

End-to-End Emotion-Cause Pair Extraction based on Sliding Window Multi-Label Learning

Zixiang Ding, Rui Xia*, Jianfei Yu

School of Computer Science and Engineering,
Nanjing University of Science and Technology, China
{dingzixiang, rxia, jfyu}@njjust.edu.cn

Abstract

Emotion-cause pair extraction (ECPE) is a new task that aims to extract the potential pairs of emotions and their corresponding causes in a document. The existing methods first perform emotion extraction and cause extraction independently, and then perform emotion-cause pairing and filtering. However, the above methods ignore the fact that the cause and the emotion it triggers are inseparable, and the extraction of the cause without specifying the emotion is pathological, which greatly limits the performance of the above methods in the first step. To tackle these shortcomings, we propose two joint frameworks for ECPE: 1) multi-label learning for the extraction of the cause clauses corresponding to the specified emotion clause (CMLL) and 2) multi-label learning for the extraction of the emotion clauses corresponding to the specified cause clause (EMLL). The window of multi-label learning is centered on the specified emotion clause or cause clause and slides as their positions move. Finally, CMLL and EMLL are integrated to obtain the final result. We evaluate our model on a benchmark emotion cause corpus, the results show that our approach achieves the best performance among all compared systems on the ECPE task.

1 Introduction

With the rapid growth of social media, emotion analysis of online text has received much attention in recent years. Unlike most studies that focused on the detection and classification of emotions, (Lee et al., 2010) first proposed the emotion cause extraction (ECE) task, which aims to extract the stimulus behind emotions. (Gui et al., 2016a) released a public corpus and defined the ECE task as a fine-grained emotion analysis task, where the goal is to judge for each clause in the document whether

*Corresponding author

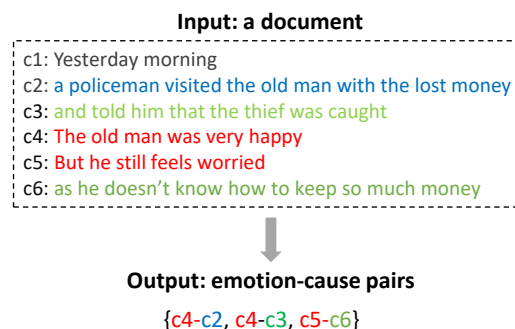


Figure 1: An example showing the emotion-cause pair extraction (ECPE) task.

it is the corresponding cause, given the annotation of emotions. This corpus has received a lot of attention in subsequent research and has become a benchmark dataset for the ECE task (Gui et al., 2017; Li et al., 2018; Yu et al., 2019; Xu et al., 2019; Ding et al., 2019; Xia et al., 2019). However, there are several inherent shortcomings in the setting of the ECE task: firstly, the need for emotion annotation greatly limits the practical applications of the ECE task; secondly, the way of annotating emotions before extracting causes ignores the fact that emotions and causes are mutually indicative. To solve these problems, we have proposed the emotion-cause pair extraction (ECPE) task in (Xia and Ding, 2019), with the goal to extract the potential pairs of emotions and corresponding causes in the document.

Figure 1 shows an example of the ECPE task. The input is a document, which has been split into six clauses. Clauses c4 and c5 are both emotion clauses, which contain the emotion expression “happy” and “worried”, respectively. C4 has two corresponding cause clauses: clause c2 (“a policeman visited the old man with the lost money”) and clause c3 (“and told him that the thief was caught”). C5 has one corresponding cause clause c6 (“as he doesn't know how to keep so much money”). The

output is a set of all emotion-cause pairs in the document: {c4-c2, c4-c3, c5-c6}.

The existing ECPE solutions use a two-step framework. Step1 independently extracts the emotion clause set and the cause clause set in the document. Step2 first applies a Cartesian product to the two sets to obtain candidate emotion-cause pairs, and then train a binary classifier to determine whether each candidate pair contains a causal relationship. However, the above methods ignore the fact that the cause and the emotion it triggers are inseparable, and the extraction of the cause without specifying the emotion is pathological, which greatly limits the performance of the above methods in the first step.

In this paper, we transform the ECPE task into the emotion-pivot cause extraction problem in the sliding window. Specifically, we assume that all clauses in the document are emotion clauses and introduce an emotion-oriented sliding window centered on each emotion clause in the document. Then we use a multi-label learning framework inside each sliding window to extract one or more cause clauses that can be paired with the current emotion clause (CMLL). Taking Figure 1 as an example, the clauses c1, c2, ..., c6 are respectively used as the centers of the emotion-oriented sliding windows, and multi-label learning is performed in each window to extract the corresponding cause clauses: the results for clause c1, c2, c3, and c6 are empty; the result for clause c4 is {c2, c3}; and the result for clause c5 is {c6}. These emotion clauses are then paired with their corresponding cause clauses to get the final emotion-cause pair set: {c4-c2, c4-c3, c5-c6}.

On the other hand, we can also transform the ECPE task into the cause-pivot emotion extraction problem in the sliding window by building a cause-oriented sliding window centered on each cause clause in the document. Then we use a multi-label learning framework inside each sliding window to extract one or more emotion clauses that can be paired with the current cause clause (EMLL).

The above two methods are dual, that is, when we perform emotion-cause pair extraction, we can use emotion clauses as the pivot to extract the corresponding cause clauses, and we can also use the cause clauses as the pivot to extract the corresponding emotion clauses. We get the final emotion-cause pairs based on the combination of these two predictions.

We evaluate our model on a benchmark emotion cause corpus, the results show that our approach achieves the best performance among all compared systems on the ECPE task.

2 Related Work

(Lee et al., 2010) first proposed the emotion cause extraction (ECE) task to extract the causes behind a given emotion expression in text. They constructed a small-scale dataset and proposed a rule-based ECE solution. Based on the same corpus, (Chen et al., 2010) proposed a multi-label approach, which can not only detect the multi-clause causes, but also capture the long-distance information. After that, many follow-up work conducted ECE exploration on their own corpus and proposed new ECE solutions based on rules (Neviarouskaya and Aono, 2013; Li and Xu, 2014; Gao et al., 2015a,b; Yada et al., 2017), traditional machine learning methods (Russo et al., 2011; Gui et al., 2014; Ghazi et al., 2015; Song and Meng, 2015), and deep learning models (Cheng et al., 2017).

(Gui et al., 2016a,b; Xu et al., 2017) constructed an open ECE dataset based on Sina City News, and redefined the ECE task as a clause level binary classification problem. They proposed to solve the ECE task by event-driven emotion cause extraction methods. In recent years, this dataset has received more and more attention and has become a benchmark dataset for ECE, and many deep learning solutions based on this corpus have been proposed in the follow-up researches (Gui et al., 2017; Li et al., 2018; Yu et al., 2019; Xu et al., 2019; Ding et al., 2019; Xia et al., 2019).

The traditional ECE tasks have several inherent shortcomings. First, the need for emotion annotations limits its practical applications. Second, annotating emotions before extracting causes ignores the mutual indication of emotions and causes. To solve these shortcomings, we proposed the emotion-cause pair extraction (ECPE) task in (Xia and Ding, 2019), with the aim to extract all pairs of emotion clauses and corresponding cause clauses in the document. We have further proposed a two-step framework to solve this new task: the first step extracts independent sets of emotion clauses and cause clauses, the second step uses the Cartesian product to obtain candidate emotion-cause pairs and introduces a binary classifier to filter out pairs that do not contain causality.

However, the above method does not directly

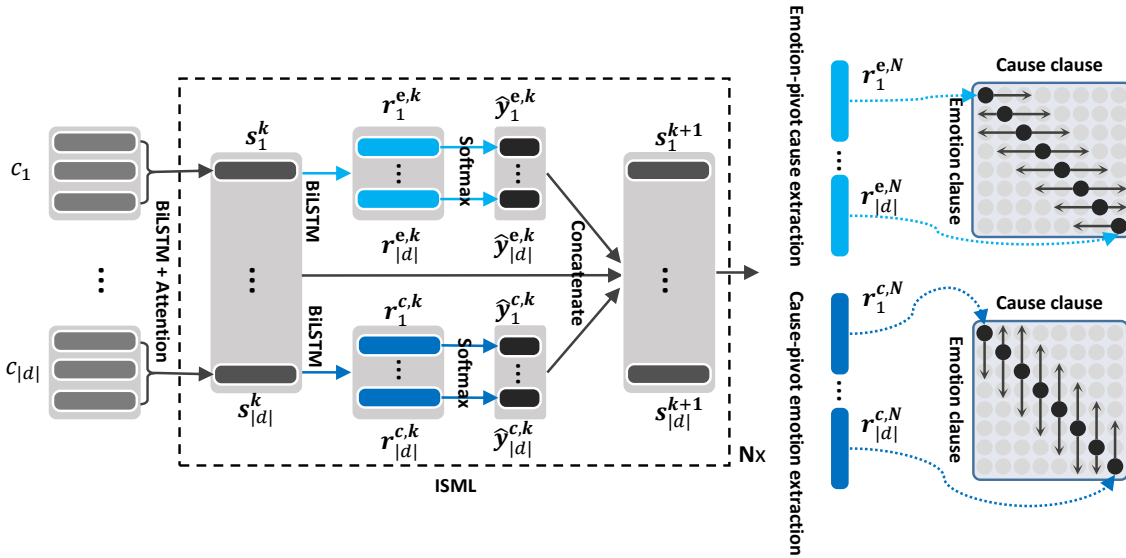


Figure 2: Overview of the ECPE-MLL model.

model the extraction of emotion-cause pairs, and the mistakes made in the first step will propagate to the second step. To deal with these problems, we have proposed an end-to-end approach which extracts emotion-cause pairs through 2D representation, interaction, and prediction (Ding et al., 2020). (Fan et al., 2020) proposed a transition-based model to transform the ECPE task into a procedure of parsing-like directed graph construction. (Wei et al., 2020) designed a joint neural approach which models the inter-clause relations with graph attention and tackles emotion-cause pair extraction from a ranking perspective. In contrast, in this paper, we propose a joint framework to solve the ECPE task by sliding window multi-label learning.

It should be noted that (Chen et al., 2018) also identifies emotions and causes together, while the task of this paper is different from ours. The main difference is that their task is a joint task of emotion cause extraction and emotion classification (rather than emotion expression extraction), which is based strictly on the condition that the emotion expressions should be annotated in advance. While ours performs automatic extractions of emotion expressions and causes, as well as their pairs.

3 Approach

3.1 Task Definition

Before describing our method, we first give the formal definition of the emotion-cause pair extraction (ECPE) task. The input is a document containing multiple clauses $d = [c_1, c_2, \dots, c_{|d|}]$, the goal of

ECPE is to extract a set of emotion-cause pairs in d :

$$P = \{\dots, c^e - c^c, \dots\}, \quad (1)$$

where c^e is an emotion clause and c^c is the corresponding cause clause.

3.2 Emotion-pivot Cause Extraction

We propose to solve the ECPE task by a sliding window multi-label learning scheme (ECPE-MLL), as shown in Figure 2. First, we assume that all clauses in the document are emotion clauses, and build an emotion-oriented sliding window centered on each emotion clause. Then, in each window, we use the emotion clause as the pivot to extract the corresponding one or more cause clauses based on multi-label learning (CMLL). As the example mentioned in section 1, the results of multi-label learning can be naturally transformed into emotion-cause pairs. Finally, by processing multiple sliding windows, we can get the emotion-cause pairs in the entire document.

3.2.1 Emotion-oriented Sliding Window Encoding

The purpose of emotion-oriented sliding window encoding is to obtain emotion-specific clause representations, which are used as features for emotion-pivot cause extraction. Specifically, we introduce the emotion extraction and cause extraction sub-tasks, and propose the iterative synchronized multi-task learning (ISML) model to solve these two sub-tasks.

The input is a document contains multiple clauses: $d = [c_1, c_2, \dots, c_{|d|}]$, and each clause also contains multiple words $c_i = [w_{i,1}, w_{i,2}, \dots, w_{i,|c_i|}]$. We use a hierarchical network with two layers to encode the document. The lower layer uses a BiLSTM to encode each clause c_i in the document independently and obtains the clause representation \mathbf{s}_i by attention mechanism.

The upper layer is composed of a module that can be iterated multiple times. We abbreviate the model in which the upper layer is iterated N times as ISML- N . In each iteration of the upper layer, we introduce two components for emotion extraction and cause extraction, and obtain the prediction results of the two subtasks respectively. The predictions of subtasks in the current iteration will assist in the prediction of subtasks in subsequent iterations to achieve deep interactions between emotions and causes.

Specifically, in the k -th iteration of the upper layer, both components for emotion extraction and cause extraction take the clause representation $(\mathbf{s}_1^k, \mathbf{s}_2^k, \dots, \mathbf{s}_{|d|}^k)$ as input (in particular, $\mathbf{s}_i^k = \mathbf{s}_i$ when $k = 1$), and use two clause level Bi-LSTM to obtain the emotion-specific representation $\mathbf{r}_i^{e,k}$ and cause-specific representation $\mathbf{r}_i^{c,k}$ for each clause c_i , respectively. The emotion distribution $\mathbf{y}_i^{e,k}$ and cause distribution $\mathbf{y}_i^{c,k}$ of the clause c_i are predicted as follows:

$$\hat{\mathbf{y}}_i^{e,k} = \text{softmax}(\mathbf{W}^e \mathbf{r}_i^{e,k} + \mathbf{b}^e), \quad (2)$$

$$\hat{\mathbf{y}}_i^{c,k} = \text{softmax}(\mathbf{W}^c \mathbf{r}_i^{c,k} + \mathbf{b}^c). \quad (3)$$

Finally, we can get the input for the next iteration $(\mathbf{s}_1^{k+1}, \mathbf{s}_2^{k+1}, \dots, \mathbf{s}_{|d|}^{k+1})$, where

$$\mathbf{s}_i^{k+1} = \mathbf{s}_i^k \oplus \hat{\mathbf{y}}_i^{e,k} \oplus \hat{\mathbf{y}}_i^{c,k}. \quad (4)$$

It should be noted that we introduce the supervision of emotion extraction and cause extraction for each iteration. Therefore, the total loss of the ISML model for a document d is the sum of the losses of N iterations:

$$L^{\text{ISML-}N} = \sum_{t=1}^N L_t, \quad (5)$$

$$L_t = - \sum_{i=1}^{|d|} \mathbf{y}_i^e \cdot \log(\hat{\mathbf{y}}_i^{e,t}) - \sum_{i=1}^{|d|} \mathbf{y}_i^c \cdot \log(\hat{\mathbf{y}}_i^{c,t}), \quad (6)$$

where \mathbf{y}_i^e and \mathbf{y}_i^c are emotion and cause annotation of clause c_i , respectively.

3.2.2 Emotion-pivot Cause Extraction based on Multi-label learning

Formally, we assume c_i is an emotion clause, and introduce a multi-label classifier to judge whether the clauses $c_{i-|w|}, \dots, c_{i-1}, c_i, c_{i+1}, \dots, c_{i+|w|}$ are corresponding cause clauses. The dimension of multi-label is $(|w| * 2 + 1)$, where $|w|$ is the size of window, which indicates the farthest distance between current emotion clause and the candidate cause clauses. Therefore, the multi-label cause annotation of each clause c_i can be represented as a $(|w| * 2 + 1)$ dimensional vector: $\mathbf{y}_i^{\text{cml}} = [y_i^{\text{cml}-|w|}, \dots, y_i^{\text{cml}0}, \dots, y_i^{\text{cml}|w|}]$, where $y_i^{\text{cml}j} = 1$ if $c_i - c_{i+j}$ is annotated as an emotion-cause pair, and $y_i^{\text{cml}j} = 0$ otherwise.

In order to implement multi-label classification of $(|w| * 2 + 1)$ candidate cause clauses for emotion clause c_i , we use emotion specific clause representation $\mathbf{r}_i^{e,N}$ as the feature and introduce $(|w| * 2 + 1)$ logistic functions to predict a probability $\{p(y_i^{\text{cml}j} = 1 | c_i), j = -|w|, \dots, 0, \dots, |w|\}$ for each candidate cause clause:

$$p(y_i^{\text{cml}j} = 1 | c_i) = \hat{y}_i^{\text{cml}j} = \frac{1}{1 + e^{\mathbf{W}^{\text{cml}j} \mathbf{r}_i^{e,N} + \mathbf{b}^{\text{cml}j}}}, \quad (7)$$

$$p(y_i^{\text{cml}j} = 0 | c_i) = 1 - \hat{y}_i^{\text{cml}j}, \quad (8)$$

where $\mathbf{W}^{\text{cml}j}$ and $\mathbf{b}^{\text{cml}j}$ are the parameters for logistic regression. It should be noted that $c_i - c_{i+j}$ is predicted as an emotion-cause pair only when $\hat{y}_i^{\text{cml}j} > 0.5$. For a given document d , the loss of CMLL is defined as the following joint binary cross entropy loss:

$$L^{\text{CMLL}} = - \sum_{i=1}^{|d|} \sum_{j=-|w|}^{|w|} (y_i^{\text{cml}j} \cdot \log \hat{y}_i^{\text{cml}j} + (1 - y_i^{\text{cml}j}) \cdot \log(1 - \hat{y}_i^{\text{cml}j})). \quad (9)$$

3.3 Cause-pivot Emotion Extraction

Dual with the method proposed in section 3.2, we can also transform the ECPE task into the cause-pivot emotion extraction problem in the sliding window. Specifically, we assume that all clauses in the document are cause clauses, and build a cause-oriented sliding window centered on each cause clause. Then, in each window, we use the cause clause as a pivot to extract the corresponding one or more emotion clauses based on multi-label learning (EMLL).

Formally, we assume c_i is a cause clause, and introduce a multi-label classifier to judge whether the clauses $c_{i-|w|}, \dots, c_{i-1}, c_i, c_{i+1}, \dots, c_{i+|w|}$ are corresponding emotion clauses. The dimension of multi-label is $(|w| * 2 + 1)$. Therefore, the multi-label emotion annotation of each clause c_i can be represented as a $(|w| * 2 + 1)$ dimensional vector: $\mathbf{y}_i^{\text{eml}} = [y_i^{\text{eml}_{-|w|}}, \dots, y_i^{\text{eml}_0}, \dots, y_i^{\text{eml}_{|w|}}]$, where $y_i^{\text{eml}_j} = 1$ if $c_{i+j}-c_i$ is annotated as an emotion-cause pair, and $y_i^{\text{eml}_j} = 0$ otherwise.

The process of multi-label classification for the EMLL model is similar as CMLL. We use cause specific clause representation $\mathbf{r}_i^{c,N}$ obtained in section 3.2.1 as the feature and introduce $(|w| * 2 + 1)$ logistic functions to predict a probability $\hat{y}_i^{\text{eml}_j} = \{p(y_i^{\text{eml}_j} = 1|c_i), j = -|w|, \dots, 0, \dots, |w|\}$ for each candidate emotion clause. And $c_{i+j}-c_i$ is predicted as an emotion-cause pair only when $\hat{y}_i^{\text{eml}_j} > 0.5$. For a given document d , the loss of EMLL is defined as the following joint binary cross entropy loss:

$$L^{\text{EMLL}} = - \sum_{i=1}^{|d|} \sum_{j=-|w|}^{|w|} (y_i^{\text{eml}_j} \cdot \log \hat{y}_i^{\text{eml}_j} + (1 - y_i^{\text{eml}_j}) \cdot \log(1 - \hat{y}_i^{\text{eml}_j})). \quad (10)$$

3.4 Dual Emotion-Cause Pair Extraction

For a candidate emotion-cause pair c_p-c_q , where $p, q \in \{1, \dots, |d|\}, p - q \in [-|w|, |w|]$, both CMLL and EMLL predict a probability of whether c_p-c_q is a valid emotion-cause pair: $\hat{y}_p^{\text{eml}_{q-p}}$ and $\hat{y}_q^{\text{eml}_{p-q}}$. We adopt three strategies to integrate these two predictions:

- Average probability. c_p-c_q is predicted as an emotion-cause pair only when $(\hat{y}_p^{\text{eml}_{q-p}} + \hat{y}_q^{\text{eml}_{p-q}})/2 > 0.5$.
- Logical AND. c_p-c_q is predicted as an emotion-cause pair only when $\hat{y}_p^{\text{eml}_{q-p}} > 0.5$ AND $\hat{y}_q^{\text{eml}_{p-q}} > 0.5$.
- Logical OR. c_p-c_q is predicted as an emotion-cause pair only when $\hat{y}_p^{\text{eml}_{q-p}} > 0.5$ OR $\hat{y}_q^{\text{eml}_{p-q}} > 0.5$.

The final loss of our model for a document d is a weighted sum of $L^{\text{ISML}}, L^{\text{CMLL}}$ and L^{EMLL} with L2-regularization term as follows:

$$L = \lambda_1 L^{\text{ISML-N}} + \lambda_2 L^{\text{CMLL}} + \lambda_3 L^{\text{EMLL}} + \lambda_4 \|\theta\|^2, \quad (11)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \in (0, 1)$ are weights, θ denotes all the parameters in this model.

The performances using different integration methods are given in the experiment section.

4 Experiments

4.1 Dataset and Metrics

We conducted experiments on the data set provided by (Xia and Ding, 2019) to verify the effectiveness of our model. For fair comparisons with (Xia and Ding, 2019), we use the same ratio to split the data that 90% are randomly selected for training and the remaining are used for testing. Similarly, we repeat the experiments 20 times and report the average result.

We use the precision, recall, and F1 score defined in (Xia and Ding, 2019) as evaluation metrics for the ECPE task as well as two sub-tasks: emotion extraction and cause extraction.

4.2 Experimental Settings

We use the same word embeddings as (Xia and Ding, 2019). The dimension of word embedding is set to 200. The number of hidden units in BiLSTM for all our models is set to 100. The size of the window in the multi-label learning is set to 3.

In the training phase, we use stochastic gradient descent (SGD) algorithm and Adam optimizer. The batch size and learning rate are set to 32 and 0.005, respectively. The weights $\lambda_1, \lambda_2, \lambda_3$ in formula 11 are all set to 1. For regularization, we apply dropout to word embeddings and perform L2 constraints over the softmax parameters. The dropout rate and L2-norm regularization are set to 0.5 and $1e-5$, respectively. The code has been made publicly available on Github¹.

4.3 Overall Performance

Table 1 shows the experimental results of our models and baseline methods on the ECPE task as well as two subtasks (emotion extraction and cause extraction).

ECPE-2Steps refers to the two-step framework proposed in our previous work (Xia and Ding, 2019), which first performs individual emotion extraction and cause extraction via multi-task learning, and then conducts emotion-cause pairing and filtering. Specifically, there are three kinds of multi-task learning settings: Indep, Inter-CE, and Inter-EC. Indep perform emotion extraction and cause

¹<https://github.com/NUSTM/ECPE-MLL>

Framework	Approach	Emotion-Cause Pair Ext.			Emotion Ext.			Cause Ext.		
		<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>
ECPE-2Steps	Indep	0.6832	0.5082	0.5818	0.8375	0.8071	0.8210	0.6902	0.5673	0.6205
	Inter-CE	0.6902	0.5135	0.5901	0.8494	0.8122	0.8300	0.6809	0.5634	0.6151
	Inter-EC	0.6721	0.5705	0.6128	0.8364	0.8107	0.8230	0.7041	0.6083	0.6507
ECPE-MLL (Ours)	Indep	0.6686	0.6260	0.6462	0.8541	0.8236	0.8378	0.7250	0.5942	0.6525
	Inter-CE	0.7117	0.6048	0.6526	0.8540	0.8287	0.8407	0.7235	0.6025	0.6565
	Inter-EC	0.6814	0.6257	0.6515	0.8494	0.8297	0.8390	0.7256	0.6417	0.6799
	ISML-2	0.6990	0.6350	0.6647	0.8609	0.8355	0.8474	0.7298	0.6417	0.6818
	ISML-3	0.7009	0.6384	0.6674	0.8569	0.8391	0.8473	0.7316	0.6473	0.6861
	ISML-4	0.7058	0.6356	0.6682	0.8649	0.8357	0.8498	0.7259	0.6540	0.6875
	ISML-5	0.7116	0.6351	0.6702	0.8601	0.8350	0.8469	0.7358	0.6547	0.6920
	ISML-6	0.7090	0.6441	0.6740	0.8582	0.8429	0.8500	0.7248	0.6702	0.6950
ISML-7	0.7068	0.6425	0.6720	0.8600	0.8376	0.8483	0.7275	0.6655	0.6933	

Table 1: Performance of our models and baseline models (Xia and Ding 2019) using precision, recall, and F1-measure as metrics on the ECPE task as well as the two sub-tasks.

extraction independently. While Inter-CE/Inter-EC uses the predictions of cause/emotion extraction to improve emotion/cause extraction.

ECPE-MLL is a joint framework proposed in this paper, which re-formalizes the ECPE task as a multi-label learning problem. In order to obtain emotion specific and cause specific clause representations for multi-label classification, we explored four multi-task learning settings. In addition to the ISML model proposed in this paper, we also explored the Indep, Inter-CE, and Inter-EC model. It should be noted that the results of our models on the ECPE task are based on Logical OR integration of EMLL and CMLL, as it performs best in all integration methods.

In order to simplify the description, we abbreviate the ECPE-2Steps and ECPE-MLL methods based on different settings as 2Steps-* and MLL-*. Among the previous methods, 2Steps-Indep is a baseline model and performs the worst on the ECPE task. 2Steps-Inter-EC uses the prediction of emotion extraction to enhance cause extraction, which improves the performance of cause extraction and ECPE tasks, and becomes the state-of-the-art model.

Compared to 2Steps-Inter-EC, our model MLL-ISML-6 has achieved substantial improvements on the ECPE, emotion extraction and cause extraction tasks (F1 scores have increased by 6.12%, 2.7%, and 4.43% on these three tasks, respectively), which proves the effectiveness of our method as a whole. We will discuss the effects of each part of our model in sections 4.4, 4.5, and 4.6.

4.4 ECPE-MLL vs. ECPE-2Steps

In order to verify the advantages of our proposed joint model ECPE-MLL compared to the

ECPE-2Steps model, we discard the ISML module and use Indep, Inter-CE, and Inter-EC for emotion/cause encoding. The experimental results are shown in Table 1.

On the ECPE task, compared with the 2Steps-* models using the same emotion/cause encoding settings, our proposed MLL-* models achieve large improvements in the recall rate (Specifically, the recall rates of MLL-Indep, MLL-Inter-CE, and MLL-Inter-EC are 11.78%, 9.13%, and 5.52% higher than 2Steps-Indep, 2Steps-Inter-CE, and 2Steps-Inter-EC, respectively), which further lead to a significant increase in F1 score. This indicates that our methods can extract more emotion-cause pairs than the previous methods.

In the emotion extraction and cause extraction subtasks, the MLL-Indep, MLL-Inter-CE and MLL-Inter-EC methods are also better than 2Steps-Indep, 2Steps-Inter-CE, and 2Steps-Inter-EC respectively. We attribute these improvements to multi-task learning, as we additionally introduced two multi-label learning tasks for emotion-cause pair extraction.

The above results show that compared with the two-step framework, our proposed multi-label learning framework for emotion-cause extraction has great advantages and potential.

4.5 The Effectiveness of Iterative Synchronized Multi-task Learning

Table 1 shows the performance of the MLL-ISML-*N* model with different number of iterations. It should be noted that the MLL-ISML-1 model and the MLL-Indep model are equivalent, so we do not repeat its results in the table.

The MLL-ISML-1 model does not model the correlation between emotion and cause, so its perfor-

Approach	CMLL			EMLL		
	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>
Indep	0.7096	0.5410	0.6128	0.7332	0.5487	0.6268
Inter-CE	0.7373	0.5850	0.6514	0.7301	0.5474	0.6244
Inter-EC	0.7090	0.5224	0.6004	0.7287	0.5986	0.6564
ISML-2	0.7301	0.5806	0.6460	0.7312	0.6056	0.6614
ISML-3	0.7291	0.6020	0.6585	0.7164	0.6181	0.6630
ISML-4	0.7343	0.6088	0.6649	0.7203	0.6223	0.6668
ISML-5	0.7185	0.6216	0.6656	0.7292	0.6200	0.6695
ISML-6	0.7306	0.6254	0.6729	0.7270	0.6252	0.6714
ISML-7	0.7297	0.616	0.6674	0.7219	0.6246	0.6694

Approach	Average Probability			Logical AND			Logical OR		
	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>
Indep	0.7674	0.5421	0.6347	0.8069	0.4635	0.5876	0.6686	0.6260	0.6462
Inter-CE	0.7514	0.5659	0.6446	0.7712	0.5235	0.6227	0.7117	0.6048	0.6526
Inter-EC	0.7546	0.5716	0.6497	0.7929	0.4888	0.6040	0.6814	0.6257	0.6515
ISML-2	0.7377	0.5953	0.6584	0.7728	0.5538	0.6446	0.6990	0.6350	0.6647
ISML-3	0.7440	0.6036	0.6652	0.7448	0.5887	0.6570	0.7009	0.6384	0.6674
ISML-4	0.7362	0.6121	0.6678	0.7566	0.5874	0.6608	0.7058	0.6356	0.6682
ISML-5	0.7348	0.6148	0.6686	0.7454	0.5991	0.6635	0.7116	0.6351	0.6702
ISML-6	0.7382	0.6202	0.6730	0.7500	0.6043	0.6680	0.7090	0.6441	0.6740
ISML-7	0.7211	0.6258	0.6694	0.7466	0.6049	0.6678	0.7068	0.6425	0.6720

Table 2: Performance of the CMLL, EMLL models and their integrated predictions on the ECPE task.

mance on the three tasks is the worst compared to other MLL-* models. As the number of iterations increases, the performance of the MLL-ISML-*N* model on all three tasks is generally getting better, especially when the number of iterations is increased from 1 to 2. One possible reason is that MLL-ISML-2 first introduced the interaction between emotion and cause. When the number of iterations is increased to 6, MLL-ISML-*N* achieves optimal performance on all three tasks.

In addition, MLL-ISML-2 achieves better performance than MLL-Inter-CE on the emotion extraction task, and better performance than MLL-Inter-EC on the cause extraction task, which shows that the single use of emotion extraction to help cause extraction and the use of cause extraction to help emotion extraction are not as effective as using them simultaneously.

The above results show that compared with the Inter-EC and Inter-CE models, our proposed ISML model can better utilize the interaction between emotion and cause, and significantly improve the performance of ECPE-MLL on three tasks.

4.6 The Effectiveness of CMLL, EMLL, and Integrated Prediction

Table 2 shows the performance of the CMLL, EMLL models and their integrated predictions on the ECPE task using different emotion/cause encoding settings. For simplicity, we refer to these models based on different settings as CMLL-*, EMLL-*,

AP-*, LAND-*, and LOR-*, for example, EMLL-Indep and LOR-Indep.

First of all, it can be seen that without integrated prediction, CMLL-* and EMLL-* alone can already perform very well. When under the same emotion/cause encoding setting, EMLL-* can obtain better performance in F1 score than CMLL-* in most cases. An interesting result is that CMLL-Inter-CE performs far better than CMLL-Inter-EC, and EMLL-Inter-EC performs far better than EMLL-Inter-CE. One possible explanation is that better emotion specific clause representation contributes to the performance of CMLL-*, better cause specific clause representation is beneficial to the performance of EMLL-*.

We now use CMLL-* and EMLL-* as the baselines and explore the effects of different integration methods under the same emotion/cause encoding setting.

- Average Probability. AP-* does not perform significantly better or worse than CMLL-* and EMLL-*.
- Logical AND. The recall rate and F1 score of LAND-* are lower than CMLL-* and EMLL-*, but the precision is higher than them.
- Logical OR. The recall rate of LOR-* is higher than that of CMLL-* and EMLL-*, but the precision score is lower than them. On the whole, the F1 score of LOR-* is mostly higher than CMLL-* and EMLL-*.

I borrowed money from all my relatives (c_1), and finally got enough money for surgery (c_2), but we can no longer afford the expensive medical expenses (c_3). He was discharged from hospital on the second day after surgery (c_4). Even the thread removal after the surgery (c_5), was all done by himself in front of the mirror (c_6). Recalling the hardest days (c_7), Xu Yongfen still couldn't help choking (c_8).				
Ground-truth: c_8-c_7	Inter-EC: Empty	EMLL: c_8-c_7	CMLL: c_8-c_7	MLL: c_8-c_7
The director told the reporter (c_1), when he heard this decision at the company's year-end banquet (c_2), he and his girlfriend were very surprised (c_3). They established a love relationship in August last year (c_4). Because they are afraid of being known by the company (c_5), they are usually sneaky (c_6), and most of their communication in the company is through QQ (c_7). They avoid saying anything intimate in front of their colleagues (c_8), and dare not make their relationship public (c_9).				
Ground-truth: c_3-c_2	Inter-EC: c_3-c_3 , c_3-c_2	EMLL: c_3-c_2	CMLL: Empty	MLL: c_3-c_2

Figure 3: Case study between Inter-EC and our models.

Approach	Emotion-Cause Pair Ext.		
	P	R	$F1$
MLL-Indep-AS	0.6762	0.6288	0.6505
MLL-Inter-CE-AS	0.6733	0.6374	0.6538
MLL-Inter-EC-AS	0.6755	0.6374	0.6540
MLL-ISML-6-AS	0.6800	0.6388	0.6578

Table 3: Performance of our models on the ECPE task when auxiliary supervisions of emotion extraction and cause extraction are removed.

The above results show that, compared to EMLL-* and CMLL-* alone, LOR-* is a better choice when we need a model with a higher recall rate and F1 score. When we need a model with a higher precision score, then LAND-* is a better choice.

4.7 The Effectiveness of Auxiliary Supervision

We explored the effectiveness of auxiliary supervisions of emotion extraction and cause extraction by removing them from our models. For simplicity, we refer to the model without auxiliary supervisions as '*-AS'. The results are shown in Table 3 (based on Logical OR integration method).

For the MLL-(Indep/Inter-CE/Inter-EC) models, removing the auxiliary supervisions did not lead to a decrease in their performance on the ECPE task, but resulted in a consistent improvement (though not much). These results indicate that our joint models do not rely on independent emotion and cause extraction. However, for the MLL-ISML-6 model, its performance is significantly reduced when the auxiliary supervisions are removed. This is intuitive because the ISML model cannot work normally without auxiliary supervisions. Nevertheless, it still outperforms MLL-Indep-AS.

4.8 Case Analysis

Figure 3 shows the advantages of our models (all using ISML-6 settings) over the Inter-EC (using 2

step framework) model through two examples in the test set. In the first example, the Inter-EC model failed to extract the emotion-cause pair c_8-c_7 . In contrast, our EMLL and CMLL models successfully identified this pair. We found this situation is quite common in the test set, which ultimately leads to a significant increase in the recall rate of our model compared to Inter-EC.

In the second example, Inter-EC extracted two emotion-cause pairs: c_3-c_3 and c_3-c_2 , where the latter is correct and the former is wrong. By observing the results of the first step in Inter-EC, we found that c_3 is predicted to be an emotion clause, c_2 and c_3 are predicted to be cause clauses. The filter ultimately failed to eliminate the invalid pair c_3-c_3 . In contrast, our models are more "cautious", EMLL did not output other pairs after predicting the correct one, CMLL did not even output any pairs. The MLL model uses the union of the output of EMLL and CMLL to identify as many emotion-cause pairs as possible. These situations make our models have higher precision performance than the Inter-EC model, and make the MLL model more robust than EMLL and CMLL models.

5 Conclusion

The emotion-cause pair extraction (ECPE) task is a new direction in emotion analysis. To overcome the shortcomings of the existing two-step approach, we propose a sliding window multi-label learning scheme. Specifically, we assume that all clauses in the document are emotion clauses, and build an emotion-oriented sliding window centered on each of them. Then, in each window, we use the emotion clause as a pivot to extract the corresponding one or more cause clauses based on multi-label learning (CMLL). At the same time, we propose the dual form of CMLL, i.e., EMLL, which uses the cause clauses as the pivot to extract the corresponding

emotion clauses. The final predictions are obtained by integrating the results of CMLL and EMLL. We evaluated our model on a benchmark emotion cause dataset, and the experimental results show that our method has achieved a substantial improvement over the state-of-the-art method.

Acknowledgments

We would like to thank three anonymous reviewers for their valuable comments. This work was supported by the Natural Science Foundation of China (No. 61672288, 62076133, and 62006117), and the Natural Science Foundation of Jiangsu Province for Young Scholars (SBK2020040749), and Distinguished Young Scholars (SBK2020010154), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX20.0298).

References

- Ying Chen, Wenjun Hou, Xiyao Cheng, and Shoushan Li. 2018. Joint learning for emotion classification and emotion cause detection. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 646–651.
- Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Computational Linguistics (COLING)*, pages 179–187.
- 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.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, pages 4171–4186.
- Zixiang Ding, Huihui He, Mengran Zhang, and Rui Xia. 2019. From independent prediction to re-ordered prediction: Integrating relative position and global label information to emotion cause identification. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 6343–6350.
- Zixiang Ding, Rui Xia, and Jianfei Yu. 2020. ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. In *Association for Computational Linguistics (ACL)*, pages 3161–3170.
- 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 *Association for Computational Linguistics (ACL)*, pages 3707–3717.
- Kai Gao, Hua Xu, and Jiushuo Wang. 2015a. Emotion cause detection for chinese micro-blogs based on e-coc model. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 3–14.
- Kai Gao, Hua Xu, and Jiushuo Wang. 2015b. A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Systems with Applications*, 42(9):4517–4528.
- Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. 2015. Detecting emotion stimuli in emotion-bearing sentences. In *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, pages 152–165.
- Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. A question answering approach to emotion cause extraction. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1593–1602.
- Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016a. Event-driven emotion cause extraction with corpus construction. In *Empirical Methods in Natural Language Processing (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.
- Lin Gui, Li Yuan, Ruifeng Xu, Bin Liu, Qin Lu, and Yu Zhou. 2014. Emotion cause detection with linguistic construction in chinese weibo text. In *Natural Language Processing and Chinese Computing (NLPCC)*, pages 457–464.
- Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. 2010. A text-driven rule-based system for emotion cause detection. In *NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 45–53.
- Weiyuan Li and Hua Xu. 2014. Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, 41(4):1742–1749.
- Xiangju Li, Kaisong Song, Shi Feng, Daling Wang, and Yifei Zhang. 2018. A co-attention neural network model for emotion cause analysis with emotional context awareness. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 4752–4757.
- Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. *Empirical Methods in Natural Language Processing (EMNLP)*, pages 3728–3738.

Alena Neviarouskaya and Masaki Aono. 2013. Extracting causes of emotions from text. In *International Joint Conference on Natural Language Processing (IJCNLP)*, pages 932–936.

Irene Russo, Tommaso Caselli, Francesco Rubino, Ester Boldrini, and Patricio Martínez-Barco. 2011. Emocause: an easy-adaptable approach to emotion cause contexts. In *Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pages 153–160.

Shuangyong Song and Yao Meng. 2015. Detecting concept-level emotion cause in microblogging. In *World Wide Web (WWW)*, pages 119–120.

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

Rui Xia and Zixiang Ding. 2019. Emotion-cause pair extraction: A new task to emotion analysis in texts. In *Association for Computational Linguistics (ACL)*, pages 1003–1012.

Rui Xia, Mengran Zhang, and Zixiang Ding. 2019. RTHN: A RNN-transformer hierarchical network for emotion cause extraction. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5285–5291.

Bo Xu, Hongfei Lin, Yuan Lin, Yufeng Diao, Liang Yang, and Kan Xu. 2019. Extracting emotion causes using learning to rank methods from an information retrieval perspective. *IEEE Access*, 7:15573–15583.

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.

Shuntaro Yada, Kazushi Ikeda, Keiichiro Hoashi, and Kyo Kageura. 2017. A bootstrap method for automatic rule acquisition on emotion cause extraction. In *IEEE International Conference on Data Mining Workshops*, pages 414–421.

Xinyi Yu, Wenge Rong, Zhuo Zhang, Yuanxin Ouyang, and Zhang Xiong. 2019. Multiple level hierarchical network-based clause selection for emotion cause extraction. *IEEE Access*, 7(1):9071–9079.

A Experimental Results with BERT

We explored the effect of using pre-trained BERT (Devlin et al., 2019) as the clause encoder in ECPE-MLL. The results are shown in Table 4.

Specifically, we follow (Liu and Lapata, 2019) to feed the entire document into pre-trained BERT. Each clause in the document is expanded to start

Approach	Emotion Ext.		
	<i>P</i>	<i>R</i>	<i>F1</i>
ECPE-MLL	0.8582	0.8429	0.8500
ECPE-MLL(BERT)	0.8608	0.9191	0.8886
Approach	Cause Ext.		
	<i>P</i>	<i>R</i>	<i>F1</i>
ECPE-MLL	0.7248	0.6702	0.6950
ECPE-MLL(BERT)	0.7382	0.7912	0.7630
Approach	Emotion-Cause Pair Ext.		
	<i>P</i>	<i>R</i>	<i>F1</i>
ECPE-MLL	0.7090	0.6441	0.6740
ECPE-MLL(BERT)	0.7700	0.7235	0.7452

Table 4: Performance of our models with/without pre-trained BERT encoder on the ECPE task as well as the two sub-tasks.

with the word “[CLS]” and end with the word “[SEP]”. We also assign interval segment embeddings E_A or E_B to each clause c_i depending on whether i is odd or even. The input representation of each word is the sum of three parts: word embedding, position embedding, and segment embedding. The representations of all clauses in the document can be obtained by gathering the representations of all corresponding “[CLS]” tokens. After that, we use two clause level transformers to obtain the emotion-specific representation and cause-specific representation for each clause, respectively.

Our model is built based on this implementation: <https://github.com/google-research/bert>, and is initialized using the pre-trained BERT model “BERT-Base, Chinese”. We apply the linear warmup and linear decay mechanism to the learning rate. The batch size and initial learning rate are set to 8 and $2e-5$, respectively. It should be noted that we use the fixed learning rate $2e-3$ for the parameters in the logistic regression layer. Readers can refer to the source code for more implementation details.

The experimental results in Table 4 show that the performance of our model on three tasks can be greatly improved by using BERT as the clause encoder. In particular, the recall rate on the cause extraction subtask has increased by 12%, indicating that the pre-trained language models have great potential in the emotion cause related tasks.