Emotion-Cause Pair Extraction as Sequence Labeling Based on A Novel Tagging Scheme

Chaofa Yuan[1]*, **Chuang Fan**[1]*, **Jianzhu Bao**[1], **Ruifeng Xu**[1,2,3]†

[1]Harbin Institute of Technology (Shenzhen), China   [2]Peng Cheng Laboratory

[3]Joint Lab of Harbin Institute of Technology and RICOH

{bruceyuan123, fanchuanghit, jianzhubao}@gmail.com   xuruifeng@hit.edu.cn

Abstract

The task of emotion-cause pair extraction deals with finding all emotions and the corresponding causes in unannotated emotion texts. Most recent studies are based on the likelihood of Cartesian product among all clause candidates, resulting in a high computational cost. Targeting this issue, we regard the task as a sequence labeling problem and propose a novel tagging scheme with coding the distance between linked components into the tags, so that emotions and the corresponding causes can be extracted simultaneously. Accordingly, an end-to-end model is presented to process the input texts from left to right, always with linear time complexity, leading to a speed up. Experimental results show that our proposed model achieves the best performance, outperforming the state-of-the-art method by 2.26% ($p < 0.001$) in $F1$ measure.

1 Introduction

Emotion-cause pair extraction (ECPE) aims to extract all potential pairs of emotions and the corresponding causes from unannotated emotion texts, such as (c_3, c_1) and (c_3, c_2) in:

Ex.1 *A policeman visited the old man with the lost money,* (c_1)| *and told him that the thief was caught.* (c_2)| *The old man was very happy,* (c_3)| *and deposited the money in the bank.* (c_4)

This task for pair extraction closely relates to the traditional emotion cause extraction task, which aims at identifying the causes for a given emotion expression. Many works (Gui et al., 2017; Li et al., 2018, 2019; Xu et al., 2019; Fan et al., 2019; Xia et al., 2019; Ding et al., 2019) related to emotion cause extraction have been published recently, and all of them are evaluated with the dataset released by Gui et al. (2016). However, it suffers

that emotions must be annotated before extracting the causes, which is labor intensive and limits the applications in real-world scenarios.

Towards this issue, Xia and Ding (2019) presents a new task, namely emotion-cause pair extraction, to extract emotions and the corresponding causes together. In comparison, it is a more challenging task due to the inherent ambiguity and subtlety of emotions, especially when there is no annotation information provided before extraction. Following this task setting, they propose a two-step approach to solve this task. However, limited by the inherent drawback of pipelined framework, error propagation may occur from the first procedure to the second. Recent studies (Song et al., 2020; Tang et al., 2020) have focused on solving this task using multi-task learning framework (Caruana, 1993) with well-designed attention mechanism (Bahdanau et al., 2015), but they extract emotion-cause pairs by calculating a *pair matrix*, which is based on the likelihood of Cartesian product among all clauses in texts, thus leading to the computational cost is expensive, that is, the time complexity is $O(n^2)$.

In this paper, we define the joint emotion-cause pair extraction as a sequence labeling problem (Eger et al., 2017; Zheng et al., 2017), so that the emotion-cause structure can be integrated into an unified framework, including representation learning, pair extraction, and causality reasoning. The challenge is to also include emotion causality into the tagging scheme, i.e., the traditional BIO tagging is not suitable for this task. Targeting this problem, we design a novel tagging scheme with multiple labels which contain the information of causes and the triggered emotions associated with these causes, and we realize it by coding the distance between linked components into the tags. Accordingly, an end-to-end model based on this tagging scheme is presented to process the input sequences from left to right, consequently, reducing the number of

---

* Equal Contributions.

† Corresponding author.

potential pairs needed to be parsed and leading to a speed up.

Specifically, BERT (Devlin et al., 2019) is trained with the objective of masked language modeling and next-sentence prediction task, therefore, we base our model on the BERT to generate powerful, general-purpose linguistic representation for each clause. Then LSTMs (Hochreiter and Schmidhuber, 1997) will be applied to capture long-range dependencies among different clauses.

To summarize, our contribution includes:

- We frame the ECPE task as a sequence labeling problem and propose an end-to-end model based on a novel tagging scheme with multiple labels, thereby the emotion-cause structure can be extracted simultaneously.

- The proposed model incrementally processes the input sequences from left to right, always with linear time complexity.

- Performance evaluation shows the superiority and robustness of the proposed model compared to a number of competitive baselines.

2 Methodology

2.1 A Tagging Problem

We define $X = (x_1, x_2, \ldots, x_n)$ as an ordered clause sequence for an emotion text. There are several emotions and at least one cause corresponds to these emotions. The goal is to output all potential pairs where exist emotion causality. Due to the difficulty describing the emotion/cause at the word or phrase level (Chen et al., 2010), in this paper, the "emotion" and "cause" are refer to "emotion clause" and "cause clause", respectively.

This research investigates such a problem by sequentially tagging each clause $x \in X$ with two-tuples label $y = (b, d) \in Y$, where $b \in \{C, O\}$ and $d \in \{-(n-1), \ldots, -1, 0, 1, \ldots, n-1, \perp\}$. Tag "C" represents the "Cause" tag which means the current clause is a cause clause, while tag "O" represent the "Other" tag, indicating the current clause is irrelevant to the final result. Moreover, d encodes the distance between the cause and the triggered emotion it relates to, e.g., "-1" denotes the **previous clause** is the related emotion, while "1" for the **subsequent clause**. The special symbol $\perp$ indicates when a particular slot is not filled, e.g., a non-cause clause (b=O) has no related emotion, thus it always associates with the symbol $\perp$. For example,
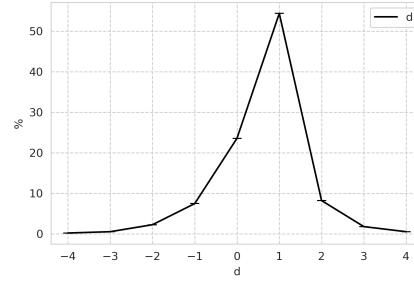

Figure 1: Distribution of distances d between emotions and causes in Gui et al. (2016) dataset.

we could incrementally label the text in **Ex.1** by sequence: $\{(C, 2), (C, 1), (O, \perp), (O, \perp)\}$.

While the total number of tags in Y is $N_t = 2 * (n - 1) + 1 + 1$, which relies on the size of text X, resulting in the inconsistency during the training stage. Empirically, for emotion events, causes generally occur on positions very close to the emotions and occur frequently. As shown in Figure 1, in the dataset released by Gui et al. (2016), there are about 55% of all emotion-cause structures have distance "1", that is, the emotions behind the causes they attach to. Overall, around 95% of all emotion-cause distances lie in {-2, -1, 0, 1, 2}. Thus, we could let hyperparameter l to denote the left and right boundary which limits the scope of emotion corresponds to the current cause, i.e., $d \in \{-l, \ldots, -1, 0, 1, \ldots, l, \perp\}$. Then, we have total $N_t = 2 * l + 1 + 1$ tags, which keep consistent in the training stage.

2.2 The End-to-End Model

In this paper, the details of our model based on the novel tagging scheme will be described.

BERT Encoder Given an emotion text $X = (x_1, x_2, \ldots, x_n)$ consisting of n clauses and each clause $x_i = (w_{i1}, w_{i2}, \ldots, w_{ik})$ contains k words. We formulate each clause as a sequence $\hat{x}_i = ([CLS], w_{i1}, \ldots, w_{ik}, [SEP])$, 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 for this task. We first obtain the hidden representation as $h_i = BERT(\hat{x}_i) \in \mathbb{R}^{d_h * |\hat{x}_i|}$ where d_h is the size of hidden dimension and $|\hat{x}_i|$ is the length of sequence $\hat{x}_i$. Then, the text X can be represented as $H_X = [h_1, h_2, \ldots, h_n]$, which will be fed into the LSTM encoder.

LSTM Encoder Based on the representation of H_X, a LSTM layer is performed to model the

context-dependent information among different clauses. To summarize the information from both directions, we use bidirectional LSTM to exploit two parallel passes, this yields:

$$[\ldots, \overrightarrow{h_i}, \ldots] = \overrightarrow{LSTM}_f([\ldots, h_i, \ldots]) \quad (1)$$

$$[\ldots, \overleftarrow{h_i}, \ldots] = \overleftarrow{LSTM}_b([\ldots, h_i, \ldots]) \quad (2)$$

where $i \in [1, n]$, both $\overrightarrow{h_i}$ and $\overleftarrow{h_i} \in \mathbb{R}^{d_r * n}$, d_r is the hidden size of LSTMs. The two directional hidden states are concatenated as the final clause representation $\hat{h}_i = [\overrightarrow{h_i}, \overleftarrow{h_i}]$.

Training The representation $\hat{h}_i$ as the final feature for tag prediction and the model is trained by minimizing the cross entropy. Specifically,

$$p_i = \text{softmax}(\text{FFN}(\hat{h}_i)) \quad (3)$$

$$\mathcal{L} = -\sum_j^{|\mathbb{D}|} \sum_i^n y_i^{(j)} \log(p_i^{(j)}) + \frac{\lambda}{2} ||\theta||^2 \quad (4)$$

where FFN is a feed-forward neural network with the parameters randomly initialized, $|\mathbb{D}|$ is the size of training set, n is the length of text X_j. $y_i^{(j)}$ is the label of clause i in text X_j and $p_i^{(j)}$ is the normalized predictive probabilities of our special tags. Besides, θ denotes all the parameters in this model and λ is the coefficient of L_2-norm regularization.

3 Experiments

3.1 Experimental Setting

Dataset The only dataset released by Gui et al. (2016) is used to evaluate our proposed model. We also pre-process the whole dataset by following Xia and Ding (2019). In detail, there are 1746 samples with one emotion-cause pair, 177 samples with two pairs, and 22 samples with more than two pairs. Besides, the quartile information about clause number of per sample is also shown in Table 1. Moreover, we stochastically divide the corpus into a training/development/test set in a ratio of 8:1:1 and evaluate our method 20 times with different data splits to obtain statistically credible results.

Evaluation We adopt standard Precision(P), Recall (R) and F-measure ($F1$) to measure the performance and report the average results over 20 runs. Note that when we extract emotion-cause pairs, we obtain the emotions and causes for each text simultaneously. Thus, we also evaluate the performance of emotion extraction and cause extraction.

# of	Samples	quartile	Clauses
All	1945	1st quartile	11
1 pair	1746	2nd quartile	14
2 pairs	177	3rd quartile	18
$\geq$ 3 pairs	22	max clauses	73

Table 1: Statistical information about the dataset.

Hyperparameters Our proposed model is trained using Adam optimizer (Kingma and Ba, 2015), and the initial learning rate is set to 3e-5. We set the batch size to 4, the coefficient of L_2 term to 1e-2, the hidden size of all LSTMs and FFN to 256. We regularize our network using dropout with rate 0.5 and adopt BERT$_{Chinese}$ as the basis in this work[1]. We perform grid search over the emotion boundary $l \in \{1,2,3,4,5,6\}$. The model is trained 10 epochs in total and the highest $F1$-measure model on the development set is used to evaluate the test set.

Baselines In this paper, we compare our model with the following methods.

- **Indep**: Emotion extraction and cause extraction are trained independently. Then pairing them and eliminating the pairs that have no emotion causality; **Inter-CE**: The predictions of cause extraction are used to improve emotion extraction; **Inter-EC**: The predictions of emotion extraction are used to improve cause extraction. All of them are pipelined framework and proposed in Xia and Ding (2019).

- **E2EECPE** (Song et al., 2020): An end-to-end multi-task learning framework for constructing *pair matrix* through biaffine attention.

- **LAE-MANN** (Tang et al., 2020): The current state-of-the-art method using a multi-level attention mechanism based on LSTM or BERT encoder, denoted as **LML** and **LMB** here.

3.2 Main analysis

The experimental results are shown in Table 2. **Indep** yields the lowest performance, because it ignores the fact that emotions and causes are usually mutually indicative. **Inter-CE** and **Inter-EC** benefit from this interaction information, thus get

---

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

Method	Emotion extraction			Cause extraction			Emotion-cause pair extraction		
	P (%)	R (%)	$F1$ (%)	P (%)	R (%)	$F1$ (%)	P (%)	R (%)	$F1$ (%)
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
LML	88.10	78.1	82.60	—	—	—	69.90	59.60	64.40
LMB	**89.90**	**80.00**	**84.70**	—	—	—	71.10	60.70	65.50
Ours†	81.96	73.29	77.39	**74.90**	**66.02**	**70.18**	**72.43**	**63.66**	**67.76**

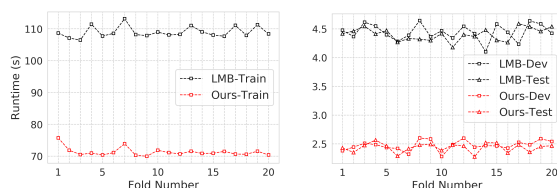Table 2: Comparison with baselines. † denotes average scores over 20 runs, and the best scores are in bold.

Scope	Extraction of. (%)		
	Emotion	Cause	Emotion-cause
1	76.62	69.52	67.29
2	**77.65**	69.87	67.38
3	77.39	**70.18**	**67.76**
4	77.14	69.68	67.28
5	76.99	69.87	67.16
6	76.11	68.95	66.42

Table 3: $F1$ scores with different emotion scope limitation over all the tasks.

(a) In training stage. (b) In inference stage.

Figure 2: Comparison of running time between LMB and our model in training/inference stage (s/epoch).

better performance. Meanwhile, the joint models (**E2EECPE**, **LML**, **LMB**) have better performance compared to the previous pipelined methods by reducing error propagation. Specifically, **LML** outperforms **E2EECPE** by capturing mutual interdependence between emotions and causes using a multi-level attention mechanism, and **LMB** further improves the performance based on BERT embeddings. Our model performs worse on emotion extraction, because it inherently learns to detect causes firstly, then identifies emotions through distance tag. Overall, our model achieves better performance compared to the previous methods, significantly improves cause extraction by 5.15% and emotion-cause pair extraction by 2.26% in $F1$-measure with $p < 0.001$. The reason may be that our model always processes the texts with linear time complexity, instead of based on Cartesian product, which the time complexity is $O(n^2)$, thereby greatly reducing the search space.

3.3 Emotion Scope Limitation Analysis

It is intuitive that the larger emotion scope is allowed, the more situations are involved. In this section, we evaluate the power of distance limitation l for this task, and set the number of l from 1 to

6. Results are shown in Table 3. With the increasing scope of emotion, the performance improves. However, when the scope of emotion is more than 3, the performance decreases, because the larger emotion scope is set, the larger search space is. As such, we choose $l = 3$ in our final model since it gives the best performance in our experiments.

3.4 Runtime Analysis

Theoretically, our proposed model always labels the input texts from left to right with linear time complexity, but all the previous end-to-end model is $O(n^2)$ time complexity. Nevertheless, we still perform a further experiment to confirm this superiority empirically. For simplicity and efficiency, we only conduct runtime analysis between **LMB** and ours, since **LMB** is also based on BERT and is the current state-of-the-art method. Figure 2 shows the running time consumed by models per epoch in different data folds. The results suggest that our model is 36% and 44% faster than **LMB** in training and inference stage respectively, indicating the efficiency of the proposed method.

4 Conclusion

In this paper, we consider the emotion-cause pair extraction as a sequence labeling problem and propose an end-to-end model based on a novel tagging scheme with multiple labels. The proposed model is capable of integrating the emotion-cause structure into a unified framework, so that emotions with the related causes can be extracted simultaneously. Moreover, the proposed model parses the input texts in order from left to right, greatly reducing the search space, leading to a speed up. Experimental results demonstrate the effectiveness and robustness of the proposed method on a benchmark dataset. In the future, we will explore the extension of this approach to achieve full coverage.

Acknowledgements

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Rich Caruana. 1993. Multitask learning: A knowledge-based source of inductive bias. In *Machine Learning, Proceedings of the Tenth International Conference, University of Massachusetts, Amherst, MA, USA, June 27-29, 1993*, pages 41–48.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *COLING 2010, 23rd International Conference on Computational Linguistics, Proceedings of the Conference, 23-27 August 2010, Beijing, China*, pages 179–187. Tsinghua University Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, 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 *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6343–6350. AAAI Press.

Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural end-to-end learning for computational argumentation mining. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 11–22.

Chuang Fan, Hongyu Yan, Jiachen Du, Lin Gui, Lidong Bing, Min Yang, Ruifeng Xu, and Ruibin Mao. 2019. A knowledge regularized hierarchical approach for emotion cause analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5613–5623. Association for Computational Linguistics.

Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. 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, September 9-11, 2017*, pages 1593–1602.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. 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, November 1-4, 2016*, pages 1639–1649.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

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

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 *Proceedings of the 2018 Conference on Empirical Methods in Natural Language*

Processing, Brussels, Belgium, October 31 - November 4, 2018, pages 4752–4757.

Haolin Song, Chen Zhang, Qiuchi Li, and Dawei Song. 2020. End-to-end emotion-cause pair extraction via learning to link. *CoRR*, abs/2002.10710.

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

Rui Xia and Zixiang Ding. 2019. Emotion-cause pair extraction: A new task to emotion analysis in texts. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 1003–1012.

Rui Xia, Mengran Zhang, and Zixiang Ding. 2019. RTHN: A rnn-transformer hierarchical network for emotion cause extraction. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5285–5291. ijcai.org.

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.

Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. 2017. Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1227–1236.