Learning Emotion-Aware Contextual Representations for Emotion-Cause Pair Extraction

Anonymous ACL submission

Abstract

Emotion Cause Pair Extraction (ECPE), the task expanded from the previous emotion cause extraction (ECE) task, focuses on extracting emotion-cause pairs in text. Two reasons have made ECPE a more challenging, but more applicable task in real world scenarios: 1) an ECPE model needs to identify both emotions and their corresponding causes without the annotation of emotions. 2) the ECPE task involves finding causes for multiple emotions in the document context, while ECE is for one emotion. However, most existing methods for ECPE adopt an unified approach that models emotion extraction and cause extraction jointly through shared contextual representations, which is suboptimal in extracting multiple emotion-cause pairs. In addition, previous ECPE works are evaluated on one ECE dataset, which exhibits a bias that majority of documents have only one emotion-cause pair. In this work, we propose a simple pipelined approach that builds on two independent encoders, in which the emotion extraction model only provide input features for the cause extraction model. We reconstruct the benchmark dataset to better meet ECPE settings. Based on experiments conducted on the original and reconstructed dataset, we validate that our model can learn distinct contextual representations specific to each emotion, and thus achieves state-of-the-art performance on both datasets, while showing robustness in analyzing more complex document context.

1 Introduction

In recent years, the task about detecting the stimuli of emotions expressed in text, has emerged in the area of text emotion analysis (Russo et al., 2011; Mohammad et al., 2014; Ghazi et al., 2015). Previous works focus on Emotion Cause Extraction (ECE), which has been proposed by (Lee et al., 2010) as a word-level sequence labeling problem. Gui et al. (2016) re-formalized ECE as a clause-level classification problem of finding cause clauses for the given emotion. They released a Chinese dataset using SINA city news which has become the benchmark dataset for the ECE task followed by many works (Gui et al., 2017; Li et al., 2018; Xia et al., 2019; Fan et al., 2019; Yan et al., 2021).

Xia and Ding (2019) pointed out that ECE task suffers from two defects: 1) The emotion must be annotated in advance. 2) The goal of ECE neglects the fact that emotions and causes are mutually indicative. They developed the task to emotion-cause pair extraction (ECPE), in which emotion clauses and their corresponding cause clauses are extracted as pairs. To solve the problem, they proposed a two-step pipelined approach that extracts all emotion clauses and cause clauses at first and then uses a filter to select emotion-cause pairs. More recently, however, the ECPE task has been dominated by end-to-end systems that model emotion-cause extraction matching jointly (Ding et al., 2020a; Cheng et al., 2020; Fan et al., 2020), with the belief that joint models capture interactions between subtasks and mitigate error propagation.

We re-investigate ECPE's motivation and observe another significant merit of ECPE over ECE, that is, the ECPE task involves the analysis of longer and more complex document context that contains multiple emotions, multiple causes and multiple semantic roles. As shown in Figure 1, the example document is divided into *two* different samples in the ECE task since two emotion clauses are annotated. An ECE model takes one annotated emotion clause as input and find corresponding causes in the context, while an ECPE model takes the whole document as input and extract all the emotion-cause pairs and therefore, ECPE model needs to tackle with richer, but more complex contextual information.

In order to utilize document context information, existing works for ECPE employ various approaches(Chen et al., 2020b; Wei et al., 2020). However, most of them apply shared context en-

coders to perform emotion-cause extraction and matching jointly. We argue that shared contextual representations lead to suboptimal results in finding causes for multiple emotions in one document since shared encoders fail to learn contextual representations specific to each emotion. For instance, the clauses `c12` and `c13` in Figure 1, "failing to cure the disease while using that much money" is crucial in detecting the causal relationship between `c14` and `c15` but not for `c18` and `c17`.

To address this problem, we propose a simple pipelined approach that builds on two independent pre-trained encoders trained separately, one for an emotion extraction model, and another for an emotion-oriented cause extraction model, with the fusion of emotion information in its input layer. Based on series of experiments, we validate that the pre-trained encoder of the cause extraction model can learn emotion-aware contextual representations and our approach outperforms all previous joint approaches.

Moreover, all previous works of ECPE are evaluated on one emotion cause corpus, which has been reconstructed by Xia and Ding (2019) based on Gui et al. (2016)'s benchmark dataset for ECE. However, we observe a bias of this dataset that only 10.23% of documents have more than one emotion-cause pair, and only 6.63% of documents have more than one emotion clause. We also find that many different documents in the dataset are actually excerpts of the same news report. Therefore, we rebuild the benchmark dataset by checking all documents manually and merging documents that are from the same original news article. Experiment conducted on the reconstructed dataset show that our approach is significantly more effective in finding multiple causes for multiple emotions than previous state-of-the-art joint approaches, meeting ECPE's motivation to analyze emotion causes in longer and more complex document context.

Our contributions can be summarized as follows:

- We realize another merit of the ECPE task over previous task, which is analyzing emotion causes in more complex document context. We argue that current joint approaches using shared contextual representations leads to suboptimal results, and propose a simple and effective pipelined approach to address the problem.

- Our approach learns two independent encoders for emotion extraction and cause information. With fusion of emotion information only at input layer of the cause extraction model, it is more effective in learning contextual representations specific to each emotion in the document.

- We also observe a bias in ECPE's benchmark dataset and reconstruct the dataset manually to better meet ECPE settings. Experimental results on both datasets show that our approach achieves state-of-the-art performance, and is more robust than previous joint approaches in extracting multiple emotion-cause pairs in more sophisticated document context.

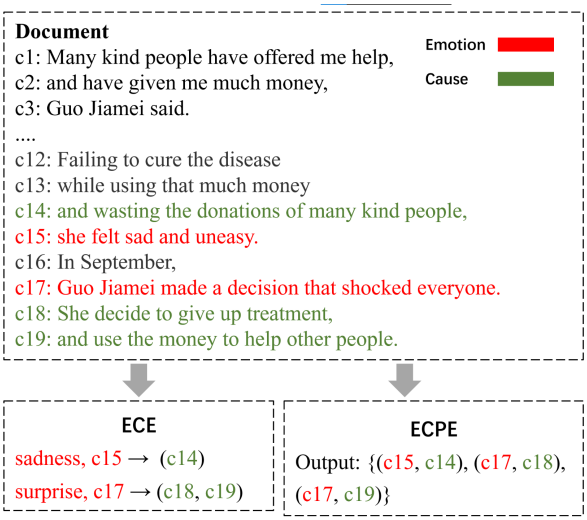


Figure 1: An Example document from Gui et al. (2016)'s dataset translated from Chinese. Texts in red and green denote the emotion clauses and cause clauses respectively.

2 Task Definition

Emotion Cause Extraction Emotion Cause Extraction (ECE) has been defined as a clause-level classification task (Gui et al., 2016) to extract the corresponding stimuli of certain given emotion in the context. Given a document $d = [c_1, ..., c_i, ..., c_{|d|}]$, where c_i is the ith clause in d, and an annotated emotion clause c^e, where $e \in E$,

$$E = \{happiness, sadness, disgust, fear, anger, surprise\} \quad (1)$$

The goal of ECE is to find all the cause clauses of the given emotion clause as $\{c^{c1}, c^{c2}, ...\}$. Note that only one emotion occur in one sample, while there may be multiple causes corresponding to it.

Emotio-Cause Pair Extraction Xia and Ding (2019) developed the ECE task to Emotion-Cause

Pair Extraction (ECPE). Given a document $d = [c_1, ..., c_i, ..., c_{|d|}]$, the goal of ECPE is to extract a set of emotion-cause pairs

$$P = \{..., (c^{emo}, c^{cau}), ...\}$$

where c^{emo} is the emotion clause and c^{cau} is its corresponding cause clause. The ECPE task deals with finding multiple causes for multiple emotions in one document.

3 Methodology

Our approach consists of two independent models, an emotion extraction model and a cause extraction model. We build both of our models on BERT (Devlin et al., 2018) as pre-trained encoders, with an multi-label learning scheme at the output layer. The emotion extraction model first takes the whole document as input and extract all possible emotion clauses. Then all the extracted emotion clauses will be used for fusing emotion information at the input layer of the cause extraction model, which we refer to as an emotion-oriented cause extraction model. We will explain the details of both models below and clarify the usage of emotion information as well as document context information in our approach.

3.1 Emotion Extraction Model

Figure 2 shows the architecture of our emotion extraction model. Given a document $d = [c_1, ..., c_i, ..., c_{|d|}]$, the model takes it as the input of a pre-trained encoder to obtain a sequence of hidden states denoted by

$$\mathbf{H_D} = (h_{[CLS]}, \mathbf{x_{c_1}}, ..., \mathbf{x_{c_i}}, ..., \mathbf{x_{c_{|d|}}}, h_{[SEP]}) \quad (2)$$

where $\mathbf{x_{c_i}} = (h_{i1}, ..., h_{ij}, ...h_{i|c_i|})$, h_{ij} is the output hidden state of j^{th} token in i^{th} clause and $|c_i|$ denotes the number of tokens in i^{th} clause. Then we apply mean pooling to build the representation of each clause, which is defined as:

$$\mathbf{h_{c_i}} = \frac{1}{|c_i|} \sum_{j=1}^{|c_i|} h_{ij} \quad (3)$$

Finally we concatenate the clause representation $\mathbf{h_{c_i}}$ with $[CLS]$ token's output hidden state, $h_{[CLS]}$, as the input of an output layer to predict the probability of the clause being an emotion clause

$$\mathbf{h^*_{c_i}} = [\mathbf{h_{c_i}}, h_{[CLS]}] \quad (4)$$

$$\hat{y}_i^{emo} = \sigma(w_{emo}^T \mathbf{h^*_{c_i}} + b_{emo}) where \quad (5)$$

where w_{emo}^T and b_{emo} are parameters of the output layer with sigmoid function $\sigma(\cdot)$.

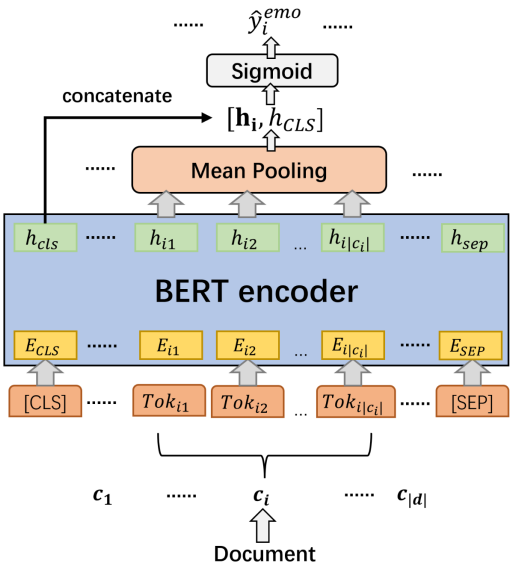


Figure 2: Emotion Extraction Model

Context Information In our emotion extraction model, we utilize the document context by concatenating each clause representation $\mathbf{h_{c_i}}$ with the output hidden state of $[CLS]$ token. In order to examine the impact of context information, we also implement a standard BERT-based sentence classification model in which each single clause is taken as the input without leveraging the context. The details are explained in section 4.3.3.

3.2 Emotion-oriented Cause Extraction Model

In the two-step model proposed by (Xia and Ding, 2019), there is an cause extraction component that extracts potential cause clauses in a document at first. We found the performance of this component unsatisfying since it ignores the fact that the identification of certain cause clause depends on its corresponding emotion clause. In our approach, we do *not* implement a model that perform cause extraction *solely*, and instead conduct emotion-oriented cause extraction.

Figure 3 displays an overview of the emotion-oriented cause extraction model. As is shown, the architecture is very similar to our emotion extraction model with BERT as the pre-trained context encoder and a multi-label learning scheme at the output layer. The difference lies in the input: we

fuse emotion information into the input sequence through various strategies.

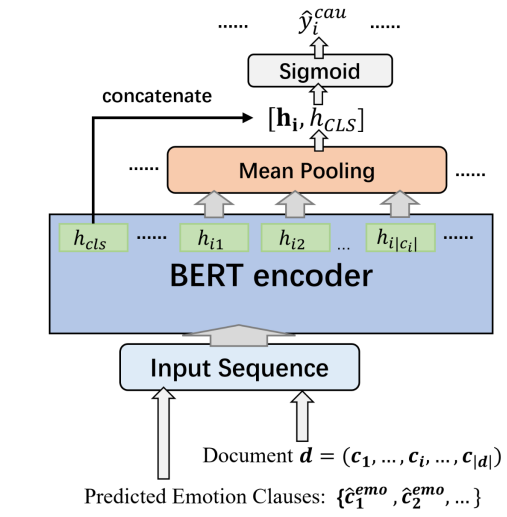
Fusing Emotion Information Previous works have attached importance to the use of emotion information in cause extraction(Tang et al., 2020; Ding et al., 2020b; Wei et al., 2020). However, all of these models use a shared LSTM layer or pre-trained encoder for contextual representations in emotion extraction and cause extraction.

We argue that shared context encoders fail to capture proper contextual information for a specific emotion clause, leading to suboptimal results in extracting multiple emotion-cause pairs in one document. Therefore, we propose three different strategies to fuse emotion information at the input layer of caue extraction model. The most direct method is to concatenate each predicted emotion clause and its document context as the input, it is denoted by EmotionalText and corresponds to the sequence at the top in figure 3b.

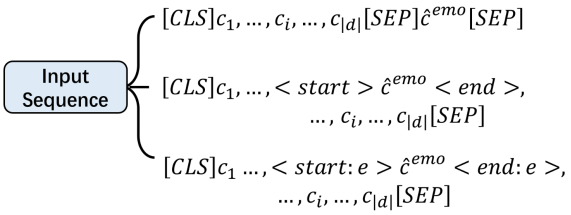
We also attempt to integrate emotion information through extra marker tokens. As shown in figure 3, we add extra text markers at the start and end position of the predicted emotion clause. In TypedMarker which corresponds to the sequence at the bottom in figure 3, we consider the emotion type of each emotion, denoted by $e \in E$ and E is defined in equation 1. However, since we cannot obtain satisfying results for emotion types, we only compare the upper bound of its effectiveness with other emotion-fusing strategies using the ground truth emotion label. For the UntypedMarker strategy denoted by the sequence in the middle, we do not consider the emotion type. The details of comparative results are elaborated in section 4.3.4.

Extracting Multiple Emotion-cause Pairs

The ECPE task involves the analysis of emotion causes in longer and more complex document context. With the fusion of emotion information, the encoder of our cause extraction model learns contextual representations specific to each emotion, leading to substantial effectiveness in tackling with document that contains multiple semantic roles and multiple emotions, while the multi-label learning scheme at the output layer enables our model to find multiple causes for one emotion. We validate that our approach is more robust than previous models that employ a shared context encoder based on evaluation results conducted on the reconstructed dataset which is closer to real-world scenarios.



(a) Emotion-oriented Cause Extraction Model



(b) Different ways of fusing emotion information. From top to bottom, they are: EmotionalText, UntypedMarker, TypedMarker. e is the emotion type defined in equation 1.

Figure 3: Emotion-oriented cause extraction model, as well as three emotion-fusing strategies.

3.3 Training and Inference

For both the emotion extraction model and cause extraction model, we fine-tune the pre-trained encoder using task-specific training objectives. Given a document d, we compute the loss for both models by:

$$L = -\frac{1}{|d|}\sum_{i=1}^{|d|} H(\hat{y}_i, y) \quad (6)$$

where $|d|$ is the number of clauses in the document, $H(\cdot)$ is the binary cross-entropy loss function. Since both models apply the same multi-label learning scheme at their output layers, $\hat{y}_i$ is $\hat{y}_i^{emo}$ defined in equation 5 in the emotion extraction model and $\hat{y}_i^{cau}$ in the cause extraction model, while y is the ground truth label of the clause.

During training, the two models are trained separately, and we use ground truth emotion labels to fuse emotion information in the cause extraction model. During inference, we first use the emo-

Item	Number
Doc. total number	1945
Doc. with one emotion cause pair	1746
Doc. with two emotion cause pairs	177
Doc. with more than two pairs	22
Doc. with more than one emotion	129

Table 1: Statistics of the original ECPE corpus.

Item	Number
Doc. total number	1679
Doc. with one emotion cause pair	1348
Doc. with two emotion cause pairs	246
Doc. with more than two pairs	85
Doc. with more than one emotion	295

Table 2: Statistics of the reconstructed ECPE corpus

tion model to extract emotion clauses in each document and fed each predicted emotion clause into the cause extraction model that learns emotion-aware contextual representations.

4 Experiments

4.1 Datasets and Evaluation Metrics

Bias in the ECPE benchmark dataset Gui et al. (2016) released an Chinese emotion cause corpus using SINA city news. This dataset has become the benchmark dataset for ECE research, which involves extraction of causes for only one annotated emotion. Thus, a large proportion of documents in this dataset have only emotion clause, and documents with multiple emotions are split to different samples.

Xia and Ding (2019) constructed an benchmark ECPE dataset based on this ECE dataset. To meet ECPE settings, they merge samples with same text content into one document. Table 1 show the statistics of the dataset, we can observe that the dataset still exhibits a bias that only 10.23% of documents contain multiple emotion-cause pairs, while only 6.63% of documents have more than one emotion.

Dataset Reconstruction Strategy By checking all the documents in the dataset, we discover that many different documents are actually excerpts from the same original news report. We manually find all such documents and merge them into one document to rebuild the ECPE benchmark dataset. As Table 1 shows, 17.57% of documents in the reconstructed dataset have multiple emotions while 19.71% of documents have multiple emotion-cause pairs. Figure 4 displays the comparision of the number of clauses in documents between the original and reconstructed dataset. As is shown, the documents in the reconstructed dataset have more clauses and thus more complex document structure. In fact, 37.42% of emotion-cause pairs are located in documents with multiple pairs, indicating that documents in the reconstructed dataset are closer to real-world scenarios.

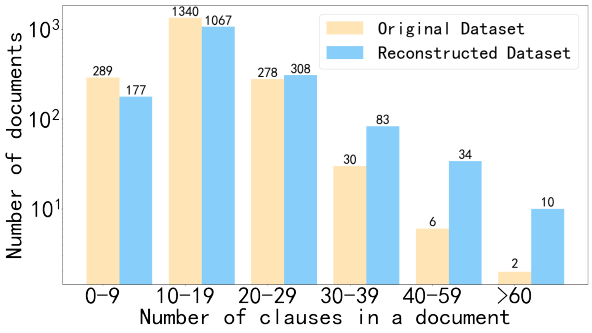


Figure 4: Statitistics on document clause number

Evaluation Metrics For evaluation metrics, precison, recall and F1 defined in (Xia and Ding, 2019) are used. Most ECPE approaches also evaluate their models on two subtasks: emotion extraction and cause extraction, yet we do this only for emotion extraction since our approach do not perform cause extraction solely.

4.2 Experimental Settings

We implement our approach based on `Pytorch` and `Transformers` and use `bert-base-chinese` as the base encoders. For both of out models, we set the random seed to 42 and use Adam optimizer for training. The learning rate is 2e-5, and the warmup ratio is 0.1. The threshold of the multi-label output layer is set to 0.5 by default.

In our experiments, we follow previous works (Xia and Ding, 2019; Wei et al., 2020; Chen et al., 2020a,b; Yuan et al., 2020; Tang et al., 2020; Cheng et al., 2020; Ding et al., 2020a) to perform 10-fold cross validation and use the same data split of the original dataset for fair comparison. In experiments conducted on the reconstructed dataset, we also follow the 10-fold cross validation setting. We have noticed there are other works (Fan et al., 2020) that randomly sampling train/validation/test sets with 8:1:1 proportion 20 times, we also evaluate our approach under this data split setting and report the comparative results in appendix A.

Model	Original Dataset			Reconstructed Dataset			Multiple Pairs		
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
Indep†(Xia and Ding, 2019)	68.32	50.82	58.18	-	-	-	-	-	-
Inter-CE†	69.02	51.35	59.01	-	-	-	-	-	-
Inter-EC†	67.21	57.05	61.28	59.02	46.48	52.00	57.37	38.35	45.97
(Chen et al., 2020a)†	71.49	62.79	66.86	-	-	-	-	-	-
(Cheng et al., 2020)†	68.36	62.91	65.45	-	-	-	-	-	-
(Tang et al., 2020)	71.10	60.70	65.50	-	-	-	-	-	-
(Yuan et al., 2020)	72.43	63.66	67.76	-	-	-	-	-	-
(Ding et al., 2020a)	72.92	65.44	68.89	-	-	-	-	-	-
(Wei et al., 2020)	71.19	76.30	73.60	77.89	49.90	60.69	76.88	53.96	63.15
(Chen et al., 2020b)	76.92	67.91	72.02	-	-	-	-	-	-
(Ding et al., 2020b)	77.00	72.35	74.52	68.46	67.06	67.65	68.84	55.90	61.55
Ours-EmotionalText	**77.83**	**76.01**	**76.81**	71.05	**74.83**	72.85	69.62	61.85	65.19
Ours-UntypedMarker	76.27	75.83	75.96	**73.78**	73.30	**73.50**	**70.94**	**62.80**	**66.50**

Table 3: Comparative results of our approach and existing ECPE models. For fair comparison, if a model has an implementation based on BERT, we report the BERT-based results, and use †to mark the models that do not have a BERT-based implementation.

4.3 Results and Analysis

4.3.1 Main Results

Table 3 displays the comparative results of our approach with the best previous works of the ECPE task on both the original dataset and reconstructed dataset. We also report the evaluation results in extracting multiple pairs in the reconstructed dataset. As is shown, by learning emotion-aware contextual representations, our approach outperforms all models using shared context encoders and achieves state-of-the-art performance in both datasets, and shows robustness in analyzing multiple emotion-cause pairs in more complex document context.

Comparative Approaches On the original dataset, we compare our approach with all existing works of ECPE, most of them are joint models using shared context encoders, except from **Indep**, **Inter-CE** and **Inter-EC**, the three variants of the two-step pipelined models proposed by (Xia and Ding, 2019) that serve as the baseline. **Rank-CP**(Wei et al., 2020) and **ECPE-MLL** (Ding et al., 2020b) are the two previous state-of-the-art approaches in the ECPE task and thus we evaluate and compare the performance of these two approaches with ours on the reconstructed dataset as well as on multiple emotion-cause pairs extraction. It should be noted that the emotion-cause pair selection of the **Rank-CP** model relies on a sentiment lexicon (Wang and Ku, 2016), which we regard as inflexible in a wider range of usage scenarios.

Results on original dataset Our approach with EmotionalText and UnTypedMarker achieves an absolute F1 improvement of 2.29% and 1.44% respectively over the best previous work(Ding et al., 2020b). For the comparison of pipelined approaches, our approach outperform the baseline Inter-EC by 15.53% and 14.68% respectively in F1 score.

Results on reconstructed dataset Our approach learns emotion-aware contextual representations by fusing emotion information as input features for the cause extraction model, and show significant effectiveness on on the reconstructed dataset, in which documents are longer and more complex. The results show that our approach with EmotionalText and UntypedMarker obtain an an absolute F1 improvement of 5.20% and 5.85% respectively over (Ding et al., 2020b)'s method.

Results on Extracting Multiple Emotion-cause Pairs

To our knowledge, there are few previous works that consider the performance of their models on multiple emotion-cause pairs extraction, except from Wei et al. (2020) that build a subset of each fold's test set by selecting documents that have more than one emotion-cause pair. Since documents that contain multiple pairs are sparse in the original dataset, we build the subset on our reconstructed dataset and evaluate the approaches. The results show that our approach with UntypedMarker outperform Wei et al. (2020)'s previous work by an absolute F1 of 3.35% on extracting multiple pairs.

Specifically, we observe that the gains of our approach mainly originate from the improvement of recall rate. The approach with EmotionalText

achieves an improvement of 7.77% in recall rate over (Ding et al., 2020b)'s method in the reconstructed dataset, while on multiple pairs extraction our approach with UntypedMarker achieve an improvement of 6.90%. We can observe that the performance of Wei et al. (2020)'s model increases on multiple pairs mainly because they apply the sentiment lexicon to filter candidate emotion-pairs and tend to select fewer pairs, resulting in high precision rate and low recall rate.

4.3.2 Importance of Emotion-aware Contextual Representations

Dataset	Model	P	R	F1
Original	EmotionalText	77.83	76.01	76.81
	UntypedMarker	76.27	75.83	75.96
	w/o emotion	69.70	71.10	70.36
Reconstructed	EmotionalText	71.05	74.83	72.85
	UntypedMarker	73.78	73.30	73.50
	w/o emotion	61.95	64.74	59.73
Multiple Pairs	EmotionalText	69.62	61.85	65.19
	UntypedMarker	70.94	62.80	66.50
	w/o emotion	41.06	42.27	41.14

Table 4: Ablation studies on the fusion of emotion information in a cause extraction model.

Our core argument is that it is crucial to build distinct contextual representations specific to each emotion clause and fuse emotion information at the input layer of the cause extraction model. Above results show that both emotion-fusing strategies achieve convincing results, and in order to further validate the importance of emotion-aware contextual representations, we conduct ablation experiments by removing emotion information in the cause extraction model.

As shown in table 4, we can observe a clear gap between our models and the model without fusion of emotion features, especially in the reconstructed dataset and on multiple emotion-cause pairs extraction. Since the classification of an emotion-cause depends on the emotion it corresponds to, it is almost meaningless to perform cause extraction without emotion information, with the decline of 18.59% F1 score in extracting multiple pairs.

Based on the ablation experiments, we validate the importance of fusing emotion information at the input layer of the cause extraction model to learn emotion-aware contextual representations, and the experimental results show the robustness of our approach in handling longer and more complex document context, which leads to wider applicability in real world scenarios of emotion cause analysis.

4.3.3 Results on Emotion Extraction

Dataset	Model	P	R	F1
Original	ECPE-MLL	86.08	91.91	88.86
	RANK-CP	91.23	89.99	90.57
	Ours	88.06	89.98	88.96
Reconstructed	ECPE-MLL	83.69	86.64	85.10
	RANK-CP	92.81	58.61	71.70
	Ours	85.58	**89.59**	**87.51**
Multiple Emotions	ECPE-MLL	82.27	74.23	77.91
	RANK-CP	94.37	62.88	75.29
	Ours	86.79	**82.67**	**84.64**
	Single Clause	79.92	90.87	84.92

Table 5: Results on Emotion Extraction

One motivation of joint approach in ECPE is that the performance of emotion extraction and cause extraction can benefit each other through joint training. Experimental results shown in Table 5 demonstrate that the usage of context brings certain positive effects, while the decline in the reconstructed dataset indicate that context information can sometimes become misleading.

We can observe that joint models outperform our emotion extraction model on the original dataset while in the reconstructed dataset, however, their performance exhibits a clear decline and is outperformed by ours, especially in extracting multiple emotions from one document. Cause information or emotion information obtained via joint training does bring some benefits, but when document context becomes more complex, shared encoders in joint models fail to capture proper context information since entangled contextual representations provide more noise than benefits for the model.

4.3.4 Upper Bound of Emotion-aware Cause Extraction

To make full use of emotion information, we also consider using emotion type information. Since we cannot obtain satisfying results in emotion type classification (see appendix A.2), we test the upper bound of emotion-aware cause extraction using ground truth emotion label in each document, consisting of both the emotion clauses and their emotion type and compare the results between different emotion-fusing strategies and other ECPE methods which also report their upper bound results.

As shown in table 6, we can observe the benefits brought by emotion type between UntypedMarker and TypedMarker, while EmotionalText obtains

Strategy	P(%)	R(%)	F1(%)
(Xia and Ding, 2019)	76.10	70.84	73.28
(Tang et al., 2020)	80.80	79.90	80.30
UntypedMarker	84.68	83.54	84.09
TypedMarker	85.44	**84.43**	84.92
EmotionalText	**85.99**	83.98	**84.95**

Table 6: Comparative results of the upper bound of emotion-aware cause Extraction

the best F1 score, indicating that it may be better to integrate emotion information through emotional text. Futhermore, there are couple of documents that exceed the max input length of BERT. We split such documents to fit our model in the experiments, but text markers cannot be used if the emotion clause is located in another part of an document. Thus, for future works, we suggest the use of emotional text, which is more flexible, as the emotion-fusing strategy.

5 Related Work

Emotion Cause Analysis Lee et al. (2010) first proposed the emotion cause extraction task and released a small scale dataset. Early works used rule-based (Neviarouskaya and Aono, 2013; Gao et al., 2015; Yada et al., 2017), machine-learning-based (Ghazi et al., 2015) methods to solve the task.

Based on analysis of linguistic features in a Chinese dataset, Chen et al. (2010) suggested that a clause may be the most proper unit for emotion cause analysis in Chinese. Gui et al. (2016) reformalized the task as clause-level binary classification and released a benchmark corpus for the ECE task, followed by many works (Gui et al., 2017; Li et al., 2018; Xia et al., 2019) and datasets (Gao et al., 2017; Fan et al., 2019).

There are works (Kim and Klinger, 2018; Oberländer et al., 2020) that study the semantic role of emotions, while Oberländer and Klinger (2020) suggested that token-level sequence labeling approaches are more appropriate for emotion stimulus detection in English based on analysis across datasets.

Emotion Cause Pair Extraction Xia and Ding (2019) expanded the task to emotion cause pair extraction and construct a benchmark ECPE corpus based the Gui et al. (2016)'s dataset. Xia and Ding (2019) proposed a two-step pipeline model to solve the task, while all of the following works employ end-to-end models (Fan et al., 2020; Tang et al., 2020; Cheng et al., 2020). Some of the models select the result from all possible pairs (Chen et al., 2020b; Ding et al., 2020a,b), and some of the models regard ECPE as a clause-level sequence labeling problem (Chen et al., 2020b; Yuan et al., 2020).

Pipeline approach vs Joint approach Disputes between joint approach and pipeline approach do not only lie in the field of ECPE. In relation extraction, many systems model entity extraction and relation classification jointly (Luan et al., 2018; Wadden et al., 2019; Lin et al., 2020) while Zhong and Chen (2021) argued that shared contextual representations are suboptimal and proposed a simple pipelined approach that reached state-of-the-art performance.

6 Conclusion

In this paper, we re-investigate the motivation of emotion-cause pair extraction (ECPE), and realize another significant merit that ECPE enables the analysis of emotion causes in richer and more complex document context.

Existing state-of-the-art works of ECPE adopts an joint approach that use shared context encoders in emotion extraction and cause extraction, leading to suboptimal results on multiple emotion-cause pair extraction in entangled document context. To address the problem, we present a simple but effective approach for ECPE that build on two independent encoders for emotion extraction and emotion-oriented cause extraction with the fusion of emotion information at its input layer.

We also find that the benchmark dataset all previous ECPE works are evaluated on exhibits a bias that many documents are actually excerpts from the same original article. We reconstruct the dataset by merging such documents and conduct series of experiments on both datasets. The results show that our approach can learn contextual representations specific to each emotion and reaches state-of-the-art performance on both datasets. Our approach is more robust in extracting multiple emotion-cause pairs among more complex document context, and thus is more applicable in real-world scenarios.

References

Xinhong Chen, Qing Li, and Jianping Wang. 2020a. A unified sequence labeling model for emotion cause

pair extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 208–218.

Ying Chen, Wenjun Hou, Shoushan Li, Caicong Wu, and Xiaoqiang Zhang. 2020b. End-to-end emotion-cause pair extraction with graph convolutional network. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 198–207.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 179–187.

Zifeng Cheng, Zhiwei Jiang, Yafeng Yin, Hua Yu, and Qing Gu. 2020. A symmetric local search network for emotion-cause pair extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 139–149.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020a. Ecpe-2d: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3161–3170.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020b. End-to-end emotion-cause pair extraction based on sliding window multi-label learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3574–3583.

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)*, pages 5614–5624.

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 *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3707–3717.

Kai Gao, Hua Xu, and Jiushuo Wang. 2015. A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Systems with Applications*, 42(9):4517–4528.

Qinghong Gao, Hu Jiannan, Xu Ruifeng, Gui Lin, Yulan He, Kam-Fai Wong, and Quin Lu. 2017. Overview of ntcir-13 eca task. In *Proceedings of the NTCIR-13 Conference*.

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*, pages 152–165. Springer.

Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. A question answering approach to emotion cause extraction. *arXiv preprint arXiv:1708.05482*.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, Yu Zhou, et al. 2016. Event-driven emotion cause extraction with corpus construction. In *EMNLP*, pages 1639–1649. World Scientific.

Evgeny Kim and Roman Klinger. 2018. Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1345–1359, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. 2010. 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*, pages 45–53.

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*, pages 4752–4757.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. *arXiv preprint arXiv:1808.09602*.

Saif Mohammad, Xiaodan Zhu, and Joel Martin. 2014. Semantic role labeling of emotions in tweets. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 32–41.

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

Laura Ana Maria Oberländer, Evgeny Kim, and Roman Klinger. 2020. Goodnewseveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1554–1566.

Laura Ana Maria Oberländer and Roman Klinger. 2020. Token sequence labeling vs. clause classification for english emotion stimulus detection. In *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*, pages 58–70.

Irene Russo, Tommaso Caselli, Francesco Rubino, Ester Boldrini, Patricio Martínez-Barco, et al. 2011. Emocause: an easy-adaptable approach to emotion cause contexts. Association for Computational Linguistics (ACL).

Hao Tang, Donghong Ji, and Qiji Zhou. 2020. Joint multi-level attentional model for emotion detection and emotion-cause pair extraction. *Neurocomputing*, 409:329–340.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. *arXiv preprint arXiv:1909.03546*.

Shih-Ming Wang and Lun-Wei Ku. 2016. Antusd: A large chinese sentiment dictionary. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2697–2702.

Penghui Wei, Jiahao Zhao, and Wenji Mao. 2020. Effective inter-clause modeling for end-to-end emotion-cause pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3171–3181.

Rui Xia and Zixiang Ding. 2019. 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*, pages 1003–1012.

Rui Xia, Mengran Zhang, and Zixiang Ding. 2019. Rthn: A rnn-transformer hierarchical network for emotion cause extraction. *arXiv preprint arXiv:1906.01236*.

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

Hanqi Yan, Lin Gui, Gabriele Pergola, and Yulan He. 2021. Position bias mitigation: A knowledge-aware graph model for emotioncause extraction. *arXiv preprint arXiv:2106.03518*.

Chaofa Yuan, Chuang Fan, Jianzhu Bao, and Ruifeng Xu. 2020. Emotion-cause pair extraction as sequence labeling based on a novel tagging scheme. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3568–3573.

Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for entity and relation extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 50–61.

A Additional Findings

A.1 Results on 8:1:1 Dataset Split Settings

Strategy	P(%)	R(%)	F1(%)
(Fan et al., 2020)	73.74	63.07	67.99
(Wei et al., 2020)	65.75	73.05	69.15
(Ding et al., 2020b)	74.88	69.76	72.20
EmotionalText	71.79	73.32	72.51
UntypedMarker	73.38	**74.19**	**73.74**

Table 7: Comparative results for the 8:1:1 data split.

A.2 Results on Emotion Classification

In order to implement TypedMarker as an emotion-fusing strategy, we also attempt to classify types of emotions to provide fine-grained information for cause extraction. As explained in Equation 1, there are six types of emotions in the benchmark dataset, so we train a multi-class classification model to classify emotion types. Unfortunately, the results of emotion type classification is not satisfying enough to avoid the error propagation issue. We report the results below.

P(%)	R(%)	F1(%)
57.68	43.82	48.50

Table 8: Results on emotion type classification.

A.3 Ethnical Consideration

The datasets on which we conducted our experiments are reconstructed from Gui et al. (2016)'s emotion cause corpus, which is selected from SINA city news. All the documents are from public news report, and during the build of the dataset, the original link information has been cleaned. Therefore, the dataset do not involves any kind of violation of individual privacy.