A Multi-turn Machine Reading Comprehension Framework with Rethink Mechanism for Emotion-Cause Pair Extraction

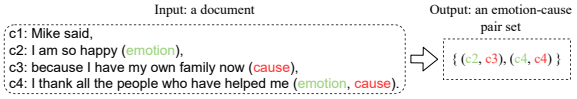
Changzhi Zhou, Dandan Song*, Jing Xu, Zhijing Wu
School of Computer Science and Technology,
Southeast Academy of Information Technology,
Beijing Institute of Technology, Beijing, China
zhou_changzhi97@163.com
sdd@bit.edu.cn, xujing@bit.edu.cn
wuzhijing.joyce@gmail.com

Abstract

Emotion-cause pair extraction (ECPE) is an emerging task in emotion cause analysis, which extracts potential emotion-cause pairs from an emotional document. Most recent studies use end-to-end methods to tackle the ECPE task. However, these methods either suffer from a label sparsity problem or fail to model complicated relations between emotions and causes. Furthermore, they all do not consider explicit semantic information of clauses. To this end, we transform the ECPE task into a document-level machine reading comprehension (MRC) task and propose a Multi-turn MRC framework with Rethink mechanism (MM-R). Our framework can model complicated relations between emotions and causes while avoiding generating the pairing matrix (the leading cause of the label sparsity problem). Besides, the multi-turn structure can fuse explicit semantic information flow between emotions and causes. Extensive experiments on the benchmark emotion cause corpus demonstrate the effectiveness of our proposed framework, which outperforms existing state-of-the-art methods.[1]

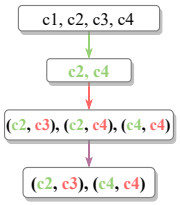
1 Introduction

Emotion cause extraction (ECE) is a classical emotion cause analysis task that aims to extract the corresponding causes of the given emotional expressions in an emotional document (Gui et al., 2016). However, the ECE task is not practical in real-world scenarios without annotated emotions. To overcome the limitation and capture mutual indications of emotions and causes together, Xia and Ding (2019) came up with a new task called emotion-cause pair extraction (ECPE), which aims to extract a potential emotion-cause pair set consisting of all emotions and their corresponding causes from a document. As shown in Figure 1(a), c2 is an emotion clause, c3 is the corresponding cause clause, and c4 is an emotion clause that is its own corresponding cause clause itself.



(a) An example of the ECPE task

(b) Pairing matrix



(c) Our approach

Figure 1: The green colour denotes an emotion clause, and the red colour denotes a cause clause. Figure (a) is an example of the ECPE task. Figure (b) is a pairing matrix generated by pair-level end-to-end approaches. Only (c2, c3) and (c4, c4) are valid pairs. Figure (c) shows the processing results by our MM-R in each turn.

In recent years, there has been a trend to use end-to-end methods to solve the ECPE task and the two sub-tasks, emotion extraction and cause extraction, which aim to extract all emotions and causes in one document. These end-to-end methods can be classified into two categories: pair-level methods by combining all clause pairs to form a clause pairing matrix; sequence labeling methods by designing novel tagging schemes. Specifically, some works (Ding et al., 2020a,b; Wei et al., 2020; Chen et al., 2020b; Wu et al., 2020) proposed pair-level methods that generate a pairing matrix by enumerating all possible combinations of clauses and then select valid emotion-cause pairs, as shown in Figure 1(b). However, the average ratio of invalid/valid pairs is more than 200:1 in the ECPE benchmark corpus (Xia and Ding, 2019), which

*Corresponding Author
[1]Data and code are available at https://github.com/zhoucz97/ECPE-MM-R

Proceedings of the 29th International Conference on Computational Linguistics, pages 6726–6735
October 12–17, 2022.

leads to a label sparsity problem. Besides, some works (Chen et al., 2020a; Yuan et al., 2020) proposed sequence labeling methods based on novel tagging schemes. However, they cannot effectively model the corresponding relation between emotions and causes. For example, Chen et al. (2020a) cannot deal with the situation where two emotion clauses of the same emotion type correspond to different cause clauses. Therefore, how to solve the label sparsity problem by avoiding the calculation of the pairing matrix while modeling the corresponding relation between emotions and causes is a crucial challenge. Furthermore, Zhong and Chen (2021) has shown that explicit semantic information can improve performance in the extraction tasks. Therefore, utilizing explicit semantic information of clauses to improve the performance in the ECPE task is also a challenge.

To address these challenges, in this study, the ECPE task is transformed into a document-level machine reading comprehension (MRC) task, and a Multi-turn MRC framework with Rethink mechanism (MM-R) is proposed. The multi-turn structure decomposes the ECPE task to model the corresponding relation between emotions and causes and avoid generating the pairing matrix. In every turn, static and dynamic queries designed manually make full use of explicit semantic information of clauses to improve the performance of emotion or cause extraction. In addition, inspired by the human two-stage reading behavior, in which search for possible answer candidates and then verify these candidates (Zheng et al., 2019), the rethink mechanism is proposed to verify candidate emotion-cause pairs further to enhance the flow of information between emotions and causes and improve the overall performance.

Specifically, in the first turn, all emotion clauses are extracted. In the second turn, cause clauses corresponding to each emotion clause are extracted with the assistance of explicit semantic information about emotions. Numerous experiments (Ding et al., 2020b; Fan et al., 2020; Ding et al., 2020a; Wei et al., 2020; Chen et al., 2020b,a; Yuan et al., 2020) have shown that emotion extraction is more reliable than cause extraction, and therefore it is reasonable to extract emotions first and then extract the corresponding causes with the assistance of emotions. Hence, a candidate emotion-cause pair set is obtained in the first two turns. In the third turn, a rethink mechanism verifies each candidate emotion-cause pair. For example, as shown in Figure 1(c), emotion clauses c2 and c4 can be obtained in the first turn, and cause clauses corresponding to each emotion clause can be obtained in the second turn. The candidate emotion-cause pair set {(c2, c3), (c2, c4), (c4, c4)} can thus be obtained without the pairing matrix. In the third turn, the rethink mechanism demonstrates that c2 is not one of the corresponding emotion clauses of c4. Therefore, the valid emotion-cause pair set becomes {(c2, c3), (c4, c4)}.

Comprehensive experiments are conducted on the ECPE benchmark datasets to verify the effectiveness of the proposed MM-R framework. The experimental results show that the proposed framework substantially outperforms the previous methods. The contributions of this research can be summarized as follows:

- The ECPE task is formalized as a document-level machine reading comprehension (MRC) task. To our best knowledge, this is the first time that the ECPE task has been transferred to the MRC task.
- Based on the MRC formalization, a multi-turn MRC framework with rethink mechanism (MM-R) is proposed. It models the corresponding relation between emotions and causes and alleviates the label sparsity problem. Furthermore, the MM-R can use explicit semantic information of clauses effectively.
- The experimental results demonstrate that the proposed framework outperforms existing state-of-the-art methods.

2 Related Work

2.1 Emotion-Cause Pair Extraction

Xia and Ding (2019) proposed the emotion-cause pair extraction (ECPE) task and released the Chinese benchmark dataset. Recently, many methods based on the pair-level end-to-end framework have been designed for the ECPE task. For example, Wei et al. (2020) obtained excellent results by modeling inter-clause relationships and using the ranking mechanism. Ding et al. (2020a) used a 2D-transformers module to model the interactions of different emotion-cause pairs. However, these pair-level end-to-end methods had a label sparsity problem caused by calculating the pairing matrix. Ding et al. (2020b) used a sliding window to restrict the amount of candidate emotion-cause pairs,

which aims to shrink the pairing matrix. However, the method still belonged to the pair-level end-to-end method because it assumed all clauses were emotion or cause clauses.

Sequence labeling was another popular method. However, these methods could not model the corresponding relation between emotions and causes. A unified tagging scheme (Chen et al., 2020a) could not be used on a document with multiple emotion clauses of the same emotion types because the tagging scheme was based on emotion types. Another joint tagging scheme (Yuan et al., 2020) based on the distance between the cause and the corresponding triggered emotion. The tagging scheme could not handle the situation in which the distance between the emotion clause and the corresponding cause clause exceeded the distance threshold. Therefore, the challenge remains of how to solve the label sparsity problem while modeling the corresponding relation between emotions and causes.

2.2 Machine Reading Comprehension

Recently, it became popular to transform many traditional natural language processing tasks into machine reading comprehension (MRC) tasks. For example, Liu et al. (2020) cast the event extraction as a machine reading comprehension problem to solve the data scarcity problem. Zhou et al. (2021) proposed a dual QA framework aiming at event argument extraction. Chen et al. (2021) and Mao et al. (2021) proposed a bidirectional MRC and a Dual-MRC frameworks to handle aspect-based sentiment analysis, respectively. In this paper, the ECPE task is formalized as a document-level MRC task, and the multi-turn MRC framework with rethink mechanism is proposed. It concatenates a query with each clause to explicitly utilize semantic information of clauses and uses a single classifier to predict whether a clause is an emotion or a cause clause in each turn.

3 Methodology

3.1 Problem Formulation

Given a document consisting of multiple clauses $D = (c_1, c_2, ..., c_{|D|})$ [2] and each clause contains multiple words $c_i = (w_{i,1}, w_{i,2}, ..., w_{i,|c_i|})$, our goal is to extract a set of emotion-cause pair in D:

$$P = \{(c_k^{e_i}, c_k^{ca_{i,j}})\}_{k=1}^{|P|}, \quad (1)$$

where the superscript e_i and $ca_{i,j}$ denote the i-th emotion clause and its corresponding j-th cause clause in D; the subscript k denotes the k-th emotion-cause pair in set P; an emotion clause corresponds to one or more cause clauses. [3]

3.2 Query Design

Static and dynamic queries are used to formalize the ECPE task to the MRC task. All queries can be formulated as follows:[4]

- **Static emotion query** $q^{se} \in Q^{se}$: The query "*Is it an emotion clause?*" is designed to extract all emotion clauses.
- **Static cause query** $q^{sc} \in Q^{sc}$: The query "*Is it a cause clause?*" is designed to extract all cause clauses.
- **Static pair query** $q^{sp} \in Q^{sp}$: The query "*Is it an emotion-cause pair?*" is designed to extract all emotion-cause pairs.
- **Dynamic emotion query** $q^{de} \in Q^{de}$: The query template "*Is it an emotion clause corresponding to c_i?*" is designed to extract emotion clauses corresponding to clause c_i.
- **Dynamic cause query** $q^{dc} \in Q^{dc}$: The query template "*Is it a cause clause corresponding to c_i?*" is designed to extract cause clauses corresponding to clause c_i.

3.3 Frameworks

The architecture of MM-R is illustrated on the left side of Figure 2. In the first turn, a static emotion query q^{se} is used to extract emotion clause set $E = \{c^{e_i}\}_{i=1}^{|E|}$. In the second turn, based on each extracted emotion clause $c^{e_i} \in E$, a dynamic cause query q^{dc} is constructed to obtain c^{e_i}'s corresponding cause clause set $C_i = \{c^{ca_{i,j}}\}_{j=1}^{|C_i|}$. In this way, the candidate emotion-cause pair set $P^{can} = \{(c_k^{e_i}, c_k^{ca_{i,j}})\}_{k=1}^{|P^{can}|}$ is obtained in the first two turns. To filter out invalid emotion-cause pairs in P^{can} further, for each cause clause $c^{ca_{i,j}}$, the framework rethinks whether the clause c^{e_i} is its corresponding emotion clause with the help of dynamic emotion query q^{de}. Finally, the valid

[2] The $|*|$ denotes the number of elements in the collection $*$.

[3] To avoid confusion, the "cause clause" is denoted to ca rather than c.

[4] Static cause query and static pair query are only used in two variants of our framework (MM-D and MRC-E2E); refer to the Experiments section for details.

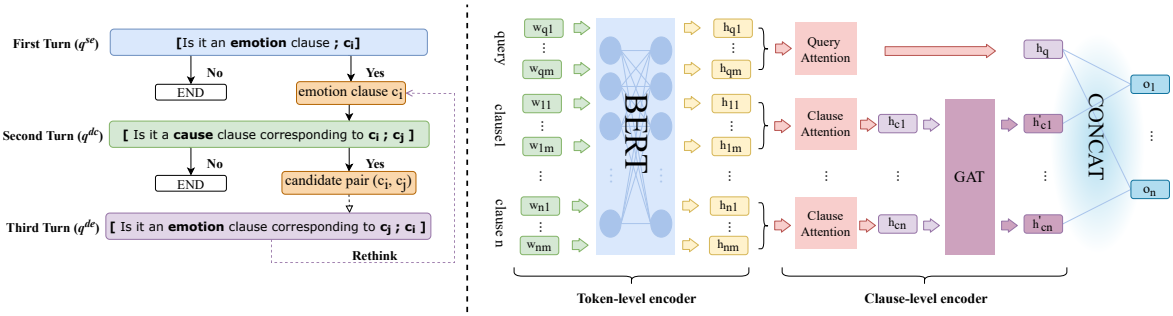


Figure 2: *Left*: The overall architecture of our MM-R framework. In each turn, the answer is yes if the probability output by the classifier is greater than 0.5, otherwise it is no. *Right*: The implementation structure of the encoding layer which includes the token-level encoder and the clause-level encoder. The token-level encoder generates the hidden representation of each token using the BERT module. The clause-level encoder provides the hidden representation of query and each clause using the attention mechanism and graph attention network. Finally, the concatenate operation (CONCAT) is executed on the hidden representations of queries and clauses.

emotion-cause pair set $P = \{(c_k^{e_i}, c_k^{ca_{i,j}})\}_{k=1}^{|P|}$ is obtained.

3.4 Encoding Layer

The structure of the encoding layer is illustrated on the right side of Figure 2. The encoding layer includes the token-level encoding layer and the clause-level encoding layer.

Token-level encoder: This study used BERT (Devlin et al., 2019) as its token-level contextualized encoder. Specifically, a document $D = (c_1, c_2, ..., c_{|D|})$ and a query q are used to construct the BERT input sequence:

$$I = \{[CLS], w_{q,1}, w_{q,2}, ..., w_{q,|q|}, [SEP], w_{1,1}, w_{1,2}, ..., w_{|D|,1}, ..., w_{|D|,|c_{|D|}|}\}, \quad (2)$$

where $q = q^{se}$ in the first turn, $q = q^{dc}$ in the second turn and $q = q^{de}$ in the third turn; $w_{q,j}$ is the j-th token of query q; $w_{i,j}$ is the j-th token of the i-th clause in the document D; and $[CLS]$, $[SEP]$ are special BERT tokens. The tokens can then be encoded into hidden representations:

$$\begin{aligned} H^I =& BERT(I) \\ =& \{h_{[CLS]}, h_{q,1}, h_{q,2}, ..., h_{q,|q|}, h_{[SEP]}, \\ & h_{1,1}, h_{1,2}, ..., h_{|D|,1}, ..., h_{|D|,|c_{|D|}|}\}, \end{aligned} \quad (3)$$

where $H^I \in \mathbb{R}^{|I| \times d}$; d is the dimension of the hidden states; and $h_{i,j}$ denotes the hidden representation of token $w_{i,j}$.

Clause-level encoder: The attention mechanism and the graph attention network (GAT) (Veličković et al., 2017) are used to obtain clause-level hidden representations.

Specifically, for clause c_i, its token-level representation set is selected from H^I:

$$S_{c_i} = \{h_{i,j}\}_{j=1}^{|c_i|} \in \mathbb{R}^{|c_i| \times d}. \quad (4)$$

The attention mechanism produces an attention weight vector α_i. The hidden representation of clause c_i is obtained by calculating the weighted sum of the hidden representations of all tokens:

$$\alpha_i = softmax(w^T S_{c_i} + b) \in \mathbb{R}^{1 \times |c_i|}, \quad (5)$$

$$h_{c_i} = sum(\alpha_i S_{c_i}) \in \mathbb{R}^{1 \times d}, \quad (6)$$

where w and b are learnable parameters. For all clauses, the attention mechanism is used to obtain the hidden representations:

$$H_C = \{h_{c_1}, h_{c_2}, ..., h_{c_{|D|}}\}. \quad (7)$$

Similarly, the attention mechanism is used to obtain the hidden representation of query q:

$$H_Q = \{h_q\}. \quad (8)$$

Furthermore, GAT models the interaction among clauses, and then the hidden representations of clauses are updated to:

$$H'_C = GAT(H_C) = \{h'_{c_1}, h'_{c_2}, ..., h'_{c_{|D|}}\}. \quad (9)$$

Finally, $h_q \in H_Q$ and $h'_{c_i} \in H'_C$ are concatenated to obtain $o_i = [h_q; h'_{c_i}]$, where $[;]$ denotes the concatenate operation. The output of the encoding layer is thus obtained:

$$O_{enc} = \{o_1, o_2, ..., o_{|D|}\}. \quad (10)$$

3.5 Answering Prediction

The output representation of the encoder layer $o_i = [h_q; h'_{c_i}]$ contains abundant semantic information, including identification information of emotion/cause, semantic information of clauses, and semantic information of the document. Therefore, only a single linear perception (SLP) is needed to obtain the answer to the query by predicting yes or no:

$$\hat{y}_i = \sigma(w_S^T o_i + b_S), \tag{11}$$

where w_S and b_S are learnable parameters of SLP; $\sigma(\cdot)$ is a logistic function; and $\hat{y}_i$ denotes the probability that the answer is yes. If $\hat{y}_i > 0.5$, the answer is judged to be yes, meaning that clause c_i is one of the answers to query q. This prediction style is similar to BoolQ (Clark et al., 2019).

3.6 Joint Training

A joint learning strategy was used in this study, the cross-entropy loss for each type of query was minimized as follows:

$$\mathcal{L}^* = -\sum_{i=1}^{N}\sum_{j=1}^{|D|}\sum_{k=1}^{|Q^*|}[p(y_{i,j,k}|c_{i,j}, q_k^*)\log\hat{p}(y_{i,j,k}|c_{i,j}, q_k^*)], \tag{12}$$

where $* \in \{se, dc, de\}$; N denotes the number of documents in the dataset; $c_{i,j}$ is the j-th clause of the i-th document; and q_k^* is the k-th query in Q^*.

Therefore, the final loss function $\mathcal{L}$ was as follows:

$$\mathcal{L} = \mathcal{L}^{se} + \mathcal{L}^{dc} + \mathcal{L}^{de}, \tag{13}$$

3.7 Inference

In the first turn, the $q^{se} \in Q^{se}$ first identifies the emotion clause set $E = \{c^{e_1}, c^{e_2}, ..., c^{e_{|E|}}\}$. In the second turn, for each predicted emotion clause c^{e_i}, the $q^{dc} \in Q^{dc}$ recognizes the corresponding cause clause set $C_i = \{c^{ca_{i,1}}, c^{ca_{i,2}}, ..., c^{ca_{i,|C_i|}}\}$ and obtains the set of candidate emotion-cause pairs $P^{can} = \{(c_k^{e_i}, c_k^{ca_{i,j}})\}_{k=1}^{|P^{can}|}$. And the probability of each candidate pair is $p(c^{e_i}, c^{ca_{i,j}}) = p(c^{e_i})p(c^{ca_{i,j}}|c^{e_i})$.

In the third turn, the $q^{de} \in Q^{de}$ is used for selecting valid emotion-cause pairs. Furthermore, the rethink mechanism is implemented through a soft selection strategy, in which the probability is adjusted and a probability threshold is used. Specifically, the probability of the candidate pair $(c^{e_i}, c^{ca_{i,j}})$ is updated by this strategy as follows:

$$p(c^{e_i}, c^{ca_{i,j}}) = \lambda p(c^{e_i})p(c^{ca_{i,j}}|c^{e_i}), \tag{14}$$

where the weight factor λ is used to adjust the probability of candidate emotion-cause pairs. λ is 1 when the predicted result of the third turn is yes, otherwise λ is a unique value between 0 and 1. The set of valid emotion-cause pairs is as follows:

$$P = \{(c^{e_i}, c^{ca_{i,j}})|(c^{e_i}, c^{ca_{i,j}}) \in P^{can}, p(c^{e_i}, c^{ca_{i,j}}) > \delta\}, \tag{15}$$

where δ is a probability threshold value.

4 Experiments

4.1 Dataset and Metrics

The benchmark dataset (Xia and Ding, 2019) was constructed based on a public Chinese emotion corpus (Gui et al., 2016) from the SINA NEWS website [5]. The dataset contains 1,945 documents and 28,727 clauses. The number of candidate clause pairs is 490,367 and the number of valid emotion-cause clause pairs is 2,167.

In the experiment, we use the two styles of data split:

- **10-fold cross validation**. Selecting 90% of the data for training and the remaining 10% for testing stochastically (as same as Xia and Ding (2019)). The most previous works use the data split style.

- **Training/Validation/Test data set**. Selecting 80% of the data for training, 10% of the data for validating and the remaining 10% for testing stochastically (as same as Fan et al. (2020)). The data split style is more plausible than the first style.

Furthermore, when we extract the emotion-cause pairs, we obtain the emotions and causes simultaneously. Thus, the performance of emotion extraction and cause extraction were also evaluated. The precision P, recall R and F1 score defined in (Gui et al., 2016; Xia and Ding, 2019) are used to evaluate the performance of the three tasks.

[5] https://news.sina.com.cn/

Model	E-C Pair Extraction			Emotion Extraction			Cause Extraction		
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
SL-NTS	72.43	63.66	67.76	81.96	73.29	77.39	74.90	66.02	70.18
TransDGC (Val)	73.74	63.07	67.99	87.16	82.44	84.74	75.62	64.71	69.74
ECPE-2D	72.92	65.44	68.89	86.27	**92.21**	89.10	73.36	69.34	71.23
PairGCN	76.92	67.91	72.02	88.57	79.58	83.75	79.07	69.28	73.75
RANKCP	71.19	76.30	73.60	91.23	89.99	90.57	74.61	77.88	76.15
ECPE-MLL	77.00	72.35	74.52	86.08	91.91	88.86	73.82	79.12	76.30
MM-R	**82.18**	**79.27**	**80.62**	**97.38**	90.38	**93.70**	**83.28**	**79.64**	**81.35**
MM-R (Val)	78.97	75.32	77.06	96.09	88.09	91.88	80.90	76.21	78.45

Table 1: Performance of our models and baselines. P, R and F1 denote precision, recall and F1-measure respectively. E-C denotes Emotion-Cause. TransDGC(Val) and MM-R(Val) use the second data split style, the rest of models use the first data split style.

4.2 Experimental Settings

We used the BERT$_{base-Chinese}$ as our encoding backbone. During training, the AdamW optimizer with a weight decay of 0.01 was used for online learning, and the initial learning rate and warmup rate are set to 1e-5 and 0.1 respectively. The batch size is set to 2. As for regularization, dropout is applied for networks and the dropout rate is set to 0.1. We trained the model 20 epochs in total and adopted early stopping strategy. In the inference stage, $\lambda \in \{0.7, 1.0\}$ and the threshold δ was set to 0.5. The model was run on a Tesla V100 GPU.

4.3 Baselines

To demonstrate the effectiveness of our method, we compare our model with the following BERT baselines.

- **SL-NTS** (Yuan et al., 2020) regards the ECPE task as a sequence labeling problem and proposes a novel tagging scheme.
- **TransDGC** (Fan et al., 2020) proposes a transition-based model to transform the task into a parsing-like directed graph construction procedure. [6]
- **ECPE-2D** (Ding et al., 2020a) uses a 2D transformers to model the interactions of different emotion-cause pairs.
- **PairGCN** (Chen et al., 2020b) constructs a pair graph convolutional network to model dependency relations among local neighboring candidate pairs.
- **RANKCP** (Wei et al., 2020) uses a graph attention network to model interactions between clauses and selects emotion-cause pairs by the ranking mechanism.
- **ECPE-MLL** (Ding et al., 2020b) extracts emotion-cause pairs based on sliding window multi-label learning. It is a state-of-the-art model of baselines.

[6]It should be noted that the TransDGC method use the second data split style, while others methods use the first data split style.

4.4 Variants

To show the effectiveness of the multi-turn structure and the rethink mechanism in the proposed method, We designed the following variants:

- **MRC-E2E** is a single-turn **MRC** framework with an **E**nd-**t**o-**E**nd style. The static queries q^{sp} is used to extract an emotion-cause pair set.
- **MM** is a simple **M**ulti-turn **M**RC framework. Compared with MM-R, it removes the rethink mechanism.
- **MM-D** is proposed inspired by (Chen et al., 2021; Mao et al., 2021), which is a **M**ulti-turn **M**RC framework with **D**ual structure. One direction sequentially recognizes emotions and causes to obtain the candidate emotion-cause pair set with the help of static emotion query q^{se} and dynamic cause query q^{dc}. In contrast, the other direction identifies causes first and then emotions to obtain another candidate set

	Natural QL	Pseudo QL	Structured QL
Q^{se}	*Is it an emotion clause?*	*emotion?*	*emotion:_;cause:None*
Q^{dc}	*Is it a cause clause corresponding to* c_i*?*	c_i*;cause?*	*emotion:*c_i*;cause:_*
Q^{de}	*Is it an emotion clause corresponding to* c_i*?*	c_i*;emotion?*	*emotion:_;cause:*c_i
MM-R	80.62 (%F1)	80.51 (%F1)	79.72 (%F1)

Table 2: The performance of different query language designs (Natural, Pseudo and Structured QL) on ECPE task. "QL" denotes "Query Language". Q^{se}, Q^{dc} and Q^{de} are static emotion query, dynamic cause query and dynamic emotion query, respectively.

with the help of static cause query q^{sc} and dynamic emotion query q^{de}. Finally, we take the intersection of the two candidate sets.

4.5 Main Results

Table 1 gives the comparative results for the ECPE task and the two sub-tasks. The proposed MM-R shows a clear advantage over other baselines, obtaining F_1 improvements of 6.10%, 3.13%, and 5.05%, respectively, over the previous baselines on the three tasks.

Specifically, for the ECPE task, the MM-R obtains 5.18%, 6.92% and 6.10% improvements in the P, R, and $F1$ measures compared to ECPE-MLL. For the emotion extraction task, the MM-R performs 3.13% better than the best-performed baseline RANKCP, and it is worth noting that the increase in precision contributed most to the boost in the F_1 score. We believe that the high precision is due to the rethink mechanism filtering out some negative samples. The cause extraction task is more difficult because the cause clauses depend heavily on emotions. However, with the help of dynamic cause queries constructed by emotion semantic information, substantial increases (+5.05% F_1) are obtained compared to ECPE-MLL.

Besides, when we use the data split as the Trans-DGC (Fan et al., 2020), our method MM-R (Val) still obtains the state-of-the-art performance over the previous baselines.

5 Further Analysis

5.1 Effect of the Multi-turn Structure

MRC-E2E has the same label sparsity problem as pair-level end-to-end baselines, and it cannot model the interaction of emotions and causes due to its single-turn structure. However, MM, the most straightforward multi-turn framework among these similar methods, achieves 2.68%, 0.02%, and 2.84% F_1 improvements over MRC-E2E as shown in Table 3, which demonstrates the effectiveness of the multi-turn structure.

Model	Extraction of. (F1 %)		
	Emotion	Cause	E-C pair
MRC-E2E	90.34	77.92	75.35
MM	93.02	77.94	78.19
MM-D	93.67	79.47	78.76
MM-R	**93.70**	**81.35**	**80.62**

Table 3: Performance of variants.

In addition, compared with end-to-end baseline ECPE-MLL, MRC-E2E still achieves 1.48%, 1.62% and 0.83% improvements of F_1 in three tasks, which demonstrates that the explicit emotion/cause semantic information played an important role.

5.2 Effect of the Rethink Mechanism

To verify the advantages of the rethink mechanism, MM-R was compared with the other two models (MM and MM-D). Specifically, compared with MM, MM-R achieved 0.68%, 3.41%, and 2.43% F1 improvements on the three tasks, demonstrating the the rethink mechanism's effectiveness. Furthermore, MM-D achieved excellent performance with its dual structure. However, its score was 1.86% lower than MM-R in the ECPE task. The dual structure treats both emotion extraction and causes extraction equally, whereas the rethink mechanism is more trusting of emotion extraction. MM-R obtains a better performance due to emotion extraction being more reliable than cause extraction[7].

5.3 Effect of the Query Design

We explore the effect of different query language design on the model. As shown in Table 2, their F1

[7]We have illustrated in the introduction that emotion extraction is more reliable than cause extraction.

“过年了 (c1)，债主把家里粮食都搬走了 (c2)，别家都在欢欢喜喜过年 (c3)，而俺家连割肉的钱都没有 (c4)，我和母亲抱头痛哭 (c5)”，陈怀军说 (c6)。					
***Translate**: "It's the New Year (c1), and the creditor looted all food of my family (c2). Other families happily celebrate the New Year (c3), but we are too poor to buy meat (c4). This makes us very sad(c5)", said Huaijun Chen (c6).*					
The first turn	The second turn	Rethink	Threshold	Valid E-C pairs	Ground-truth
c3, c5	(c3, c4), $p_{(3,4)}$ = 0.5135 (c5, c4), $p_{(5,4)}$ = 0.6313	(c3, c4), $p_{(3,4)}$ = 0.3595 (c5, c4), $p_{(5,4)}$ = 0.6313	0.5	(c5, c4)	(c5, c4)

Figure 3: An example in the test set. The emotion clause set {c3, c5} was obtained in the first turn and the candidate emotion-cause pair set {(c3, c4), (c5, c4)} in the second turn. After using the rethink mechanism, the valid emotion-cause pair set was identified as {(c5, c4)}.

scores all exceeded 79% on the ECPE task. Concretely, Natural QL performs best in the ECPE task. Compared with Pseudo and Structured QL, Natural QL makes it easier for our model to understand the meaning of the queries, because BERT's pre-trained corpus are all natural languages. The Pseudo QL can be regarded as a concise expression of Natural QL. Hence our model can still understand the meaning of queries. Structured QL is the most abstract of three QLs, transforming the three queries ($Q^{se,dc,de}$) into a unified paradigm. It gains the lowest F1 score, which indicates that the inductive ability of our model is not enough to understand abstract structured query language.

5.4 Case Analysis

Figure 3 shows the inference process of MM-R and the advantage of the rethink mechanism through an example in the test set. The candidate emotion-cause pair set {(c3, c4), (c5, c4)} were obtained in the first two turns, and their probabilities were $p_{3,4} = 0.5135$ and $p_{5,4} = 0.6313$. After this, the proposed model adjusted the probabilities by rethinking whether each candidate emotion-cause pair was valid. The bottom of Figure 3 shows that the adjusted probabilities were $p_{3,4} = 0.3595$ and $p_{5,4} = 0.6313$, and the threshold value was 0.5. This means that the answer to the dynamic emotion query *"Is it an emotion clause corresponding to 'but we are too poor to buy meat'?"* was no for clause $c3$ in the third turn. Therefore, candidate emotion-cause pair (c3, c4) was filtered out, while (c5, c4) was preserved.

5.5 Results for Emotion Cause Extraction

The proposed framework is also applied to the emotion cause extraction (ECE) task. Compared with the ECPE task, the ECE task has emotion annotations, and therefore in the MM-R framework, the emotion extraction of the first turn is omitted, and the emotion annotations are used to extract the corresponding cause clauses directly in the second turn. Finally, each candidate emotion-cause pair was verified by the rethink mechanism.

Methods	Emotion Cause Extraction		
	P(%)	R(%)	F1(%)
RTHN	76.97	76.62	76.77
KAG	79.12	75.81	77.43
RHNN	81.12	77.25	79.14
2-step RANKING	80.76	78.45	79.59
MM-R	**83.59**	**83.47**	**83.48**

Table 4: Results on the Emotion Cause Extraction task.

Comparisons are also conducted with recently proposed methods for the ECE task: RTHN (Xia et al., 2019), RHNN (Fan et al., 2019), KAG (Yan et al., 2021) and 2-step RANKING (Xu et al., 2021). Table 4 clearly demonstrates that our proposed framework achieves the state-of-the-art performance on the ECE task.

6 Conclusions

This study transforms the emotion-cause pair extraction (ECPE) task into the machine reading comprehension (MRC) task and proposes a multi-turn MRC with rethink mechanism (MM-R). This structure, which extracts emotions and causes in turn, avoids the label sparsity problem and models the complicated corresponding relations between emotions and causes. In every turn, explicit semantic information can be used effectively. Furthermore, the rethink mechanism verifies each candidate emotion-cause pair by modeling the flow of information from causes to emotions. Experimental results on the ECPE corpus demonstrated the effectiveness of the proposed model.

Acknowledgements

This work was supported by National Key Research and Development Program of China (Grant No. 2020YFC0833402), National Natural Science Foundation of China (Grant Nos. 61976021, U1811262), and Beijing Academy of Artificial Intelligence (BAAI).

References

Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12666–12674.

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.

Christopher Clark, Kenton Lee, Ming-Wei Chang, Tom Kwiatkowski, Michael Collins, and Kristina Toutanova. 2019. Boolq: Exploring the surprising difficulty of natural yes/no questions. *arXiv preprint arXiv:1905.10044*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

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.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-driven emotion cause extraction with corpus construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1639–1649.

Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. Event extraction as machine reading comprehension. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1641–1651.

Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13543–13551.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

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.

Sixing Wu, Fang Chen, Fangzhao Wu, Yongfeng Huang, and Xing Li. 2020. A multi-task learning neural network for emotion-cause pair extraction. In *ECAI 2020*, pages 2212–2219. IOS Press.

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*.

Bo Xu, Hongfei Lin, Yuan Lin, and Kan Xu. 2021. Two-stage supervised ranking for emotion cause extraction. *Knowledge-Based Systems*, page 107225.

Hanqi Yan, Lin Gui, Gabriele Pergola, and Yulan He. 2021. Position bias mitigation: A knowledge-aware graph model for emotion cause 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.

Yukun Zheng, Jiaxin Mao, Yiqun Liu, Zixin Ye, Min Zhang, and Shaoping Ma. 2019. Human behavior inspired machine reading comprehension. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 425–434.

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.

Yang Zhou, Yubo Chen, Jun Zhao, Yin Wu, Jiexin Xu, and Jinlong Li. 2021. What the role is vs. what plays the role: Semi-supervised event argument extraction via dual question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14638–14646.