Conversational Emotion-Cause Pair Extraction with Guided Mixture of Experts

DongJin Jeong
Sungkyunkwan University
Department of Artificial Intelligence
South Korea
jdjin3000@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. We employ a new ECPE dataset, which is an English dialogue dataset, with more emotion-cause pairs in documents than news articles. We also propose Cause Type Classification that classifies emotion-cause pairs according to the types of the cause of a detected emotion. For reproducing the results, we make available all our code and data[1].

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).

# of Emotion-Cause Pairs	ECPE-news (Xia and Ding, 2019)	ECPE-D
1	1,746 (**89.77%**)	9 (**0.80%**)
2	177 (**9.10%**)	19 (**1.69%**)
≥ 3	22 (**1.13%**)	1,094 (**97.50%**)

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.

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

[1] https://github.com/jdjin3000/PRG-MoE

Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pages 3288–3298

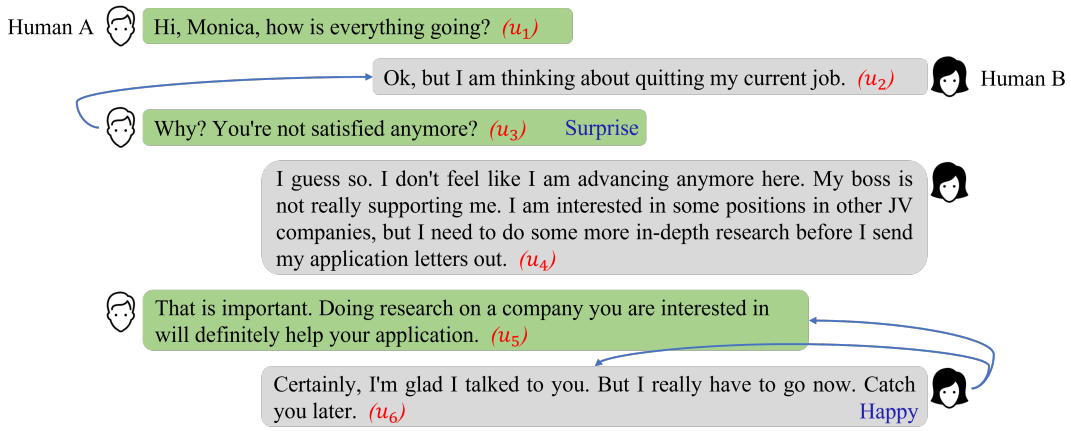

Figure 1: Example of ECPE task. u_3 has a *surprise* emotion because Human A learns that his colleague is thinking of quitting her job (u_2). Human B feels happy at u_6 because she thinks her conversation with Human A is interesting (u_6) and has obtained good advice from Human A (u_5). We can extract 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) into 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 considers 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].

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

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 $\hat{e}_i$.

$$\hat{e}_i = Softmax(W^e h_i + b^e), \tag{2}$$

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.

$$u_i = h_i \oplus \hat{e}_i \oplus s_i \tag{3}$$

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.

$$x_{ij} = u_i \oplus u_j, \tag{4}$$

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 $D = \{..., u_i, ...\}$, where u_i is a non-neutral emotion utterance, u_i's cause candidates $\{u_{i-|w|+1}, ..., u_i\}$ 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

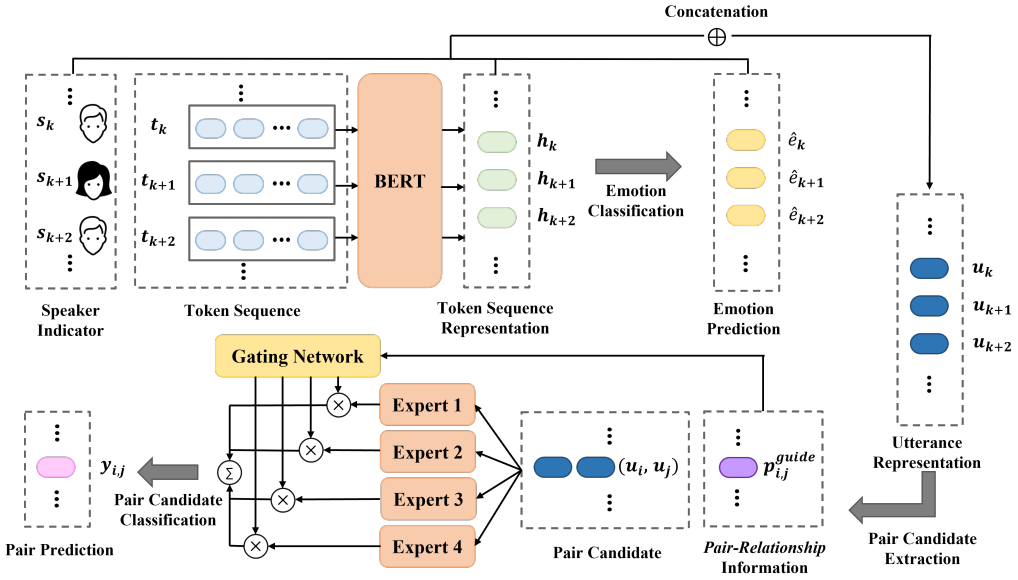

Figure 2: Overall Architecture of PRG-MoE consists of three parts: utterance representation construction, emotion-cause pair candidate extraction, and mixture-of-experts based pair classification. Token sequences in dialogue are converted into semantic representation by BERT. The concatenation of speaker 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_θ and *pair-relationship* probability p^{guide}.

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.

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

pair-relationship for routing emotion-cause pair candidates to proper experts.

MoE consists of k experts $\{f_\theta^1, ..., f_\theta^k\}$ and gating network g_θ. Experts and gating network get a set of emotion-cause pair candidates $\{..., x_{ij}, x_{ii}, ...\}$ as input. Experts return the emotion-cause pair classification prediction $f_\theta(x_{ij})$, respectively. Gating network g_θ returns the routing probability $g_\theta(x_{ij})$, where $g_\theta(x_{ij})$ 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 $g_\theta(x_{ij})$ and p_{ij}^{guide}. For combination, the number of experts should be the same as the number of *pair-relationship*.

$$p_{ij} = (1 - \lambda) \times g_\theta(x_{ij}) + \lambda \times p_{ij}^{guide}, \quad (5)$$

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

The output of PRG-MoE is as follows:

$$y_{ij} = \sum_{n=1}^{k} p_{ij}^n f_\theta^n(x_{ij}), \quad (6)$$

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.

Approach	ECPE-D-DD	ECPE-D-IE
# of Dialogues	1,106	16
Avg. of Dialogue length	10	42
# of *no-context* pair	3,370	243
# of *inter-personal* pair	3,796	365
# of *self-contagion* pair	1,958	445
Avg. of emotion-cause pairs	8	66

Table 2: Characteristics of the ECPE-D Dataset. ECPE-D-IE is uses 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 huggingface[2] (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

[2] https://github.com/huggingface

Test Dataset	ECPE-D-DD						ECPE-D-IE					
Approach	Emotion-Cause Pair Extraction			Emotion Extraction			Emotion-Cause Pair Extraction			Emotion Extraction		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
ECPE-2D	49.34	47.37	48.34**	71.91	**77.16**	**74.44**	45.17	12.37	19.42**	78.16	**49.16**	**60.36**
ECPE-MLL (avg-pair)	53.58	39.58	45.53**	68.97	73.64	71.23	47.44	1.63	3.15**	83.82	6.85	12.66
ECPE-MLL (or-pair)	50.71	43.56	46.86**	68.97	73.64	71.23	43.41	2.17	4.13**	83.82	6.85	12.66
Rank-Emotion-Cause	58.09	10.49	17.77**	**75.85**	20.17	31.87	**65.91**	1.10	2.09**	**91.75**	3.24	6.26
RECCON-BERT	49.31	33.19	39.68*	-	-	-	46.52	4.33	7.92*	-	-	-
PRG-MoE	**58.95**	**55.67**	**57.26**	71.76	76.09	73.86	51.95	**20.02**	**28.90**	85.58	43.06	57.29

Table 3: Performance of PRG-MoE and baseline models for ECPE-D. All models are trained with only ECPE-D-DD. PRG-MoE outperforms all other models in the emotion-cause pair extraction. We also test models in ECPE-D-IE to validate the models in an environment different from the train dataset. Despite the differences, PRG-MoE outperforms all other models. We run the statistical significance test for the F1-Score in the Emotion-Cause Pair Extraction task. PRG-MoE shows statistically significant difference in performance than baselines (**: $p < 0.0001$, *: $p < 0.001$).

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 c_1 may identify the emotion utterance e_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 (e_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-

λ	Speaker+Emotion Guide			Emotion Guide			Speaker Guide		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
0.0 (Pure MoE)	51.63	55.98	53.72	52.07	54.79	53.40	52.07	54.79	53.40
0.2	53.25	54.98	54.10	52.13	**55.33**	53.68	53.78	56.28	55.00
0.4	56.49	55.20	55.84	54.09	53.92	54.00	54.40	**58.95**	**56.58**
0.6	**58.95**	55.67	**57.26**	54.09	53.56	53.82	56.31	56.34	56.32
0.8	55.53	**57.14**	56.32	53.92	54.61	**54.27**	**58.53**	53.99	56.17
1.0 (Pure guide)	58.07	54.64	56.30	**54.16**	52.91	53.52	54.77	58.40	56.53

Table 4: A study on mixing ratio for routing probability and *pair-relationship* information in guided-MoE method. When λ is 0, it means the pure Mixture-of-Experts method and when λ 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.

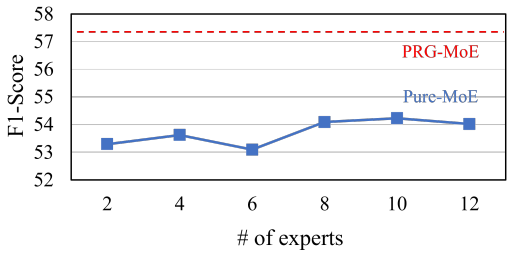

Figure 3: Ablation study for the number of experts in the mixture-of-experts method. We set PRG-MoE with $\lambda = 0$ for experimenting pure MoE method. Above experiments show no meaningful difference without the *pair-relationship* probability p^{guide}.

Concatenated element	Emotion-Cause Pair Extraction		
	Precision	Recall	F1-Score
h	58.27	55.82	57.02
$h \oplus \hat{e}$	56.91	55.88	56.39
$h \oplus s$	55.41	**59.04**	57.16
$h \oplus \hat{e} \oplus s$	**58.95**	55.67	**57.26**

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.

Approach	Emotion-Cause Pair Extraction		
	Precision	Recall	F1-Score
w/o window-constraint	43.91	43.08	43.49
with window-constraint	**58.95**	**55.67**	**57.26**

Table 6: Ablation study of window-constraint method.

terances in ECPE-D. This method is unfavorable for extracting multiple pairs.

RECCON argues that the language model with a simple classification layer outperforms earlier ECPE models. They experiment with emotion labels, candidate pairs, and dialogue context as input. However, there are two problems here: First, they exclude a scenario in which emotional information is incorrect. So, the comparison is not fair because ECPE approaches use emotional information that they predict. Second, while ECPE should consider all possible pairs for a given text, RECCON is tested using a dataset in which positive and negative samples are appropriately mixed. We re-evaluate the model by presenting the same emotion label predicted by PRG-MoE to RECCON for accurate comparison. RECCON-BERT shows low performance compared to when it has true emotion labels. This means that RECCON-BERT, unlike other ECPE models, does not have a structure that operates robustly with an inaccurate emotion.

Table 3 also shows the experimental results for ECPE-D-IE. All models are learned with ECPE-D-DD. This experiment is conducted to evaluate how they perform on out-of-domain data. Besides, as shown in Table 2, ECPE-D-IE has about eight times more emotion-cause pairs in a dialogue than the trained data, so the difficulty of pair extraction becomes extremely high. These adverse conditions make the models have poor performances. PRG-MoE shows robust performance compared to other models.

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 λ in Guided-MoE, we set λ from 0 (pure Mixture-of-Experts) to 1 (pure guiding *pair-relationship* information). Table 4 shows the proper

Approach	no-context			inter-personal			self-contagion			weighted average		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
ECPE-2D	53.23	60.88	56.80	51.70	42.18	46.46**	36.10	14.72	20.92	48.98	43.42	46.03**
PRG-MoE	**54.66**	**62.30**	**58.23**	**62.22**	**59.11**	**60.63**	**39.99**	**17.19**	**24.05**	**55.43**	**51.34**	**53.31**

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 (**: $p < 0.0001$).

mixing ratio in the guided-MoE. PRG-MoE has the highest performance when λ 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 multi-class 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

JinYeong Bak and Alice Oh. 2020. Speaker sensitive response evaluation model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6376–6385, Online. Association for Computational Linguistics.

Yinan Bao, Qianwen Ma, Lingwei Wei, Wei Zhou, and Song Hu. 2022. Speaker-guided encoder-decoder framework for emotion recognition in conversation. *ArXiv*, abs/2206.03173.

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

Narayanan. 2008. Iemocap: interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4):335–359.

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, Barcelona, Spain (Online). International Committee on Computational Linguistics.

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, Barcelona, Spain (Online). International Committee on Computational Linguistics.

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, Barcelona, Spain (Online). International Committee on Computational Linguistics.

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, Minneapolis, Minnesota. Association for Computational Linguistics.

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, Online. Association for Computational Linguistics.

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, Online. Association for Computational Linguistics.

David Eigen, Marc'Aurelio Ranzato, and Ilya Sutskever. 2014. Learning factored representations in a deep mixture of experts. *CoRR*, abs/1312.4314.

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, Online. Association for Computational Linguistics.

Jun Gao, Yuhan Liu, Haolin Deng, Wei Wang, Yu Cao, Jiachen Du, and Ruifeng Xu. 2021. Improving empathetic response generation by recognizing emotion cause in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 807–819, Punta Cana, Dominican Republic. Association for Computational Linguistics.

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, Austin, Texas. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Shivani Kumar, Anubhav Shrimal, Md. Shad Akhtar, and Tanmoy Chakraborty. 2022. Discovering emotion and reasoning its flip in multi-party conversations using masked memory network and transformer. *Knowl. Based Syst.*, 240:108112.

Jeff Larsen and A. Peter Mcgraw. 2011. Further evidence for mixed emotions. *Journal of personality and social psychology*, 100:1095–110.

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, Los Angeles, CA. Association for Computational Linguistics.

Jiangnan Li, Fandong Meng, Zheng Lin, Rui Liu, Peng Fu, Yanan Cao, Weiping Wang, and Jie Zhou. 2022a. Neutral utterances are also causes: Enhancing conversational causal emotion entailment with social commonsense knowledge. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4209–4215. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Jingye Li, Donghong Ji, Fei Li, Meishan Zhang, and Yijiang Liu. 2020. HiTrans: A transformer-based context- and speaker-sensitive model for emotion detection in conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4190–4200, Barcelona, Spain (Online). International Committee on Computational Linguistics.

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

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings*

of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 986–995, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007.

Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

Soujanya Poria, Navonil Majumder, Devamanyu Hazarika, Deepanway Ghosal, Rishabh Bhardwaj, Samson Yu, Romila Ghosh, Niyati Chhaya, Alexander F. Gelbukh, and Rada Mihalcea. 2021. Recognizing emotion cause in conversations. *Cogn. Comput.*, 13:1317–1332.

Elsbeth Turcan, Shuai Wang, Rishita Anubhai, Kasturi Bhattacharjee, Yaser Al-Onaizan, and Smaranda Muresan. 2021. Multi-task learning and adapted knowledge models for emotion-cause extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3975–3989, Online. Association for Computational Linguistics.

Cheng Wang, Georges Balazs, Gyuri Szarvas, Patrick Ernst, Lahari Poddar, and Pavel Danchenko. 2022. Calibrating imbalanced classifiers with focal loss: An empirical study. In *EMNLP 2022*.

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, Online. Association for Computational Linguistics.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

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, Florence, Italy. Association for Computational Linguistics.

Hanqi Yan, Lin Gui, Gabriele Pergola, and Yulan He. 2021. Position bias mitigation: A knowledge-aware graph model for emotion cause extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3364–3375, Online. Association for Computational Linguistics.

Dong Zhang, Liangqing Wu, Changlong Sun, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5415–5421. International Joint Conferences on Artificial Intelligence Organization.

A Implementation Details

We use pretrained *bert-base-cased* from huggingface (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 λ_{emo} and λ_{pair} 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.