Multi-Task Sequence Tagging for Emotion-Cause Pair Extraction Via Tag Distribution Refinement

Chuang Fan, Chaofa Yuan, Lin Gui, Yue Zhang, *Member, IEEE*, and Ruifeng Xu, *Member, IEEE*

***Abstract*—The task emotion-cause pair extraction deals with finding all emotions and the corresponding causes from emotion texts. Existing joint methods solve it as multi-task learning, which introduces two auxiliary tasks (i.e., emotion extraction and cause extraction) to make use of task correlations for their mutual benefits. However, these methods focus on capturing such correlations by sharing parameters in an implicit way, not only have a limitation of cannot explicitly model their information interaction, but also suffer from low interpretability. Towards these issues, we propose a multi-task sequence tagging framework, which can extract emotions with the associated causes simultaneously by encoding their distances into a novel tagging scheme. In addition, the output of both auxiliary tasks can be directly used as inductive bias, to refine the tag distribution for benefiting emotion-cause pair extraction, so that the information exchange between them can be more explicit and interpretable. Results show that our model achieves the best performance, outperforming a number of competitive baselines by at least 1.03% ($p < 0.01$) in F_1 score. The comprehensive analysis further confirms the superiority and robustness of our model.**

***Index Terms*—Emotion-cause pair extraction, sequence tagging, multi-task learning, tag distribution refinement.**

I. Introduction

EMOTIONS play an important role in human communication and decision making [1]. Previous studies on emotion analysis focus on emotion classification [2]–[6], which aims to infer polarities and/or retrieve opinions from texts. However, sometimes, we may care more about the stimuli, or the cause why the people hold or change the emotion rather than a simple category label. Accordingly, emotion cause extraction [7], which aims at inferring the reason behind an emotion expression, has attracted increasing interest. Traditional approaches for emotion cause extraction depend on linguistic-based rules [8]–[11] or feature engineering [12]–[16], which are time-consuming and labor-intensive. Recent studies have solved this task using neural models with well-designed attention mechanisms [17]–[19]. Gui *et al.* [17] proposed a convolution-based attention mechanism to store relevant context in different memory slots to better capture word-level sequence features. Transformer based models [18] achieved its superior performance to the multi-layer and multi-head self-attention architecture. Also, emotion cause extraction can benefit from the use of external sentiment resources and prior knowledge for parameter constraints [19].

TABLE I
An Example Adopted From the Dataset Proposed by Gui *et al.* [15]. Darker Color Indicates Emotions While Lighter Color for the Associated Causes. Superscripts Denote Clause IDs and "PER" Replaces the Person Name

Example
[Through the efforts of PER]^1, [he and his several like-minded friends established this photography studio last year]^2. [With more and more orders from customers] ^{3}, [although PER often stays up late to fix films]^4, [he enjoys it] ^{5}, [but this happiness is always with a layer of worry] ^{6}, [his parents still do not know about his resignation] ^{7}, [and he does not know how to speak to his parents]^8.

Although much progress has been made in the theories, methods and experiments that support emotion cause extraction, existing works require that emotions must be annotated before extracting the causes, and this ignores the mutual benefits of emotion-cause structure, restricting the range of applications in real-world scenarios. Towards these issues, Xia and Ding [20] presented a new task named emotion-cause pair extraction (ECPE), aiming to extract all potential pairs of emotions and the corresponding causes from unannotated texts. Consider the emotion text shown in Table I. Here are two emotion-cause pairs: (*he enjoys it*, *with more... customers*) and (*but this... worry*, *his parents... resignation*). We can observe that even in the same emotion text, there may be multiple opposite sentiment polarities expressed by different affective words (e.g., enjoys and worry), and associated with different causes. This suggests that we need a comprehensive understanding of text content and structure to perform causal reasoning and identify emotion-cause pairs from negative ones.

In general, emotion-cause pair extraction is a more challenging task due to the inherent ambiguity and subtlety of emotions.

Manuscript received October 30, 2020; revised February 15, 2021 and May 7, 2021; accepted May 31, 2021. Date of publication June 16, 2021; date of current version July 21, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 61632011, 61876053, and 62006062, in part by Shenzhen Foundational Research Funding under Grants JCYJ20180507183527919, JCYJ20180507183608379, and in part by Guangdong Province COVID-19 Pandemic Control Research Funding under Grant 2020KZDZX1224. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zoraida Callejas.
(Corresponding author: Ruifeng Xu.)

Chuang Fan and Chaofa Yuan are with the School of Computer Science, and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China, and also with the Joint Lab of HITSZ, China Merchants Securities Company, Ltd, Shenzhen 518000, China (e-mail: fanchuanghit@gmail.com; bruceyuan123@gmail.com).

Lin Gui is with the Department of Computer Science, University of Warwick, Coventry CV47AL, U.K (e-mail: lin.gui@warwick.ac.uk).

Yue Zhang is with the School of Engineering, Westlake University, Hangzhou 310024, China (e-mail: yue.zhang@wias.org.cn).

Ruifeng Xu is with the School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China, and also with Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: xuruifeng@hit.edu.cn).

Digital Object Identifier 10.1109/TASLP.2021.3089837

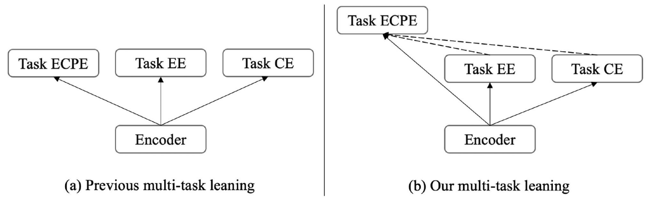


Fig. 1. Previous multi-task learning vs. Our multi-task learning. EE and CE are abbreviations of emotion extraction and cause extraction, respectively.

Existing two-step solutions have shown its effectiveness on this task [20], which involve two procedures: 1) Extracting emotions/causes from texts; 2) Pairing them and filtering out negative pairs from all possible pairs. However, such pipelined framework is suboptimal, leading to a drawback that mutually beneficial of related tasks cannot be fully exploited. Besides, error propagation may occur. As shown in Fig. 1(a), given the strong correlation between emotion/cause extraction and ECPE, multi-task learning with a shared encoder has been investigated for modeling their mutual benefits [21]–[23]. Though achieving the promising performance, we argue that it is risky to simply rely on the parameter sharing to summarize or memorize such multi-aspect correlations, e.g., the influence of predicted results of EE/CE for ECPE and the relative distance information between emotions and causes. In addition, the interpretability of how the emotion and cause information guide the emotion-cause pair extraction is still weak, since the information exchange between them is implicit.

In this paper, we propose a multi-task sequence tagging framework with tag distribution refinement to address the above issues. First, we regard emotion-cause pair extraction as a sequence tagging problem. The challenge is to also include causal reasoning into the tagging scheme, which means that the traditional BIO tagging is not suitable for this task. Accordingly, we design a tagging scheme with multiple labels, in which the information of both emotions and the associated causes can be contained by coding the distance between linked components. In this way, the emotion-cause structure is integrated into a unified framework, including representation learning, components extraction, and causal reasoning.

Second, as shown in Fig. 1(b), the proposed framework also solves ECPE with the help of two auxiliary tasks (i.e., EE and CE), enabling the relevance of these tasks to be exploited. On the top of the multi-task framework, the predicted tag distribution of each clause is refined through an offset vector, which is produced by directly using the output of the two auxiliary tasks and the relative distance between two clauses. This operation is performed explicitly on the outputs instead of only sharing information through learning common feature representations. As a result, our approach not only allows shared features, but also models the relevance of different tasks explicitly and reduces the predicted variance to improve the performance of each other, thereby maximizing the mutual benefits and making the interaction procedure more interpretable.

We conduct experiments on the benchmark dataset [15]. Results show that our approach achieves the best performance, outperforming a number of competitive baselines by a large margin on emotion-cause pair extraction (at least +1.03% in F_1 measure). Extensive analysis further confirms the superiority and robustness of our approach. The source code is avaiable at https://github.com/HLT-HITSZ/MTST-ECPE.

II. Related Work

A. Emotion Classification

In the field of affective computing, emotions have been described as discrete and consistent responses to events with significance for the organism [24]. There are two different views on emotion representation [25]. The first indicates that basic emotions have evolved through natural selection. Plutchik [26] proposed eight basic emotions: sadness, anger, disgust, fear, surprise, curiosity, acceptance, and joy. Following this study, Ekman further derived six primary emotions, namely happiness, sadness, fear, anger, disgust and surprise [27]. In the second view, according to cognition, the emotions can be mapped into the valence, arousal, and dominance (VAD) dimensions. Valence goes from very positive feelings to very negative; Arousal indicates states like sleepy to excited; and dominance relates to the strength of emotions [28], [29].

Various studies have been conducted to identify different categories of emotions. For example, Xu *et al.* [30] proposed a coarse-to-fine analysis strategy considering the similarity and adjacency of sentences. Beck *et al.* [31] proposed a low-rank coregionalization approach, which combines a vector-valued Gaussian Process with a rich parameterisation scheme to learn correlations and anti-correlations between emotions. Li *et al.* [32] converted sentence-level emotion classification into a factor graph inferring problem in which the label and context dependence are both modeled as factor functions. Relations of different emotions are also incorporated into the learning algorithm to improve the accuracy of emotion classification [33]. Chang *et al.* [6] proposed a principle-based approach to learn emotion templates. Such feature-based methods have been investigated not only for text-based data, but also for voice [34], [35] and facial imagery [36]–[41]. Recent studies have paid attention to solving the task using deep neural models with well designed attention mechanisms [42]–[46]. These models are powerful to learn relevant and complex feature representations without using any traditional hand-crafted features. However, classification-based emotion analysis focuses on the emotion expressions that have been observed, and may ignore the evolvement of human emotions, such as the provocation, evolution, and aftermath of emotions.

B. Emotion Cause Extraction

To capture fine-grained information concerning emotions, more studies have sought to extract key elements for the provoked emotions, such as discovering the cause or the stimuli behind an emotion expression. Lee *et al.* [7] first proposed emotion cause extraction task and defined it as a word-level extraction task. They manually constructed a dataset from the Academia Sinica Balanced Chinese Corpus with cause event annotation, and generalized a series of linguistics rules for this task. Machine learning methods with hand-crafted features and classifiers such as support vector machines (SVMs) and conditional random fields (CRFs) were also adopted to detect emotion causes [9], [13]. In addition, there are also some works on cause detection for Chinese microblogs using a multiple-user structure dataset and formalized two cause detection tasks (current-subtweet-based cause detection and original-subtweet-based cause detection) [47]–[50].

Chen *et al.* [51] converted the task from word-level to clause-level and extracted causes using six groups of manually constructed linguistic cues. Following this task setting, Gui *et al.* [15] released a Chinese corpus collected from SINA city news,[1] which was followed by most recent studies in this field. Gui *et al.* further [17] proposed a convolutional attention mechanism to store relevant context in different memory slots to model context information of words. The context around the emotion words was also considered to better model the mutual impacts between candidate emotions and the associated causes [52]. Li *et al.* [53] proposed a multi-attention-based neural model to capture the mutual influences of emotion-cause structure to generate better representations. Xu *et al.* [16] proposed a learning to re-rank method involving both emotion-dependent and emotion-independent features to detect causes. Fan *et al.* [19] incorporated sentiment- and position-based regularization to restrain the parameter learning. The hierarchical network architecture [54] and Transformer based model [55] were also explored for emotion cause extraction.

C. Emotion-Cause Pair Extraction

Existing approaches on emotion cause extraction rely on emotion annotations, which are time-consuming and expensive, limiting the applications in real-word scenarios. Targeting this issue, Xia and Ding [20] proposed a novel task based on ECE, namely emotion-cause pair extraction (ECPE), the goal is to extract emotions and the corresponding causes from unannotated emotion texts. Accordingly, they tackled this task in two subtasks: 1) Extracting emotion and cause clauses separately; 2) Training a classifier to filter out negatived pairs. However, due to inherent drawback of the pipelined framework, error propagation may occur from the first procedure to the second. Recent studies have attempted to solve this task using a unified framework. For instance, [56] regarded the task as a link prediction problem and learned to link from emotions to causes by using a vanilla multi-task framework. [57] explored a multi-level attention mechanism to model the relationship between two clauses in emotion-cause structure. Fan *et al.* [22] converted the task into a procedure of parsing-like directed graph construction and designed a novel transition-based system to incrementally generate the directed graph with labeled edges, from which they can recognize emotions and the corresponding causes simultaneously. Ding *et al.* [21] represented emotion-cause pairs by using a 2D representation scheme and integrated the 2D emotion-cause pair representation, interaction, and prediction into a joint framework. Wei *et al.* [23] tackled emotion-cause pair extraction from a ranking perspective, applying graph attention to learn the feature representations by considering the interrelations between clauses, and enhanced the representations with kernel-based relative position embedding for effective ranking.

Our model differs from existing works in two main aspects. First, we formulate ECPE as a sequence tagging problem and design a novel tagging scheme accordingly, so that each input can be parsed with linear time complexity. Second, we achieve information exchange between ECPE and EE/CE using a tag distribution refinement strategy, such explicit modeling between related tasks has been shown useful in various NLP researches [58]–[61], which can not only improve the performance, but also make the improvements more interpretable.

[1][Online]. Available: http://news.sina.com.cn/society/

III. METHOD

A. Task Definition

A formal definition of emotion-cause pair extraction is described in [20]. Briefly, given an emotion document $X = (x_1, x_2, \ldots, x_n)$ consisting of n manually segmented clauses, with several emotions and at least one cause corresponding to each emotion. The goal of ECPE is to output all potential clause pairs where hold a causal relation:

$$P = \{\cdots, (x_i^e, x_j^c), \cdots\} \quad (1 \le i, j \le n) \tag{1}$$

where x_i^e is an emotion clause, and x_j^c is the corresponding cause clause. As shown in Table I, this task is defined at the clause level. That is, in this paper, the “emotion” and “cause” are refer to “emotion clause” and “cause clause,” respectively.

B. Model Overview

As shown in Fig. 2, our model receives a document as input at each time and assigns a task-specific tag to every clause, from which emotions with the associated causes can be extracted simultaneously. To be specific, we first design a novel tagging scheme (Section III-C) and then use BERT [62] with a bidirectional LSTM [63] as the encoding layer, to extract both the word- and clause-level sequential context. The outputs of encoding layer, along with the predicted tags are fed into a unidirectional LSTM to generate the final hidden states, which are used to sequentially predict the distribution of emotion, cause, and tag for each clause, respectively (Section III-D). We further refine the tag distribution of each clause by directly using the output of emotion extraction and cause extraction, to model their mutual benefits more explicit (Section III-E).

C. The Tagging Scheme

Given a document $X = (x_1, x_2, \ldots, x_n)$ with n clauses and corresponding linguistic labels $Y = (y_1^t, y_2^t, \ldots, y_n^t)$ with equal length, our goal is to learn a parameterized mapping function $f_\theta : X \to Y$ from input clauses to ECPE-specific tags. The traditional BIO tags are not suitable for this task, since we need to identify the emotion causality between two clauses which may be discontinuous in an emotion text.

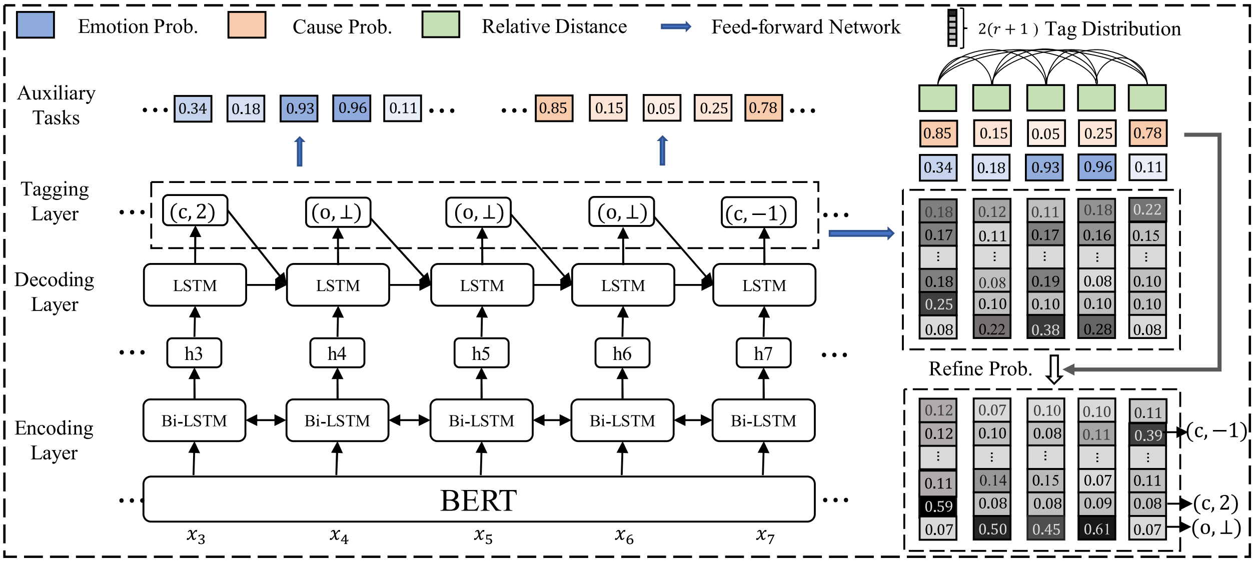


Fig. 2. The architecture of our model (take the example in Table I as an input for ease of illustration). `Prob.` is the abbreviation for `Probability`.

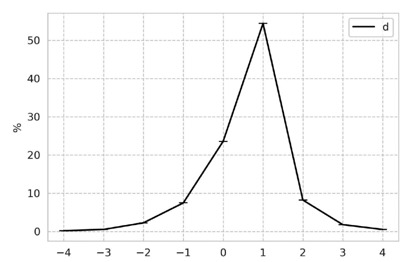


Fig. 3. Distribution of distances d between emotion-cause structure in the dataset released by Gui *et al.* [15].

We tackle this challenge by focusing on the identification of causes[2] and designing a novel tagging scheme to include the emotion causality into the tags. In particular, we sequentially tagging each clause $x_i \in X$ with a two-tuple label $y_i^t = (b, d) \in Y$, where $b \in \{\mathrm{C}, \mathrm{O}\}$ and $d \in \{-(n-1), \ldots, -1, 0, 1, \ldots, n-1, \perp\}$. Tag "C" represents the "cause" tag, which means that the current clause is a cause, while the tag "O" represent the "other" categories. d encodes the distance between a cause and its triggered emotion, e.g., "-1" denotes that the **previous clause** is the corresponding emotion, while "1" the **succeeding clause**. The special symbol $\perp$ indicates when a particular slot is not filled (e.g., a non-cause clause $b = \mathrm{O}$ has no related emotion, thus it always associates with the symbol $\perp$).

For each emotion text X, the total number of tags in Y is $N_t = 2 * (n-1) + 1 + 1$, which relies on the length of X, resulting in inconsistency during training. Empirically, in emotion events, causes usually occur at positions which are very close to the emotions. As shown in Fig. 3, ~55% of emotion-cause pairs have a distance tag "1," that is, emotions are behind the causes they attach to. Overall, ~95% of emotion-cause distances lie in {-2, -1, 0, 1, 2}. Thus, we could use a hyperparameter r to limit the range of emotion that is associated with the current cause (i.e., $d \in \{-r, \ldots, -1, 0, 1, \ldots, r, \perp\}$). Then, we have a total number of $N_t = 2(r+1)$ tags, which can keep consistent during training.

For example, in Table I, we label the third clause with a tag $(C, 2)$, given it is the cause of the fifth clause which **behind** it with a distance of 2. Similarly, the seventh clause is assigned to $(C, -1)$, since its emotion clause **before** it with a distance of 1. Other clauses are assigned to $(O, \perp)$. That is, the whole texts are labeled as: $[(O, \perp)^1, \ldots, (C, 2)^3, \ldots, (C, -1)^7, (O, \perp)^8]$.

[2]If we focus on the identification of emotions, we cannot assign all rational tags when an emotion is associated with multiple causes, because different causes have different relative distances to the emotion.

D. Multi-Task Sequence Tagging Framework

Sequence Tagging Encoder. Given an emotion text $X = (x_1, x_2, \ldots, x_n)$ consisting of n clauses with each clause $x_i = (w_{i1}, w_{i2}, \ldots, w_{im})$ containing m words, we first formulate each clause as a sequence $\tilde{x}_i = ([\mathrm{CLS}], w_{i1}, \ldots, w_{im}, [\mathrm{SEP}])$, where [CLS] is a special token that the final hidden state is used as the aggregate sequence features and [SEP] is a dummy token not used in this work. We then obtain a hidden representation $\mathbf{x}_i = \mathrm{BERT}(\tilde{x}_i) \in \mathbb{R}^{d_h}$ where d_h is the hidden dimension of the pre-trained BERT model. In this way, the input clauses are encoded into a sequence of distributed vectors $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}]$. Subsequently, we use a bidirectional LSTM to encode the $\mathbf{X}$ into more context-sensitive representations, which yields:

$$\overrightarrow{\mathbf{h}}_i = \overrightarrow{\mathbf{LSTM}}_{en}(\mathbf{x}_i, \overrightarrow{\mathbf{h}}_{i-1}; \theta_{\overrightarrow{\mathbf{LSTM}}_{en}}) \quad (2)$$

$$\overleftarrow{\mathbf{h}}_i = \overleftarrow{\mathbf{LSTM}}_{en}(\mathbf{x}_i, \overleftarrow{\mathbf{h}}_{i+1}; \theta_{\overleftarrow{\mathbf{LSTM}}_{en}}) \quad (3)$$

where $\overrightarrow{\mathbf{h}}_i \in \mathbb{R}^{d_r}$, $\overleftarrow{\mathbf{h}}_i \in \mathbb{R}^{d_r}$, and d_r is the hidden size of the both LSTMs. The two directional hidden states are concatenated as $\mathbf{h}_i = [\overrightarrow{\mathbf{h}}_i, \overleftarrow{\mathbf{h}}_i]$, which is fed into our decoder.

Sequence Tagging Decoder. The LSTM-Softmax approach [64] is used to model the output distribution over the class tags and to output the sequence of predicted tags. Specifically, considering the i-th clause in this document, the output of *sequence tagging encoder* $\mathbf{h}_i$ along with the last tag embedding $\mathbf{y}_{i-1}^t$ are fed into the *decoder*. Subsequently, the output of *decoder* $\mathbf{s}_i$ is transformed through a multi-layer perceptron (MLP) for the softmax layer over tag vocabularies. Formally, the label of x_i is predicted as the (Eq.5).

$$\mathbf{s}_i = \mathbf{LSTM}_{de}(\mathbf{h}_i, \mathbf{s}_{i-1}, \mathbf{y}_{i-1}^t; \theta_{\mathbf{LSTM}_{de}}) \tag{4}$$

$$p(y_i^t = l_t | x_{1:i}) = \frac{\mathbf{exp}(\mathbf{w}_t^\top \mathbf{s}_i + \mathbf{b}_t)}{\sum_{t'=1}^{N_t} \mathbf{exp}(\mathbf{w}_{t'}^\top \mathbf{s}_i + \mathbf{b}_{t'})} \tag{5}$$

where $\mathbf{w}_t$, $\mathbf{w}_{t'}$ and $\mathbf{b}_t$, $\mathbf{b}_{t'}$ denote weight vectors and bias vectors, respectively, l_t is the ground truth of clause x_i for tag prediction, and N_t is the total number of tags. We train our model to maximize the log-likelihood of the data, and the objective function for tag prediction is defined as:

$$\mathcal{O}_{tag} = \sum_{1}^{|\mathbb{D}|} \sum_{i}^{n} \log p(y_i^t = l_t | x_{1:i}; \theta_{tag}) \tag{6}$$

where $|\mathbb{D}|$ is the size of training set, n is the length of document X, and $p(y_i^t = l_t | x_{1:i}; \theta_{tag})$ is the normalized predictive probability of the tag l_t.

Multi-task Learning. Intuitively, the ECPE task can benefit from the detection of emotion and cause. For instance, if we know that a clause is an emotion, the corresponding causes have a high probability of being in its surrounding clauses. Considering the strong correlations among EE, EC, and ECPE, we extend our model to a multi-task architecture to improve the generalization by taking advantage of the inductive bias in training signals of related tasks. Specifically, we feed $\mathbf{s}_i$ into two output layers to predict whether a clause is an emotion/cause or not. Firstly, a MLP is used to predict the probability distribution of x_i being an emotion:

$$p(y_i^e = l_e | x_i) = \frac{\mathbf{exp}(\mathbf{w}_e^\top \mathbf{s}_i + \mathbf{b}_e)}{\sum_{e'=1}^{2} \mathbf{exp}(\mathbf{w}_{e'}^\top \mathbf{s}_i + \mathbf{b}_{e'})} \tag{7}$$

$$\tag{8}$$

where $\mathbf{w}_e$, $\mathbf{w}_{e'}$ and $\mathbf{b}_e$, $\mathbf{b}_{e'}$ denote weight vectors and bias vectors, respectively. l_e is the label of x_i for emotion extraction. Thus, the log-likelihood objective function for emotion extraction can be defined as:

$$\mathcal{O}_{emo} = \sum_{1}^{|\mathbb{D}|} \sum_{i}^{n} \log p(y_i^e = l_e | x_i; \theta_{emo}) \tag{9}$$

Similarity, the probability distribution of x_i being a cause and the related objective function can be obtained by:

$$p(y_i^c = l_c | x_i) = \frac{\mathbf{exp}(\mathbf{w}_c^\top \mathbf{s}_i + \mathbf{b}_c)}{\sum_{c'=1}^{2} \mathbf{exp}(\mathbf{w}_{c'}^\top \mathbf{s}_i + \mathbf{b}_{c'})} \tag{10}$$

$$\mathcal{O}_{cau} = \sum_{1}^{|\mathbb{D}|} \sum_{i}^{n} \log p(y_i^c = l_c | x_i; \theta_{cau}) \tag{11}$$

where $\mathbf{w}_c$, $\mathbf{w}_{c'}$ and $\mathbf{b}_c$, $\mathbf{b}_{c'}$ denote weight vectors and bias vectors, respectively. l_c is the label of x_i for cause extraction.

In this work, we employ the sum of the above three terms as the final objective function:

$$\mathcal{O} = \mathrm{Max}_\theta \left\{ \lambda \mathcal{O}_{tag} + \frac{1-\lambda}{2} \mathcal{O}_{emo} + \frac{1-\lambda}{2} \mathcal{O}_{cau} \right\} \tag{12}$$

where θ denotes all the parameters in this model and λ guides the model to achieve the best trade-off among the three types of objective functions.

E. Tag Distribution Refinement

The goal of refining tag distribution is to leverage the correlations between different tasks explicitly and exploit information interaction for improvements. Specifically, the tag distribution of each clause is refined by an offset vector $\mathbf{v}$, which is computed by directly using the output of two auxiliary tasks (i.e., emotion extraction and cause extraction). In this way, we can make it easy to show how the two auxiliary tasks guide the emotion-cause pair extraction intuitively.

Formally, for each $x_i \in X$, we first consider a clause window consisting of x_i and r surrounding clauses located immediately to the left and right of x_i. This gives us a window of clauses in the size of $2r + 1$, named as "clause chunk" and denoted as K. We define the tag distribution refinement strategy over such a clause chunk, which we believe convey the most relevant information compared to other distant clauses.

Considering a clause chunk centered on x_i, according to our multi-task sequence tagging scheme, we have $\mathbf{p}_k^t \in \mathbb{R}^{N_t}$, $\mathbf{p}_k^e \in \mathbb{R}^2$, $\mathbf{p}_k^c \in \mathbb{R}^2$ to denote the predictive distribution of tag, emotion, and cause for each $x_k \in K$ ($k = i$ is allowed), respectively. We use $p(y_k^e = 1 | x_k)$ and $p(y_k^c = 1 | x_k)$ to denote the probability of x_k being an emotion and a cause, respectively. In particular, $p(y_k^t = j | x_k)$ can indicate the probability that x_i and x_k forms an emotion-cause pair (where $j \in [-r, r]$ encodes the distance between x_k and x_i) or indicate the probability that x_k is not a cause ($j = \perp$). Intuitively, if $p(y_i^e = 1 | x_i)$ increases, $p(y_k^t = j | x_k)$ should also increase. We model this correlation explicitly by:

$$\tilde{\mathbf{p}}_k^t = \begin{cases} \mathbf{p_k^t} + \mathbf{v_k} & p(y_i^e = 1 | x_i) > 0.5 \\ \mathbf{p}_k^t - \mathbf{v}_k & p(y_i^e = 1 | x_i) \leq 0.5 \end{cases} \tag{13}$$

where $\tilde{\mathbf{p}}_k^t$ is the refined tag distribution of x_k, $\mathbf{v}_k \in \mathbb{R}^{N_t}$ is an offset vector which is calculated by considering $p(y_i^e = 1 | x_i)$, $p(y_k^c = 1 | x_k)$, and the relative distance between x_i and x_k. According to the cohesion and coherence of discourse [65], the probability of two close clauses holding a causal relation is relatively higher than those have a long distance. Thus, to measure the importance of this distance information, we have:

$$w_k^i = 1 - \frac{|pos_k - pos_i| + \gamma}{l + 2\gamma} \tag{14}$$

where $|pos_k - pos_i|$ is the absolute value of distance between x_k and x_i. γ is a factor for smoothing purposes (we set $\gamma = 0.5$ in this work) and l is the number of clauses. In this way, the closer the distance between x_k and x_i, the greater the weight

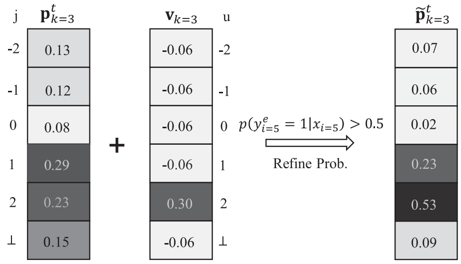


Fig. 4. An example of tag refinement about the third clause ($k = 3$) in Table I. For ease of understanding, assume $r = 2$ and $i = 5$. In this setting, $x_{i=5}$ and $x_{k=3}$ form an emotion-cause pair. During tagging, we want the model to assign the highest probability to $p^t_{k=3} = 2$, since $x_{i=5}$ behind $x_{k=3}$ with a distance of 2, but our model fail to do this. However, if we know $x_{k=3}$ and $x_{i=5}$ located in a relatively small distance, and they have higher probabilities to be a cause and an emotion, respectively, we then have a strong posterior belief to improve the probability of $p^t_{k=3} = 2$ and reduce those of others (i.e., $j = -2, -1, 0, 1, \perp$). According to Eq.14-16, we can obtain an offset vector $\mathbf{v}_{k=3}$ to refine its distribution and then may generate a better distribution $\tilde{\mathbf{p}}^t_{k=3}$ with the highest probability is assigned to $\tilde{p}^t_{k=3} = 2$.

of w^i_k. Regarding each element in $\mathbf{v}_k$, there are two situations based on our assumption. If $p(y^e_i = 1|x_i) > 0.5$,

$$v^u_k = \begin{cases} \alpha_k \cdot (1 - p(y^t_k = j|x_k)) & u = j \\ -\alpha_k \cdot (1 - p(y^t_k = j|x_k))/(N_t - 1) & u \neq j \end{cases} \quad (15)$$

where α_k is a factor to control how much probability should be transfered to tag j from other tags. Given the intuition that $p(y^e_i = 1|x_i), p(y^c_k = 1|x_k)$ and w^i_k should directly proportional to $p(y^t_k = j|x_k)$, α_k can be calculated by:

$$\alpha_k = w^i_k \cdot p(y^e_i = 1|x_i) \cdot p(y^c_k = 1|x_k) \quad (16)$$

In contrast, when $p(y^e_i = 1|x_i) \leq 0.5$, the probability of tag j should be transfered to other tags, thus:

$$v^u_k = \begin{cases} \alpha_k \cdot p(y^t_k = j|x_k) & u = j \\ -\alpha_k \cdot p(y^t_k = j|x_k)/(N_t - 1) & u \neq j \end{cases} \quad (17)$$

$$\alpha_k = (1 - w^i_k) \cdot (1 - p(y^e_i = 1|x_i)) \cdot (1 - p(y^c_k = 1|x_k)) \quad (18)$$

Based on the above procedures, the tag distribution of each $x_k \in K$ will be updated incrementally. Since $\sum \tilde{\mathbf{p}}^t_k = 1$ is guaranteed by our refinement strategy, the revised distribution $\tilde{\mathbf{p}}^t_k$ can be directly applied to extract emotion-cause pairs. Fig. 4 shows an illustration of tag distribution refinement.

For efficient decoding, we employ a greedy search algorithm to choose the maximum probability tags during testing. Besides, our refinement strategy is only performed in the inference stage, which do not introduce any additional parameters, and only influence the training of standard model parameters.

IV. Experiments

A. Experimental Settings

Dataset. We use the dataset released by Gui *et al.* [15] to conduct our experiments. To better meet the ECPE task setting, we merge the samples with same text content into one sample and label each emotion-cause pair in this sample, which is consistent with [20]–[22]. The details are listed in Table II.

TABLE II
Statistical Information of the Dataset for Evaluation

Item	Number	(%)
Texts with one emotion-cause pair	1,746	89.77
Texts with two emotion-cause pairs	177	9.10
Texts with more than two emotion-cause pairs	22	1.13
Average of clause per text	14.77	—
Max of clause per text	73	—

There are two different data split strategies in previous studies, Fan *et al.* [22] stochastically divides the corpus into a training/development/test set in a ratio of 8:1:1 and evaluates their method 20 times with different data splits, while Ding *et al.* [20] adopts 10-fold cross-validation for evaluation. In this paper, we perform our model using the both data splits to obtain comprehensive and statistically credible results.

Evaluation. We choose standard Precision(P), Recall (R) and F-measure (F_1) as the evaluation metrics and report the average results over 20/10 runs: When extracting emotion-cause pairs, we obtain emotions and causes for each text simultaneously. Thus, we also evaluate the performance of emotion extraction and cause extraction using P, R and F_1.

Hyperparameters. We use BERT$_{Chinese}$ as the basis[3] and train the model for 10 epochs in total using Adam [66] with 1e-5 learning rate for BERT parameters and 1e-3 learning rate for the rest. Grid search is performed over $r \in \{1, 2, 3, 4, 5, 6\}$ and $\lambda \in \{0.15, 0.30, 0.45, 0.60, 0.75, 0.90\}$. The hidden size of LSTMs and MLP layers is set to 256. We set the mini-batch size to 3 and set the coefficient of L_2 term to 1e-5. We also add dropout with a rate of 0.5 for each MLP layer and adopt early stopping to avoid overfitting.

Baselines. We compare our model with the below methods.

- **Indep**: Emotion extraction and cause extraction are trained independently. Then they pair them and eliminate the pairs that have no emotion causality;
- **Inter-CE**: Different from Indep, the predictions of cause extraction are used to improve emotion extraction;
- **Inter-EC**: It is similar to Inter-CE except that the predictions of emotion extraction are used to improve cause extraction. The above three models follow pipeline framework and were proposed in [20].
- **E2EECPE** [56]: A multi-task learning method, which regards emotion-cause pair extraction as a link prediction task [67] and learns to link from emotions to causes.
- **LAE-MANN** [57]: A joint model with a multi-level attention mechanism, which can capture both the word- and clause-level dependency relations for emotion extraction and emotion-cause pair extraction.
- **ECPE-2D** [21]: This method represents each emotion-cause pair by using a 2D representation scheme, and integrates the 2D emotion-cause pair representation, interaction, and prediction into a joint model.

[3][Online]. Available: https://github.com/huggingface/transformers

TABLE III
EXPERIMENTAL RESULTS ON BOTH DATA SPLITS. RESULTS ARE AVERAGES OVER 10/20 RUNS. * DENOTES THE RESULTS ARE IMPLEMENTED IN THIS PAPER ACCORDING TO THEIR RELEASED CODE, AND † DENOTES THE BASELINES ARE ALSO USING BERT AS THE BASIC ENCODER ($p < 0.01$)

Method	Emotion Extraction (%)			Cause Extraction (%)			Emotion-Cause Pair Extraction (%)		
	P	R	F_1	P	R	F_1	P	R	F_1
(a) 10-fold data splits									
Indep	83.75	80.71	82.10	69.02	56.73	62.05	68.32	50.82	58.18
Inter-CE	84.94	81.22	83.00	68.09	56.34	61.51	69.02	51.35	59.01
Inter-EC	83.64	81.07	82.30	70.41	60.83	65.07	67.21	57.05	61.28
E2EECPE	85.95	79.15	82.38	70.62	60.30	65.03	64.78	61.05	62.80
LAE-MANN†*	65.49	66.07	65.71	—	—	—	62.23	55.14	64.31
ECPE-2D†	86.27	**92.21**	89.10	73.36	69.34	71.23	72.92	65.44	68.89
TransECPE†*	88.79	83.15	85.88	78.74	66.89	72.33	77.08	65.32	70.72
RankCP†	**91.23**	89.99	**90.57**	74.61	**77.88**	76.15	71.19	**76.30**	73.60
Our-baseline	86.20	81.57	83.82	76.18	71.92	73.99	74.06	69.74	71.84
+Refinement	87.11	81.78	84.36	**79.47**	74.04	**76.66**	**77.46**	71.99	**74.63**
(b) 20-fold data splits									
Indep*	83.14	77.54	80.20	65.84	54.09	59.22	64.34	46.97	54.04
Inter-CE*	82.05	78.45	80.16	65.08	55.39	59.62	62.70	48.88	54.80
Inter-EC*	81.17	77.74	79.39	65.82	57.25	61.14	61.43	50.95	55.56
E2EECPE*	82.83	74.72	78.53	65.89	54.76	59.71	58.95	56.00	57.32
LAE-MANN†	**89.90**	80.00	84.70	—	—	—	71.10	60.70	65.50
ECPE-2D†*	83.86	88.49	86.11	71.02	64.45	67.57	66.02	63.96	64.97
TransECPE†	87.16	82.44	84.74	75.62	64.71	69.74	73.74	63.07	67.99
RankCP†	89.36	**89.48**	**89.42**	69.40	**74.71**	71.91	65.75	**73.05**	69.15
Our-baseline	84.93	78.23	81.44	73.97	67.73	70.71	71.38	65.20	68.15
+Refinement	85.93	79.93	82.82	**76.14**	70.39	**73.15**	**73.77**	68.02	**70.78**

- **TransECPE** [22]: A transition-based system which recasts emotion-cause pair extraction as a procedure of directed graph construction, from which emotions and the corresponding causes can be extracted simultaneously through a sequence of actions.
- **RankCP** [23]: The current state-of-the-art method, which solves emotion-cause pair extraction from a ranking perspective and emphasizes inter-clause modeling to perform end-to-end extraction.[4]
- **Our-baseline**: Our multi-task sequence tagging framework without using the tag distribution refinement, which is described in Section III-D.

B. Main Results

Table III reports the results on the two data splits. Regarding pipelined methods, we can observe that **Indep** yields the lowest performance, because it trains the model individually and ignores the interactive information between emotions and causes. **Inter-CE** and **Inter-EC** obtain better results by exploiting this relevance. The joint models consistently outperform pipelined methods on both data splits, demonstrating the superiority of reducing error propagation and capturing the interdependence of related tasks through parameter sharing. Among them, **E2EECPE** is a vanilla multi-task framework and performs worst. **LAE-MANN** introduces a multi-level attention mechanism to model the correlations of different task, thus obtains better results. **ECPE-2D** learns features by using a 2D representation scheme, can achieve better information interaction to improve the performance. **TransECPE** uses a well-designed transition system which can capture rich non-local features for prediction, thus outperforming the previous baselines. **RankCP** learns clause pair representations using graph attention and further enhances the representations with kernel-based relative position embedding. Besides, it leverages an external sentiment lexicon to assist emotion-cause pair extraction, thus achieving the current best performance.

Our-baseline produces comparable or better results compared to the above baselines (except **RankCP**). Since previous baselines mostly rely on Cartesian product to compute the likelihood of clause pair candidates. **Our-baseline** is designed to assign a tag to each clause, which can greatly reduce negative samples, thus yield better performance. With the tag distribution refinement, our full model can further improve the performance over all the tasks, especially for emotion-cause pair extraction (+2.79% and +2.63% in F_1 respectively). This shows the superior of exploiting the mutual benefits of related tasks explicitly. Moreover, our full model significantly exceeds **RankCP** with p less than 0.01 in t-test. Specifically, on 10-fold data splits, our full model boosts cause extraction (by 4.86% in precision and 0.51% in F_1) and emotion-cause pair extraction (by 6.27% in precision and 1.03% in F_1). While on 20-fold data splits, the results of most baselines have different decreases. Our full model still achieves higher performance for cause extraction and emotion-cause pair extraction (+1.24% and +1.63% in F_1 respectively). This again shows the clear advantage of modeling their mutual benefits in an explicit manner. An interesting observation is that the tag distribution refinement strategy can also improve emotion/cause extraction. One possible reason is that, once our model corrects two clauses as an emotion-cause pair by refining the tag distribution, which means one of the both is an emotion while the other is a cause, so that the predicted results of emotion/cause extraction can also have an opportunity to be corrected.

On the other hand, the performance of emotion extraction is worse than **RankCP**, which requires ANTUSD [68] as a

[4] [Online]. Available: https://github.com/Determined22/Rank-Emotion-Cause

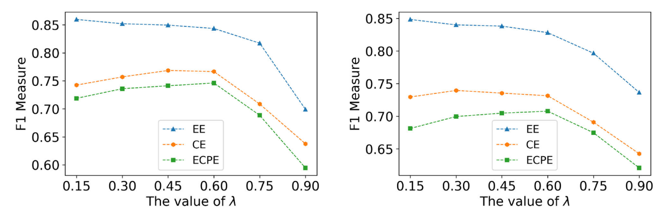
TABLE IV
EXPERIMENTAL RESULTS ON BOTH DATA SPLITS WITH DIFFERENT OBJECTIVE FUNCTIONS. RESULTS ARE AVERAGES OVER 10/20 RUNS

Method	Emotion Extraction (%)			Cause Extraction (%)			Emotion-Cause Pair Extraction (%)		
	P	R	F_1	P	R	F_1	P	R	F_1
(a) 10-fold data splits									
$\mathcal{O}_{tag}$	—	—	—	—	—	—	70.93	60.94	65.30
$\mathcal{O}_{tag} + \mathcal{O}_{emo}$	86.68	79.75	83.07	—	—	—	75.78	68.15	71.76
$\mathcal{O}_{tag} + \mathcal{O}_{cau}$	—	—	—	74.44	65.74	69.82	71.03	62.51	66.50
$\mathcal{O}_{tag} + \mathcal{O}_{cau} + \mathcal{O}_{emo}$	**87.11**	**81.78**	**84.36**	**79.47**	**74.04**	**76.66**	**77.46**	**71.99**	**74.63**
(b) 20-fold data splits									
$\mathcal{O}_{tag}$	—	—	—	—	—	—	62.40	52.06	56.77
$\mathcal{O}_{tag} + \mathcal{O}_{emo}$	84.48	77.77	80.99	—	—	—	70.90	64.52	67.56
$\mathcal{O}_{tag} + \mathcal{O}_{cau}$	—	—	—	72.75	62.06	66.98	68.35	58.10	62.81
$\mathcal{O}_{tag} + \mathcal{O}_{cau} + \mathcal{O}_{emo}$	**85.93**	**79.93**	**82.82**	**76.14**	**70.39**	**73.15**	**73.77**	**68.02**	**70.78**

sentiment lexicon to assist prediction. In contrast, our model does not require any additional resources.

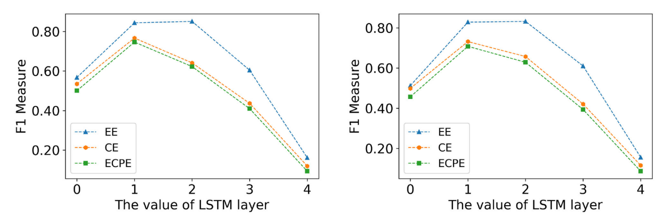
C. Ablation for Objective Functions

Our model is trained with a mixture of three supervision signals: $\mathcal{O}_{tag}$, $\mathcal{O}_{emo}$, and $\mathcal{O}_{cau}$ (see Eq.12). To verify the effect of different supervision, we train our model using different combinations of objective functions. The results are given in Table IV. When only using $\mathcal{O}_{tag}$, the performance of emotion-cause pair extraction drops heavily on both data splits (at least -6.53% in precision, -11.55% in recall, and -9.33% in F_1). However, when $\mathcal{O}_{emo}$ and $\mathcal{O}_{cause}$ are considered, we can observe significant improvements for the emotion-cause pair extraction task. This indicates that introducing the two supervision (emotion and cause) signals can allow our model to learn better clause representations. We also note that our full model further boosts the performance by at least 1.29%, 6.17%, and 2.87% in F_1 measure over all the tasks, respectively. These results show the importance of using interactive information contained in the training signals of related tasks. Our full model can learn such joint features more effective with the help of tag distribution refinement.



(a) Results on 10-fold data splits. (b) Results on 20-fold data splits.

Fig. 5. Experimental results over the three tasks with different λ values on both data splits. Results are averages over 10/20 runs.



(a) Results on 10-fold data splits. (b) Results on 20-fold data splits.

Fig. 6. Experimental results over the three tasks with different LSTM layers on both data splits. Results are averages over 10/20 runs.

D. More Insights Into Our Model

To gain better insights into our model, we conduct further experiments to understand the effect on performance by using: 1) different values of λ; 2) different values of LSTM layer; 3) different values of r;

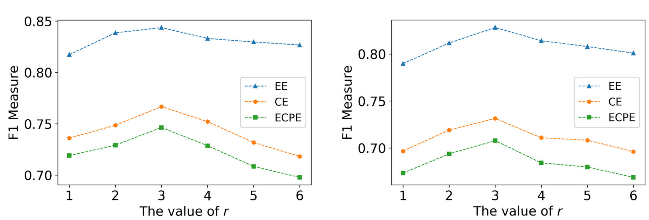
1) Effect of the Value of λ: During the training stage, the threshold λ guides the model to achieve the best trade-off among the three objective functions. In general, it is intuitive to set 0.5 as the default threshold, but the optimal values for different tasks could be different. Therefore, we conduct experiments to explore the impact of λ on test set over the both data splits. We set $\lambda \in (0.15, 0.30, 0.45, 0.60, 0.75, 0.90)$ and show the results in Fig. 5. It can be seen that for our full model, the optimal thresholds for these three tasks are 0.15, 0.30/0.45, 0.60, respectively. The lower value of λ is, the higher weight for training emotion extraction and cause extraction. Thus the performance of these two tasks obtain the best results when the value of λ is lower. However, when $\lambda > 0.60$, although it is beneficial for emotion-cause pair extraction, the poor performance of both emotion extraction and cause extraction lowered the overall performance, because the output of the two tasks is directly used to refine the tag distribution. We choose $\lambda = 0.60$ as the final value since it gives the highest results for emotion-cause pair extraction.

2) Effect of Different Values of LSTM Layer: To analyze the effect of using different values of LSTM layer, we set the numbers of LSTM layer from 0 to 4 for both the encoding layer and decoding layer, results are shown in Fig. 6. We can see that the performance trend on both data splits is similar. The single layer model achieves the highest results. When removing the LSTM layer from our model, the results of all the tasks reduces heavily, since such model is unable to capture the contextual information at clause-level. However, with the increasing number of layers, the performance also decreases seriously, especially when the value of layer is 4. The reason may be that the dataset for this task is small, and more parameters will lead to overfitting. Thus, we choose single LSTM layer in our final model because it performs best.

3) Effect of Different Values of r: We further design a series of experiments to explore the effect of the range of emotion (r) on both data splits. The results are shown in Fig. 7. We can observe

TABLE V
EXAMPLES OF PREDICTED EMOTION-CAUSE PAIRS. DARKER COLOR INDICATES THE EMOTION WHILE LIGHTER COLOR FOR THE ASSOCIATED CAUSE. † DENOTES A CLAUSE HAS EMOTION-CAUSE STRUCTURE ITSELF, RED COLOR DENOTES INCORRECT PREDICTIONS. "PER" REPLACES THE PERSON NAME.

Id	Example	Predicted Pairs	
		Our full model	Our-baseline
1	[Before the incident]^{1}, [$\cdots$], [PER kindly agreed but he was unexpectedly attacked by a crocodile]^{3}. [This sudden situation shocked the tourists] ^{4†}, [$\cdots$], [to prevent similar tragedies from happening again]^{9}.	[(4,4)]	[(4, 4)]
2	[At 15:00 yesterday]^{1}, [PER took the old man to do a brain CT]^{2}, [$\cdots$], [She heard that there are so many kind people helping her] ^{9}, [she was very grateful] ^{10}, [$\cdots$].	[(10, 9)]	[(10, 9)]
3	[Although her life has been rough]^{1}, [PER is very optimistic]^{2}, [$\cdots$], [she is still worried] ^{7}, [but not because of her injuries]^{8}, [she is afraid that her daughter will know about it] ^{9}, [$\cdots$].	[(7, 9)]	[(8, 8)]
4	[$\cdots$], [The chili sauce spilled out]^{9}, [but they did not care]^{10}.[They thought it might be that the jar containing the chili sauce was broken]^{11}, [but what made the old man angry is that] ^{12}, [a piece of bacon in the box was cut off] ^{13}. [$\cdots$].	[(12, 13)]	[(12, 11)]
5	[At that time]^{1}, [maybe the girl felt that the music in her ears was gone]^{2}, [and touched her phone subconsciously]^{3}. [After finding that her iPhone was missing] ^{4}, [the girl didn't know what to do]^{5}, [and was especially anxious] ^{6}. [$\cdots$].	[(6, 4)]	NULL
6	[After PER's divorce]^{1}, [under the arrangement of his parents]^{2}, [he had 3 girlfriends]^{3}. [$\cdots$], [The passing of his parents] ^{9}, [and the thoughts of his ex-wife and children] ^{10}, [making him often feel desperate and depressed] ^{11}.	[(11, 9), (11, 10)]	[(11, 10)]
7	[PER and her husband are already in their 30s]^{1}. [Under the persuasion of relatives and friends]^{2}, [they decided to have a baby]^{3}. [Happily] ^{4}, [PER is pregnant with twins] ^{5}, [$\cdots$], [and she feels guilty for not having more time to take care of their children] ^{14†}.	[(4, 5), (14, 14)]	[(4, 3), (14, 14)]
8	[$\cdots$], [She heard that her husband was chatting with his friends and said that he wanted to marry another woman after the divorce.] ^{6}, [she felt very angry] ^{7}. [$\cdots$], [her husband complained that she was not considerate enough] ^{12}, [which made her both sad and angry] ^{13}.	[(7, 6), (13, 12)]	[(7, 6)]



(a) Results on 10-fold data splits. (b) Results on 20-fold data splits.

Fig. 7. Experimental results over the three tasks with different ranges of emotion on both data splits. Results are averages over 10/20 runs.

that the F_1 measure of emotion extraction, cause extraction, and emotion-cause pair extraction improves with increasing values of r and peaks when $r = 3$. Further increasing the value of r results in worse performance. Intuitively, a small value of r leads to a small clause chunk to be used for tag distribution refinement, making the model ignore some crucial cues during information interaction. However, the excessive value of r will enlarge the search space, may introduce irrelevant information. In our final experiments, we set the value of r to 3, since it gives the best performance.

E. Comparison on Extracting Multiple Pairs

In this section, we compare the results for extracting multiple pairs from one text. We divide each fold's test set into two subsets: one subset contains texts having only one emotion-cause pair, and the other subset contains texts having two or more emotion-cause pairs. We conduct this experiment on 10-fold data splits, which is consistent with **RankCP**. The results on two subsets are listed in Table VI. It can be seen that our full model achieves comparable results on the first subset (only -0.26% in F_1). However, our full model improves the results by 8.18% in F_1 on the second subset, thereby giving higher overall performance for emotion-cause pair extraction. This demonstrates that our full model is relatively more robust for texts having more than one emotion-cause pairs.

TABLE VI
COMPARATIVE RESULTS FOR TEXTS WITH ONLY ONE AND MORE THAN ONE EMOTION-CAUSE PAIR

# Pairs	Method	P (%)	R (%)	F_1 (%)
One per text	RankCP	72.03	**81.23**	**76.33**
	Our full model	**76.53**	75.61	76.07
Two or more per text	RankCP	75.08	43.90	55.31
	Our full model	**81.75**	**51.89**	**63.49**

F. Case Study

We present case studies with some typical examples selected from the test set to better analyze how our model works in extracting emotion-cause pairs with/without using the tag distribution refinement. The results are shown in Table V. To make it easier to visualize the results, we highlight the emotion-cause structure with different colors and distinguish emotion/cause with different color intensities. **Example 1** and **Example 2** are two simple texts with one emotion and one associated cause. Besides, the relative distance between them is small (0 and 1, respectively), thus it is easy for our full model and **Our-baseline** to extract correct pairs. For **Example 3**, we can see that our full model is able to extract correct pairs. However, **Our-baseline** shows a incorrectly predicted pair (8, 8). One possible reason is that there is an affective word "`injuries`" and a causal conjunction "`because`" in this clause, **Our-baseline** may focus on these informative words but ignore the transition word "`but`"and the negation word "`not`," thus regards the clause as an emotion-cause pair.

TABLE VII
ERROR ANALYSIS. "PER" REPLACES THE PERSON NAME

Id	Sample	Ground Truth	Predicted Pairs
1	[She was in poor health][1], [and the baby had to be taken care of][2]. [When PER learned that his wife was going to donate a kidney to him][3], [he was touched][4], [but was also very worried][5].	[(5, 3), (4, 3)]	[(4, 3)]
2	[· · ·], [Now the criminal sees no hope for the complaint][12], [and believes that some witnesses have died one after another][13]. [If he waits][1]4, [there will be no one to testify for him][15], [so he is very depressed][16]. [· · ·].	[(16, 12), (16, 15)]	[(16, 15)]
3	[· · ·], [With the help of her adoptive father's family][9], [PER led a happy life][10]. [· · ·]. [After PER was abducted][13], [her mother regretted it so much][14]. [· · ·]. [But there was no news][16], [her mother fell into endless miss for her daughter][17].	[(10, 9), (14, 13), (17, 16)]	[(14, 13)]

Similarly, in **Example 4**, there is a transition word "but" in the emotion "but what made the old man angry is that". Our full model can correct the prediction made by **Our-baseline**. The two cases above show that our full model is more robust for transitions and negations. This is intuitively reasonable, since it can capture more effective cues for supporting its tag distribution refinement. **Example 5** is still with one emotion and one associated cause but with nonadjacent emotion-cause structure, compared to **Our-baseline**, which extracts nothing in this example, our full model still outputs the correct pair. This shows that the tag distribution refinement strategy has capacity to extract the emotion-cause pairs which have long-range dependencies by using rich emotion- and cause-related features explicitly.

Example 6 to **Example 8** are texts with multiple emotion-cause pairs, which are more challenging for prediction. **Example 6** contains one emotion associated with two different causes. **Our-baseline** can only detect one of the both pairs, the reason may be that it has a poor performance on such situations of which the relative distance more than 1, and our full model can solve it better through refined tag distribution. **Example 7** contains two emotion-cause pairs, and one of them sharing the same clause. **Our-baseline** performs causal reasoning incorrectly and predicted an incorrect pair (4, 3). **Example 8** is a more complex situation that contains multiple emotions with different causes in different clauses. Our full model still works better than **Our-baseline**. The examples above show that the output of emotion extraction and cause extraction can be directly used as inductive bias to refine the tag distribution for each clause. With such explicit information interaction, our full model is able to obtain better results.

G. Error Analysis

We find that most of the errors can be broadly represented by three types of examples, as shown in Table VII. The first type of errors is shown in **Example 1**, where the same cause becomes the stimulus of multiple emotions, and these emotions express opposite sentiment polarity. Our model has difficulty in dealing with such cases. Another type of errors occurr when there are multiple emotion-cause structure but some of them have a very long distance, as shown in **Example 2**. Our model has no capacity to extract the pair (16, 12), since the relative distance between them is 4, which exceeds the range of emotion ($r = 3$ in this work). It is worth investigating how to achieve full coverage in future work. **Example 3** is a very complex text with intricate emotion-cause structure. Hence our model is unable to detect all the emotions and causes with inferring the causal relation between them correctly. It would be interesting to see if incorporating discourse features could assist our model to learn latent relations between emotions and causes better and thus lead to the improvement of the performance.

V. CONCLUSION

In this paper, we proposed method to improve emotion-cause pair extraction on two dimensions: a) tackling the task from a sequence tagging perspective and designing a novel tagging scheme accordingly, so that emotion-cause pair extraction with two auxiliary tasks (i.e., EE and CE) can be integrated into a unified framework through multi-task learning, and b) refining tag distribution for each clause by directly using the output of the two auxiliary tasks, thereby maximizing the mutual benefits between different tasks and making the information interaction more interpretable. On the benchmark dataset, our method improves emotion-cause pair extraction by at least 1.03% ($p < 0.01$) in F_1 measure. Extensive analysis further confirms the effectiveness and robustness of our method, especially in the condition of extracting multiple pairs in one texts.

REFERENCES

[1] T. Brosch, K. R. Scherer, D. M. Grandjean, and D. Sander, "The impact of emotion on perception, attention, memory, and decision-making," *Swiss Med. Weekly*, vol. 143, 2013, Art. no. w 13786.
[2] A. Qadir and E. Riloff, "Learning emotion indicators from tweets: Hashtags, hashtag patterns, and phrases," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 1203–1209, doi: 10.3115/v1/d14-1127.
[3] G. Ou *et al.*, "Exploiting community emotion for microblog event detection," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 1159–1168, doi: 10.3115/v1/d14-1123.
[4] K. Luo, Z. Deng, H. Yu, and L. Wei, "JEAM: A novel model for cross-domain sentiment classification based on emotion analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 2503–2508, doi: 10.18653/v1/d15-1297.
[5] Y. Tang and H. Chen, "Mining sentiment words from microblogs for predicting writer-reader emotion transition," in *Proc. 8th Int. Conf. Lang. Resour. Evaluation*, 2012, pp. 1226–1229. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2012/summaries/117.html
[6] Y. Chang, C. Chen, Y. Hsieh, C. C. Chen, and W. Hsu, "Linguistic template extraction for recognizing reader-emotion and emotional resonance writing assistance," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process. Asian Federation Natural Lang. Process.*, 2015, pp. 775–780, doi: 10.3115/v1/p15-2127.
[7] S. Y. M. Lee, Y. Chen, and C.-R. Huang, "A text-driven rule-based system for emotion cause detection," in *Proc. NAACL HLT 2010 Workshop Comput. Approaches Anal. Gener. Emotion Text*. Association for Computational Linguistics, 2010, pp. 45–53.
[8] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1742–1749, 2014.

[9] L. Gui, L. Yuan, R. Xu, B. Liu, Q. Lu, and Y. Zhou, "Emotion cause detection with linguistic construction in Chinese Weibo text," in *Proc. Natural Lang. Process. Chin. Comput.*, Shenzhen, China, 2014, pp. 457–464, doi: 10.1007/978-3-662-45924-9_42.

[10] S. Yada, K. Ikeda, K. Hoashi, and K. Kageura, "A bootstrap method for automatic rule acquisition on emotion cause extraction," in *Proc. IEEE Int. Conf. Data Mining Workshops*, New Orleans, LA, USA, 2017, pp. 414–421, doi: 10.1109/ICDMW.2017.60.

[11] K. Gao, H. Xu, and J. Wang, "A rule-based approach to emotion cause detection for chinese micro-blogs," *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4517–4528, 2015.

[12] I. Russo, T. Caselli, F. Rubino, E. Boldrini, and P. Martínez-Barco, "Emocause: An easy-adaptable approach to extract emotion cause contexts," in *Proc. 2nd Workshop Comput. Approaches Subjectivity Sentiment Anal.*, 2011, pp. 153–160. [Online]. Available: https://www.aclweb.org/anthology/W11-1720/

[13] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Detecting emotion stimuli in emotion-bearing sentences," in *Proc. Comput. Linguistics Intell. Text Process.*, vol. 9042, 2015, pp. 152–165, doi: 10.1007/978-3-319-18117-2_12.

[14] S. Song and Y. Meng, "Detecting concept-level emotion cause in microblogging," in *Proc. 24th Int. Conf. World Wide Web Companion*, 2015, pp. 119–120, doi: 10.1145/2740908.2742710.

[15] L. Gui, D. Wu, R. Xu, Q. Lu, and Y. Zhou, "Event-driven emotion cause extraction with corpus construction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Austin, Texas, USA, 2016, pp. 1639–1649. [Online]. Available: http://aclweb.org/anthology/D/D16/D16-1170.pdf

[16] B. Xu, H. Lin, Y. Lin, Y. Diao, L. Yang, and K. Xu, "Extracting emotion causes using learning to rank methods from an information retrieval perspective," *IEEE Access*, vol. 7, pp. 15 573–15 583, 2019.

[17] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, and J. Du, "A question answering approach for emotion cause extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Copenhagen, Denmark, 2017, pp. 1593–1602. [Online]. Available: https://aclanthology.info/papers/D17-1167/d17-1167

[18] R. Xia, M. Zhang, and Z. Ding, "RTHN: A RNN-transformer hierarchical network for emotion cause extraction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 5285–5291, doi: 10.24963/ijcai.2019/734.

[19] C. Fan *et al.*, "A knowledge regularized hierarchical approach for emotion cause analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process.*, 2019, pp. 5613–5623, doi: 10.18653/v1/D19-1563.

[20] R. Xia and Z. Ding, "Emotion-cause pair extraction: A new task to emotion analysis in texts," in *Proc. 57th Conf. Assoc. Comput. Linguistics*, Florence, Italy, 2019, pp. 1003–1012. [Online]. Available: https://www.aclweb.org/anthology/P19-1096/

[21] Z. Ding, R. Xia, and J. Yu, "ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3161–3170. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.288/

[22] C. Fan, C. Yuan, J. Du, L. Gui, M. Yang, and R. Xu, "Transition-based directed graph construction for emotion-cause pair extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3707–3717. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.342/

[23] P. Wei, J. Zhao, and W. Mao, "Effective inter-clause modeling for end-to-end emotion-cause pair extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3171–3181. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.289/

[24] E. Fox, *Emotion Science Cognitive and Neuroscientific Approaches to Understanding Human Emotions*. Basingstoke, U.K.: Palgrave Macmillan, 2008.

[25] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul.–Sep. 2019.

[26] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[27] P. Ekman, "Expression and the nature of emotion," *Approaches Emotion*, vol. 3, no. 19, pp. 319–344, 1984.

[28] P. Ekman, "Basic emotions," *Handbook Cogn. Emotion*, vol. 98, no. 16, pp. 45–60, 1999.

[29] P. J. Lang, "The emotion probe: Studies of motivation and attention," *Amer. Psychol.*, vol. 50, no. 5, pp. 372–385, 1995.

[30] J. Xu, R. Xu, Q. Lu, and X. Wang, "Coarse-to-fine sentence-level emotion classification based on the intra-sentence features and sentential context," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 2455–2458, doi: 10.1145/2396761.2398665.

[31] D. Beck, T. Cohn, and L. Specia, "Joint emotion analysis via multi-task Gaussian processes," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 1798–1803, doi: 10.3115/v1/d14-1190.

[32] S. Li, L. Huang, R. Wang, and G. Zhou, "Sentence-level emotion classification with label and context dependence," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process. Asian Federation Natural Lang. Process.*, 2015, pp. 1045–1053, doi: 10.3115/v1/p15-1101.

[33] D. Zhou, X. Zhang, Y. Zhou, Q. Zhao, and X. Geng, "Emotion distribution learning from texts," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 638–647, doi: 10.18653/v1/d16-1061.

[34] L. Chen, X. Mao, P. Wei, and A. Compare, "Speech emotional features extraction based on electroglottograph," *Neural Comput.*, vol. 25, no. 12, pp. 3294–3317, 2013.

[35] L. Chen, X. Mao, P. Wei, and A. Compare, "Speech synthesis research based on EGG," in *Proc. IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput.*, Beijing, China, 2013, pp. 1799–1803, doi: 10.1109/GreenCom-iThings-CPSCom.2013.332.

[36] X. Wu, X. Mao, L. Chen, and A. Compare, "Combined motion and region-based 3D tracking in active depth image sequence," in *Proc. IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput.*, Beijing, China, 2013, pp. 1734–1739, doi: 10.1109/GreenCom-iThings-CPSCom.2013.319.

[37] X. Wu, X. Mao, L. Chen, Y. Xue, and A. Compare, "View-invariant gesture recognition using nonparametric shape descriptor," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, 2014, pp. 544–549, doi: 10.1109/ICPR.2014.104.

[38] J. Yi, X. Mao, L. Chen, Y. Xue, and A. Compare, "Facial expression recognition considering individual differences in facial structure and texture," *IET Comput. Vis.*, vol. 8, no. 5, pp. 429–440, 2014.

[39] J. Yi, X. Mao, L. Chen, Y. Xue, and A. Compare, "Illuminant direction estimation for a single image based on local region complexity analysis and average gray value," *Appl. Opt.*, vol. 53, no. 2, pp. 226–236, 2014.

[40] J. Yi, X. Mao, Y. Xue, and A. Compare, "Facial expression recognition based on t-SNE and adaboostM2," in *Proc. IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput.*, Beijing, China, 2013, pp. 1744–1749, doi: 10.1109/GreenCom-iThings-CPSCom.2013.321.

[41] S. Yuan, X. Mao, Y. Xue, L. Chen, Q. Xiong, and A. Compare, "SQR: A simple quantum representation of infrared images," *Quantum Inf. Process.*, vol. 13, no. 6, pp. 1353–1379, 2014.

[42] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015. [Online]. Available: http://arxiv.org/abs/1409.0473

[43] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 1746–1751. [Online]. Available: http://aclweb.org/anthology/D/D14/D14-1181.pdf

[44] Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol.*, San Diego, CA, USA, Jun. 12-17, 2016, pp. 1480–1489. [Online]. Available: http://aclweb.org/anthology/N/N16/N16-1174.pdf

[45] M. Abdul-Mageed and L. H. Ungar, "Emonet: Fine-grained emotion detection with gated recurrent neural networks," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 718–728, doi: 10.18653/v1/P17-1067.

[46] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1615–1625, doi: 10.18653/v1/d17-1169.

[47] K. Gao, H. Xu, and J. Wang, "Emotion cause detection for chinese microblogs based on ECOCC model," in *Proc. 19th Pacific-Asia Conf. Adv. Knowl. Discov. Data Mining*, 2015, pp. 3–14, doi: 10.1007/978-3-319-18032-8_1.

[48] X. Cheng, Y. Chen, B. Cheng, S. Li, and G. Zhou, "An emotion cause corpus for chinese microblogs with multiple-user structures," *ACM Trans. Asian Low Resour. Lang. Inf. Process.*, vol. 17, no. 1, pp. 6: 1–6:19, 2017.

[49] Y. Chen, W. Hou, X. Cheng, and S. Li, "Joint learning for emotion classification and emotion cause detection," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Brussels, Belgium, 2018, pp. 646–651, doi: 10.18653/v1/d18-1066.

[50] Y. Chen, W. Hou, and X. Cheng, "Hierarchical convolution neural network for emotion cause detection on microblogs," in *Proc. 27th Int. Conf. Artif. Neural Netw. Artif. Neural Netw. Mach. Learn.*, 2018, pp. 115–122, doi: 10.1007/978-3-030-01418-6_12.

[51] Y. Chen, S. Y. M. Lee, S. Li, and C. Huang, "Emotion cause detection with linguistic constructions," in *Proc. 23rd Int. Conf. Comput. Linguistics, Conf.*, Beijing, China, 2010, pp. 179–187. [Online]. Available: https://www.aclweb.org/anthology/C10-1021/

[52] X. Li, K. Song, S. Feng, D. Wang, and Y. Zhang, "A co-attention neural network model for emotion cause analysis with emotional context awareness," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Brussels, Belgium, 2018, pp. 4752–4757. [Online]. Available: https://aclanthology.info/papers/D18-1506/d18-1506

[53] X. Li, S. Feng, D. Wang, and Y. Zhang, "Context-aware emotion cause analysis with multi-attention-based neural network," *Knowl.-Based Syst.*, vol. 174, pp. 205–218, 2019.

[54] X. Yu, W. Rong, Z. Zhang, Y. Ouyang, and Z. Xiong, "Multiple level hierarchical network-based clause selection for emotion cause extraction," *IEEE Access*, vol. 7, pp. 9071–9079, 2019.

[55] Z. Ding, H. He, M. Zhang, and R. Xia, "From independent prediction to reordered prediction: integrating relative position and global label information to emotion cause identification," in *Proc. 33rd AAAI Conf. Artif. Intell., 31st Innov. Appl. Artif. Intell. Conf., 9th AAAI Symp. Educ. Adv. Artif. Intell.*, Honolulu, Hawaii, USA, AAAI Press, 2019, pp. 6343–6350, doi: 10.1609/aaai.v33i01.33016343.

[56] H. Song, C. Zhang, Q. Li, and D. Song, "End-to-End emotion-cause pair extraction via learning to link," 2020, *arXiv:2002.10710*.

[57] H. Tang, D. Ji, and Q. Zhou, "Joint multi-level attentional model for emotion detection and emotion-cause pair extraction," *Neurocomputing*, vol. 409, pp. 329–340, Oct. 2020.

[58] L. Qin, W. Che, Y. Li, H. Wen, and T. Liu, "A stack-propagation framework with token-level intent detection for spoken language understanding," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process.*, 2019, pp. 2078–2087, doi: 10.18653/v1/D19-1214.

[59] C. Goo *et al.*, "Slot-Gated modeling for joint slot filling and intent prediction," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol.*, 2018, pp. 753–757, doi: 10.18653/v1/n18-2118.

[60] L. Qin, W. Che, Y. Li, M. Ni, and T. Liu, "DCR-Net: A deep Co-interactive relation network for joint dialog ACT recognition and sentiment classification," in *Proc. 34th AAAI Conf. Artif. Intell., 32nd Innov. Appl. Artif. Intell. Conf., 10th AAAI Symp. Educ. Adv. Artif. Intell.*, New York, NY, USA, 2020, pp. 8665–8672. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/6391

[61] Z. Chen *et al.*, "Co-attentive multi-task learning for explainable recommendation," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 2137–2143, doi: 10.24963/ijcai.2019/296.

[62] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol.*, Minneapolis, MN, USA, 2019, pp. 4171–4186. [Online]. Available: https://www.aclweb.org/anthology/N19-1423/

[63] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[64] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu, "Joint extraction of entities and relations based on a novel tagging scheme," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, Vancouver, Canada, 2017, pp. 1227–1236, doi: 10.18653/v1/P17-1113.

[65] R.-A. De Beaugrande and W. U. Dressler, *Introduction to Text Linguistics*. vol. 1, London, U.K.: Longman, 1981.

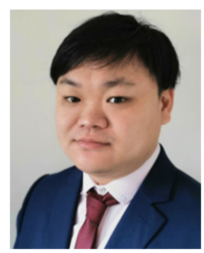
[66] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, 2015, pp. 1–15. [Online]. Available: http://arxiv.org/abs/1412.6980

[67] M. S. Schlichtkrull *et al.*, "Modeling relational data with graph convolutional networks," in *Proc. 15th Int. Conf. Semantic Web*, 2018, pp. 593–607, doi: 10.1007/978-3-319-93417-4_38.

[68] S. Wang and L. Ku, "ANTUSD: A large chinese sentiment dictionary," in *Proc. 10th Int. Conf. Lang. Resour. Eval.*, 2016, pp. 2697–2702. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2016/summaries/450.html

Chuang Fan is currently working toward the Ph.D. degree with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. His current research interests include natural language processing and information extraction, and he also focuses on several application fields, including emotion-cause pair extraction and event detection.

Chaofa Yuan is currently working toward the master's degree with the School of Computer Science and Technology, Harbin Institution of Technology, Shenzhen, China. His current research interests include natural language processing, text classification, and sentiment analysis.



Lin Gui received the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, Shenzhen, China. He is currently a Marie Curie Research Fellow with the Department of Computer Science, University of Warwick, Coventry, U.K. His specific research interests include the development of text classification algorithms and the models for natural language understanding, sentiment analysis, stance detection, emotion cause detection, and topic modeling.

Yue Zhang (Member, IEEE) received the Ph.D. degree from the University of Oxford, Oxford, U.K, working on statistical Chinese processing. He is currently an Associate Professor with Westlake University, Hangzhou, China. From July 2012 to August 2018, he was an Assistant Professor with the Singapore University of Technology and Design, Singapore. His research interests include natural language parsing and generation, and machine translation.

Ruifeng Xu (Member, IEEE) received the Ph.D. degree in computer science from The Hong Kong Polytechnic University, Hong Kong, China. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. He has authored or coauthored more than 100 papers in natural language processing, sentiment analysis, and social media analysis.