A New Component of Interactive Multi-task Network Model for Emotion-Cause Pair Extraction

Jingzhe Shan and Min Zhu

School of Data Science and Engineering, East China Normal University, Shanghai, 200062, China
51184507001@stu.ecnu.edu.cn

Abstract. The task emotional-cause pair extraction (ECPE) is proposed based on the emotion cause extraction (ECE) task. The ECPE task proposes a new two-step frame method: first step is to extract emotion and cause information separately, then the second step is to match the emotional cause pair. Compared to the ECE task, this new research no longer relies on manual annotation of documents. However, the Inter-EC model component for the ECPE task does not fully consider the relevance of sentimental clause and causal clause, and the available context is limited. In this paper, we propose a new extraction component with self-attention for the Inter-EC model, optimizing the interactive multi-task network model. Our approach achieved a good recall score on the ECPE dataset, and the F1 value is slightly better than the baseline.

1. Introduction
With the explosive growth of text data on the Internet, the field of natural language processing pays close attention to emotional texts. The research on emotions can be divided into two categories: classification and extraction. The former tries to distinguish the positive and negative polarity of the emotion, and the latter focuses on various information related to the emotion. The task of extracting emotion cause is one of them. The definition of emotion cause extraction task is based on the emotional information in the document, looking for its possible inducements [1]. The ECE task has practical significance and can be applied to public opinion mining, so this research is meaningful. As the specification of this task requires numerous emotional annotations, this increases the labor cost of this work and hinders the further practical application of ECE tasks. To overcome these shortcomings, the ECPE task proposed a new extraction method that does not require data annotation, which enhances the application value of this work. The ECPE task is divided into two steps. The first step uses a multi-task model to find the emotion clause and the cause clause in the document respectively. The second step is to get the emotion-cause pair through filtering and matching algorithms.

As shown in Fig.1, the text "... The reporter followed Kaiqing Hu to …, he is very worried. ... " is the input document, which contains 5 clauses. The ECPE task aims to Look for all the emotion-cause pairs in the document. The picture shows a case with two pairs, (he is very worried - that now his father is in the late stage of tuberculosis) and (he is very worried - and his mother is injured again). The model found two clauses that cause "worried" in the text.

However, we found two key problems in this work. One is that the model does not consider the relationship between clauses, and the other is that the cause component has limited information coding ability and fails to make full use of emotional tags. Therefore, based on the original model, we combine the transformer encoder frame to improve the effect of cause extraction component, use the
multi-head self-attention mechanism to get the hidden state of the clause, filter and transmit
information through the feed forward layer, the way of the self-attention encoding can obtain global
context information in a document, which achieve a better performance to understand the cause
clauses, and ultimately affects the accuracy of the results of this task.

Fig.1 An example of the ECPE task

2. Related Work
Emotion cause extraction (ECE) is one of the concerned issues in the field of natural language
processing, with the development of Internet technology, more and more researchers have tried to use
different methods to solve it. In these studies, how to construct a suitable model to accurately extract
the cause information of emotion from the document is a key to solving the problem of ECE task.

The original emotion cause extraction task mainly used a rule-based method. Researchers usually
label a small-scale corpus for analysis and statistics by themselves, extract suitable rules for this task,
and then use the rules they found to extract causes from unlabeled text. The advantage is that the rules
are uncomplicated and easy to add, understand and accept. However, this method also has the
disadvantages such as numerous rules and low accuracy, and with the update of the Internet
buzzwords, the rules also need to be updated continuously, so this kind of method has certain
limitations. The ECE task was first proposed and defined by Lee et al. [1]. The author summarized the
relationship between emotional words and reason clauses, constructed a small emotional reason data
set, and proposed corresponding evaluation specifications. Based on the corpus designed by Lee et al.
[1], Chen designed a multi-category model that can detect texts with multiple reasons [2]. In addition,
Li and Xu et al. [3], Gao et al. [4] and others extended the rule-based method to informal Weibo texts.

With the rise of neural network methods, ECE task was regarded as a classification or sequence
labeling task. At this stage, many neural network-based methods have been proposed. Sequence
labeling problems often consider more part-of-speech features and the relationship between words,
which are very effective for optimizing global label tasks. The classification problem is to classify all
sentences and find out the cause sentences. In 2014, Gui et al. manually extracted 25 rules suitable for
the ECE task, and then used SVM-based classification algorithms and CRF-based sequence annotation
algorithms to learn emotional text and achieved good results [5]. Similar studies include the ECOCC
model [6], Ghazi's CRF-based model [7] and so on. However, the research work at this stage requires
that the cause and emotion information of the text are in a same sentence, which limits the extraction
of information from longer distance.

In order to solve the above problems, in 2016, Gui et al. released a new Chinese corpus, which
regarded the ECE task as a classification problem, and used the SVM algorithm based on the
convolution kernel to extract the causes [8]. This corpus becomes the benchmark dataset for this task.
With the development of deep learning and attention mechanisms, methods for extracting emotional
reasons based on deep network models have been widely proposed. This sort of research includes the
new deep network model proposed by regarded ECE task as a QA task [9], a CANN model based on
attention mechanism and recurrent neural network [10], and a hierarchical network model based on CNN and attention mechanism [11], etc.

In 2019, Xia proposed a new task: extracting emotion-cause pairs. The goal of this new task is to match related emotion and cause from given documents. However, the traditional emotion cause extraction task relies on sentiment-annotated text, which separates the extraction task. Xia solved these shortcomings and reconstructed a dataset suitable for the ECPE tasks [12], which is also an important benchmark dataset in this new field. The contribution of this new task is to make research in this field no longer rely on artificial tags, but to automatically extract emotion and cause labels through a multi-task model, and then pair and filter the obtained labels to get possible emotions and related reason information. However, we found that the Inter-EC model proposed in the research above only relies on emotional labels for the extraction of cause clause, and fails to fully consider the relevance between clauses. There is still room for optimization in the calculation of sentence weight. Inspired by models in the field of machine translation [13], we designed a new extraction component, which uses the self-attention mechanism in Transformer to encode the clause, and the more distant context information is integrated into the sentence representation, this reasonable weight calculation method improves the extraction result.

3. Model

3.1. Overall structure

This paper designs a new cause extraction component, and optimizes the interactive multi-task information extraction model. The model is shown in Fig.2.
The input of the model is a document containing multiple clauses \( c_1, c_2, \ldots, c_n \), and each document contains at least one pair of emotion and related cause clauses \((c^e, c^c)\). In order to comprehend clauses, the lower layer model is multiple Bi-LSTMs with attention. The upper layer model consists of two components, one component uses the bi-directional LSTM network to gain emotion labels, and the other is improved based on the transformer mechanism, its input is combined with the output result of first component, and the sentence is encoded by the multi-layer transformer encoder, which can receive long distance context, improve the recall rate of the cause clause extraction by about 2%, and make the final effect of ECPE task improved.

3.2. Lower layer model

This part consists of a group of Bi-LSTMs and an attention layer. The input is word vectors dealt with segmentation and pre-processing. Each Bi-LSTM module corresponds to a clause. The word vector \( w_i \) is calculated by RNN model to obtain the hidden state \( h_i \), which collect the context information. Then the \( h_i \) is sent to the attention layer to obtain the sentence representation \( s_i \) that contains context content for subsequent calculations. The structure of this section refers to paper [12], which will not be described in detail here.

3.3. Upper layer structure

The first component of the upper layer in this paper is the emotion extraction component, which is used to obtain the emotion label \( y^e_i \) in the document. Its input is the sentence representation \( s_i \) calculated by the lower layer network, and the hidden state \( r^e_i \) of the clause is obtained through multiple clause-level Bi-LSTM networks, those vectors is regarded as the context information of the current clause. Then the network transfer \( r^e_i \) to the softmax layer, through which the sentiment labels of sentences are received:

\[
\hat{y}^e_i = \text{softmax}(W^e r^e_i + b^e)
\]

The second part is the cause extraction component. Inspired by the transformer encoder model [13], we redesign this component to gain more exhaustive context content in consideration of the correlation between emotion and reason clauses. Each encoder module based on Transformer consists of two parts: The first part is the multi-head self-attention mechanism. The clause context information calculated by this mechanism is based on the entire document, which means that it can be more distantly related. Compared with the original Bi-LSTM component, it can get more complete global information. The second part is the fully connected feed forward networks. This part is used to transmit sentence information and alleviate the problem of gradient disappearance. The input of this component is \( x^c_i \):

\[
x^c_i = \hat{y}^e_i + s_i + p_i
\]

Where \( p_i \) is the positional embedding which presents the absolute position of words in a clause, we calculate the query vector \( q^c_i \), the key vector \( k^c_i \) and the value vector \( v^c_i \) of each clause \( s_i \) in the document as follows:

\[
q^c_i = \text{ReLU}(x^c_i w_Q)
\]
\[
k^c_i = \text{ReLU}(x^c_i w_K)
\]
\[
v^c_i = \text{ReLU}(s_i w_V)
\]

Then a group of weight \( \beta_i = \{\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,n}\}\) of clause \( s_i \) is learned by self-attention with query and key vector, \( n \) is the number of clauses in a document and the final output of this layer is weighted sum:

\[
\beta^c_{i,j} = \frac{\exp(q^c_i \cdot k^c_j)}{\sum_{j} \exp(q^c_i \cdot k^c_j)}
\]
\[
x^c_i = \sum_{j} \beta^c_{i,j} v^c_j
\]

The output from the self-attention layer is sent to the next feed forward layer, which is a fully connected layer, ReLU is used as the activation function, and then normalize the result:
The normalization layer is also used after getting $z_{ij}$. Finally, repeat the above calculation steps by a couple of transformer encoder layers, the softmax function is used to predict the cause label:

$$
\hat{y}_i^c = \text{softmax}(W_i^c o_i^c + b_i^c)
$$

Meanwhile, we choose the cross-entropy loss function as the loss function of this model, the total loss is calculated by the sum of two components:

$$
L = \lambda L_e + (1 - \lambda) L_c
$$

3.4. Pairing step
Up to this step, we have extracted all the possible emotion and reason clauses from document, and then the emotion-cause pairs can be matched by a series of functions. Please refer to ECPE task [12] for the specific algorithm.

4. Experiment and analysis

4.1. Dataset and verification
This paper uses the dataset proposed by Xia et al. [12]. The data is Chinese news text. Each document is required to contain at least a pair of sentiment and cause clause. The dataset is divided into 1945 documents, and the proportions of documents containing different emotion-cause pairs in the dataset are shown in Table 1.

| Emotion-cause Pairs               | Number | Percentage |
|-----------------------------------|--------|------------|
| With a pair of emotion-cause      | 1746   | 89.77%     |
| With two pairs of emotion-cause   | 177    | 9.10%      |
| More than two pairs in a document | 22     | 1.13%      |
| Counts                            | 1945   | 100%       |

We randomly divide the dataset, select 10% of it as the test set, and use the rest data as the training set. We will repeat the experiment 5 times and take the average of all results to ensure the actual effect of the model. The experiment uses the value of precision, recall and F1 to evaluate the effect of the model, you can refer to the work of Gui et al. [8] and Xia et al. [12] for specific calculation methods. Meanwhile, the training results of our optimized model will be compared with the three baseline models mentioned in the ECPE task:

Indep: This is an independent multi-task network model, which is a two-layer network. The lower layer is the bi-directional LSTM, which is used to extract word-level features of clauses. The upper layer contains two sub-tasks. One task is emotion extraction, and the other is cause extraction. Each task uses the RNN model to receive the clause vectors obtained from the lower layer to classify and distinguish the emotion and cause tags independently.

Inter-EC: This is an enhanced network. Based on the Indep model, the emotion label obtained in the first sub-task is integrated into the cause extraction task, and used as the input of the second task to predict the cause label. This is how this model implement interaction.

Inter-CE: The model structure is similar to Inter-EC, but the execution order is reversed. First, the cause label is extracted, then the cause label is integrated into the emotion extraction sub-task, as interactive parameters to find emotion clauses.

4.2. Experiment setting
The dataset was pre-trained by word2vec toolkit [14], and the dimension of word embedding is set to 200. The hidden unit of the lower layer network is 200. The number of multi-head is 5 and the dimension of the hidden unit of transformer is 200. We set the batch size to 32, the learning rate was
finally fixed to be 0.005 and Adam is used. Reasonable parameter selection can achieve better calculation results and ensure that the model is easier to converge.

### 4.3. Results analysis

The comparison of experimental results is shown in table 2. We use Inter-ECNC presents the Inter-EC model with the new component in the table 2.

|                | Cause Extraction | Emotion-Cause pair Extraction |
|----------------|------------------|-------------------------------|
|                | P    | R    | F1   | P    | R    | F1   |
| Indep          | 0.6902| 0.5673| 0.6205| 0.6832| 0.5082| 0.5818|
| Inter-CE       | 0.6809| 0.5634| 0.6151| 0.6902| 0.5135| 0.5901|
| Inter-EC       | **0.7041**| 0.6083| 0.6507| 0.6721| 0.5705| 0.6128|
| Inter-ECNC     | 0.6863| **0.6254**| **0.6544**| 0.6601| **0.5734**| **0.6138**|

From table 2, we can notice that compared with those baseline models in cause extraction sub-task, the new component we designed for Inter-EC model has achieved the best recall value, and the F1 score is slightly higher than the baseline. Although the Inter-EC model build up based on the Indep model and takes into account the interaction of emotion and cause clauses, it failed to make good use of the correlation between those clauses to predict the cause label. Therefore, many correct cause sentences were not successfully labeled. The components we designed fully compensated for this defect and achieved a better recall value. And the slightly higher F1 score indicates that our method is effective. Meanwhile, the improvement of the training effect brought by our components finally makes the ECPE task perform better.

### 5. Conclusion

The proposal of ECPE task is an improvement in the direction of ECE research. It is an ingenious idea with simple method. This paper is inspired by the transformer encoder framework that has achieved good results in machine translation task, so we apply it to the clause encoding of Inter-EC model, redesigns and optimizes the cause extraction component. The model has achieved good results on related dataset after reasonable experiments, which verify that this method is effective. However, we also notice some shortcomings: The algorithm of the pairing step fails to consider the interactivity of clauses. Future research can be tried in these aspects.

### Acknowledgments

Thanks for the discussion and support from DMN team and Miss Gao.

### References

[1] Lee, Sophia Yat Mei, Ying Chen and Chu-Ren Huang. “A Text-driven Rule-based System for Emotion Cause Detection.” HLT-NAACL 2010 (2010).
[2] Chen, Ying, Sophia Yat Mei Lee, Shoushan Li and Chu-Ren Huang. “Emotion Cause Detection with Linguistic Constructions.” COLING (2010).
[3] LiWeiyuan and Xu-hua. “Text-based emotion classification using emotion cause extraction.” Expert Systems With Applications (2014): n. pag.
[4] GaoKai, Xu-hua and WangJiushuo. “A rule-based approach to emotion cause detection for Chinese micro-blogs.” Expert Systems With Applications (2015): n. pag.
[5] Gui, Lin, Li Yuan, Ruifeng Xu, Bin Liu, Qin Lu and Yu Zhou. “Emotion Cause Detection with Linguistic Construction in Chinese Weibo Text.” NLPCC (2014).
[6] Gao, Kai, Huasong Xu and Jiushuo Wang. “Emotion Cause Detection for Chinese Micro-Blogs Based on ECOCC Model.” PAKDD (2015).
[7] Ghazi, Diman, Diana Inkpen and Stan Szpakowicz. “Detecting Emotion Stimuli in Emotion-Bearing Sentences.” CICLLing (2015).
[8] Gui, Lin, Dongyin Wu, Ruifeng Xu, Qin Lu and Yu Zhou. “Event-Driven Emotion Cause Extraction with Corpus Construction.” EMNLP (2016).

[9] Gui, Lin, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu and Jiachen Du. “A Question Answering Approach for Emotion Cause Extraction.” EMNLP (2017).

[10] Li, Xiangju, Kaisong Song, Shi Feng, Daling Wang and Yifei Zhang. “A Co-attention Neural Network Model for Emotion Cause Analysis with Emotional Context Awareness.” EMNLP (2018).

[11] Yu, Xinyi, Wenge Rong, Zhuo Zhang, Yuanxin Ouyang and Zhang Xiong. “Multiple Level Hierarchical Network-Based Clause Selection for Emotion Cause Extraction.” IEEE Access 7 (2019): 9071-9079.

[12] Xia, Rui and Zixiang Ding. “Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts.” ACL (2019).

[13] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, L. Kaiser and Illia Polosukhin. “Attention is All you Need.” ArXiv abs/1706.03762 (2017): n. pag.

[14] Mikolov, Tomas, Ilya Sutskever, Kai Chen, G. S. Corrado and J. Dean. “Distributed Representations of Words and Phrases and their Compositionality.” ArXiv abs/1310.4546 (2013): n. pag.