UECA-Prompt: Universal Prompt for Emotion Cause Analysis

Xiaopeng Zheng[1], Zhiyue Liu[1,2], Zizhen Zhang[1], Zhaoyang Wang[1], Jiahai Wang[1]*

[1]School of Computer Science and Engineering, Sun Yat-sen University
Guangzhou, China
[2]School of Computer, Electronics and Information, Guangxi University
Nanning, China
{zhengxp26,liuzhy93,wangzhaoy22}@mail2.sysu.edu.cn,
{wangjiah,zhangzzh7}@mail.sysu.edu.cn

Abstract

Emotion cause analysis (ECA) aims to extract emotion clauses and find the corresponding cause of the emotion. Existing methods adopt fine-tuning paradigm to solve certain types of ECA tasks. These task-specific methods have a deficiency of universality. And the relations among multiple objectives in one task are not explicitly modeled. Moreover, the relative position information introduced in most existing methods may make the model suffer from dataset bias. To address the first two problems, this paper proposes a universal prompt tuning method to solve different ECA tasks in the unified framework. As for the third problem, this paper designs a directional constraint module and a sequential learning module to ease the bias. Considering the commonalities among different tasks, this paper proposes a cross-task training method to further explore the capability of the model. The experimental results show that our method achieves competitive performance on the ECA datasets.

1 Introduction

Recently, emotion cause analysis (ECA) has obtained increasing attention. As a classic task of ECA, emotion cause extraction (ECE) was first proposed and defined as a clause-level classification problem (Gui et al., 2016a). ECE aims to explore the potential causes behind a certain emotional expression in a clause. However, the applications of ECE are limited in real-world scenarios because the emotion must be annotated before cause extraction. Therefore, emotion-cause pair extraction (ECPE) was proposed to identify all emotions and their corresponding causes from an unannotated text at the same time (Xia and Ding, 2019). Chen et al. (2020a) proposed conditional causal relationships classification (CCRC) to explore the relation of emotion-cause pairs and contexts. Although there are other emerging tasks proposed (Li et al.,
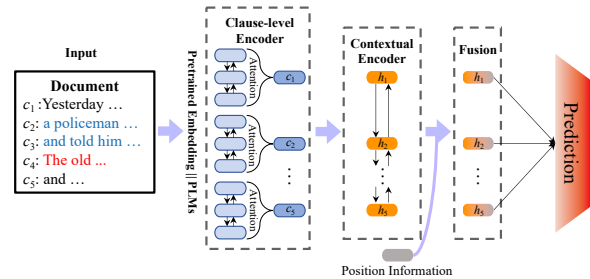

Figure 1: Fine-tuning for emotion cause analysis.

2021b; Bi and Liu, 2020) to expand the applications of ECA, this paper focuses on the most complex and challenging three clause-level sub-tasks: ECE, ECPE, and CCRC.

In literature, ECA tasks have been widely studied. On the ECE task, Gui et al. (2016a) released a benchmark Chinese emotion cause dataset and proposed a multi-kernel based method. Several statistical learning (Gui et al., 2016b; Xu et al., 2017) and deep learning methods (Cheng et al., 2017; Hu et al., 2021b) have been applied to ECE, which show competitive performance on emotion cause prediction. On the ECPE task, Xia and Ding (2019) proposed a two-step method to extract emotions and corresponding causes at the same time. This method first individually extracts the emotion set and cause set. Then, it gets the emotion-cause pairs by applying a Cartesian product to these two sets and trains a filter to remove the invalid pairs. However, the error may propagate from the first procedure to the second because of the inherent drawback of the pipelined framework. To address this issue, several works adopted end-to-end architecture. Part of these works (Tang et al., 2020; Wu et al., 2020) focused on the multi-task learning (Caruana, 1997) of the ECPE task with the joint learning framework (Ding et al., 2020a,b). Some other works represented the relation between emotion and cause with graph construction (Wei et al., 2020; Chen et al., 2020c). In addition, transition-

---

*Corresponding author

based parsing (Fan et al., 2020) and unified sequence labeling (Chen et al., 2020b; Cheng et al., 2021) are also employed. On the CCRC task, the relationship between emotion and cause clauses in different contexts is studied (Chen et al., 2020a).

The existing methods could be divided into multi-stage framework (Xia and Ding, 2019; Xu et al., 2021) and end-to-end framework (Ding et al., 2020a; Wei et al., 2020). Both frameworks adopt the fine-tuning paradigm. As shown in Fig. 1, these fine-tuning methods firstly obtain clause-level features from the pre-trained embedding or pre-trained language models (PLMs). Next, a contextual encoder is designed to yield contextual representations of clauses. Then, the contextual representation is merged with clause-level position information. Finally, the output is used for task prediction. In general, the existing methods have three obvious shortcomings: firstly, the contextual encoder has the deficiency of universality because they are designed for specific task objectives. Secondly, the existing methods implicitly learn the relationships among multiple task objectives, rather than explicitly model their relations. Thirdly, position information would make the model sensitive to data distribution and lack robustness.

Prompt (Liu et al., 2021; Han et al., 2021; Houlsby et al., 2019) is a new paradigm that can be traced back to GPT-3 (Brown et al., 2020). This paradigm could overcome the difficulty of uniformly solving various tasks by transforming specific fine-tuning tasks into the form of pre-training tasks. The universality of prompt has been widely validated in various tasks of natural language processing (Schick et al., 2020; Ma et al., 2021; Li et al., 2021a; Li and Liang, 2021), especially in few-shot (Gao et al., 2021; Zhang et al., 2021) or zero-shot (Sanh et al., 2021; Wei et al., 2021) scenarios. Multiple predictions should be performed for one sample with several clauses in ECA. And ECA is composed based on multiple task objectives (e.g. emotion extraction, cause extraction, emotion-cause pair extraction). Appropriate prompt construction could complete multiple predictions simultaneously. Constructing special prompts could explicitly model the relations among multiple objectives in one task. In general, prompt could make up for the shortcomings of existing methods.

Inspired by this, this paper proposes a universal prompt-based method for ECA tasks (UECA-Prompt). UECA-Prompt first modifies each task ob-
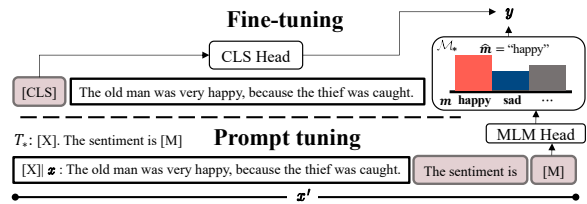

Figure 2: Different task forms for prompt tuning and fine-tuning.

jective into a sub-prompt after decomposing ECA tasks into multiple task objectives. Then, it explicitly models the relations among different task objectives by combining different sub-prompts into a composite prompt. As most of the works (Petroni et al., 2020; Schick and Schütze, 2020a,b), UECA-Prompt is manually constructed to solve different ECA tasks in a unified framework. As far as we know, UECA-Prompt is the first attempt at multi-task multiple predictions with prompt.

Some previous works (Ding et al., 2020b; Chen et al., 2020c) artificially introduced position information into the model. Such characteristic would make the model extremely sensitive to the distribution of positions of cause clauses relative to their corresponding emotion clauses in the dataset. Specifically, existing methods with position information may not generalize well to de-bias dataset (Ding and Kejriwal, 2020). Because the cause clause of most samples in the de-bias dataset is not in proximity to the emotion clause. This paper designs a directional constraint module and a sequential learning module to better identify the emotion-cause pairs. These two modules could ease the bias by discarding position information between clauses.

There are commonalities among different tasks. For example, CCRC focuses more on context information, and context information is also crucial for ECE and ECPE. This paper proposes a cross-task training method. The model would be able to learn commonalities among tasks with the cross-task training method.

We evaluate our method on three benchmark Chinese emotion cause datasets. The experimental results show that our method can obtain better results than state-of-the-art methods on three ECA tasks solely based on the BERT.

Our contributions are summarized as follows:

- We propose a universal prompt method for a variety of ECA tasks, such as ECE, ECPE, and
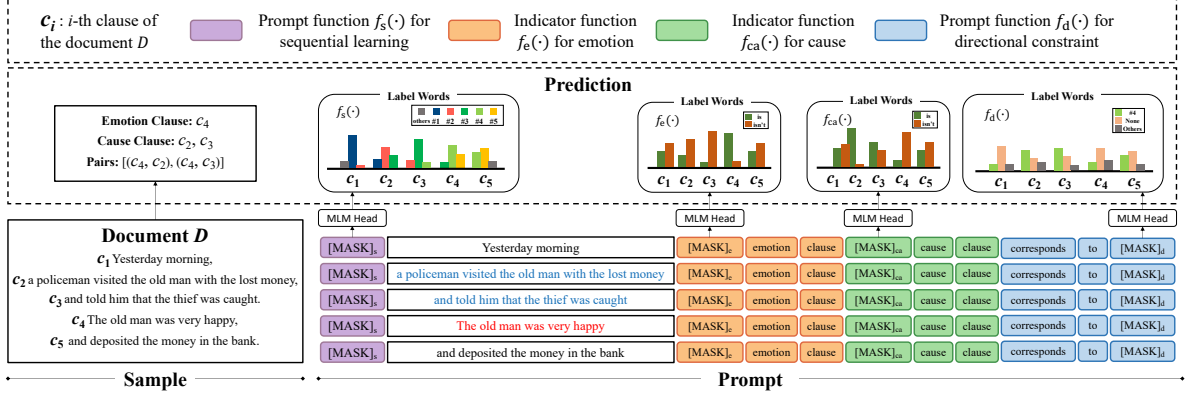
Figure 3: Overview of UECA-Prompt. The answer slot $[\text{M}]$ of template is replaced by the token $[\text{MASK}]$. Subscripts are added for token $[\text{MASK}]$ to distinguish different sub-prompt. At the bottom of the figure, input x is presented on the left, and prompt x' constructed from the input text is on the right. The predictions of each sub-prompt are presented above the corresponding module.

Function $f_*(\cdot)$	Template $T_*(\cdot)$	Label Words $\mathcal{M}_*(\cdot)$
$f_e(\cdot)$	$T_e(\cdot) = $ "$[\text{X}]$ $[\text{M}]$ emotion clause"	$\mathcal{M}_e(\cdot) = \{$"is", "isn't"$\}$
$f_{ca}(\cdot)$	$T_{ca}(\cdot) = $ "$[\text{X}]$ $[\text{M}]$ cause clause"	$\mathcal{M}_{ca}(\cdot) = \{$"is", "isn't"$\}$
$f_d(\cdot)$	$T_d(\cdot) = $ "$[\text{X}]$ corresponds to $[\text{M}]$"	$\mathcal{M}_d(\cdot) = \{$"None", "1", "2", ..., "$n$"$\}$
$f_s(\cdot)$	$T_s(\cdot) = $ "$[\text{M}]$ $[\text{X}]$"	$\mathcal{M}_s(\cdot) = \{$"1", "2", ..., "$n$"$\}$

Table 1: Different sub-prompt in UECA-Prompt. $[\text{X}]$ is input slot filled with text x and $[\text{M}]$ is answer slot for prediction. We translate the original Chinese words into English for better illustration.

CCRC. It is the first attempt to solve multi-task multiple predictions with prompt.

- We design the directional constraint and sequential learning module to ease the bias effect caused by position information, making UECA-Prompt more robust toward the de-bias dataset.

- Experimental results show that UECA-Prompt outperforms the state-of-the-art methods. And the cross-task training method further improves the performance of UECA-Prompt.

2 Preliminary

2.1 Task Definition

Given a document $D = \{c_1, ..., c_n\}$ with n clauses, where c_i is the i-th clause. ECE is to determine the cause clause set $C_{ca} \subseteq D$ according to a given emotion clause $c_e \in D$. The cause clause set C_{ca} may have more than one clause. ECPE is to identify each emotion-cause pair $\{c_e, c_{ca}\}$ in the document. CCRC determines whether the emotion-cause pair $\{c_e, c_{ca}\}$ still has a causal relationship under different context within a set $\mathcal{T} = \{t_1, ..., t_j\}$, where context t_j is the residual document except for the

emotion clauses c_e and cause clauses c_{ca}.

2.2 Prompt Tuning

As shown in Fig. 2, a typical prompt consists of a template $T_*(\cdot)$ (e.g. "$[\text{X}]$. The sentiment is $[\text{M}]$") and a set of permissible values $\mathcal{M}_*(\cdot)$ (e.g. "happy", "sad"). Firstly, a prompt function $f_*(\cdot)$ fills the input slot $[\text{X}]$ with original input x to get x' (e.g. "The old man was very happy, because the thief was caught. The sentiment is $[\text{M}]$"). Secondly, an argmax is used to search for the highest-scoring intermediate result $\hat{m}$ from a set of label words $\mathcal{M}_*(\cdot)$. $\mathcal{M}_*(\cdot)$ is a set of potential answers for answer slot $[\text{M}]$. Finally, $\hat{m}$ would be mapped into final result y.

3 Method

UECA-Prompt is a BERT-based method constructed as the form of the MLM task. This paper would discuss UECA-Prompt from two perspectives, sub-prompt for task decomposition and composite prompt for multiple predictions.

3.1 Sub-prompt for Task Decomposition

ECA involves multi-task learning and could be decomposed into emotion extraction, cause extrac-

tion, and emotion-cause pair extraction. This paper designs a sub-prompt for each objective. The sub-prompts include indicator functions, directional constraint module, and sequential token module.

Indicator Function In ECA, we are requested to find all the emotion-cause pairs in the text. Empirically, we should first search the text for all the emotion clauses. Then the cause clauses could be identified according to those emotion clauses. Finally, the causal relation between emotion and cause clauses can be checked depending on the context.

Two clause-level indicator functions, $f_e(\cdot)$ and $f_{ca}(\cdot)$, are designed to extract emotion and cause. Specifically, the indicator function $f_e(c_i)$ determines whether a clause c_i is an emotion clause, while $f_{ca}(c_i)$ determines the cause clause. The orange rectangles and green rectangles in Fig. 3 show these two sub-prompts, and the answer slot [M] is replaced by token [MASK]. The sub-prompt templates and label words for $f_e(\cdot)$ and $f_{ca}(\cdot)$ are formalized as first and second rows of Table 1.

The candidate label words for predicting answer [M] include "is" and "isn't". "is" represents the current clause belonging to the target set, and "isn't" represents there is no subordinate relationship.

A directional constraint module and sequential module are proposed to further extract the emotion-cause pair. Different from the usual sub-prompt, these two modules could be regarded as a pointer structure when combined. It effectively alleviates bias caused by position information.

Directional Constraint As the blue rectangles shown in Fig. 3, the template and label words of $f_d(\cdot)$ can be formalized as third row of Table 1. In the label words set $\mathcal{M}_d(\cdot)$, "None" indicates that there is no clause associated with the current clause. And numeric token "i" represents the current clause associated with the i-th clause. n represents the number of clauses in the document.

The numeric tokens in $\mathcal{M}_d(\cdot)$ indicate the sequence information of clauses. A sequential learning module is designed to facilitate the model acquiring such knowledge.

Sequential Learning In sequential learning, we set a prefix answer slot [M] for each clause. The numeric token of the answer slot for each clause is a unique identifier. The model could learn the unique identifier with sequential learning. As the purple rectangles shown in Fig. 3, a sequential function

$f_s(\cdot)$ is designed for sequential learning. The template and label words could be formalized as the last rows of Table 1. Intermediate result $\hat{m}$ for i-th clause is "i". Label words for sequential learning are also used in directional constraint function (i.e., $\mathcal{M}_s(\cdot) \subseteq \mathcal{M}_d(\cdot)$). Sequential learning is only used in the training stage.

3.2 Composite Prompt for Multiple Predictions

A composite prompt comprised of all the sub-prompt (Liu et al., 2021) is defined to address different multiple prediction tasks in a unified prompt.

Composite Prompt Composite prompt explicitly models the relation among different task objectives. The template of composite prompt $f_{CP}(\cdot)$ is formalized as:

$$T_{CP}(\cdot) = \text{"Clause [M] [X] [M] emotion clause} \atop \text{[M] cause clause corresponds to [M]"}, \quad (1)$$

where the label words for each answer slot [M] in composite prompt template $T_{CP}(\cdot)$ correspond to the label words of each sub-prompt, respectively.

Filling the composite template with document D to form prompted document $\boldsymbol{D'}$ could be formalized as:

$$\boldsymbol{D'} = [f_{CP}(c_1); ...; f_{CP}(c_n)], \quad (2)$$

where c_i is the i-th clause in document D, and $[\cdot; \cdot]$ is the concatenation operation.

Multiple Predictions In ECA tasks, multiple predictions should be performed for each clause. UECA-prompt is capable of multiple predictions. As shown in Fig. 3, firstly, we convert the input doc into a set of clauses. Secondly, the template of composite prompt is applied to each clause in the input document D. Thirdly, the intermediate answer for each sub-prompt is searched to separately predict the answer of slot [M]. Finally, the intermediate answers are aggregated and mapped to the final result. This paper explores composite prompt (See Eq. (1)) for three ECA tasks (i.e., ECPE, ECE, and CCRC). Since the operations on each clause are the same, this paper will discuss the prompt on a single clause in different tasks.

Implementation for ECPE The composite prompt function instantiated for the ECPE task is given in Eq. (3). Unrelated prompt tokens are omitted for clarity. Subscripts are added for the answer

Method	ECPE			Emotion Extraction			Cause Extraction		
	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)
Inter-EC†	61.28	67.21	57.05	82.30	83.64	81.07	65.07	70.41	60.83
TransECPE†	67.99	73.74	63.07	84.74	87.16	82.44	69.74	75.62	64.71
UTOS-BERT†	72.03	73.89	70.62	85.56	88.15	83.21	74.71	76.71	73.20
PairGCN-BERT†	72.02	76.92	67.91	83.75	88.57	79.58	73.75	79.07	69.28
RANK-CP-bert†	73.60	71.19	76.30	**90.57**	**91.23**	89.99	76.15	74.61	77.88
ECPE-MLL-bert†	74.52	77.00	72.35	88.86	86.08	91.91	76.30	73.82	79.12
MTST+Refinement†	74.63	**77.46**	71.99	84.36	87.11	81.78	76.66	**79.47**	74.04
UECA-Prompt	**74.70***	71.82	**77.99***	88.16	84.75	**91.95***	**77.55***	76.24	**79.16***
UECA-Prompt (m2m)	73.41	70.19	77.19	87.81	85.45	90.43	75.84	74.13	77.97

Table 2: Experimental results on the ECPE task. UECA-Prompt (m2m) is the result of UECA-Prompt with the M2M module. The best result is marked in **bold**. † indicates the results are reported in the original paper. * indicates statistically significant improvement ($p < 0.01$) over the best baseline.

Method	F_1 (%)	P (%)	R (%)
UECA-Prompt	61.14	69.52	54.66
UECA-Prompt (m2m)	**63.45**	**73.92**	**56.30**

Table 3: Experimental results on multi-emotion samples of the ECPE task.

Method	F_1 (%)	P (%)	R (%)
Multi-kernel†	67.52	65.88	69.72
MANN†	77.06	78.43	75.87
RTHN (Layer 3)†	76.77	76.97	76.62
FSS-GCN†	78.61	75.72	77.14
EF-BHA†	78.68	79.38	78.08
RHNN†	79.14	81.12	77.25
UECA-Prompt	**84.40***	**84.57***	**84.27***

Table 4: Experimental results on the ECE task. * indicates statistically significant improvement ($p < 0.001$) over the best baseline.

slot [M] to distinguish different sub-prompt:

$$f_{\text{ECPE}}(c_i) = \text{"}[\text{M}]_s\ c_i\ [\text{M}]_e\ [\text{M}]_{\text{ca}}\ [\text{M}]_d\text{"}. \quad (3)$$

The prediction of the emotion and cause could be mapped from the intermediate result of slot $[\text{M}]_e$ and $[\text{M}]_{\text{ca}}$ in indicator functions, formalized as:

$$P_e(c_i) = \begin{cases} 1, & \hat{m}_e = \text{"is"} \\ 0, & \hat{m}_e = \text{"isn't"} \end{cases}, \quad P_{\text{ca}}(c_i) = \begin{cases} 1, & \hat{m}_{\text{ca}} = \text{"is"} \\ 0, & \hat{m}_{\text{ca}} = \text{"isn't"} \end{cases}, \quad (4)$$

where $\hat{m}_e$ and $\hat{m}_{\text{ca}}$ are highest-scoring outputs of search the function for slot $[\text{M}]_e$ and $[\text{M}]_{\text{ca}}$, respectively. The searching function is formalized as:

$$\hat{m} = \underset{m \in \mathcal{M}}{\arg\max}\ p(f(x', m); \theta) \quad (5)$$

where function $f(x', m)$ is used to fill the answer slot [M] in prompt x' with potential answer m, and θ represents the parameters of PLMs.

Method	F_1 (%)	P (%)	R (%)
BiLSTM+Concatenation†	61.27	54.12	71.19
BiLSTM+BiLSTM †	69.76	66.06	74.00
BiLSTM+Self-Attention†	66.05	57.66	77.70
UECA-Prompt	**81.18***	**76.35***	**86.76***

Table 5: Experimental results on the CCRC task. * indicates statistically significant improvement ($p < 0.001$) over the best baseline.

By combining the intermediate results of $[\text{M}]_{\text{ca}}$ and $[\text{M}]_d$, the prediction of emotion-cause pairs could be formalized as:

$$P_{\text{pair}}(c_i) = \begin{cases} (i,\ j), & \hat{m}_d = \text{"}j\text{" and } \hat{m}_{\text{ca}} = \text{"is"} \\ \text{null, otherwise,} \end{cases} \quad (6)$$

where $(i,\ j)$ represents that the i-th and j-th clauses constitute an emotion-cause pair (the former is the cause and the latter is the emotion), and "null" represents there is no clause associated with the current clause.

In some cases, multiple emotions correspond to one cause. This paper further designs a many-to-many (M2M) module to deal with this situation. The variant prompt template with the M2M module is as follows:

$$f_{\text{ECPE}}^{\text{m2m}}(c_i) = \text{"}[\text{M}]_s\ c_i\ [\text{M}]_e\ [\text{M}]_{\text{ca}}\ [\text{M}]_{d_1}\ ...\ [\text{M}]_{d_M}\text{"}, \quad (7)$$

where M is the maximum number of pairs in one document, and M is set to 3 in the experiment.

Implementation for ECE The emotion clause is annotated in the ECE task. This paper replaces the slot $[\text{M}]_e$ in Eq. (3) with a specific token ("is" for the emotion clause and "isn't" for others). The composite prompt function instantiated for ECE is as follows:

$$f_{\text{ECE}}(c_i) = \text{"}[\text{M}]_s\ c_i\ \text{is/isn't}\ [\text{M}]_{\text{ca}}\ [\text{M}]_d\text{"}, \quad (8)$$

Method	ECPE			Emotion Extraction			Cause Extraction		
	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)
UTOS-BERT	34.14	42.76	28.95	66.45	**83.69**	56.08	36.29	45.83	30.55
ECPE-MLL-bert	45.57	**61.53**	36.39	**80.33**	77.15	84.04	**54.68**	51.40	**59.01**
MTST+Refinement	44.93	51.99	40.34	64.10	77.65	55.46	48.84	**56.25**	43.96
UECA-Prompt	**49.37**	46.30	**53.22**	76.15	68.68	**85.67**	50.64	45.75	57.71

Table 6: Experimental results on the ECPE task under few-shot setting.

Method	F_1 (%)	P (%)	R (%)
PADGL	66.16	66.16	67.52
RTHN (Layer 3)	62.49	61.72	63.48
UECA-Prompt	**72.21**	**72.10**	**72.54**

Table 7: Experimental results on the ECE task under few-shot setting.

Method	F_1 (%)	P (%)	R (%)
BiLSTM+Concatenation	66.13	49.53	**99.53**
BiLSTM+BiLSTM	62.41	52.00	78.17
BiLSTM+Self-Attention	62.45	51.54	79.40
UECA-Prompt	**67.56**	**55.46**	86.70

Table 8: Experimental results on the CCRC task under few-shot setting.

where is/isn't is determined by whether the current sentence is an emotion clause, and slot $[\text{M}]_\text{d}$ is replaced by the numeric token of the emotion clause in the testing stage.

The prediction of cause is the same as ECPE (see Eq. (4)).

Implementation for CCRC The emotion-cause pair is annotated in the CCRC task. Thus, the slot $[\text{M}]_\text{ca}$ in Eq. (8) is also replaced by a specific token ("is" for cause clauses and "isn't" for others) as the ECE task. The composite prompt function instantiated for CCRC is as follows:

$$f_\text{CCRC}(c_i) = \text{"}[\text{M}]_\text{s} \, c_i \text{ is/isn't is/isn't } [\text{M}]_\text{d}\text{"} , \quad (9)$$

Different from ECE and ECPE, we are required to tell the causal relationship between an emotion clause and multiple cause clauses in CCRC. Thus, a voting mechanism is proposed. The final result is co-determined by the intermediate result of answer slot $[\text{M}]_\text{d}$ in Eq. (9) for all the cause clauses. Based on the voting mechanism, the prediction formula of CCRC is as follows:

$$\mu_\text{vote}(D) = \frac{1}{|C_\text{ca}|} \sum_{c_i \in C_\text{ca}} p(\hat{\boldsymbol{m}}_{\text{ca},c_i} = \hat{\boldsymbol{m}}_{\text{d},c_e}), \quad (10)$$

$$P_\text{ccrc}(D) = \begin{cases} 1, & \mu_\text{vote}(D) > 0.5 \\ 0, & \mu_\text{vote}(D) \le 0.5, \end{cases} \quad (11)$$

where c_e is the emotion clause, C_{ca} is the cause clause set, $\hat{\boldsymbol{m}}_{\text{ca},c_i}$ is the intermediate result of slot $[\text{M}]_\text{ca}$ for clause c_i, and $\hat{\boldsymbol{m}}_{\text{d},c_e}$ is the intermediate result of slot $[\text{M}]_\text{d}$ for emotion clause c_e.

4 Cross-task Training Method

Empirically, there are commonalities among different ECA tasks. These commonalities would further improve the the model performance. This paper proposes a cross-task training method to make the model better adapt knowledge from one domain to another. The following section will introduce this training method through an example. The initial pre-trained model is defined as $\mathcal{A}$. The first step, the model $\mathcal{A}$ is trained in the ECE task to obtain the trained model $\mathcal{B}$. Second step, the trained model $\mathcal{B}$ is used to train for the ECPE task to obtain the trained model $\mathcal{C}$. Finally, model $\mathcal{C}$ is used to measure the performance of the UECA-Prompt on ECPE. Any two tasks of ECA could perform the training steps mentioned above. The commonality among tasks could be learned by the model in the process of cross-task training.

5 Experiments

This paper conducts experiments on the ECE, ECPE, and CCRC tasks to evaluate our approach. We release our code at https://github.com/yajus/UECA-Prompt.

5.1 Datasets

The experiments are conducted on three public datasets. The **ECE dataset** (Gui et al., 2016a) is collected from SINA city news and contains 2105 instances. Its document has only one emotion word and one or more emotion causes. The **ECPE dataset** (Xia and Ding, 2019) is constructed based on the ECE dataset. It aggregates the instances containing the same text and different emotion cause labels. The **CCRC dataset** (Chen et al., 2020a) is also built based on the ECE dataset. It is constructed following two steps: manual annotation

	ECPE			Emotion Extraction			Cause Extraction			ECE			CCRC		
Method	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)
UECA-Prompt	73.39	70.01	77.26	87.64	84.26	91.43	75.35	72.06	79.21	83.19	82.74	83.73	80.74	77.07	84.87
w/o f_s	73.24	69.31	77.96	87.38	83.93	91.24	74.95	71.89	78.80	82.82	82.89	82.81	80.39	75.29	86.38
w/o f_e	73.56	70.93	76.59	-	-	-	76.16	75.22	77.21	73.11	72.26	74.18	-	-	-
w/o f_{ca}	71.43	68.91	74.46	87.51	83.90	91.57	-	-	-	82.72	80.46	85.25	-	-	-
w/o f_d	-	-	-	87.21	83.75	91.09	75.44	73.07	78.13	83.23	82.74	83.95	-	-	-

Table 9: Ablation study on ECPE, ECE, and CCRC tasks.

and negative sampling. Each dataset is randomly split into ten folds for cross validation.

5.2 Implementation Details

The optimizer is AdamW (Loshchilov and Hutter, 2017). The batch size and learning rate are set to 8 and 1e-5, respectively. M is set to 3. The weight decay is set to 0.01 while other parameters of $\beta 1$, $\beta 2$ and ϵ are set to 0.9, 0.999 and 1e-8 by default. The dropout rate of the attention layer and hidden layer in BERT are both slightly modified to 0.2. The prompt method is implemented based on the BERT initialized with "BERT-Base, Chinese"[1] to achieve a fair comparison since the selected baselines are mostly based on BERT. All experiments are run on the machine containing a piece of RTX 3090 GPU.

During testing, to get a fair comparison on the original bias dataset, the indicate token in the word label set for the prediction of the directional constraint module is restricted to a smaller boundary, which achieves better performance. Specifically, for the i-th clause, the prediction result of the associated clause will be between $i - l$ and $i + l$. The experiment shows that our method achieves the best results when $l = 2$.

5.3 Baselines and Evaluation Metrics

The proposed UECA-Prompt is compared with several state-of-the-art methods for different ECA tasks.

Baselines for ECE. Baselines include statistical learning methods, **Multi-kernel** (Gui et al., 2016b); And deep learning methods, **MANN** (Li et al., 2019), **RHNN** (Fan et al., 2019), **RTHN** (Xia et al., 2019), **FSS-GCN** (Hu et al., 2021b) , and **EF-BHA** (Hu et al., 2021a).

Baselines for ECPE. Baselines include two-step methods, **Inter-EC** (Xia and Ding, 2019); And end-to-end methods, **UTOS** (Cheng et al., 2021), **PairGCN** (Chen et al., 2020c), **TransECPE** (Fan

et al., 2020), **MTST** (Fan et al., 2021), **RANK-CP** (Wei et al., 2020), and **ECPE-MLL** (Ding et al., 2020b).

Baselines for CCRC. Baselines include **BiL-STM + Concatenation** (Chen et al., 2020a), **BiL-STM + BiLSTM** (Chen et al., 2020a), and **BiL-STM + Self-Attention** (Chen et al., 2020a).

Evaluation Metrics. Following the previous works (Xia and Ding, 2019; Fan et al., 2020), this paper adopts the precision (P), recall (R), and F1 score (F1) as the metrics for evaluation. The final results are obtained by averaging the ten-fold results.

5.4 Main Results

UECA-Prompt produces competitive results when compared with the other baselines on three tasks.

Results on ECPE. Table 2 reports the results of three task objectives of ECPE. The competitive performance of our method is mainly attributed to the significant improvement of the recall. This is because the sub-prompt modules in UECA-Prompt pay more attention to global information of the entire document rather than the local prediction for a single clause. However, this will lead to the loss of precision to a certain extent.

The comparison between UECA-Prompt and UECA-Prompt (m2m) shows that the performance of our method with the M2M module is not always superior. This is because multi-emotion samples only account for a small proportion of the dataset. Additional experiments on those multi-emotion samples are conducted. The results in Table 3 show that the method with M2M module obtains 2.3% improvements on ECPE. This indicates that the incorporation of the M2M module helps better handle the multi-emotion instances.

Results on ECE. The result in Table 4 demonstrates that UECA-Prompt obtains better results than RHNN (+5.26% in F_1). This shows the clear advantage of modeling the emotional causality through constructing prompt. Different from

---

[1]https://github.com/huggingface/transformers

Method	ECPE			Emotion Extraction			Cause Extraction			Method	ECE		
	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)		F_1 (%)	P (%)	R (%)
UTOS-BERT	34.14	42.76	28.95	66.45	83.69	56.08	36.29	45.83	30.55	PAE$^\diamond$	38.51	55.11	30.78
ECPE-MLL-bert	45.57	**61.53**	36.39	**80.33**	77.15	84.04	54.68	51.40	59.01	PAEDGL$^\diamond$	40.96	55.25	32.79
MTST+Refinement	44.93	51.99	40.34	64.10	**77.65**	55.46	48.84	**56.25**	43.96	RTHN$^\diamond$	54.45	54.67	54.66
UECA-Prompt	49.37	46.30	**53.22***	76.15	68.68	**85.67***	50.64	45.75	57.71	EF-BHA$^\diamond$	60.61	56.40	65.49
UECA-Prompt (m2m)	**49.93***	48.60	52.11	75.66	70.66	81.68	**51.28***	44.76	**61.23***	UECA-Prompt	**65.62***	**63.67***	**68.12***

Table 10: Experimental results on the de-bias dataset for ECE and ECPE. $\diamond$ indicates the results are reported in (Hu et al., 2021a). $*$ indicates statistically significant improvement ($p < 0.01$) over the best baseline.

Method	ECPE			Emotion Extraction			Cause Extraction			ECE			CCRC		
	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)
UECA-Prompt	73.39	70.01	77.26	87.64	84.26	91.43	75.35	72.06	79.21	83.19	82.74	83.73	81.18	76.35	86.76
UECA-Prompt+ECE	74.86	72.59	77.44	88.23	84.94	91.90	76.80	75.06	78.78	-	-	-	81.18	76.35	86.76
UECA-Prompt+ECPE	-	-	-	-	-	-	-	-	-	83.46	84.32	82.70	80.98	76.08	86.78
UECA-Prompt+CCRC	74.70	71.82	77.99	88.16	84.75	91.95	77.55	76.24	79.16	84.40	84.57	84.27	-	-	-

Table 11: The experimental results of UECA-Prompt under cross-task training and non-cross-task training. UECA-Prompt+ECE, UECA-Prompt+ECPE and UECA-Prompt+CCRC are the results with cross-taks training.

ECPE, ECE is essentially single-task learning. The improvement of our method in precision on the ECE task verifies our conjecture that multi-task learning will lead to a decrease in precision.

Results on CCRC. Table 5 shows that UECA-Prompt significantly outperforms BiL-STM+BiLSTM (+11.42% in F_1). This indicates that UECA-Prompt can capture global context information, which is essential in emotion cause analysis.

5.5 Experimental Results Under Few-shot Setting Scenario

To further explore the potential of our method, few-shot setting experiments with 10% of training data are conducted. This section reports the experimental results under few-shot setting scenario. We compare the proposed UECA-Prompt with some state-of-the-art methods under the same experimental setting.

Results on ECPE. The results on ECPE task are shown in Table 6. UECA-Prompt obtains the best result toward the emotion cause pair extraction objective. It further proves that UECA-Prompt has the advantage in modeling the relation between emotion and cause. On the emotion extraction and cause extraction objectives, UECA-Prompt is slightly inferior to ECPE-MLL. This may be due to the data bias brought to ECPE-MLL.

Results on ECE. As shown in the last row of Table 7, UECA-Prompt achieves the best results on all evaluation metrics. Our method even approximates or exceeds the model trained under complete training sets (See Table 4). Compared with other

methods, UECA-Prompt can better capture the association between emotion and cause with a few training samples.

Results on CCRC. UECA-Prompt also outperforms the state-of-the-art methods on the CCRC task. We observe that the simpler approach works better. For example, simply feature concatenation is more excellent than encoding features with BiL-STM or self-attention.

5.6 Ablation Study

UECA-Prompt is comprised of four sub-prompt components: emotion indicator function f_e, cause indicator function f_{ca}, directional constraint f_s, and sequential learning f_d. To verify the effect of different components, ablation experiments are carried out for different modules. The results are given in Table 9.

The performance of UECA-Prompt on the ECPE task integrally declines without any of the four components. This indicates that each of the four sub-prompt components plays distinct roles in feature learning, thereby proving the effectiveness of the four components. Specifically, the performance drops sharply without the cause indicator function. This observation indicates that the extraction of the cause clause is predominant in the ECPE task.

The influence of the emotion indicator function is significant on ECE tasks because the emotion clause is the most important information in ECE. Furthermore, UECA-Prompt can still achieve better performance than Multi-Kernel without emotional information on ECE tasks. This means UECA-Prompt is competent to obtain useful information from the context.

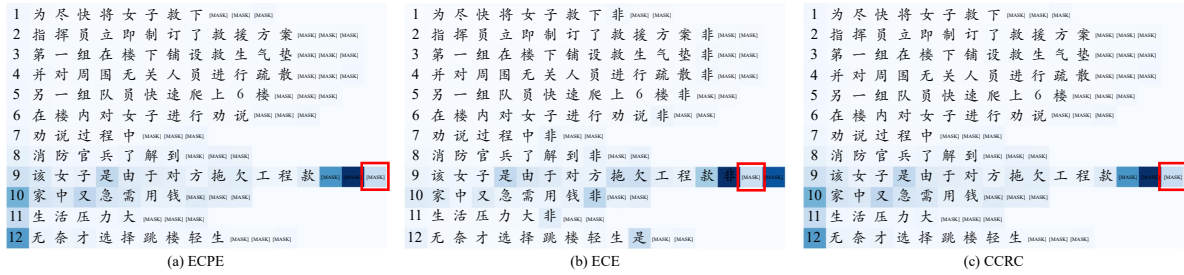The result for the CCRC task shows that the

Figure 4: Visualization of the token for different tasks. The Chinese token "是" means "is", while "非" means "isn't". The intensity of color is proportional to the weight value, which is the mean result of all heads and all layers in BERT.

performance of UECA-Prompt slightly degrades without the component of sequential token learning. This is mainly due to the change in precision. This indicates that eliminating the learning effect of the sequential learning module would impair the directional constraint module to extrapolate the correct results.

5.7 Analysis

Results on De-bias Data To verify the ability of UECA-Prompt to ease the bias caused by relative position, this paper conducts the experiments on de-bias datasets (Ding and Kejriwal, 2020). The results are shown in Table 10. The experiments are only conducted on ECPE and ECE because the CCRC task is position-irrelevant. Our method gains at least 5.01% improvement of F_1 on the ECE task and attains a 3.8% improvement of F_1 on the ECPE task. The results show that UECA-Prompt is more robust than baselines.

Commonalities Among Tasks To further explore the universality of UECA-Prompt, we train the model with the cross-task training method. The experimental results are shown in Table 11. It could be observed that the cross-task training method improves the performance of UECA-Prompt on three tasks. This indicates that UECA-Prompt can learn the commonalities among ECA tasks. It is noteworthy that the model firstly trained on the CCRC task (UECA-Prompt+CCRC) achieves the best performance. This result shows the importance of contextual information and demonstrates that UECA-Prompt could discriminate different contexts.

6 Case Study

To further understand the operation principle of UECA-Prompt, Fig. 4 visualizes the attention for different tokens in the constructed prompt text of a Chinese text. The subgraph (a), (b), and (c) in Fig. 4 represent the attention weights in ECPE, ECE, and CCRC, respectively. The query token is marked with the red box.

It can be observed that the attention weights are mainly concentrated in the answer slot [M] and the context in emotion and cause clauses. This indicates that different sub-prompt modules in UECA-Prompt could capture the key information of emotion and cause clauses as well as cooperate.

7 Conclusions

This paper proposes a universal prompt method for emotion cause analysis tasks. UECA-Prompt could uniformly model different ECA tasks by decomposing ECA tasks into multiple objectives and converting these objectives into sub-prompts. Meanwhile, the proposed directional constraint module and sequential learning module could effectively ease the bias caused by position information. Moreover, the cross-task training method further improves the performance of UECA-Prompt. The ability of UECA-Prompt to learn commonalities and contextual knowledge from different tasks is verified. The experimental results on three ECA tasks demonstrate the effectiveness of the proposed method.

This work chooses the general model, BERT, for a fair comparison. The sentiment-related PLM, such as SKEP (Tian et al., 2020), may further improve the performance of UECA-Prompt, which will be explored in our future work.

References

Hongliang Bi and Pengyuan Liu. 2020. ECSP: A new task for emotion-cause span-pair extraction and classification. *arXiv preprint arXiv:2003.03507*.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amode. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Rich Caruana. 1997. Multitask learning. *Machine learning*, 28(1):41–75.

Xinhong Chen, Qing Li, and Jianping Wang. 2020a. Conditional causal relationships between emotions and causes in texts. In *Proceedings of EMNLP*, pages 3111–3121.

Xinhong Chen, Qing Li, and Jianping Wang. 2020b. A unified sequence labeling model for emotion cause pair extraction. In *Proceedings of CCL*, pages 208–218.

Ying Chen, Wenjun Hou, Shoushan Li, Caicong Wu, and Xiaoqiang Zhang. 2020c. End-to-end emotion-cause pair extraction with graph convolutional network. In *Proceedings of CCL*, pages 198–207.

Xiyao Cheng, Ying Chen, Bixiao Cheng, Shoushan Li, and Guodong Zhou. 2017. An emotion cause corpus for chinese microblogs with multiple-user structures. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(1):1–19.

Zifeng Cheng, Zhiwei Jiang, Yafeng Yin, Na Li, and Qing Gu. 2021. A unified target-oriented sequence-to-sequence model for emotion-cause pair extraction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2779–2791.

Jiayuan Ding and Mayank Kejriwal. 2020. An experimental study of the effects of position bias on emotion causeextraction. *arXiv preprint arXiv:2007.15066*.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020a. ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. In *Proceedings of ACL*, pages 3161–3170.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020b. End-to-end emotion-cause pair extraction based on sliding window multi-label learning. In *Proceedings of EMNLP*, pages 3574–3583.

Chuang Fan, Hongyu Yan, Jiachen Du, Lin Gui, Lidong Bing, Min Yang, Ruifeng Xu, and Ruibin Mao. 2019. A knowledge regularized hierarchical approach for emotion cause analysis. In *Proceedings of EMNLP-IJCNLP*, pages 5614–5624.

Chuang Fan, Chaofa Yuan, Jiachen Du, Lin Gui, Min Yang, and Ruifeng Xu. 2020. Transition-based directed graph construction for emotion-cause pair extraction. In *Proceedings of ACL*, pages 3707–3717.

Chuang Fan, Chaofa Yuan, Lin Gui, Yue Zhang, and Ruifeng Xu. 2021. Multi-task sequence tagging for emotion-cause pair extraction via tag distribution refinement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2339–2350.

Tianyu Gao, Adam Fisch, and Danqi Chen. 2021. Making pre-trained language models better few-shot learners. In *ACL/IJCNLP (1)*.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016a. Event-driven emotion cause extraction with corpus construction. In *Proceedings of EMNLP*, pages 1639–1649.

Lin Gui, Ruifeng Xu, Qin Lu, Dongyin Wu, and Yu Zhou. 2016b. Emotion cause extraction, a challenging task with corpus construction. In *Chinese National Conference on Social Media Processing*, pages 98–109.

Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, et al. 2021. Pre-trained models: Past, present and future. *AI Open*, 2:225–250.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799. PMLR.

Guimin Hu, Guangming Lu, and Yi Zhao. 2021a. Bidirectional hierarchical attention networks based on document-level context for emotion cause extraction. In *Findings of EMNLP*, pages 558–568.

Guimin Hu, Guangming Lu, and Yi Zhao. 2021b. Fss-gcn: A graph convolutional networks with fusion of semantic and structure for emotion cause analysis. *Knowledge-Based Systems*, 212:106584.

Chengxi Li, Feiyu Gao, Jiajun Bu, Lu Xu, Xiang Chen, Yu Gu, Zirui Shao, Qi Zheng, Ningyu Zhang, Yongpan Wang, et al. 2021a. Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis. *arXiv preprint arXiv:2109.08306*.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*.

Xiangju Li, Shi Feng, Daling Wang, and Yifei Zhang. 2019. Context-aware emotion cause analysis with multi-attention-based neural network. *Knowledge-Based Systems*, 174:205–218.

Xiangju Li, Wei Gao, Shi Feng, Daling Wang, and Shafiq Joty. 2021b. Span-level emotion cause analysis by bert-based graph attention network. In *Proceedings of CIKM*, pages 3221–3226.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Ruotian Ma, Xin Zhou, Tao Gui, Yiding Tan, Qi Zhang, and Xuanjing Huang. 2021. Template-free prompt tuning for few-shot ner. *arXiv preprint arXiv:2109.13532*.

Fabio Petroni, Patrick Lewis, Aleksandra Piktus, Tim Rocktäschel, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2020. How context affects language models' factual predictions. *arXiv preprint arXiv:2005.04611*.

Victor Sanh, Albert Webson, Colin Raffel, Stephen H Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, et al. 2021. Multitask prompted training enables zero-shot task generalization. *arXiv preprint arXiv:2110.08207*.

Timo Schick, Helmut Schmid, and Hinrich Schütze. 2020. Automatically identifying words that can serve as labels for few-shot text classification. In *Proceedings of CCL*, pages 5569–5578.

Timo Schick and Hinrich Schütze. 2020a. Exploiting cloze questions for few shot text classification and natural language inference. *arXiv preprint arXiv:2001.07676*.

Timo Schick and Hinrich Schütze. 2020b. It's not just size that matters: Small language models are also few-shot learners. *arXiv preprint arXiv:2009.07118*.

Hao Tang, Donghong Ji, and Qiji Zhou. 2020. Joint multi-level attentional model for emotion detection and emotion-cause pair extraction. *Neurocomputing*, 409:329–340.

Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4067–4076, Online. Association for Computational Linguistics.

Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2021. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*.

Penghui Wei, Jiahao Zhao, and Wenji Mao. 2020. Effective inter-clause modeling for end-to-end emotion-cause pair extraction. In *Proceedings of ACL*, pages 3171–3181.

Sixing Wu, Fang Chen, Fangzhao Wu, Yongfeng Huang, and Xing Li. 2020. A multi-task learning neural network for emotion-cause pair extraction. In *ECAI 2020*, pages 2212–2219. IOS Press.

Rui Xia and Zixiang Ding. 2019. Emotion-cause pair extraction: A new task to emotion analysis in texts. In *Proceedings of ACL*, pages 1003–1012.

Rui Xia, Mengran Zhang, and Zixiang Ding. 2019. Rthn: A rnn-transformer hierarchical network for emotion cause extraction. *arXiv preprint arXiv:1906.01236*.

Bo Xu, Hongfei Lin, Yuan Lin, and Kan Xu. 2021. Two-stage supervised ranking for emotion cause extraction. *Knowledge-Based Systems*, 228:107225.

Ruifeng Xu, Jiannan Hu, Qin Lu, Dongyin Wu, and Lin Gui. 2017. An ensemble approach for emotion cause detection with event extraction and multi-kernel svms. *Tsinghua Science and Technology*, 22(6):646–659.

Ningyu Zhang, Luoqiu Li, Xiang Chen, Shumin Deng, Zhen Bi, Chuanqi Tan, Fei Huang, and Huajun Chen. 2021. Differentiable prompt makes pre-trained language models better few-shot learners. *arXiv preprint arXiv:2108.13161*.