Learning Emotion-Aware Contextual Representations for Emotion Cause Analysis

Anonymous ACL submission

Abstract

Emotion Cause Analysis has been a key topic in natural language processing. Previous works focus on Emotion Cause Extraction (ECE), a clause-level classification task aimed at extracting causes of certain given emotion in text. The task has been expanded to Emotion Cause Pair Extraction (ECPE) that focus on extracting both emotions and corresponding causes in the context. Most existing methods for the ECPE task implement a joint model that performs extracting and matching of emotion and cause clauses simultaneously. However, we argue that different input features are needed for the two subtasks, thus sharing contextual representations may be suboptimal. In this work, we propose a pipelined approach that builds on two independent pre-trained encoders, in which the emotion extraction model only provide input features for the cause extraction model. Based on a series of careful experiments, we validate that our model can create distinct contextual representations according to specific emotional texts, and thus achieve state-of-the-art performance in both ECE and ECPE tasks, with the absolute F1 improvements of 1.5% and 4.72% over best previous works respectively. Besides, we apply a set of simple clause selection rules to extract multiple pairs in the document, strengthening the applicability of our model in real world scenarios.

1 Introduction

Emotion Cause Analysis, the task about detecting the stimuli of emotions expressed in text, has attracted incresing attention from both academia and industry in recent years(Russo et al., 2011; Ghazi et al., 2015; Chen et al., 2021).

Previous works focus on Emotion Cause Extraction (ECE), which has been proposed by (Lee et al., 2010) as a word-level sequence labeling problem. (Chen et al., 2010) suggested that clause-level extraction may better fit the task of finding causes for certain emotions. (Gui et al., 2016) re-formalized ECE as a clause-level extraction problem and released a new dataset for the task. Figure 1 displays an example document in the corpus. As is shown, the goal of ECE is to extract all cause clauses corresponding to the given emotion. This framework has been followed by many works (Gui et al., 2017; Li et al., 2018; Xia et al.) and the dataset has become the benchmark dataset for ECE task.

Xia and Ding (2019) pointed out that ECE task suffers from two shortcomings: 1) The emotion must be annotated in advance. 2) The goal of ECE neglects the fact that emotions and causes are mutually indicative. Therefore, Xia and Ding (2019) developed the task to emotion cause pair extraction (ECPE), in which emotion clauses and their corresponding cause clauses are extracted as pairs, as Figure 1 shows. They proposed a two-step piplined approach to solve the ECPE problem. The model first extract all the emotion clauses and cause clauses respectively, and then feed all possible pairs to a filter and select the final result.

More recently, however, the ECPE task has been dominated by end-to-end systems that model emotion/cause clause extraction and matching jointly (Ding et al., 2020a; Cheng et al., 2020; Fan et al., 2020). It is believed that joint models are able to capture interactions between subtasks of emotion/cause extraction and emotion-cause pairing while mitigating error propagation.

In this work, we re-investigate the good and bad of emotion information in cause extraction. We argue that emotions that are expressed by certain emotional words can be extracted independently while cause clauses must depend on certain emotion clauses to *be* a cause. For the example showed in Figure 1, c18 is a cause because "surprise" is expressed in c17 and if we remove c17 and add a clause like "Guo Jiamei was very happy for their help" then c1 becomes a cause while c17 is not. Therefore, document context containing emotion

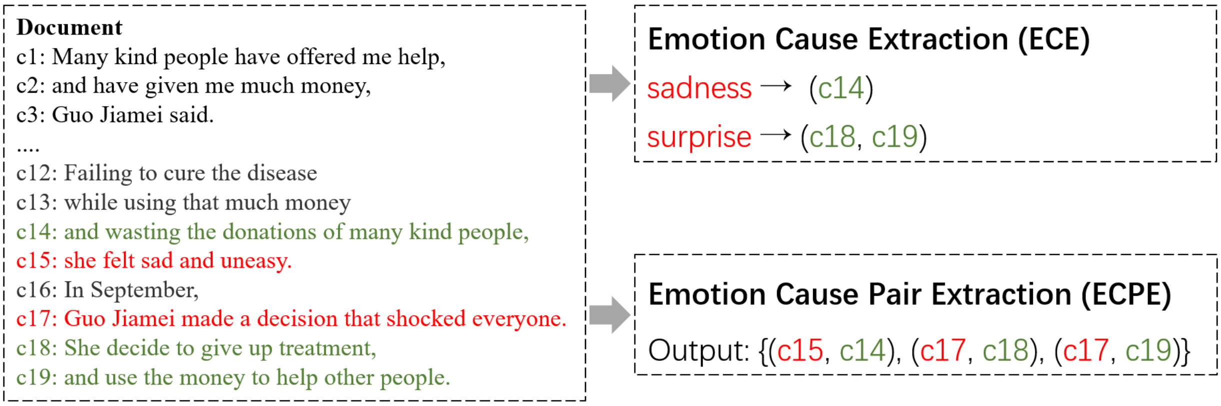


Figure 1: An Example document that occur in both ECE and ECPE benchmark dataset. Note that this document is regarded as *two different* samples in the ECE corpus since they are annotated with two differnet types of emotion, *sadness* and *surprise*.

information is so crucial for cause extraction that it is almost meaningless to perform cause extraction solely without any emotion information.

To better utilize emotion information, our approach builds on two independent pre-trained encoders, one for emotion extraction and one for emotion-oriented cause extraction, and they are trained separately. The emotion model only provide input features for the cause extraction model by encoding text of extracted emotion clause with its document context or by inserting text markers at the start and end position of the predicted emotion clause. The pre-trained encoder of the cause extraction are able to create emotion-aware contextual representations and is surprisingly effective in finding the corresponding causes of the extracted emotions. Moreover, we apply a set of clause selection rules to deal with extracting multiple causes for one emotion, and obtain even more inspiring results in multiple pair extraction.

Our contributions can be summarized as follows:

- We propose a simple but effective approach for emotion cause analysis based on two independent pre-trained encoders, and by fusing emotion information obtained from the emotion model, our emotion-oriented cause extraction model can create emotion-aware contextual representations and thus reaches state-of-the-art performance in both ECE and ECPE tasks. And we conduct extensive analysis to better understand the effectiveness.
- In order to deal with longer and more complex documents, we apply a set of simple rules for clause selection on the cause span predicted by our span-level cause extraction model. Experimental results show that our model is very suitable for extracting multiple emotion-cause pairs, strengthening the applicability of our approach in real world scenarios.

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 $e \in E$,

$$\begin{aligned} E = \{&happiness, sadness, disgust, \\ &fear, anger, surprise\} \end{aligned} \tag{1}$$

an ECE model are able to find all the cause clauses of the given emotion clause as $\{c^{c1}, c^{c2}, ...\}$. Note that the emotion clause must be annotated in advance and only one emotion occur in one document, while there may be multiple causes corresponding to it.

Emotion Cause Pair Extraction Xia and Ding (2019) developed the ECE task to Emotion Cause Pair Extraction (ECPE), in which all emotion clauses coupled with their corresponding causes will be extracted simultaneously. 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

As Figure 2 shows, our approach consist of two independent models, an emotion extraction model and a cause extraction model. The emotion extraction model first takes a clause as input and predict whether it is an emotion clause or not. Then all the extracted emotion clauses will be used as input features for the cause extraction model with their contexts in a document. By encoding the context with specific emotion text, the cause extraction model can create emotion-aware contextual representations for the document and find a span that stimulate the emotion. We select all the possible cause clauses based on the span and form emotion-cause pairs as results. Both of our models build on pre-trained encoders such as BERT (Devlin et al., 2018). We will explain the details of both models below and demonstrate why we choose independent pre-trained encoders to generate contextual representations as input features to the model.

3.1 Emotion Extraction Model

Our emotion model is a standard sentence classification model based on pre-trained encoders. Previous works take the whole document as input for emotion extraction and put emphasis on capturing context information. However, we argue that in the emotion extraction task, context information is not always helpful since there may be multiple emotion in one document and sharing contextual representations may be suboptimal for extracting *different* emotions. We train an sentence-level binary classification model for sentiment classification.

The error propagation issue is an important shortcoming that limits the performance of pipeline models. To address this problem, we hypothesize that compared to inter-clause information, inner-clause information is more useful because emotion is always expressed by certain emotional words in a clause. We implement a lexicon-based scheme to calibrate the sentiment classification result for each document and use `ANTUSD` (Wang and Ku, 2016) as the sentiment lexicon. The scheme consists of two simple rules: 1) For each emotion clause extracted by our model, if it is not predicted by the lexicon-based method, delete it. 2) If our model extract no emotion clauses for a document, we use the emotion clauses predicted by the lexicon-based method instead. The scheme greatly improves both the precision rate and recall rate of emotion extraction, with the overall $F1 > 0.95$, and thus mitigate the error propagation issue.

3.2 Emotion-oriented Cause Extraction Model

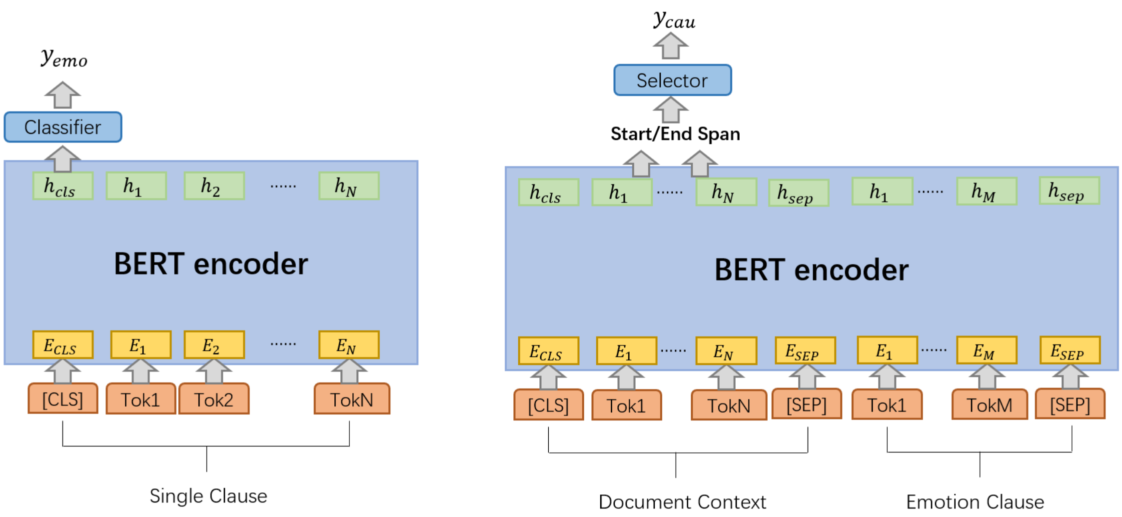
In the two-step model proposed by (Xia and Ding, 2019), there is an cause extraction component which extract potential cause clauses in a document at first. We found the performance of this independent model unsatisfying since it ignores the fact that certain cause clause depends on its corresponding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

Previous works have attached importance to the use of emotion information in cause extraction. Tang et al. (2020) point out that emotion information help the model capture inner relationship between emotion clauses and cause clauses. Xia and Ding (2019) propose a variant two-step model that use emotion classification output to improve cause extraction, while (Ding et al., 2020b) combine the result of emotion-pivot cause extraction and cause-pivot emotion extraction. However, all of these models share one LSTM layer or pre-trained encoder for contextual representations.

We argue that sharing contextual encoders lead to suboptimal results since there may be mutiple emotions and multiple causes in one document and the shared representations fail to capture proper contextual information for a specific emotion clause. For example, the 14^{th} clause in Figure 1 is crucial in finding the cause of *sadness* expressed in the 15^{th} clause but not, or even misleading in finding the cause of *surprise* expressed in the 17^{th} clause

3.2.1 Cause Extraction based on Span-Level Extraction

Inspired by BERT-based question answering models (Devlin et al., 2018), we build our emotion-oriented cause extraction model on a pre-trained encoder. The model takes each extracted emotion and its context as input and create emotion-aware



(a) Emotion Extraction Model (b) Emotion-oriented Cause Extraction Model

Figure 2: Overall architecture of our approach, consisting of two independent models based on BERT encoders.

contextual representations for the document context. Given a document $d = [c_1, ..., c_i, ..., c_{|d|}]$ and extracted emotion clause c^e, we use D denote the input for BERT encoder:

$$D = [CLS]\ c_1, ..., c_i, ..., c_{|d|}\ [SEP]\ c^e\ [SEP]$$

As Figure 2 shows, the architecture of our model is very similar to the input of a bert-based question answering model, while we replace the query with the emotion clause predicted by our emotion model. Note that the emotion model may extract multiple emotion clauses, and each cause is processed independently with its document context. We use $\mathbf{H_D}$ to denote the sequence of hidden states produced by the BERT encoder.

$$\begin{aligned} \mathbf{H_D} =& (h_{[CLS]}, \mathbf{x_{c_1}}, ..., \mathbf{x_{c_i}}, \\ & ..., \mathbf{x_{c_{|d|}}}, h_{[SEP]}, \mathbf{x_{c^e}}, h_{[SEP]}) \end{aligned} \tag{2}$$

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.

The output layer of our cause extraction model is also implemented in the similar way with BERT-based question answering model. The model predict the start and end position of a cause span by selecting tokens of maximum probabilities. For each token in the document context, its probability of being a start or an end position is computed by a softmax layer on top of the output hidden states.

Then we use the predicted cause span to select cause clauses as final result of the cause extraction model. Given the predicted cause span s, we calculate two values for each clause c in the document, $ratio_c$ and $Absolute\_length_c$. We define an indicater function

$$\mathbb{I}(w_i, s) = \begin{cases} 1 & \text{if } w_i \text{ is covered by } s \\ 0 & otherwise \end{cases} \tag{3}$$

where w_i is the i^{th} token in clause c, then the values of clause c can be calculated as

$$Absolute\_length_c = \sum_{w_i \in c} \mathbb{I}(w_i, s) \tag{4}$$

$$ratio_c = \frac{Absolute\_length_c}{\#number\ of\ tokens\ in\ clause\ c} \tag{5}$$

We select the cause clause by applying two simple rules: 1) Extract the clause of maximum $ratio$ value. 2) If there are more than one candidate clauses, we select the one with larger $Absolute\_length$ as the result.

3.2.2 Extracting Multiple Causes

By processing each predicted emotion clause independently, our model can deal with finding causes for multiple emotions in one document. However, there are some documents in which *one* emotion corresponds to *multiple* causes.

We find that causes of the same emotion are always adjacent, as shown in Figure 1, so we can utilize the predicted cause span to extract multiple causes simply by adding another rule: 3) For each clause c, if $ratio_c > 0.90$, select it as a cause clause.

3.2.3 Fusing Emotion Information through Extra Marker Tokens

In the relation extraction task, there has been an idea of using additional markers to highlight the subject and object (Peters et al., 2019; Soares et al., 2019; Sun et al., 2019). Zhong and Chen (2021) use typed markers to fuse entity information at the input layer of their relation model.

Similarly, we define untyped markers as $<\texttt{start}>$, $<\texttt{end}>$ and typed markers $<e:\texttt{start}>$, $<e:\texttt{end}>$ where $e \in E$ to mark the start and end of an emotion clause. The input of BERT encoder can be modified as:

$$[CLS]\ c_1,..., <\texttt{start}>, c^e, <\texttt{end}>, \\ ..., c_i, ..., c_{|d|}\ [SEP] \tag{6}$$

for untyped markers and

$$[CLS]\ c_1,..., <\text{start}:e>, c^e, <\text{end}:e>, \\ ..., c_i, ..., c_{|d|}\ [SEP] \tag{7}$$

for typed markers.

We use the modified input sequence to replace the input shown in Figure 2, and use the same output layer and clause selection rules to extract cause clauses. For the model with untyped markers, the performance is only slightly lower than the model based on question answering. For typed markers, since we can not obtain convincing results for emotion types, we only evaluate the model on ECE task. The details are elaborated in section 4.

3.3 Training and Optimization

For both the emotion extraction model and cause extraction model, we fine-tune the pre-trained encoder using task-specific training objectives.

For the emotion model, we use cross entropy loss.

$$L_e = -\frac{1}{N}\sum_{i=1}^{N} e_i \cdot log(p(e_i)) + (1-e_i) \cdot log(1-p(e_i)) \tag{8}$$

where N is the number of clauses in the dataset, e_i is the label (1 for emotion clase and 0 for others) and $p(e_i)$ is the probability of the clause being an emotion clause predicted by the model.

For the cause extraction model based on span-level extraction, we compute the cross entropy loss for both start and end positions and take the average value as the final loss.

$$L_{start} = -\frac{1}{|D|}\sum_{i=1}^{|D|} s_i \cdot log(p(s_i)) \tag{9}$$

Item	Number
Samples total number	2105
Documents total number	1945
Doc. with one emotion	1816
Doc. with two emotion	118
Doc. with three emotion	11
Ave. of Clause per document	14.77
Max. of Clause per document	73

Table 1: Statistics of the ECE corpus

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 emotion corres-ponding to multiple causes	66

Table 2: Statistics of the ECPE corpus.

$$L_{end} = -\frac{1}{|D|}\sum_{i=1}^{|D|} t_i \cdot log(p(t_i)) \tag{10}$$

$$L_c = \frac{1}{2}(L_{start} + L_{end}) \tag{11}$$

where $|D|$ is the number of documents in the dataset, s_i and t_i are one hot vectors (0, ..., 1, 0,...) used to denote ground truth position of start and end tokens of cause clauses. $p(s_i)$ and $p(t_i)$ are the predicted probabilities of the start and end position.

4 Experiments

4.1 Datasets and Evaluation Metrics

We conduct our experiments on two benchmark datasets in Emotion Cause Analysis: the ECE corpus and ECPE corpus.

Emotion Cause Extraction (ECE) Gui et al. (2016) released a Chinese emotion cause dataset using SINA city news and this dataset has become the benchmark dataset for ECE research. Table 1 shows the statistics of the ECE dataset. Note that documents with more than one emotions are split to several samples, thus every sample in the dataset contains only one emotion, while it may have multiple causes. We use precision, recall and F1 score as metrics for evaluation, which can be computed as:

$$P = \frac{\sum correct\_cause}{\sum proposed\_cause} \tag{12}$$

$$R = \frac{\sum correct\_cause}{\sum annotated\_cause} \quad (13)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (14)$$

Emotion Cause Pair Extraction (ECPE) Xia and Ding (2019) construct an benchmark ECPE corpus based on the ECE corpus. To meet the ECPE task settings, samples with same text content are merged into one document with multiple emotion-cause pairs. Table 2 shows statistics of the ECPE corpus. For evaluation metrics, precison, recall and F1 defined in (Xia and Ding, 2019) are used. Besides, most ECPE approaches also evaluate their models on two sub-tasks: emotion extraction and cause extraction. We do this only for emotion extraction since our approach do not conduct cause extraction solely.

Following the previous work (Xia and Ding, 2019; Wei et al., 2020; Ding et al., 2020a), we perform 10-fold cross validation and use the same data split for ECPE task. For ECE task, we also perform 10-fold cross validation by stochastically selecting 90% of the data for training and the remaining 10% for testing, repeating the experiments 10 times and reporting the average result.

4.2 Experimental Settings

We implement our model based on `Pytorch` and `Transformers` and use $\texttt{bert} - \texttt{base} - \texttt{chinese}$ as the base encoders. We use ground truth emotion labels to train our cause extraction model and during inference, we first use the emotion model to extract emotion clauses in each document and calibrate the result with the sentiment lexicon. We use `ANTUSD` (Wang and Ku, 2016) as the sentiment lexicon, and following previous work (Wei et al., 2020), the hyperparameter N is set to 3. The predicted emotion clauses are fed into the emotion-oriented cause extraction model. We set the random seed to 42 and all of our models are trained using Adam optimizer. The learning rate for emotion model is 1e-5, and the warmup ratio is 0.2. For the cause extraction models, the learning rate is 2e-5 while warmup ratio for span-level cause extraction models is 0.2.

4.3 Results and Analysis

4.3.1 Results on the ECPE Task

Table 3 compares our approach to the best previous works on the ECPE task. As is shown, our approach outperform all existing joint models and achieve state-of-the-art performance, with an absolute F1 improvement of 4.72% over the best previous work(Ding et al., 2020b). **Indep**, **Inter-CE** and **Inter-EC** are the two-step pipelined models proposed by (Xia and Ding, 2019), and our pipelined approach achieve an absolute F1 improvment of 17.76% over Inter-EC.

We can also observe the effectiveness of the sentiment lexicon, with the significant improvement of 4.86% in absolute F1 score over (Wei et al., 2020) on the evaluation of emotion extractioin subtask.

Our cause extraction model with untyped markers also achieve convincing results, with the overall F1 only slightly lower (-0.09%) than the QA-based model with lexicon and even 0.06% higher without lexicon, indicating that the use of text markers are able to substitute encodings of the emotion text as query.

Specifically, we observe that the gains of our approach mainly originate from the improvement of recall rate. Our QA-based model achieve an improvement of 3.84% in recall rate over the model proposed by Wei et al. (2020), which is also lexicon-based. Even without lexicon, our approach achieve competitive results, with an 3.21% improvement for the QA-based model and 3.26% improvement for the untyped-marker model in recall rate over (Ding et al., 2020b).

4.3.2 Results on Extracting Multiple Pairs

Since we create emotion-aware contextual representations by encoding the context with aspecific emotion clause, our approach is very suitable for finding causes for multiple emotions in one document. Following settings of previous work Wei et al. (2020), we build a subset of each fold's test set by selecting documents that have more than one emotion-cause pair and use the QA-based cause extraction model to conduct our experiments.

Table 4 reports the comparative results on the subsets. In order to process douments with emotion that have multiple corresponding causes, we apply several clause selction rules as explained in section 3.2.2. We evaluate our approach without applying rule 3. As is shown, our approach outperforms existing ECPE models on multiple pairs extraction considerably by an absolute F1 of 19.41% for the model with lexicon and 14.24% for the one without lexicon. We also find that applying rule3 improve model performance by boosting the recall rate, indicating that our approach is able to extract

Model	Emotion Extraction			Cause Extraction			Pair extraction		
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
Indep†(Xia and Ding, 2019)	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
(Chen et al., 2020a)†	86.14	78.11	81.88	73.48	58.41	64.96	71.49	62.79	66.86
(Cheng et al., 2020)†	84.06	79.80	81.81	69.92	65.88	67.78	68.36	62.91	65.45
(Tang et al., 2020)	89.90	80.00	84.70	-	-	-	71.10	60.70	65.50
(Yuan et al., 2020)	81.96	73.29	77.39	74.90	66.02	70.18	72.43	63.66	67.76
(Ding et al., 2020a)	86.27	92.21	89.10	73.36	69.34	71.23	72.92	65.44	68.89
(Wei et al., 2020)	91.23	89.99	90.57	74.61	77.88	76.15	71.19	76.30	73.60
(Chen et al., 2020b)	88.57	79.58	83.75	79.07	69.28	73.75	76.92	67.91	72.02
(Ding et al., 2020b)	86.08	91.91	88.86	73.82	79.12	76.30	77.00	72.35	74.52
Ours(marker) (w/o lexicon)	79.92	90.87	84.92	-	-	-	66.96	75.61	70.91
Ours(marker)	94.09	96.85	95.43	-	-	-	78.30	80.05	79.15
Ours(QA) (w/o lexicon)	79.92	90.87	84.92	-	-	-	66.91	75.56	70.85
Ours(QA)	**94.09**	**96.85**	**95.43**	-	-	-	**78.39**	**80.14**	**79.24**

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.

Model	P(%)	R(%)	F1(%)
Inter-EC	59.12	33.02	42.06
(Wei et al., 2020)	75.08	43.90	55.31
Ours-rule3-lexicon	68.45	58.18	62.53
Ours -lexicon	70.41	69.21	69.55
Ours -rule3	77.02	59.37	66.85
Ours	**79.28**	**70.89**	**74.72**

Table 4: Results on extracting multiple pairs. "-lexicon" means thatwe do not use the lexicon, and "-rule3" means that we do not apply rule3 explained in section 3.2.2

multiple causes for one emotion in a document.

Existing ECPE models usually suffer from performance degration when extracting multiple pairs (-19.22% for the Inter-EC model proposed by (Xia and Ding, 2019) and -10.79% for Wei et al. (2020)'s model). This greatly limits their applications in real world scenarios when document context is longer and more complex. However, by creating emotion-aware contextual representations and applying cause selection rules, the performance of our approach does not heavily depend on the number of emotion-cause pairs in a document, with the degration of -4.52% for the model with lexicon and -1.30% for the one without lexicon on extracting multiple pairs. The results shows that our approach is more robust in dealing with more complex articles more suitable for real world scenarios of emotion cause analysis.

4.3.3 Results on the ECE Task

As shown in table 5, our approach achieve state-of-the-art results in the ECE task. While the QA-based model obtain higher recall rate, the model with typed markers reaches the best F1 by advancing the previous best by 1.5% in absolute F1. Compared to untyped markers, typed markerss bring an improvement of 1.2%, indicating that emotion types benifits the extraction of cause clause.

Model	P(%)	R(%)	F1(%)
(Gui et al., 2017)	70.7	68.4	69.6
(Li et al., 2018)	77.2	68.9	72.7
(Xia et al.)	76.9	76.6	76.7
(Tang et al., 2020)	**80.8**	79.9	80.3
Ours(untyped marker)	78.2	83.1	80.6
Ours(QA)	78.5	**83.7**	81.0
Ours(typed marker)	80.2	83.5	**81.8**

Table 5: Results on the ECE task.

4.3.4 Importance of Emotion-aware Contextual Representations

Our core argument is that it is crucial to build distinct contextual representations according to specific emotion text by encoding the text of predicted emotion clauses with their document contexts or by inserting untyped markers to the start and end positions of predicted emotion clause in the context.

Above results show that both methods achieve convincing results, and in order to further validate the importance of emotion-aware contextual representations, we conduct ablation experiments by removing emotion text and position markers in the context.

	Pair Extraction		
Model	P(%)	R(%)	F1(%)
-emotion -lexicon	61.14	69.06	64.75
-emotion	71.09	72.68	71.86
-emotion -lexicon†	53.53	52.80	52.94
-emotion†	59.90	54.19	56.80

Table 6: Results of ablation experiments. We use †to mark the models that are evaluated on extracting multiple pairs, following settings explained in section 4.3.2

From Table 6, we can observe a clear gap between the model without fusion of emotion features and our models. Compared to our QA-based model, the model suffers from the degration of -7.38% with lexicon and -6.1% without lexicon. The gap is more obvious for the results on extracting multiple pairs, with the degration of -17.92% with lexicon and -16.61% without lexicon. In other words, we validate that the performance of cause extraction can boost simply by encoding emotion text with the context, or by inserting two marker tokens.

5 Related Work

Emotion Cause Extraction Lee et al. (2010) first proposed the emotion cause extraction task and realeased a small scale dataset. Early works used rule-based (Chen et al., 2010), machine learning based (Ghazi et al., 2015) to solve the task.

Gui et al. (2016) re-formalized the task as clause-level binary classification and realeased a benchmark corpus for the ECE task. This framework is followed by many works (Gui et al., 2017; Li et al., 2018; Xia et al.).

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, all of the following works employs 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 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), with the belief that joint models can capture the interactions between entities and relations and avoid error propagation. However, 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 present a simple but effective approach for emotion cause analysis, including both emotion cause extraction (ECE) and emotion cause pair extraction (ECPE) tasks. Unlike most of the existing ECPE systems that employ a joint model and conduct end-to-end training, our approach build on two independent encoders for emotion extraction and emotion-oriented cause extraction. Experiments show that our approach reach state-of-the-art performance in both ECE and ECPE task and is even more suitable for extracting multiple emotion-cause pairs in a document, making it more applicable in real world scenarios. We conduct extensive ablation experiments to show the importance of creating emotion-aware contextual representations and the effectiveness of applying a set of simple rules for cause clause selection.

In the future work, we will try to utilize emotion type information and deal with longer and more complex documents, in which causes of the same emotion may not be adjacent. It is also necessary to compare the role of cause clause with cause span in emotion cause analysis.

References

Jie Chen, Jingying Yu, Shu Zhao, and Yanping Zhang. 2021. User's review habits enhanced hierarchical neural network for document-level sentiment classification. *Neural Processing Letters*, 53(3):2095–2111.

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, 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.

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.

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*.

Matthew E Peters, Mark Neumann, Robert L Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A Smith. 2019. Knowledge enhanced contextual word representations. *arXiv preprint arXiv:1909.04164*.

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).

Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. *arXiv preprint arXiv:1906.03158*.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

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. Rthn: A rnn-transformer hierarchical network for emotion cause extraction.

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 Direct Clause-Level Extraction

We also attempt to extract cause clauses directly by applying a emotion-oriented multi-label learning model. The model is also based on a BERT encoder and we add a multi-label output layer on top of the sequence of hidden states $\mathbf{H_D}$. We construct the representation of each clause r_{c_i} by computing the average of all hidden states of tokens in clause c_i and concatenate it with the context vecotor.

$$\mathbf{h_{c_i}} = \frac{1}{|c_i|}\sum_{j=1}^{|c_i|} h_{ij} \tag{15}$$

$$\mathbf{r_{c_i}} = [\mathbf{h_{c_i}}, h_{[CLS]}] \tag{16}$$

where h_{ij} and $h_{[CLS]}$ are the hidden states obtained by Equation 2. Then we use sigmoid function to compute the output label $\mathbf{y_{c_i}}$ where $\mathbf{y_{c_i}} \in \{0, 1\}$, and 1 means that clause c_i is a cause clause of the predicted emotion.

During training, we use the cross entropy loss that can be calculated as:

$$\begin{aligned} L_e = -\frac{1}{|D|}\sum_{i=1}^{|D|}\frac{1}{|c_i|}\sum_{j=1}^{|c_i|} c_{ij} \cdot log(p(c_{ij})) \\ + (1 - c_{ij}) \cdot log(1 - p(c_{ij})) \end{aligned} \tag{17}$$

where $|c_i|$ is the number of clauses in i^{th} document, c_{ij} is the label of j^{th} clause in i^{th} document (1 for cause clase and 0 for others) and $p(c_{ij})$ is the probability of the clause being an cause predicted by the model.

However, the performance of clause-level multi-label learning model underperform the span-level model. We obtain more negative results when we insert $[CLS]$ and $[SEP]$ tokens into start and end positions of clauses and use $h_{[CLS_i]}$ to replace $\mathbf{r_{c_i}}$ in Equation 16. These results indicates that token-level information is crucial in finding cause clauses, thus we use the span-level model for emotion-oriented cause extraction in our approach.

A.2 Results on Emotion Classification

The results on the ECE task show that emotion type benifits cause extraction. We also attempt to classify types of emotions to provide fine-grained information for cause extraction in ECPE task. 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 7: Results on emotion classification.