



# An End-to-End Multi-task Learning Network with Scope Controller for Emotion-Cause Pair Extraction

Rui Fan<sup>1,2,4</sup>, Yufan Wang<sup>1,3,4</sup>, and Tingting He<sup>1,2,4</sup>(✉)

<sup>1</sup> Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, China Central Normal University, Wuhan 430079, China

{fanrui, yufan\_wang}@mails.ccnu.edu.cn, tthe@mail.ccnu.edu.cn

<sup>2</sup> School of Computer, Central China Normal University, Wuhan 430079, China

<sup>3</sup> National Engineering Research Center for E-Learning, China Central Normal University, Wuhan 430079, China

<sup>4</sup> National Language Resources Monitor & Research Center for Network Media, Wuhan 430079, China

**Abstract.** Emotion-cause pair extraction (ECPE) aims to extract all potential pairs of emotions and corresponding causes in a document. It has an advantage over traditional emotion cause extraction (ECE) that it does not require annotating emotions manually. Existing methods for ECPE task are based on two-step framework. However, they ignore the fact that the emotion-cause pair is regarded as a whole unit and there are cascading errors in two-step framework. In this paper, we propose an end-to-end hierarchical neural network model, which directly extracts emotion-cause pairs and enhances mutual interaction between emotions and causes via multi-task learning. In addition, we introduce a scope controller to constrain the emotion-cause pair predictions in a high probability area, according to the position correlation between emotions and causes. The experimental results demonstrate that our method achieves the state-of-the-art performance and improves F-measure by 2.24%.

**Keywords:** Emotion-cause pair extraction · Multi-task learning · Emotion cause analysis · Scope controller

## 1 Introduction

Emotion cause analysis is an important task in natural language processing. It has gained increasing attention in recent years and been applied to many real-world applications. Emotion cause extraction (ECE) is a fine-grained task of emotion cause analysis, which intends to discover the possible causes behind emotions. However, [1] mentioned that the ECE task has two flaws: 1) its application in real world is limited because of few documents with the annotation of emotions, 2) the way of first annotating emotions and

---

R. Fan and Y. Wang—contributed equally to this work and should be regarded as co-first authors.

© Springer Nature Switzerland AG 2020

X. Zhu et al. (Eds.): NLPCC 2020, LNAI 12430, pp. 764–776, 2020.

[https://doi.org/10.1007/978-3-030-60450-9\\_60](https://doi.org/10.1007/978-3-030-60450-9_60)

then detecting causes ignores the intrinsic relationship between them. Therefore, they proposed a new task of emotion-cause pair extraction (ECPE) and constructed a new corpus based on the ECE benchmark corpus [2]. The object of the ECPE task is to extract all emotion-cause pairs of emotions and their corresponding causes in a document. The Fig. 1 shows the difference between ECE and ECPE. In ECPE, emotion-cause pairs  $(c_4, c_2)$  and  $(c_4, c_3)$  in this document are extracted directly without providing the annotation of emotion clause.

Document	ECE
Clause $c_1$ : Yesterday morning,	$c_4$ (label) $\rightarrow$ $c_2, c_3$
Clause $c_2$ : a policeman visited the old man with the lost money.(cause)	
Clause $c_3$ : and told him that the thief was caught.(cause)	<b>ECPE</b> Pair: $(c_4, c_2), (c_4, c_3)$
Clause $c_4$ : The old man was very happy.(emotion)	
Clause $c_5$ : and deposited the money in the bank.	

**Fig. 1.** Examples of the ECE task and the ECPE task.

The previous studies of emotion cause analysis are generally rule-based methods [3–8] and machine learning methods [2, 9–11]. They need a lot of manual work to define linguistic rules or features. In recent years, due to the powerful learning ability of neural networks, many deep learning methods are proposed to solve the ECE task [12–17]. For solving the ECPE task, [1] proposed methods based on two-step framework, which first complete emotion extraction and cause extraction individually and then conduct pairing and filtering.

Although this two-step strategy performs well in the ECPE task, there are some disadvantages. At first, this method is not natural enough. Humans usually get emotions and their causes according to the correspondence between them, but in the two-step strategy, it first extracts emotions and causes and then pairs them, which ignores the intrinsic relationship of emotions and causes as a whole unit. Secondly, the false prediction of the first step will directly affect the performance of the second step.

Emotion is the demonstration of the potential cause and cause is the source of the emotion expression. Cause and emotion should be regarded as a whole unit, rather than being separated into two independent parts. Consequently, we propose an end-to-end neural network model, which directly extracts all emotion-cause pairs in a document. Besides the main task emotion-cause pair extraction, we also introduce emotion extraction and cause extraction as the auxiliary tasks. They not only increase the interaction between emotions and causes, but also provide a more appropriate semantic representation for the main task. In addition, the causes usually occur on positions very close to the emotion word for an emotion event [17]. Therefore, we propose the scope controller based on the relative position of emotions and causes, which can capture the position of emotion-cause pairs more accurately.

There is not much research on the ECPE task because it is a new task. We take the methods of two-step strategy as our baselines, one of which is the state-of-the-art. Compared with them, the F-measure of our method on the emotion-cause pair extraction task is increased by 2.24% at least.

In summary, the main contributions of our work are as follows:

- We propose a novel end-to-end neural network model for ECPE task. It effectively solves the issues of two-step framework.
- Based on the position correlation between emotions and causes, we propose a scope controller to control the scope of emotion-cause pair predictions.
- Considering the mutual interaction between emotions and causes. We introduce two auxiliary tasks: emotion extraction and cause extraction and demonstrate the efficiency of multi-task learning on emotion-cause pair extraction.

## 2 Related Work

As a new task developed by ECE, ECPE has a strong correlation with ECE. Following, we describe the study of ECE and ECPE. Since [3] proposed the emotion cause extraction task, it has been extensively studied in recent years. They defined this task as a word-level sequence labeling problem. [4] designed two sets of linguistic patterns to extract emotion cause, including manually generalized patterns and automatically generalized patterns. [18] proposed a rule-based emotion cause detection method which used 25 manually compiled rules. [9] built a conditional random fields (CRFs) learner to detect the emotion stimulus spans in emotion sentences. [2, 19] released a Chinese emotion cause corpus and redefined the ECE task as a clause-level binary-classification problem. [11] proposed to identify causes by ranking candidate clauses from the perspective of information retrieval.

With the rise of neural network, deep learning models and attention mechanisms have been widely applied in ECE task. [20] used long short-term memory for emotion cause extraction. [12] regarded emotion cause extraction as a question-answering task and used the memory network to store context information. [13, 16] presented the attention-based neural network to capture the mutual influences between emotions and causes. [15] proposed a three-level hierarchical network to model clause representation. From discourse and some prior knowledge on text understanding, [17] proposed a method based on knowledge regularization for emotion cause prediction.

All of the above methods are used to solve the traditional ECE task, which is to extract the potential causes for the given annotated emotions. However, the emotions of a real-world document are generally not labeled. Therefore, [1] proposed the emotion-cause pair extraction (ECPE) task and a two-step strategy to address this task.

In order to satisfy real-world scenarios, model the mutual interactions between emotions and causes, and avoid the cascading errors from two-step models, we propose an end-to-end model for ECPE task, which directly extracts emotion-cause pairs.

## 3 Approach

### 3.1 Task Definition

Firstly, we describe the definition of the ECEP task. Given a document  $d = \{c_1, c_2, \dots, c_n\}$  consisting of  $n$  clauses, where each clause  $c_i = \{w_{i,1}, w_{i,2}, \dots, w_{i,m}\}$

consists of  $m$  words, the goal of ECPE is to extract a pair set  $P = \{\dots, (c^e, c^c), \dots\}$  containing all emotion-cause pairs in the document  $d$ , where  $c^e$  is an emotion clause and  $c^c$  is a cause clause.

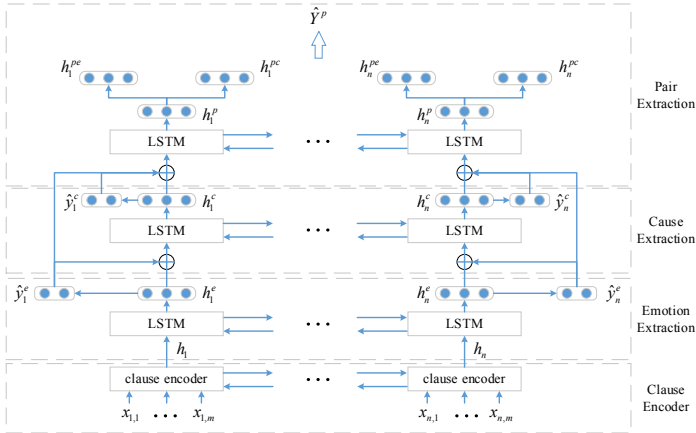


Fig. 2. The overall architecture of RHNSC.

### 3.2 Overall Architecture

The RNN-Based Hierarchical Network with Scope Controller (RHNSC) model will be described in this section. The overall architecture of RHNSC is shown in Fig. 2. It includes four parts: clause encoder, emotion extraction, cause extraction and pair extraction. Firstly, we obtain the clause representation and then we introduce emotion extraction and cause extraction to learn more informative representation and enhance the mutual influence between emotions and causes. Lastly, the goal of emotion-cause pair extraction is achieved by pair extraction and the scope controller further constrains the result. Following, we will introduce each part in detail.

### 3.3 Clause Encoder

In the ECPE task, a document consists of several clauses, which is the basic unit. The significance of each word to a clause semantic representation is different. So we adopt an attention-based network to obtain an effective clause representation. Specifically, we first feed embedding vector of each word into a Bi-LSTM and concatenate the hidden states of the two directions as the word representation. Then, we assigning different weights to word representations by the attention mechanism and finally the weighted sum of them is the clause representation.

### 3.4 Two Auxiliary Tasks

We have mentioned that emotions and causes have strong correlation. To utilize the influence between them for the later emotion-cause pair extraction task, we add two

auxiliary tasks into the model: emotion extraction and cause extraction. By these two auxiliary tasks, we can get a clause representation combining interactive information between emotions and causes. In the final main task, we further explore the impact of them for emotion-cause pairs.

**Emotion Extraction.** The goal of emotion extraction is to identify each clause whether it is an emotion clause in a document. This is a simple binary classification problem. Formally,

$$h_i^e = BiLSTM_e(h_i), \quad (1)$$

$$\hat{y}_i^e = \text{softmax}(W_e h_i^e + b_e), \quad (2)$$

where  $h_i$  is the clause representation of clause  $c_i$ .  $h_i^e$  is concatenation from the hidden states of the two directions,  $W_e$  and  $b_e$  are learning parameters. Then we get the emotion prediction  $\hat{y}_i^e$  by the softmax function.

**Cause Extraction.** In the same way, we introduce the cause extraction, which is to detect whether the clause  $c_i$  is the cause clause. It is different from emotion extraction that we expect that the predictions of emotion can help the cause detection and the results of the cause prediction influence emotion identification. Therefore, we use the prediction of emotion extraction as a part of the input in cause extraction. Formally,

$$h_i^c = BiLSTM_c([h_i^e, \hat{y}_i^e]), \quad (3)$$

$$\hat{y}_i^c = \text{softmax}(W_c h_i^c + b_c), \quad (4)$$

where the two directional hidden states are concatenated as  $h_i^c$ ,  $W_c$  and  $b_c$  are learning parameters,  $\hat{y}_i^c$  is the predicted cause class distribution.

### 3.5 Emotion-Cause Pair Extraction

Emotion-cause pair extraction is our main task. We use the representation after two auxiliary tasks mining more semantic expression as the input. Emotion-cause pair is regarded as a whole unit and we cannot ignore its correlation with emotion and cause. Analogously, we append the prediction results of two auxiliary tasks as the input.

It is natural that an emotion-cause pair contains two factors: one is the emotion factor, the other is the cause factor. By two auxiliary tasks learning, we have obtained the representation containing emotion and cause features. But we need separate it into the emotion factor and the cause factor for pairing task. Formally,

$$h_i^p = BiLSTM_p([h_i^c, \hat{y}_i^e, \hat{y}_i^c]), \quad (5)$$

$$h_i^{pe} = W_{pe} h_i^p + b_{pe}, \quad (6)$$

$$h_i^{pc} = W_{pc} h_i^p + b_{pc}, \quad (7)$$

where  $W_{pe}, W_{pc}, b_{pe}, b_{pc}$  are parameters,  $h_i^p$  is the concatenation of two directional hidden states at time step  $i$ ,  $h_i^{pe}$  and  $h_i^{pc}$  are obtained through two linear layers. They represent the emotion factor and cause factor for each clause.

Then  $h_i^{pe}$  and  $h_i^{pc}$  at time step  $i (1 \leq i \leq n)$  are concatenated to get emotion factor matrix  $H^{pe} = [h_1^{pe}, h_2^{pe}, \dots, h_n^{pe}]$  and cause factor matrix  $H^{pc} = [h_1^{pc}, h_2^{pc}, \dots, h_n^{pc}]$ ,

**Table 1.** The proportion of emotion-cause pairs with different relative positions in the dataset.

Relative position	Number	Percentage
0	511	23.58%
$\leq 1$	1853	85.51%
$\leq 2$	2027	95.85%
$\leq 3$	2127	98.15%
$\leq 4$	2142	98.85%
All	2167	100%

where  $n$  is the total number of clauses in a document. Finally, we multiply the two matrices and pass the result through the sigmoid function to gain the prediction of each pair:

$$\hat{Y}^p = \text{sigmoid}(H^{pe}H^{pcT}), \quad (8)$$

where  $\hat{Y}^p = \{\dots, \hat{y}_{i,j}^p, \dots\}$ ,  $\hat{Y}^p \in R^{n \times n}$ ,  $\hat{y}_{i,j}^p = \text{sigmoid}(h_i^{pe}h_j^{pcT})$  indicates that the probability of pair  $(c_i, c_j)$  is an emotion-cause pair. In other words, it represents the pairing score of the emotion factor of clause  $c_i$  and the cause factor of clause  $c_j$ .

### 3.6 Scope Controller

We have mentioned that the cause usually occurs near the emotion expression in a document. In Table 1, we report the proportion of emotion-cause pairs with different relative position in the dataset. We can observe that when the relative position is less than 2, it already contains 95.85% emotion-cause pairs. However, unlike the relative position in the ECE task, we cannot determine the relative position by the emotion labels. Hence, we define the scope as the description of likely position for emotion-cause pairs. If the scope equals to 2, we expect that our model pays more attention to the predicted pairs with the relative position of two clause less than 2 and ignores other predicted pairs with larger relative position.

After observation, we find an issue that the predicted distribution of  $\hat{Y}^p$  may not be concentrated enough. Therefore, we want to take advantage of such a position correlation to address this issue. We propose a scope controller to narrow the difference between the predicted distribution and the true distribution of emotion-cause pairs. We assume that the expected pair position value is 1 and the uncorrelated position is 0. Then, we define

a matrix  $P = \{\dots, p_{i,j}, \dots\}$ ,  $P \in R^{n \times n}$ , that is the same size as  $\hat{Y}^p$ . The definition of  $p_{i,j}$  is expressed as follows:

$$p_{i,j} = \begin{cases} 1 & \text{if } |i - j| \leq s \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

where  $s$  is the scope of the expected pairs, that is, the relative position  $|i - j|$  of the pair  $(c_i, c_j)$  is less than scope  $s$  and the  $p_{i,j}$  equals 1 in the matrix  $P$ . The matrix  $P$  and  $\hat{Y}^p$  are regarded as the target and prediction respectively.

When training, we use cross entropy to measure the difference between them:

$$L_{pc} = - \sum_i \sum_j p_{i,j} \cdot \log(\hat{y}_{i,j}^p), \quad (10)$$

the decrease of  $L_{pc}$  make the position of predicted emotion-cause pairs in  $\hat{Y}^p$  more concentrated near the diagonal, in other words, the pairs in this area with a small relative position. Under the restriction of the matrix  $P$ , the predicted emotion-cause pairs of our model are more concentrated in the high probability area.

## 4 Experiments and Results

### 4.1 Datasets and Metrics

We select the ECPE corpus [1] which is constructed based on the benchmark ECE corpus [2]. The ECPE corpus consist of 1945 documents, where each document contains one or more emotion-cause pairs. We use 10-fold cross-validation to evaluate our model and report the average result of 10 times. The precision (P), recall (R) and F1 score are employed to measure the performance in ECPE task.

### 4.2 Experiments Setting

The size of the vocabulary is 24165, where 12151 word vectors are pre-trained on the corpus from Chinese Weibo with Word2vec [21]. The remaining word vectors, all parameter matrices and bias use uniform distribution of random initialization on  $(-0.01, 0.01)$ . The word embedding dimension is 200 and all Bi-LSTM hidden state dimension in the model is 100. When training the model, we adopt the Adam optimizer [22]. Learning rate and batch size are set to 0.005 and 32 respectively. In addition, we add L2 regularization on the output layer. For all three tasks, we use the cross entropy as a loss function during model training. We first train two auxiliary tasks and then train all loss together, including  $L_{pc}$ .

### 4.3 Evaluation of the ECPE Task

The number of researches on the ECPE task is not extensive because it is a new task proposed recently. We select three methods: Indep, Inter-CE, Inter-EC [1] as our base-lines, which are based on two-step framework. In the first step, these methods regard emotion-cause pair extraction as two individual tasks:

- **Indep** completing emotion extraction and cause extraction independently.
- **Inter-CE** using the predictions of cause extraction to improve emotion extraction.
- **Inter-EC** using the predictions of emotion extraction to improve cause extraction.

In the second step, these methods share the same strategy, which first apply Cartesian product to the emotion set and the cause set from the first step to get a pair set, and then filter the pair set by a classifier to get the final emotion-cause pairs.

As shown in Table 2, we compare our RHN and RHNSC model with baselines, where RHN is RHNSC without scope controller. Obviously, our models outperform these two-step methods in precision, recall and F1 score. In particular, the performance of baselines increases significantly after filtering pairs, while our models are almost no improvement. This phenomenon proves the efficiency of our end-to-end model. The reason will be detailed in the next subsection. Here we only focus on the performance of RHNSC without pair filter, which still outperforms baselines. In addition, we notice that Inter-CE and Inter-EC perform better than Indep. As mentioned earlier, emotions and causes are not completely independent. Using emotion or cause predictions to help identify another one can enhance interaction between them.

**Table 2.** Experimental results of models on the ECPE with or without the pair filter.

	Without pair filter			With pair filter			
	P	R	F1	P	R	F1	<i>Keep_rate</i>
Indep	0.5894	0.5114	0.5451	0.6832	0.5082	0.5818	0.8507
Inter-CE	0.5883	0.5192	0.5500	0.6902	0.5135	0.5901	0.8412
Inter-EC	0.6019	0.5775	0.5842	0.6721	0.5705	0.6128	0.8889
RHN	0.6742	0.5354	0.5958	0.6773	0.5354	0.5969	0.9952
<b>RHNSC</b>	<b>0.6944</b>	<b>0.5871</b>	<b>0.6352</b>	<b>0.6956</b>	<b>0.5871</b>	<b>0.6357</b>	<b>0.9983</b>

Comparing Inter-CE and Inter-EC, the result of Inter-EC is better. The emotion extraction task is less difficult than the cause extraction task, hence emotion predictions can provide more improvement to the cause extraction task. That is why we select emotion extraction as lower-level auxiliary task.

#### 4.4 Effect of End-to-End Approach

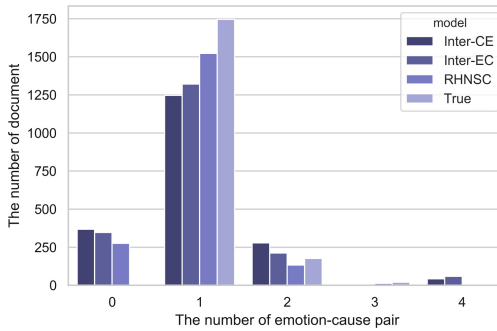
Inter-CE model and Inter-EC model are based on two-step framework. To prove the effectiveness of end-to-end approach, we take the same filtering method for RHN and RHNSC model. In Table 2, we display the result of emotion-cause pair extraction with and without pair filter, where *keep\_rate* indicates that the final result retains the proportion of emotion-cause pairs after filtering. Obviously, the performance of RHN and RHNSC is almost not impacted by filtering, and the *keep\_rate* reaches 99.83% for RHNSC. The F1 score and precision of these two-step models are improved significantly, but the recall is slightly reduced. Because applying the Cartesian product will contain all possible



pairs, which may include some extra pairs that are not correct, especially the predicted emotions and causes are not unique in the first step.

For example, there are five clauses. We first get one predicted emotion clause  $c_1$  and two predicted cause clauses  $c_2$  and  $c_3$ , while  $c_3$  is a wrong prediction. We get two emotion-cause pairs  $(c_1, c_2)$  and  $(c_1, c_3)$  by applying Cartesian product, but  $(c_1, c_3)$  is a wrong result. Similarly, as long as the emotion clause and cause clause in the first step are not unique or wrong, extra pairs will exist in the second step. This is the two-step framework shortcoming: the result of first step will affect the next step performance and the final result cannot guide the previous learning.

In Fig. 3, we categorize documents into 5 groups on the number of emotion-cause pairs in each document. The number of documents in other cases is too few to consider. Comparing the number of documents in each group using the results of Inter-CE, Inter-EC, RHNSC prediction and ground truth. The display of the chart is consistent with our analysis. Inter-CE model and Inter-EC model predict more two-pair and four-pair documents than reality. Therefore, they need filter the extra pairs in the second step. However, our RHNSC model is closer to the true distribution. In addition, in the zero-pair group, the number of documents is low than the two-step methods. It is the advantage of the end-to-end approach that our model directly extracts emotion-cause pairs, which tackle the problem of predicting extra pairs to a certain extent.



**Fig. 3.** The number of documents with different number of emotion-cause pairs in Inter-CE, Inter-EC, RHNSC prediction and ground truth.

#### 4.5 Effect of Multi-task Learning

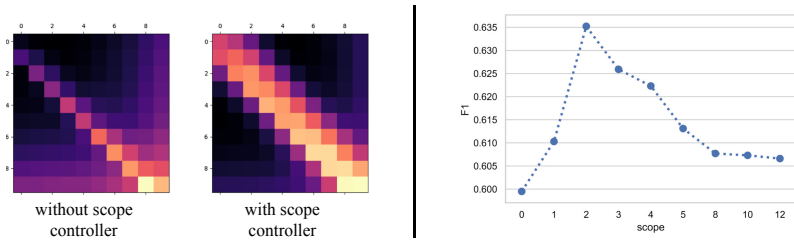
In this section, we describe the effectiveness of two auxiliary tasks in detail. We compare the performance of different models: RHNSC(P), RHNSC(CEP), RHNSC(ECP), where CEP means cause extraction as the lower layer, emotion extraction as the middle layer, and emotion-cause pair extraction as the upper layer. ECP can be interpreted as the same way.

In Table 3, compared with RHNSC(P), RHNSC(CEP) and RHNSC(ECP) get significant improvement. We are amazed at the impact of the two auxiliary tasks on the pairing results. When observing the training process, we find that there are always one or

**Table 3.** The result of models. P, E, C represent emotion-cause pairs extraction, emotion extraction and cause extraction.

	P	R	F1
RHNSC(P)	0.5789	0.4474	0.4497
RHNSC(CEP)	0.6915	0.5642	0.6189
<b>RHNSC(ECP)</b>	<b>0.6944</b>	<b>0.5871</b>	<b>0.6352</b>

two results that are abnormally low among the ten results using 10-fold cross-validation method. After many experiments, we believe that this is not an accidental phenomenon, because this phenomenon disappears completely after adding the two auxiliary tasks. We guess that the clause representation trained by the two auxiliary tasks implies more useful information for the pairing task. As we mentioned before, emotion factor and cause factor are required for pairing task. The learning of these two auxiliary tasks can extract these two corresponding factors better. In addition, the mutual interaction of these three tasks is fully utilized in a holistic model. Therefore, the model can learn more rapidly.



**Fig. 4.** The impact of scope controller. The left part is the prediction matrices of emotion-cause pair. The right part is the F1 in different scopes.

Comparing RHNSC(CEP) and RHNSC(ECP), we observe that if the lower layer is emotion extraction task and the middle layer is cause extraction task, the performance of emotion-cause pair extraction task is better, because the emotion prediction is simpler than the cause prediction. In ECP, the three tasks are sorted from simple to difficult, which is in line with the characteristics of the learning process.

### 4.6 Effect of Scope Controller

We explore the performance of our model with or without using the scope controller. In Table 2, RNH is a hierarchical neural network without a scope controller. Compared with RHN, the scope controller brings 3.94% improvement in F1 score. Indeed, the relative position of emotions and causes is a very important factor in the emotion-cause pair extraction task. We add a scope controller to capture the correlation, which can effectively narrow the predicted emotion-cause pairs in a high probability scope.

The left part of Fig. 4 displays the matrices of prediction with or without scope controller. After using the scope controller, the predicted emotion-cause pairs are concentrated near the diagonal of the matrix, which is an area where emotion-cause pairs frequently occur. Contrastively, the matrix of not using scope controller is less prominent near the diagonal. In particular, the upper right near the diagonal is usually of a high probability, but there is almost the darkest color. This explains why using the scope controller can gain better performance.

We further investigate the effect of different scopes. The result is reported in the right part of Fig. 4. The performance is best when the scope is 2. Since it has almost covered all the emotion-cause pairs. If enlarging the scope, the performance decreases. When the scope is 12, which means containing all emotion-cause pairs in the dataset, the result still does not increase. Because the area with wide scope contains more irrelevant pairs, we should control the predicted emotion-cause pairs in the high probability area. A large scope allows the model to learn predictions of pairs farther away from the high probability area.

## 5 Conclusions

In this paper, we propose an end-to-end RNN-Based Hierarchical Network with Scope Controller (RHNSC) model for emotion-cause pair extraction. Our model takes emotion-cause pair extraction as main tasks and adds two auxiliary tasks: emotion extraction, cause extraction for mutual interaction between emotions and cause. The experimental results demonstrate that our method achieves state-of-the-art performance and addresses issue of the two-step framework to a certain extent. We examine and analyze the effectiveness of end-to-end approach for ECPE tasks.

**Acknowledgments.** This research is supported by the National Natural Science Foundation of China (61532008), the National Natural Science Foundation of China (61572223), the National Key Research and Development Program of China (2017YFC0909502), and Wuhan Science and Technology Program (2019010701011392).

## References

1. Xia, R., Ding, Z.: Emotion-Cause pair extraction: a new task to emotion analysis in texts. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics ACL 2019, Florence, Italy, 28 July–2 August 2019 (2019)
2. Gui, L., Wu, D., Xu, R., Lu, Q., Zhou, Y.: Event-driven emotion cause extraction with corpus construction. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing EMNLP 2016, Austin, Texas, USA, 1–4 November 2016, pp. 1639–1649 (2016)
3. Lee, S.Y.M., Chen, Y., Huang, C.R.: A text-driven rule-based system for emotion cause detection. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pp. 45–53. Association for Computational Linguistics (2010)

4. Chen, Y., Lee, S.Y.M., Li, S., Huang, C.R.: Emotion cause detection with linguistic constructions. In: Proceedings of the Conference 23rd International Conference on Computational Linguistics COLING 2010, 23–27 August 2010, Beijing, China, pp. 179–187 (2010)
5. Neviarouskaya, A., Aono, M.: Extracting causes of emotions from text. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing (IJCNLP), pp. 932–936 (2013)
6. Gao, K., Xu, H., Wang, J.: Emotion cause detection for chinese micro-blogs based on ECOCC model. In: Cao, T., Lim, E.-P., Zhou, Z.-H., Ho, T.-B., Cheung, D., Motoda, H. (eds.) PAKDD 2015. LNCS (LNAI), vol. 9078, pp. 3–14. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-18032-8\\_1](https://doi.org/10.1007/978-3-319-18032-8_1)
7. Gao, K., Xu, H., Wang, J.: A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Syst. Appl.* **42**(9), 4517–4528 (2015)
8. Yada, S., Ikeda, K., Hoashi, K., Kageura, K.: A bootstrap method for automatic rule acquisition on emotion cause extraction. In: IEEE International Conference on Data Mining Workshops, pp. 414–421 (2017)
9. Ghazi, D., Inkpen, D., Szpakowicz, S.: Detecting emotion stimuli in emotion-bearing sentences. In: Gelbukh, A. (ed.) CILing 2015. LNCS, vol. 9042, pp. 152–165. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-18117-2\\_12](https://doi.org/10.1007/978-3-319-18117-2_12)
10. Song, S., Meng, Y.: Detecting concept-level emotion cause in microblogging. In: Proceedings of the 24th International Conference on World Wide Web (WWW), pp. 119–120 (2015)
11. Xu, B., Lin, H., Lin, Y., Diao, Y., Yang, L., Xu, K.: Extracting emotion causes using learning to rank methods from an information retrieval perspective. *IEEE Access* **7**, 15573–15583 (2019)
12. Gui, L., Hu, J., He, Y., Xu, R., Lu, Q., Du, J.: A question answering approach for emotion cause extraction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing EMNLP 2017, Copenhagen, Denmark, 9–11 September 2017, pp. 1593–1602 (2017)
13. Li, X., Song, K., Feng, S., Wang, D., Zhang, Y.: A co-attention neural network model for emotion cause analysis with emotional context awareness. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31–October–4 November 2018, pp. 4752–4757 (2018)
14. Chen, Y., Hou, W., Cheng, X., Li, S.: Joint learning for emotion classification and emotion cause detection. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018, pp. 646–651 (2018)
15. Yu, X., Rong, W., Zhang, Z., Ouyang, Y., Xiong, Z.: Multiple level hierarchical network-based clause selection for emotion cause extraction. *IEEE Access* **7**, 9071–9079 (2019)
16. Li, X., Feng, S., Wang, D., Zhang, Y.: Context-aware emotion cause analysis with multi-attention-based neural network. *Knowl. Based Syst.* **174**, 205–218 (2019)
17. Fan, C., et al.: A knowledge regularized hierarchical approach for emotion cause analysis. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 5618–5628 (2019)
18. Gui, L., Yuan, L., Xu, R., Liu, B., Lu, Q., Zhou, Yu.: Emotion cause detection with linguistic construction in chinese weibo text. In: Zong, C., Nie, J.-Y., Zhao, D., Feng, Y. (eds.) NLPCC 2014. CCIS, vol. 496, pp. 457–464. Springer, Heidelberg (2014). [https://doi.org/10.1007/978-3-662-45924-9\\_42](https://doi.org/10.1007/978-3-662-45924-9_42)
19. Gui, L., Xu, R., Lu, Q., Wu, D., Zhou, Yu.: Emotion cause extraction, a challenging task with corpus construction. In: Li, Y., Xiang, G., Lin, H., Wang, M. (eds.) SMP 2016. CCIS, vol. 669, pp. 98–109. Springer, Singapore (2016). [https://doi.org/10.1007/978-981-10-2993-6\\_8](https://doi.org/10.1007/978-981-10-2993-6_8)

20. Cheng, X., Chen, Y., Cheng, B., Li, S., Zhou, G.: An emotion cause corpus for chinese microblogs with multiple-user structures. *ACM Trans. Asian Low-Resource Lang. Inf. Process.* **17**(1), 6 (2017)
21. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 3111–3119 (2013)
22. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: *Conference Track Proceedings 3rd International Conference on Learning Representations ICLR 2015, San Diego, CA, USA, 7–9 May 2015*