Conversational Emotion-Cause Pair Extraction with Guided Mixture of Experts

DongJin Jeong
Sungkyunkwan University
Department of Artificial Intelligence
South Korea
jdjim3000@g.skku.edu

JinYeong Bak
Sungkyunkwan University
Department of Artificial Intelligence
South Korea
jy.bak@skku.edu

Abstract

Emotion-Cause Pair Extraction (ECPE) task aims to pair all emotions and corresponding causes in documents. ECPE is an important task for developing human-like responses. However, previous ECPE research is conducted based on news articles, which has different characteristics compared to dialogues. To address this issue, we propose a Pair-Relationship Guided Mixture-of-Experts (PRG-MoE) model, which considers dialogue features (e.g., speaker information). PRG-MoE automatically learns relationship between utterances and advises a gating network to incorporate dialogue features in the evaluation, yielding substantial performance improvement.

Table 1: The amount of emotion-cause pairs in a document compared with the ECPE-news corpus (Xia and Ding, 2019) and ECPE-D corpus. ECPE-D is a dialogue dataset reconstructed based on RECCON (Poria et al., 2021). On average, ECPE-D corpus has more emotion-cause pairs than ECPE-news corpus.

1 Introduction

With increased interest in developing human-like responses, it is crucial to determine the cause of a given emotion. As part of such interest, there is a surge of research activities that analyze the cause of emotions (Yan et al., 2021; Turcan et al., 2021; Li et al., 2022a). Recently, Poria et al. (2021) presents RECCON, a new dataset for Emotion Cause Extraction (ECE) task in dialogue. ECE is a task to find a clause that contains the cause of an annotated emotion in a clause of a given document. However, ECE is limited in that the model requires manually annotated emotions.

To overcome the limitation of ECE (Lee et al., 2010), Xia and Ding (2019) suggest Emotion Cause Pair Extraction (ECPE) task, which automatically predicts emotion clauses in a given document and identifies their corresponding causes. They also build a new ECPE corpus from Chinese news articles. Since ECPE and the dataset were proposed, it has attracted the interest of numerous researchers Ding et al. (2020a,b); Wei et al. (2020); Fan et al. (2020); Cheng et al. (2020); Chen et al. (2020a,b).

However, ECPE in dialogues is different from ECPE in news articles. A dialogue is an interaction between two or more people, while a news article describes a fact. So, dialogues contains meta information such as the speakers, which is one of the most important information in understanding dialogues. In addition, dialogues contain more diverse and emotional expressions, and emotions change as the dialogue progresses, creating even more emotion-cause pairs. This makes the task of ECPE even more challenging. Table 1 shows that most documents in the current ECPE news corpus have only one emotion-cause pair per document, whereas most of the dialogues have multiple emotion-cause pairs. So, we employ RECCON, an English dialogue dataset (Poria et al., 2021) as a new ECPE dataset. We reconstruct RECCON suitable for the ECPE task and we call this dataset ECPE-D. Figure 1 shows an example of ECPE-D.

In this paper, we propose a Pair-Relationship Guided Mixture-of-Experts (PRG-MoE) model, which considers dialogue features (e.g., speaker information) in ECPE. We employ Mixture-Of-Experts (MoE) (Eigen et al., 2014) to customize
Figure 1: Example of ECPE task. \( u_3 \) has a surprise emotion because Human A learns that his colleague is thinking of quitting her job \((u_2)\). Human B feels happy at \( u_6 \) because she thinks her conversation with Human A is interesting \((u_6)\) and has obtained good advice from Human A \((u_5)\). We can extract the emotion-cause pair \((u_3, u_2), (u_6, u_5), (u_6, u_6)\) from this dialogue. Blue line indicates the cause of the emotion utterance.

the experts in the relationship between utterances. It is important to consider the relationship between utterances, since it helps us grasp the emotional flow from the conversation history or understand emotions (or its causes) through other speakers. PRG-MoE automatically learns the relationship between utterances and advises gating networks to incorporate dialogue features, which yields excellent performance. We evaluate PRG-MoE and other models using ECPE-D dataset and show that PRG-MoE outperforms other models.

Furthermore, we propose a multi-class classification task that identifies cause types of emotion-cause pairs in a dialogue - ECPE-CT. An understanding of the cause types is beneficial especially for empathetic response generation (Gao et al., 2021). ECPE-CT helps to generate more specific empathetic responses rather than simple reactions for all kinds of context such as "good luck" by understanding several cause types. Knowing the cause types of the last utterance by the user can help the agent to comprehend the context and obtain the implicit feedback from the user. There are three cause types in ECPE-D: no-context, interpersonal, and self-contagion. Type is categorized depending on 1) from which speaker the cause of the emotion originated, and 2) whether the cause appears in the current utterance or not.

We performed ECPE-CT tasks under various models; PRG-MoE outperformed other baselines in identifying not only emotion-cause pairs but also types of causes in the pairs.

Our contributions are as follows:

- We propose a new ECPE task in dialogues, and provide related dataset - ECPE-D.
- We present PRG-MoE, a new approach that outperforms other models in ECPE-D.
- We propose a new Cause Type Classification task (ECPE-CT) that helps categorizing the type of a cause for an emotion.

2 Related Work

Xia and Ding (2019) propose the ECPE task, which predicts an emotion clause and extracts a corresponding cause in a given document. The authors construct a new ECPE corpus from Chinese news ECE corpus (Gui et al., 2016). They also propose a two-step approach: a pipeline structure consisting of emotion/cause clause extraction and an emotion-cause pairing. However, the two-step approach in a pipeline structure has limitation in that errors cannot be propagated to the entire model. PRG-MoE performs end-to-end pair extraction so it avoids inaccurate inference originating from a pipeline structure.

Subsequent ECPE research suggests 2D transformer (Ding et al., 2020a), sliding window (Ding et al., 2020b) and graph neural network (Chen et al., 2020b; Wei et al., 2020). However, these approaches do not consider meta information such as speakers. Speaker information is a factor that improves performance in dialogue-related tasks. Several studies report improved performance in emotion recognition by taking into account speaker information (Zhang et al. (2019), Li et al. (2020),
Bao et al. (2022)). In addition, methods that consider speakers obtain higher performance in conversation related tasks (Li et al. (2022b), Bak and Oh (2020)). PRG-MoE adopts Mixture-of-Experts for incorporating speaker information as a feature in pair extraction. Also, previous ECPE research considers only the existence of emotion in a clause. PRG-MoE suggests a method where a combination of speaker information and type of emotion in an utterances is used.

3 Task Definition

This section describes the definition of ECPE task in a dialogue. The input is a dialogue $D = \{u_1, ..., u_n\}$ that contains multiple utterances between two people, where $u_i$ is $i$-th utterance in a dialogue and consists of token sequence $t_i$, speaker indicator $s_i \in \{0, 1\}$ and emotion information $e_i$.

Objective of the task is to extract a set of emotion-cause utterance pairs $(..., (u_i, u_j), ...)$, where $u_i$ is an emotion utterance and $u_j$ is a cause utterance in a pair.

4 Pair-Relationship Guided Mixture-of-Experts

We propose a Pair-Relationship Guided Mixture-of-Experts (PRG-MoE) model that adopts a Mixture-of-Experts module. Figure 2 shows the overall architecture. PRG-MoE consists of three modules: utterance representation construction (§4.1), emotion-cause pair candidate extraction (§4.2) and mixture-of-experts based emotion-cause pair classification (§4.3).

4.1 Utterance Representation Construction

First, PRG-MoE creates a representation of each utterance. Input is a dialogue $D = \{u_1, ..., u_n\}$, where $i$-th utterance $u_i = (t_i, s_i)$ contains token sequence $t_i$ and speaker indicator $s_i$. We use BERT (Devlin et al., 2019) for constructing token sequence representation. We surround each token sequence with pre-defined special tokens ([CLS], [SEP]), $t'_i = \{[CLS], w_{i1}, ..., w_{ik}, [SEP]\}$, where $w_{ik}$ is $k$-th token in $i$-th utterance’s token sequence. [CLS] token is used for generating representation for classification tasks. [SEP] token is used to denote the end of a sentence. We obtain the utterance’s representation $h_i$ via BERT, which is the final hidden state of [CLS].

$$h_i = BERT(t'_i)$$  \hspace{1cm} (1)

Emotion Classification PRG-MoE performs emotion classification not only to obtain emotion utterance candidates for emotion-cause pairs, but also to convey emotion information in utterance representation. Emotion classification is performed by feeding a token sequence representation $h_i$ into a Feed-Forward Neural Network (FFNN) layer. We can get the emotion prediction $\hat{e}_i$.

$$\hat{e}_i = \text{Softmax}(W^e h_i + b^e),$$  \hspace{1cm} (2)

where $W^e$ is a weight and $b^e$ is a bias of the emotion classification layer, respectively.

Utterance Representation We use the concatenation of token sequence representation, emotion prediction and speaker information as utterance representation.

$$u_i = h_i \oplus \hat{e}_i \oplus s_i$$  \hspace{1cm} (3)

4.2 Emotion-Cause Pair Candidate Extraction

To extract emotion-cause pair candidates, PRG-MoE needs to pair emotion utterance candidates and cause utterance candidates. The candidate pair $x_{ij}$ is created by concatenating two utterances.

$$x_{ij} = u_i \oplus u_j,$$  \hspace{1cm} (4)

where $u_i$ is a non-neutral emotion utterance representation and $u_j$ is a cause utterance candidate representation ($j \in \{1, ..., i\}$).

We assume two properties of the pairs. First, only non-neutral emotion utterances can be emotion utterance candidates. This is because ECPE tries to find the cause of emotions that occurred during a conversation. Second, PRG-MoE assumes that the cause of an emotion exists in previous or present utterances, since future utterances are not known to the speakers.

Window-constrained Strategy For computing efficiency, we adopt a window-constrained strategy in ECPE task (Ding et al., 2020a). In a given dialogue $D = \{..., u_i, \ldots\}$, where $u_i$ is a non-neutral emotion utterance, $u_j$’s cause candidates $\{u_{i-[w]+1}, ..., u_j\}$ are selected up to the predefined window size $|w|$ distance.

4.3 Mixture-of-Experts based Pair Classification

Pair-Relationship in ECPE In emotion-cause pairs, a specific relationship is formed depending
Figure 2: Overall Architecture of PRG-MoE consists of three parts: utterance representation construction, emotion-cause pair candidate extraction, and mixture-of-experts based pair classification. Token sequences in dialogue are converted into semantic representation by BERT. The concatenation of speaker indicator, token sequence representation, and emotion prediction information is used as an utterance representation. In the emotion-cause pair candidate extraction, only non-neutral utterances are considered to be emotion utterances. Pair candidates are routed to proper experts using routing probability. The routing probability consists of the gating network probability $g_\theta$ and pair-relationship probability $p_{\text{guide}}$.

Guided-MoE Method

We are inspired by the mixture-of-experts (MoE) method to consider pair-relationship. MoE is the process of utilizing multiple experts for a specific task. The expert is a trainable neural network. The gating network determines which expert is suitable for a given input, and this mechanism automatically enhances the expertise of experts through learning.

However, there is no guarantee that the pure MoE learns pair-relationship. So, we guide each expert to have expertise in pair-relationship. We combine the decision of the gating network and on speaker or emotion relationship. We call these relationship pair-relationship. Depending on who the speaker is of an utterance $u_i$, $u_j$ in an emotion-cause utterance pair $x_{ij}$ and what emotion the utterance contains, we can identify the following four categories.

- **Same speaker - Same emotion**: This is a case where the utterances in a pair belong to the same speaker and have the same emotions, such as maintaining emotional state.

- **Same speaker - Different emotion**: This is a case where the utterances in a pair belong to the same speaker but have different emotions. For example, an emotion can appear in the second utterance, following the first utterance where the speaker talks neutrally about what could be the cause of the emotion that arises in the second utterance. Also, the speaker can have multiple emotions occur simultaneously in one utterance, such as ambivalence (Larsen and Mcgraw, 2011). So, the cause of one emotion can trigger different emotions.

- **Different speaker - Same emotion**: This is a case where the utterances in a pair belong to different speakers but share the same emotion, such as sharing the emotion of empathy.

- **Different speaker - Different emotion**: This is a case where the utterances in a pair belong to different speakers and each have different emotions. For example, a speaker’s utterance triggers an emotion in the other speaker’s utterance. This case is similar to the Same speaker - Different emotion case, but the subject of the cause utterance is another speaker.

Pair-relationship is constructed based on the predicted emotions, not the ground-truth emotion information.
pair-relationship for routing emotion-cause pair candidates to proper experts.

MoE consists of $k$ experts $\{f_1, ..., f_k\}$ and gating network $g_0$. Experts and gating network get a set of emotion-cause pair candidates $\{..., x_{ij}, x_{iii}, \ldots\}$ as input. Experts return the emotion-cause pair classification prediction $f_0(x_{ij})$, respectively. Gating network $g_0$ returns the routing probability $g_0(x_{ij})$, where $g_0(x_{ij})$ is a distribution over $k$ experts that sums to 1.

To guide the routing probability to consider pair-relationship, PRG-MoE first creates one-hot label that represents the category of pair-relationship $p_{ij}^{guide}$. PRG-MoE routes an input pair representation $x_{ij}$ by combining $g_0(x_{ij})$ and $p_{ij}^{guide}$. For combination, the number of experts should be the same as the number of pair-relationship.

$$p_{ij} = (1 - \lambda) \times g_0(x_{ij}) + \lambda \times p_{ij}^{guide},$$

where $p_{ij}$ is a distribution over $k$ experts that sums to 1.

The output of PRG-MoE is as follows:

$$y_{ij} = \sum_{n=1}^{k} p_{ij}^{n} f_0^n(x_{ij}),$$

where $k$ is the number of the pair-relationship information as one expert is assigned to each category.

The loss function of PRG-MoE consists of emotion classification loss and emotion-cause pair classification loss that are focal loss (Lin et al., 2017). Essentially, ECPE task faces the challenge of class imbalance since it has a few positive samples among pair candidates. The adoption of the focal loss alleviates this issue by balancing the weight assigned to minority classes, facilitating the learning process (Wang et al., 2022).

5 Experiments

5.1 Settings

Dataset. We use the RECCON dataset (Poria et al., 2021) for experiments. RECCON is a dataset for ECE task that finds a corresponding cause for an utterance with a given emotion. It consists of DailyDialog (Li et al., 2017) and IEMOCAP (Busso et al., 2008), and the authors additionally annotated the cause of an emotion and type of the cause.

We reconstruct a corpus for ECPE in dialogues, named ECPE-D from RECCON. We also call DailyDialog data in ECPE-D as ECPE-D-DD, and IEMOCAP data in ECPE-D as ECPE-D-IE. RECCON has several cause type classes. Among them, there is a cause type called "hybrid" which encompasses both "inter-personal" and "self-contagion" causes. We separate "hybrid" type into "inter-personal" and "self-contagion" and reannotate in dialogues to clarify information associated with the pair label by making multiple single pairs for cause type classification. In addition, RECCON has annotated cause information for each utterance. To facilitate the ECPE task, we add the emotion-cause pair label per dialogue.

Table 2 shows basic statistics of ECPE-D. We split the ECPE-D-DD as 80/10/10 for training/validation/test. We use ECPE-D-IE as test data only. This is because it has fewer dialogues than ECPE-D-DD and we can show the robustness of ECPE models on different domain dataset. IEMOCAP has frustration and excited emotion labels that are not in DailyDialog. So, we map frustration and excited to sad and happy, respectively.

For statistically significant results, we conduct a total of five experiments using randomly split data, and report the average result.

<table>
<thead>
<tr>
<th>Approach</th>
<th>ECPE-D-DD</th>
<th>ECPE-D-IE</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Dialogues</td>
<td>1,106</td>
<td>16</td>
</tr>
<tr>
<td>Avg. of Dialogue length</td>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td># of no-context pair</td>
<td>3,370</td>
<td>243</td>
</tr>
<tr>
<td># of inter-personal pair</td>
<td>3,796</td>
<td>365</td>
</tr>
<tr>
<td># of self-contagion pair</td>
<td>1,958</td>
<td>445</td>
</tr>
<tr>
<td>Avg. of emotion-cause pairs</td>
<td>8</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 2: Characteristics of the ECPE-D Dataset. ECPE-D-IE is used only as test dataset. The two datasets differ significantly in their properties.

Baselines. We select following models as a compared approach. For a fair comparisons, language models of all models are fixed with a bert-base-cased from huggingface2 (Wolf et al., 2020).

- **ECPE-2D** (Ding et al., 2020a) proposes a method of expressing the emotion-cause pairs by a two-dimensional representation scheme. They use a window-constrained method to restrict the scope of the search for extracting emotion-cause pairs.

- **ECPE-MLL** (Ding et al., 2020b) defines an ECPE task as a multi-label learning problem. ECPE-MLL first assumes that all utterances

2https://github.com/huggingface
are emotion utterances, and finds corresponding cause utterances; then, assumes that all utterances are cause utterances, and finds corresponding emotion utterances. There are two ways for the ECPE-MLL to identify emotion-cause pairs; avg-pair and or-pair. The avg-pair method identifies a match as a pair when both the cause and emotion utterances select each other as their match. The or-pair method identifies an emotion-cause pair even if only one side selects the other as their match (i.e. the cause utterance \(e_1\) may identify the emotion utterance \(c_1\) as its match, while \(e_1\) selects a different cause utterance \(c_2\) as its match. There can be two cause-emotion pairs identified through the or-pair method; pair \((e_1, c_1)\) and pair \((c_1, c_2)\).

- **Rank-Emotion-Cause** (Wei et al., 2020) is a method that ranks candidates for emotion-cause pairs and filters them using a sentiment word lexicon. Since the prior lexicon is developed for Chinese data, we adapt it using the Loughran-McDonald sentiment lexicon (Loughran and McDonald, 2011) for testing ECPE-D.

- **RECCON** (Poria et al., 2021) uses RoBERTa with a classification layer for ECPE. They claim that the simple language model outperforms other ECPE models. For fair comparisons, we set the language model to BERT. We denote the model RECCON-BERT.

**Evaluation Metrics** We follow the evaluation metrics from previous research (Xia and Ding, 2019); we use precision, recall, and F1 score as metrics.

5.2 Results

Table 3 shows experimental results for the ECPE task of PRG-MoE and baseline methods with ECPE-D-DD. There is no difference between the PRG-MoE and other models in terms of Emotion Extraction performance. This is because Emotion extraction is performed using only BERT. However, even though it shows similar emotion extraction performance, PRG-MoE outperforms in the pair extraction performance, which is the main goal of our study. The usage of speaker information and type of emotion makes PRG-MoE more suitable to extract emotion-cause pairs in a dialogue than other baselines.

ECPE-MLL predicts emotion-cause pairs utilizing two methods; avg-pair and or-pair. Or-pair shows better recall and f1-score than avg-pair. This is because or-pair predicts the emotion-cause pair optimistically, whereas avg-pair satisfies two indicators for predicting pairs. It gives or-pair a wider search space than avg-pair, making the or-pair more advantageous in finding pairs in emotionally-rich environment and thus to have better performance than avg-pair.

Rank-Emotion-Cause shows low performance. It has two stages for extracting pairs. First, Rank-Emotion-Cause ranks pair candidates and chooses the first pair. Second, the model determines if there is a sentiment word among the other unselected candidate pairs using a sentiment word dictionary and when found, selects it. However, there are many cases where emotions are expressed without explicit expression of emotions in an utterance. Dictionary-based emotion detection makes it hard to capture implicit expression. Dictionary-based emotion detection of Rank-Emotion-Cause captures 1,764 utterances out of 6,384 emotion ut-
Table 4: A study on mixing ratio for routing probability and *pair-relationship* information in guided-MoE method. When $\lambda$ is 0, it means the pure Mixture-of-Experts method and when $\lambda$ is 1, it means the pure guide information from *pair-relationship*. Above experiments prove that guided-MoE method is superior to pure MoE or pure guide for experts.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Speaker+Emotion Guide</th>
<th>Emotion Guide</th>
<th>Speaker Guide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>F1-Score</td>
</tr>
<tr>
<td>0.0 (Pure MoE)</td>
<td>51.63</td>
<td>55.98</td>
<td>53.72</td>
</tr>
<tr>
<td>0.2</td>
<td>53.25</td>
<td>54.98</td>
<td>54.10</td>
</tr>
<tr>
<td>0.4</td>
<td>56.49</td>
<td>55.20</td>
<td>55.84</td>
</tr>
<tr>
<td>0.6</td>
<td>58.95</td>
<td>55.67</td>
<td><strong>57.26</strong></td>
</tr>
<tr>
<td>0.8</td>
<td>55.53</td>
<td><strong>57.14</strong></td>
<td>56.32</td>
</tr>
<tr>
<td>1.0 (Pure guide)</td>
<td>58.07</td>
<td>54.64</td>
<td>56.30</td>
</tr>
</tbody>
</table>

Table 5: Ablation study for utterance representation components. It is performed by PRG-MoE. $h$, $\hat{e}$ and $s$ mean token sequence representation, emotion prediction and speaker indicator, respectively. Above results show that giving information through concatenation has a positive effect.

Table 6: Ablation study of window-constraint method.

6 Discussion

6.1 Effects on Pair-Relationship Information

The main idea of PRG-MoE is to combine the decision of the gating network and *pair-relationship* information. For evaluating effects on the mixing ratio $\lambda$ in Guided-MoE, we set $\lambda$ from 0 (pure Mixture-of-Experts) to 1 (pure guiding *pair-relationship* information). Table 4 shows the proper

![Figure 3: Ablation study for the number of experts in the mixture-of-experts method. We set PRG-MoE with $\lambda = 0$ for experimenting pure MoE method. Above experiments show no meaningful difference without the *pair-relationship* probability $p^{guide}$.](image-url)

![Table 5: Ablation study for utterance representation components. It is performed by PRG-MoE. $h$, $\hat{e}$ and $s$ mean token sequence representation, emotion prediction and speaker indicator, respectively. Above results show that giving information through concatenation has a positive effect.](image-url)

![Table 6: Ablation study of window-constraint method.](image-url)
Table 7: ECPE-CT performance of models for ECPE-D-DD. The performance between PRG-MoE and ECPE-2D is insignificant in no-context. However, there are notable advances in inter-personal and self-contagion, which relate to the interaction between the two utterances. Through the above performances, we can prove that the Guided-MoE method is helpful in judging the relationship between different utterances. We run the statistical significance test for the F1-Score in the ECPE-CT task. PRG-MoE shows statistically significant difference in “inter-personal” performance than ECPE-2D (**: p < 0.0001).

<table>
<thead>
<tr>
<th>Approach</th>
<th>no-context</th>
<th>inter-personal</th>
<th>self-contagion</th>
<th>weighted average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>F1-Score</td>
<td>Precision</td>
</tr>
<tr>
<td>ECPE-2D</td>
<td>53.23</td>
<td>60.88</td>
<td>56.80</td>
<td>51.70</td>
</tr>
<tr>
<td>PRG-MoE</td>
<td>54.66</td>
<td>62.30</td>
<td>58.23</td>
<td>62.22</td>
</tr>
</tbody>
</table>

mixing ratio in the guided-MoE. PRG-MoE has the highest performance when $\lambda$ is 0.6.

Furthermore, we experiment ablation study of pair-relationship information; emotion-guide and speaker-guide. Emotion-guide method constructs pair-relationship into same emotion and different emotion. Speaker-guide method constructs pair-relationship into same speaker and different speaker.

In the comparison of emotion-guide and speaker-guide, speaker information has better guidance for pair extraction than emotion information. All three pair-relationship experiments show similar tendency that the combination of MoE and pair-relationship information performs better than pure methods.

We also test the effects on the number of experts. We set four experts and assigned one category each to learn the pair-relationship information. Note, however, pure MoE can have multiple experts. Figure 3 shows that even when we change the number of experts, PRG-MoE outperforms all pure MoE models.

6.2 Effects on Elements of Utterance Representation

Table 5 shows the performance of PRG-MoE trained with various cases of utterance representation concatenation. The components for concatenation are token sequence representation, emotion prediction, and speaker indicator.

It shows the best performance when all components are concatenated. Providing extra information allows experts to learn more about features between utterances.

6.3 Effects on Window-Constrained method

Table 6 shows the ablation study of the window-constrained method. In natural conversation, the cause of an emotion generally exists near the emotion utterance (Kumar et al., 2022). PRG-MoE focuses on utterances near an emotion utterance through the window strategy, and does not consider utterances far from the emotion utterance because the farther away from the emotion utterance, the less likely an utterance becomes a cause utterance for that emotion utterance. PRG-MoE shows significant improvement in performance with the window-constrained method.

7 Cause Type Classification in ECPE

This section describes the experiments and results of classifying the cause type of emotions - ECPE-CT. The cause type in ECPE-CT is categorized based on from which speaker the cause is generated and where the cause is found in the paired utterances. ECPE-CT enhances understanding of the cause of the emotion, enabling more effective use of the cause. For example, most chatbots used in different settings generate responses based on the found cause (Gao et al., 2021). However, ECPE-CT could assist to generate more empathetic responses by incorporating other cases, as categorized below.

Cause types in ECPE-D are as follows:

- **No-context** indicates that an emotion and its cause are found in one utterance.
- **Inter-personal** signifies that the cause exists in the other person’s utterances.
- **Self-contagion** refers to the situation where the cause exists in the prior utterance of the same speaker.
- **Latent** refers to a scenario in which the cause does not exist or may occur in the future. The latent type naturally has no pair information, so we classified it as having no pair.

We test PRG-MoE and ECPE-2D, which have best performances among the baselines. For multiclass classification, we modify the output layer of
models to be able to output multi-class prediction. Table 7 shows the performance for each cause type. The performance difference between PRG-MoE and ECPE-2D is not significant in "no-context" and "self-contagion". However, there is a significant advance in "inter-personal", which relates to the interaction between the two different speakers. We believe that the Guided-MoE approach performs the function of an identifier, confirming the speaker of a paired utterance.

8 Conclusion

In this paper, we present PRG-MoE, a novel approach for extracting emotion-cause pairs from a dialogue by considering speaker and emotion information. We guide the mixture-of-experts module to consider the relationship between utterances in pairs. To guide mixture-of-experts, we define pair-relationship, which is the relationship between utterances. We combine the decision of the gating network and pair-relationship information for routing the emotion-cause pairs to proper experts. We also propose a new task, ECPE-CT, which classifies emotion-cause pair by cause type. We evaluate the task with ECPE-D, a dialogue dataset with more emotion-cause pairs than other benchmark ECPE datasets. With ECPE-D, we show that PRG-MoE outperforms other ECPE models in ECPE and Multi-class ECPE tasks.

Limitations

First, we limit the scope of cause to be found in one conversation. However, the actual cause of an emotion may come from other sources outside the given conversation, such as news, weather, and the speakers’ previous conversations. But, the ECPE-D dataset does not have such external information, and there are no multiple conversations by the same speaker pairs. Second, we encode the speakers as 0 or 1 since we do not know about the speakers and their relationships. However, emotional conversations would occur more frequently in close relationships such as between family members and friends. Third, we do not test with multi-party conversations. We will experiment with multi-party conversations by annotating the emotion-cause label to another multi-party conversation dataset (e.g., MELD (Poria et al., 2019)). Lastly, we do not consider the order of emotion-cause pairs in a conversation. The order might be helpful in modeling the emotion-cause pairs. For example, if a speaker’s emotional state remains unchanged throughout a conversation, a previous pair can help predict a future pair.

Ethics Statement

This paper presents a new ECPE method, PRG-MoE, which extracts emotion and their corresponding cause in a dialogue through the relationship between utterances in pairs. PRG-MoE shows high performance in extracting emotion-cause pairs in a conversation. In this regard, PRG-MoE could be deployed in cause extraction in a dialogue and other real-world applications. We do not report any data collection process in this paper, as we experiment with an open-domain dataset. We experiment with the dialogue dataset based on RECCON (Poria et al., 2021). RECCON is publicly available, and there is no ethical issue.

Acknowledgments

We would like to thank the anonymous reviewers for helpful questions and comments. This work was partly supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.2019-0-00421, AI Graduate School Support Program(Sungkyunkwan University)), the Technology Innovation Program (or Industrial Strategic Technology Development Program-Source Technology Development and Commercialization of Digital Therapeutics) (20014967, Development of Digital Therapeutics for Depression from COVID19) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Ministry of Science and ICT(MSIT) to the National Research Foundation of Korea (NRF) [NRF-2021R1A4A3033128].

References


Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S.


Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. 2022b. Knowledge bridging for empathetic dialogue generation.


A Implementation Details

We use pretrained bert-base-cased from hugging-face (Wolf et al., 2020) as a language model. We train PRG-MoE using Adam optimizer (Kingma and Ba, 2015) for 40 epochs and decay the learning rate exponentially for each epoch. The decay rate is 0.05. The batch contains 5 dialogue documents, and the learning rate is set to 5e-5. Dropout is applied to utterance representation with a 0.5 rate. We set the window size as 3 since we follow the previous work for fair comparisons (Ding et al., 2020a). The final loss weight \( \lambda_{emo} \) and \( \lambda_{par} \) are set to 0.2 and 0.8, respectively. We choose the hyperparameters by manual tuning. We select the hyperparameter based on the f1 score performance. We selected parameter related to data characteristics such as window-size through experiments, and compared baseline models using the same parameter.

Our hardware setting is Intel(R) Xeon(R) Gold 5218R CPU @ 2.10GHz (CPU), and NVIDIA RTX A6000 (GPU). The average running time of PRG-MoE per one epoch is 3min 20s. The inference time per one batch is 1.2 sec. The number of parameters of PRG-MoE is 110M.