). In addition, our model has also achieved advantages in few-shot scenarios answering **Q3**.

We plan to investigate the correlations and differences between different affective tasks to enhance the effectiveness of multi-affection joint detection. We aim to explore more effective methods to capture and utilize the inter-task dependencies, thereby improve the model's performance across different tasks in the future.

Limitations. There are also a few limitations. While this paper focuses on proposing a new learning framework, there is limited discussion on how the generated prompts by the framework impact the interpretability and explainability of model predictions. Future research can explore methods to enhance and explain the generated prompts to improve users' understanding and trust in the model predictions.

Acknowledgment

This paper was partly supported by Hong Kong RGC Theme-based Research Scheme (No. T45-401/22-N), the National Natural Science Foundation of China (No. 62006212), Fellowship from the China Postdoctoral Science Foundation (No. 2023M733907), Natural Science Foundation of Hunan Province of China (No. 242300421412), Foundation of Key Laboratory of Dependable Service Computing in Cyber-Physical-Society (Ministry of Education), and Chongqing University (No. CPSDSC202103).

References

[1] J. Wang, S. Wang, M. Lin, Z. Xu, and W. Guo, Learning speaker-independent multimodal representation for sentiment analysis, *Inf. Sci.*, vol. 628, pp. 208–225, 2023.

[2] M. Wang, L. Zhou, X. Huang, and W. Zheng, Towards federated learning driving technology for privacy-preserving micro-expression recognition, *Tsinghua Sci. Technol.*, vol. 30, no. 5, pp. 2169–2183, 2025.

[3] Q. Zhang, H. Zhang, K. Zhou, and L. Zhang, Developing a physiological signal-based, mean threshold and decision-level fusion algorithm (PMD) for emotion recognition, *Tsinghua Sci. Technol.*, vol. 28, no. 4, pp. 673–685, 2023.

[4] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies* (*NAACL-HLT*), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[5] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, ALBERT: A lite BERT for self-supervised learning of language representations. in *Proc. 8th Int. Conf. Learning Representations*, Addis Ababa, Ethiopia, 2020. https://dblp.uni-trier.de/db/conf/iclr/iclr2020.html#LanCGGSS20.

[6] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv: 1907.11692, 2019.

[7] K. Maity, P. Jha, S. Saha, and P. Bhattacharyya, A multitask framework for sentiment, emotion and sarcasm aware cyberbullying detection from multi-modal code-mixed memes, in *Proc. 45th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Madrid, Spain, 2022, pp. 1739–1749.

[8] Y. Zhang, A. Jia, B. Wang, P. Zhang, D. Zhao, P. Li, Y. Hou, X. Jin, D. Song, and J. Qin, M3GAT: A multi-modal, multi-task interactive graph attention network for conversational sentiment analysis and emotion recognition, *ACM Trans. Inf. Syst*., vol. 42, no. 1, p. 13, 2024.

[9] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, in *Proc. 34th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2020, p. 159.

[10] K. Zhou, J. Yang, C. C. Loy, and Z. Liu, Conditional prompt learning for vision-language models, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, LA, USA, 2022, pp. 16795–16804.

[11] G. Jiang, S. Liu, Y. Zhao, Y. Sun, and M. Zhang, Fake news detection via knowledgeable prompt learning, *Inf. Process. Manage.*, vol. 59, no. 5, p. 103029, 2022.

[12] K. Qi, H. Wan, J. Du, and H. Chen, Enhancing cross-lingual natural language inference by prompt-learning from cross-lingual templates, in *Proc. 60th Annu. Meeting of the Association for Computational Linguistics*, Dublin, Ireland, 2022, pp. 1910–1923.

[13] J. Fu, S. K. Ng, and P. Liu, Polyglot prompt: Multilingual multitask prompt training. in *Proc. 2022 Conf. Empirical Methods in Natural Language Processing*, Abu Dhabi, United Arab Emirates, 2022, pp. 9919–9935.

[14] A. Asai, M. Salehi, M. Peters, H. Hajishirzi, ATTEMPT: Parameter-efficient multi-task tuning via attentional mixtures of soft prompts, in *Proc. 2022 Conf. Empirical Methods in Natural Language Processing* (*EMNLP*), Abu Dhabi, United Arab Emirates, 2022, pp. 6655–6672.

[15] Y. Liu, Y. Lu, H. Liu, Y. An, Z. Xu, Z. Yao, B. Zhang, Z. Xiong, and C. Gui, Hierarchical prompt learning for multi-task learning, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, BC, Canada, 2023, pp. 10888–10898.

[16] Y. Zhang, D. Song, P. Zhang, P. Wang, J. Li, X. Li, and B. Wang, A quantum-inspired multimodal sentiment analysis framework, *Theor. Comput. Sci.*, vol. 752, pp. 21–40, 2018.

[17] Y. Liu, Q. Li, B. Wang, Y. Zhang, and D. Song, A survey of quantum-cognitively inspired sentiment analysis models, *ACM Comput. Surv.*, vol. 56, no. 1, p. 15, 2024.

[18] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, and E. Cambria, DialogueRNN: An attentive RNN for emotion detection in conversations, in *Proc. AAAI Conf. Artificial Intelligence*, Honolulu, HI, USA: AAAI Press, 2019, pp. 6818–6825.

[19] H. Zhang, Z. Li, H. Xie, R. Y. K. Lau, G. Cheng, Q. Li, and D. Zhang, Leveraging statistical information in fine-grained financial sentiment analysis, *World Wide Web*, vol. 25, no. 2, pp. 513–531, 2022.

[20] L. Xiao, X. Wu, W. Wu, J. Yang, and L. He, Multi-channel attentive graph convolutional network with sentiment fusion for multimodal sentiment analysis, in *Proc. 2022 IEEE Int. Conf. Acoustics, Speech and Signal Processing* (*ICASSP*), Singapore, 2022, pp. 4578–4582.

[21] A. Khare, A. Gangwar, S. Singh, and S. Prakash, Sentiment analysis and sarcasm detection of Indian general election tweets. arXiv preprint arXiv: 2201.02127, 2022.

[22] J. Chen, Z. Yao, S. Zhao, and Y. Zhang, Fusion pre-trained emoji feature enhancement for sentiment analysis, *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 22, no. 4, p. 99, 2023.

[23] Y. Cai, H. Cai, and X. Wan, Multi-modal sarcasm detection in Twitter with hierarchical fusion model, in *Proc. 57th Annu. Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019, pp. 2506–2515.

[24] B. Liang, C. Lou, X. Li, L. Gui, M. Yang, and R. Xu, Multi-modal sarcasm detection with interactive in-modal and cross-modal graphs, in *Proc. 29th ACM Int. Conf. Multimedia*, Virtual Event, 2021, pp. 4707–4715.

[25] Y. Liu, Y. Zhang, Q. Li, B. Wang, and D. Song, What does your smile mean? Jointly detecting multi-modal sarcasm and sentiment using quantum probability, in *Proc. Findings of the Association for Computational Linguistics*: *EMNLP 2021*, Punta Cana, Dominican Republic, 2021, pp. 871–880.

[26] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, Sentiment and sarcasm classification with multitask learning, *IEEE Intell. Syst.*, vol. 34, no. 3, pp. 38–43, 2019.

[27] L. Zhang, X. Zhao, X. Song, Y. Fang, D. Li, and H. Wang, A novel Chinese sarcasm detection model based on retrospective reader, in *Proc. 28th Int. Conf. MultiMedia Modeling*, Phu Quoc, Vietnam, 2022, pp. 267–278.

[28] C. Wen, G. Jia, and J. Yang, DIP: Dual incongruity perceiving network for sarcasm detection, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, Canada, 2023, pp. 2540–2550.

[29] M. S. Akhtar, D. Chauhan, D. Ghosal, S. Poria, A. Ekbal, and P. Bhattacharyya, Multi-task learning for multi-modal emotion recognition and sentiment analysis, in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, Minneapolis, MN, USA, 2019, pp. 370–379.

[30] Y. Zhang, L. Rong, X. Li, and R. Chen, Multi-modal sentiment and emotion joint analysis with a deep attentive multi-task learning model, in *Proc. 44th European Conf. Advances in Information Retrieval*, Stavanger, Norway, 2022, pp. 518–532.

[31] D. S. Chauhan, S. R. Dhanush, A. Ekbal, and P. Bhattacharyya, Sentiment and emotion help sarcasm? A multi-task learning framework for multi-modal sarcasm, sentiment and emotion analysis, in *Proc. 58th Annu. Meeting of the Association for Computational Linguistics*, Online, 2020, pp. 4351–4360.

[32] K. Maity, P. Jha, S. Saha, and P. Bhattacharyya, A multitask framework for sentiment, emotion and sarcasm aware cyberbullying detection from multi-modal code-mixed memes, in *Proc. 45th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Madrid, Spain, 2022, pp. 1739–1749.

[33] Y. Liu, Y. Zhang, Q. Li, B. Wang, and D. Song, What does your smile mean? Jointly detecting multi-modal sarcasm and sentiment using quantum probability, in *Proc. Findings of the Association for Computational Linguistics*: *EMNLP 2021*, Punta Cana, Dominican Republic, 2021, pp. 871–880.

[34] T. Tang, J. Li, W. X. Zhao, and J. R. Wen, Context-tuning: Learning contextualized prompts for natural language generation, in *Proc. 29th Int. Conf. Computational Linguistics*, Gyeongju, Republic of Korea, 2022, pp. 6340–6354.

[35] X. Han, W. Zhao, N. Ding, Z. Liu, and M. Sun, PTR: Prompt tuning with rules for text classification. *AI Open*, vol. 3, pp. 182–192, 2022.

[36] H. Wu and X. Shi, Adversarial soft prompt tuning for cross-domain sentiment analysis, in *Proc. 60th Annu. Meeting of the Association for Computational Linguistics*, Dublin, Ireland, 2022, pp. 2438–2447.

[37] Y. He, H. S. Zheng, Y. Tay, J. P. Gupta, Y. Du, V. Aribandi, Z. Zhao, Y. Li, Z. Chen, and D. Metzler, et al., Hyperprompt: Prompt-based task-conditioning of transformers, in *Proc. 39th Int. Conf. Machine Learning*, Baltimore, MD, USA, 2022, pp. 8678–8690.

[38] R. Herzig, O. Abramovich, E. Ben Avraham, A. Arbelle, L. Karlinsky, A. Shamir, T. Darrell, and A. Globerson, PromptonomyViT: Multi-task prompt learning improves video transformers using synthetic scene data. in *Proc. 2024 IEEE/CVF Winter Conf. Applications of Computer Vision*, Waikoloa, HI, USA, 2024, pp. 6789–6801.

[39] A. F. Akyürek, S. Paik, M. Kocyigit, S. Akbiyik, S. L. Runyun, and D. Wijaya, On measuring social biases in prompt-based multi-task learning, in *Proc. Findings of the Association for Computational Linguistics*: *NAACL 2022*, Seattle, WA, USA, 2022, pp. 551–564.

[40] T. Schick and H. Schütze, Few-shot text generation with natural language instructions, in *Proc. 2021 Conf. Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, 2021, pp. 390–402.

[41] J. Zeng, Y. Jiang, S. Wu, and M. Li, Enhanced few-shot learning with multiple-pattern-exploiting training, in *Proc. 10th CCF Int. Conf. Natural Language Processing and Chinese Computing*, Qingdao, China, 2021, pp. 388–399.

[42] X. Liu, Y. Zheng, Z. Du, M. Ding, Y. Qian, Z. Yang, and J. Tang, GPT understands, too, *AI Open*, vol. 5, pp. 208–215, 2024.

[43] X. L. Li and P. Liang, Prefix-tuning: Optimizing continuous prompts for generation, in *Proc. 59th Annu. Meeting of the Association for Computational Linguistics and the 11th Int. Joint Conf. Natural Language Processing*, Online, 2021, pp. 4582–4597.

[44] X. Liu, K. Ji, Y. Fu, W. Tam, Z. Du, Z. Yang, and J. Tang, P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks, in *Proc. 60th Annu. Meeting of the Association for Computational Linguistics*, Dublin, Ireland, 2022, pp. 61–68.

[45] Y. S. Choi and J. H. Lee, CodePrompt: Task-agnostic prefix tuning for program and language generation, in *Proc. Findings of the Association for Computational Linguistics*: *ACL 2023*, Toronto, Canada, 2023, pp. 5282–5297.

[46] C. Sharma, D. Bhageria, W. Scott, S. Pykl, A. Das, T. Chakraborty, V. Pulabaigari, and B. Gambäck, SemEval-2020 Task 8: Memotion analysis – the visuo-lingual metaphor, in *Proc. 14th Workshop on Semantic Evaluation*,

Barcelona, Spain, 2020, pp. 759–773.
[47] R. A. Potamias, G. Siolas, and A. G. Stafylopatis, A transformer-based approach to irony and sarcasm detection, *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17309–17320, 2020.
[48] S. Shah, S. Reddy, and P. Bhattacharyya, Emotion enriched retrofitted word embeddings, in *Proc. 29th Int. Conf. Computational Linguistics*, Gyeongju, Republic of Korea, 2022, pp. 4136–4148.
[49] G. A. Vlad, G. E. Zaharia, D. C. Cercel, C. Chiru, and S. Trausan-Matu, UPB at SemEval-2020 task 8: Joint textual and visual modeling in a multi-task learning architecture for memotion analysis, in *Proc. 14th Workshop on Semantic Evaluation*, Barcelona, Spain, 2020, pp. 1208–1214.
[50] I. Chaturvedi, C. L. Su, and R. E. Welsch, Fuzzy aggregated topology evolution for cognitive multi-tasks, *Cognit. Comput.*, vol. 13, no. 1, pp. 96–107, 2021.
[51] A. Asai, M. Salehi, M. E. Peters, and H. Hajishirzi, Attentional mixtures of soft prompt tuning for parameter-efficient multi-task knowledge sharing. arXiv preprint arXiv: 2205.11961, 2022.
[52] Y. Zhang, Y. Yu, Q. Guo, B. Wang, D. Zhao, S. Uprety, D. Song, Q. Li, and J. Qin, CMMA: Benchmarking multi-affection detection in Chinese multi-modal conversations, in *Proc. 37th Int. Conf. Neural Information Processing Systems*, New Orleans, LA, USA, 2023, p. 824.

Yang Yu received both the B.S. and M.S. degrees in Software Engineering from the Software Engineering College, Zhengzhou University of Light Industry. He is currently pursuing the PhD degree at Chongqing University. His current research interests include sentiment analysis and sarcasm detection.

Hui Liang received the PhD degree from Communication University of China. He is currently the PhD supervisor and a professor with Zhengzhou University of Light Industry. He has published more than 30 research papers in some journals, such as *IEEE TKDE*, *IEEE TCYB*, *ACM TKDD*, *ACM TMIS*, *SCIS*, *INS*, *JCST*, *KBS*, *ESWA*, *JIS*, and *APIN*. He is a distinguished member of CCF and a senior member of IEEE. His current research interests include data mining and machine learning.

Prayag Tiwari received the PhD degree from the University of Padova, Italy. He is currently working as an assistant professor at Halmstad University, Sweden. Previously, he worked as a postdoctoral researcher at the Aalto University, Finland, and Marie Curie Researcher at the University of Padova, Italy. His name has appeared in the World's Top 2% of Scientists released by Stanford University and Elsevier. He received the 2022 Best Paper Award for a paper published in *Neural Networks*, Elsevier. His papers were also awarded as one of the Highly Cited Papers published in journals like *Applied Sciences*, *Measurements*, etc. He has several publications in top journals and conferences, including *Neural Networks*, *Journal of Physics A*, *Information Fusion*, *Knowledge-Based Systems*, *International Journal of Computer Vision*, *IEEE TNNLS*, *IEEE TFS*, *IEEE JBHI*, *IEEE TAI*, *IEEE IoTJ*, *IEEE BIBM*, *ACM TOIT*, *ACM CSUR*, *CIKM*, *SIGIR*, *AAAI*, *ECIR*, *ECML*, *ACL*, etc. His research interests include artificial intelligence, quantum machine learning, deep learning, healthcare, bioinformatics, NLP, computer vision, intelligent transport, and IoT.

Yazhou Zhang received the PhD degree in college of intelligence and computing from Tianjin University in 2020. He is currently an associate researcher in Tianjin University, and also a postdoctoral researcher in School of Nursing at The Hong Kong Polytechnic University, China. His research interests include opinion mining (or sentiment analysis), data fusion, and quantum cognition. He is currently working on developing quantum inspired sentiment analysis models and their application to problems like conversational sentiment analysis, information fusion, and evolution of user emotional state.

Xiang Li received the PhD degree in Computer Science from College of Intelligence and Computing, Tianjin University, currently serves as an assistant researcher at Shandong Computing Center (National Supercomputing Jinan Center) of Qilu University of Technology (Shandong Academy of Sciences). He has published 30 papers in important academic journals and conferences (including one highly cited paper in ESI). His research interests include intelligent analysis of time series data, brain information processing, and affective computing.

Xiangkai Wang is currently the director of Artificial Intelligence Laboratory in Shandong Zhengyun Information Technology Limited Company, and has published more than 20 academic papers, all of which are SCI/EI retrieval. He has accumulated 20 invention patents. His main research interests are computer vision and intelligent information processing. He is also a senior member of CCF and AAAI, and has obtained 162 scientific and technological achievements and awards in artificial intelligence.