Global-Local Modeling with Prompt-Based Knowledge Enhancement for Emotion Inference in Conversation

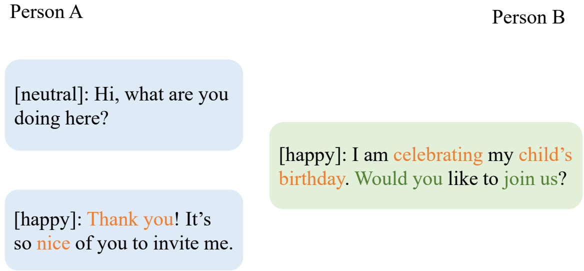
Renxi Wang and **Shi Feng**
School of Computer Science and Engineering
Northeastern University
Shenyang, China
`realreasonwang@gmail.com`, `fengshi@cse.neu.edu.cn`

Abstract

The ability to recognize emotions in conversations is necessary and important for the online chatbot to do tasks such as empathetic response generation and emotional support. Present researches mainly focus on recognizing emotions through a speaker's utterance, while research on emotion inference predicts emotions of addressees through previous utterances. Because of the lack of the addressee's utterance, emotion inference is more challenging than emotion recognition. In this paper, we propose a global-local modeling method based on recurrent neural networks (RNN) and pre-trained language models (PLM) to do emotion inference, which utilizes the sequence modeling ability of RNNs and abundant knowledge from PLMs. Moreover, we take the whole dialogue history as input of PLM to generate knowledge by in-context learning. Experimental results show that our model with knowledge enhancement achieves state-of-the-art performance on all three datasets.[1]

1 Introduction

The task of emotion recognition in conversation (ERC) (Poria et al., 2019b) aims to identify emotion labels of an utterance, where the whole dialogue history along with the current utterance is given. However, in emotion inference in conversation (EIC) the current utterance is lacking but the dialogue history and the current addressee are known (Li et al., 2021a). For example, Figure 1 shows a conversation between A and B. In the third turn, ERC detects A's emotion using all available information while EIC predicts A's emotion using all the information except the last utterance. ERC is a popular task that has been explored widely and deeply, while EIC is a new task that measures the emotion understanding ability of models from a different perspective.



[neutral]: Hi, what are you doing here?

[happy]: I am celebrating my child's birthday. Would you like to join us?

[happy]: Thank you! It's so nice of you to invite me.

Figure 1: A conversation between A and B. The text in orange might be helpful for ERC. The text in green might be helpful for EIC.

In ERC, some previous works utilize sequence-based neural networks to model the context and speaking parties (Majumder et al., 2019; Hu et al., 2021a; Li et al., 2021a). These approaches first finetune a model on utterances to classify emotions. Then this model is used to extract features of utterances. As shown in Figure 1, B makes A happy because he/she is inviting A to join in a party. However, B feels happy because he/she is celebrating his/her child's birthday. The finetuning tends to keep the semantics that is helpful to classify the current emotion and the feature extraction compresses the utterance's information, which may cause information loss that is valuable to infer A's emotion. Some works model dialogues at utterance-level with graph-based models (Ghosal et al., 2019, Shen et al., 2021). They have the same problem as sequence-based models. Also, it becomes difficult for them to distinguish similar emotions as their layers deepen (Li et al., 2022). Pre-trained language models do not need to do the feature extraction process and they contain knowledge suitable for EIC. However, these models can not naturally process sequential utterances from different parties. Motivated by this, we propose global-local modeling method to combine different abilities from these models. Specifically, we use a sequence-based model to get the representation

[1] The code is available at `https://github.com/Reason-Wang/DialogueGLP`.

Findings of the Association for Computational Linguistics: EACL 2023, pages 2120–2127
May 2-6, 2023 ©2023 Association for Computational Linguistics

of the dialogue history and a pre-trained model to process utterances that are close to the addressee's turn in which his/her emotion is to be inferred. In our framework, the global representation and local utterances can attend to each other, which we believe is helpful for EIC.

Some researchers introduce external knowledge to improve the performance of emotion detection (Ghosal et al., 2020, Li et al., 2021b). They generate commonsense knowledge using COMET (Bosselut et al., 2019) which is trained on ATOMIC (Sap et al., 2019). However, this knowledge is limited to certain event types. Also, it is generated based on a single utterance instead of the whole dialogue, which further limits the quality of the knowledge. Recent advancements of in-context learning (Liu et al., 2022) show that it is possible to generate high-quality knowledge when language models are provided with appropriate examples. Based on the above analysis, we propose a knowledge generation method specially designed for EIC task based on prompt learning. Specifically, we use templates to obtain two kinds of knowledge: I. We let GPT fill the dialogue, thus we get pseudo utterances that may be spoken by the addressee and take them as knowledge. II. We ask GPT how the addressee feels and take generated texts as knowledge. Our knowledge is more precise since we take the whole dialogue history as input. Also, it is more diverse because GPT is trained on a large number of texts in different fields.

2 Methods

2.1 Problem Definition

Given a dialogue $D = [\ (U_1, p_1)\ ,\ (U_2, p_2)\ ,\ \cdots\ ,\ (U_m, p_m)\ ,\ p_{m+1}]$, where U_i is the utterance in i-th turn and p_i is the participant in i-th turn. For $i = m + 1$, p_i is the addressee, otherwise p_i is the speaker. The task is to predict the addressee's emotion e using D.

2.2 Global Model

We use DialogueInfer (Li et al., 2021a) as our global model. We first finetune a RoBERTa-Large (Liu et al., 2019) model to predict the emotion label of utterances as ERC task. Then we use the finetuned model to extract features of utterances and get a 1024-dimensional vector u_i for each utterance U_i. These representations of utterances are then put into DialogueInfer to get the representation of the dialogue. DialogueInfer is a model designed for the EIC task. It adopts addressee-aware modules to capture the persistence and contagiousness of utterances. Formally, the output of the global model can be defined as:

$$\begin{aligned} h_t, c_t = &\mathbb{1}\{p_t = p_{m+1}\} LSTM_a(u_t, (h_{t-1}, c_{t-1})) \\ &+ \mathbb{1}\{p_t \neq p_{m+1}\} LSTM_o(u_t, (h_{t-1}, c_{t-1})) \end{aligned}$$

$$h_g = h_{m+1} \tag{1}$$

where $t = 1, 2, \cdots, m$ is the turn step, $\mathbb{1}\{\texttt{condition}\}$ is the indicator function and returns 1 if the condition is true otherwise 0, $h_t \in \mathbb{R}^{d_1}$ and $c_t \in \mathbb{R}^{d_1}$ are hidden state and cell state respectively, d_1 is the hidden dimension in LSTM unit, h_g is the global representation of the dialogue. The final output $h_g \in \mathbb{R}^{d_1}$ is then fed into the local model.

2.3 Local Model

We employ RoBERTa (Liu et al., 2019) as the local model. RoBERTa shares the same architecture as BERT (Devlin et al., 2019) and is trained with masked language modeling objective function. We concatenate the last k utterances to form the input. To make the local model addressee-aware as global model, we prepend a speaker prefix to indicate whether the utterance comes from the addressee. The final text input is:

$$\begin{aligned} U_t = &\text{prfix}(p_{m-k+1}) U_{m-k+1} \texttt{</s>} \text{prefix}(p_{m-k+2}) \\ &U_{m-k+2} \texttt{</s>} \cdots \text{prfix}(p_m)\, U_m \end{aligned} \tag{2}$$

$$\text{prfix}(p_i) = \begin{cases} \texttt{"I:"}, & p_i = p_{m+1} \\ \texttt{"Other:"}, & p_i \neq p_{m+1} \end{cases} \tag{3}$$

where `</s>` is the special token that indicates the separation of utterances.

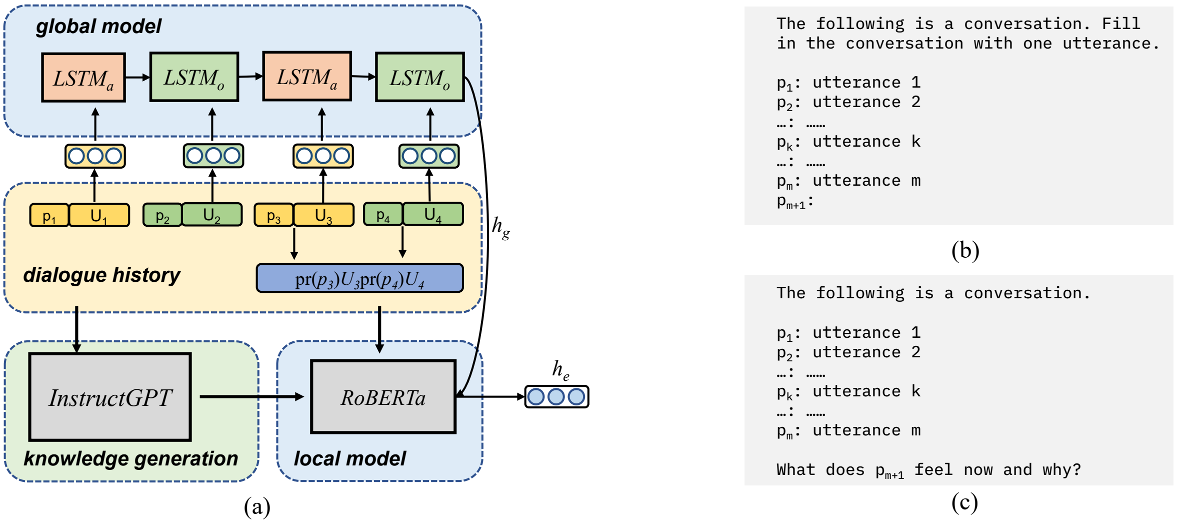
To fuse the global information, we add the global representation h_g to the first token's embedding of the text input. The whole process can be formulated as:

$$\hat{h}_g = W^T h_g + b \tag{4}$$

$$H = [\text{Emb}(U_t[0]) + \hat{h}_g; \text{Emb}(U_t[1:])] \tag{5}$$

$$h_e = \text{RoBERTa-Model}(H) \tag{6}$$

where $W \in \mathbb{R}^{d_1 \times d_2}$ is the matrix to project dimensions, d_2 is the hidden dimension in RoBERTa model, Emb is the embedding layer of RoBERTa.



```
The following is a conversation. Fill
in the conversation with one utterance.

p_1: utterance 1
p_2: utterance 2
…: ……
p_k: utterance k
…: ……
p_m: utterance m
p_m+1:
```

(b)

```
The following is a conversation.

p_1: utterance 1
p_2: utterance 2
…: ……
p_k: utterance k
…: ……
p_m: utterance m

What does p_m+1 feel now and why?
```

(c)

Figure 2: (a) Our framework to infer the emotion. (b) The template to generate pseudo utterances. (c) The template to generate feelings and corresponding reasons.

2.4 Prompt Based Knowledge Generation

GPT-3 (Brown et al., 2020) is a powerful model which generates informative and accurate texts when provided with appropriate examples. The model is further finetuned to align with users so the outputs are more truthful and less toxic (Ouyang et al., 2022). We use the resulting model, called InstructGPT, to generate two kinds of knowledge.

Pseudo Utterances We take the dialogue history as input and let InstructGPT generate the utterance that might be spoken by the addressee. Figure 2 shows the template to generate pseudo utterances. After obtaining these knowledge texts, we first prepend the addressee prefix to them. Then we append them to the text input in the local model.

Feelings and Corresponding Reasons Since InstructGPT is able to do many tasks, we ask InstructGPT directly about the addressee's emotions and corresponding reasons. The output from the model is taken as knowledge and is used the same way as pseudo utterances. Figure 2 shows the template to generate this kind of knowledge.

2.5 Classifier

We use the first token's representation h_e as the final output. A softmax layer is employed after a linear projection layer:

$$p_e = \mathrm{softmax}(W^T h_e + b) \tag{7}$$

where $W \in \mathbb{R}^{d_2 \times c}$ is the projection matrix, c is the number of emotions, $p_e \in \mathbb{R}^c$ is the probability distribution over different emotions.

3 Experiments

3.1 Implementation

We train our model on three datasets: DailyDialog (Li et al., 2017), MELD (Poria et al., 2019a) and EmoryNLP (Zahiri and Choi, 2018). We first finetune a RoBERTa-Large model on the training set of each dataset. The batch size is set to 16 and the model with the best performance on the development set is saved. We then use this model to extract features of the datasets. For emotion inference, we set the learning rate to 1e-5. AdamW is used as the optimizer to update parameters. In the first two epochs, we only update the global model and freeze the local model. After that, we finetune the whole model. We find this updating scheme makes training more stable.

For other baselines, we adapted their official codes to make them applicable to EIC task. The parameters we used in training refered to their original paper. We select all the models based on their best performance on the development set. We use cross entropy as the loss function.

3.2 Main Results

Table 1 shows the main results of our experiments. Our base model without knowledge augmentation already performs best on DailyDialog and MELD. In most cases, the generated knowledge improves the performance. However, knowledge U (pseudo utterances) decreases the performance measured by macro F1 on DailyDialog. COMET decreases the performance measured by macro F1 on DailyDia-

Model	DailyDialog		MELD		EmoryNLP	
	macro F1	weighted F1	macro F1	weighted F1	macro F1	weighted F1
DialogueRNN*	36.28	71.67	16.77	33.96	17.62	20.49
DialogueCRN*	33.82	72.23	16.00	35.44	16.92	21.29
DialogueInfer*	34.43	71.00	17.06	35.41	17.64	20.28
DialogueGCN†	37.62	70.89	16.15	34.59	16.86	20.39
DAG†	34.87	71.93	18.27	34.94	17.88	21.86
CoG-BART‡	35.51	72.10	17.15	34.60	17.46	21.35
CoMPM‡	37.67	68.60	17.53	34.67	17.70	21.21
DialogueGLP	40.64	73.55	18.66	37.08	17.35	21.37
DialogueGLP(C)	39.64	73.97	17.46	37.14	16.69	19.97
DialogueGLP(U)	39.30	74.83	19.02	37.39	17.97	21.41
DialogueGLP(F)	40.79	75.10	19.65	37.32	17.70	21.84
DialogueGLP(F+U)	**40.93**	**75.11**	**20.88**	**38.42**	**19.13**	**22.08**

Table 1: Comparison of our models and sequence-based (*), graph-based (†) and transformer-based (‡) models. DialogueInfer and our models are designed for EIC. Others are designed for ERC. (C) denotes knowledge enhancement with COMET, (U) and (F) denote knowledge enhancement with pseudo utterances and feelings respectively. We report the mean score over 5 random seeds.

Model	DailyDialog	MELD	EmoryNLP
DialogueGLP	73.55	37.08	21.37
w/o global model	72.74	36.66	20.60
w/o local model	71.37	35.44	20.93
w/o addressee-aware	73.51	36.58	20.07

Table 2: Ablation studies on three datasets.

log and MELD. Generally, knowledge F (feelings and corresponding reasons) are better than knowledge U. Our prompt-based knowledge generation method is better than COMET. We also concatenate the two generated prompt-based knowledge (U+F). The performance is further improved compared to single knowledge augmentation.

Knowledge F consists of emotions and reasons for those emotions. To explore whether only InstructGPT is enough to predict the emotions, we let it directly infer the emotions of addressees in DailyDialog. The resulting weighted F1 is 34.65, which shows that it is not good at inferring emotions and the main performance boost of DialogueGLP(F) comes from the part of reasons.

Ablation Analysis To explore the effectiveness of different modules in our model, we also do ablation studies on the three datasets. To remove the addressee information, we simply replace the global model with a single LSTM and the addressee prefix with the speaker's name. The results show that the local model is generally more important than other modules. Since DailyDialog is dyadic, the second to last utterance in our input texts must be from the addressee. Therefore, the addressee information is less important in DailyDialog.

4 Related Work

Emotion recognition in conversation has been a popular area where different models have been proposed. We divide them into three categories: sequence-based, graph-based and transformer-based. DialogueRNN (Majumder et al., 2019) models different parties and global state by different recurrent neural networks. DialogueInfer (Li et al., 2021a) adopts two LSTMs to process utterances by whether they are from addressees. DialogueCRN (Hu et al., 2021b) iteratively does retrieving and reasoning process to extract and integrate emotional clues. DialogueGCN (Ghosal et al., 2019) utilizes graph neural networks to connect utterances with surrounding utterances. DAG (Shen et al., 2021) uses a directed acyclic graph network to gather information over long distances. CoG-BART (Li et al., 2022) adopts supervised contrastive learning and response generation as auxiliary tasks. CoMPM (Lee and Lee, 2022) combines speaker's memory using a pre-trained model as an extractor.

Some works focus on introducing knowledge to help detect emotions. KET (Zhong et al., 2019) retrieve commonsense knowledge from ConceptNet (Speer et al., 2017) and NRC_VAD (Mohammad, 2018). COSMIC (Ghosal et al., 2020), DialogueInfer (Li et al., 2021b) and ToDKAT (Zhu et al., 2021) incorporates commonsense knowledge generated by COMET (Bosselut et al., 2019). GKP (Liu et al., 2022) generates knowledge from language models with prompt learning to do commonsense reasoning.

5 Conclusions

In this paper we combine the ability of sequence models and pre-trained models and propose global-local modeling method to do emotion inference in conversation. Moreover, we take the whole dialogue as input and generate knowledge with prompt learning. Experiments show that our model has achieved state-of-the-art performance on three datasets. Ablation studies show the effectiveness of different modules in our model.

Limitations

Since in our framework the global model needs to first compute the global representation then the local model outputs the emotion distribution, it takes longer time to train and inference than other models. We utilize a pre-trained model in our framework, which requires large GPU memory.

Acknowledgements

The work was supported by National Natural Science Foundation of China (62272092, 62172086).

References

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4762–4779, Florence, Italy. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901.

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.

Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. COSMIC: COmmonSense knowledge for eMotion identification in conversations. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 2470–2481, Online. Association for Computational Linguistics.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. 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 154–164, Hong Kong, China. Association for Computational Linguistics.

Dou Hu, Lingwei Wei, and Xiaoyong Huai. 2021a. DialogueCRN: Contextual reasoning networks for emotion recognition in conversations. 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 7042–7052, Online. Association for Computational Linguistics.

Zhe Hu, Zuohui Fu, Yu Yin, and Gerard de Melo. 2021b. Context-aware interaction network for question matching. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 3846–3853, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Joosung Lee and Wooin Lee. 2022. CoMPM: Context modeling with speaker's pre-trained memory tracking for emotion recognition in conversation. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5669–5679, Seattle, United States. Association for Computational Linguistics.

Dayu Li, Xiaodan Zhu, Yang Li, Suge Wang, Deyu Li, Jian Liao, and Jianxing Zheng. 2021a. Emotion inference in multi-turn conversations with addressee-aware module and ensemble strategy. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 3935–3941, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Dayu Li, Xiaodan Zhu, Yang Li, Suge Wang, Deyu Li, Jian Liao, and Jianxing Zheng. 2021b. Enhancing emotion inference in conversations with commonsense knowledge. Knowledge-Based Systems, 232:107449.

Shimin Li, Hang Yan, and Xipeng Qiu. 2022. Contrast and generation make bart a good dialogue emotion recognizer. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 11002–11010.

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.

Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. 2022. Generated knowledge prompting for commonsense reasoning. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 3154–3169, Dublin, Ireland. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. CoRR, abs/1907.11692.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 6818–6825.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 174–184, Melbourne, Australia. Association for Computational Linguistics.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. arXiv preprint arXiv:2203.02155.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019a. 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, Rada Mihalcea, and Eduard Hovy. 2019b. Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access, 7:100943–100953.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 3027–3035.

Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. 2021. Directed acyclic graph network for conversational emotion recognition. 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 1551–1560, Online. Association for Computational Linguistics.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In Thirty-first AAAI conference on artificial intelligence.

Sayyed M Zahiri and Jinho D Choi. 2018. Emotion detection on tv show transcripts with sequence-based convolutional neural networks. In Workshops at the thirty-second aaai conference on artificial intelligence.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. 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 165–176, Hong Kong, China. Association for Computational Linguistics.

Lixing Zhu, Gabriele Pergola, Lin Gui, Deyu Zhou, and Yulan He. 2021. Topic-driven and knowledge-aware transformer for dialogue emotion detection. 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 1571–1582, Online. Association for Computational Linguistics.

A Case Study

Figure 3 shows a dialogue example between A and B from DailyDialog (Li et al., 2017) dataset and generated knowledge by COMET (Bosselut et al., 2019) and prompt learning. The task is to infer A's emotion. For COMET, we take each utterance as input and generate three types of knowledge as Li et al. (2021b), which are oReact, oWant and oEffect. For prompt-based knowledge generation, we formulate the input by the template and dialogue history and input it into InstructGPT. As a result, we get a pseudo utterance that may be spoken by A and feeings of A.

As Figure 3 shows, our knowledge summarizes the dialogue well and is much more human-readable. This property makes our knowledge more suitable to concatenate with text inputs. COMET trained on ATOMIC (Sap et al., 2019) generates knowledge based on events. Therefore only one utterance can be taken as input instead of longer contexts. If an utterance contains multiple events, the generated knowledge may not be accurate. Also, the knowledge generated by COMET often repeats.

B Datasets Preprocessing

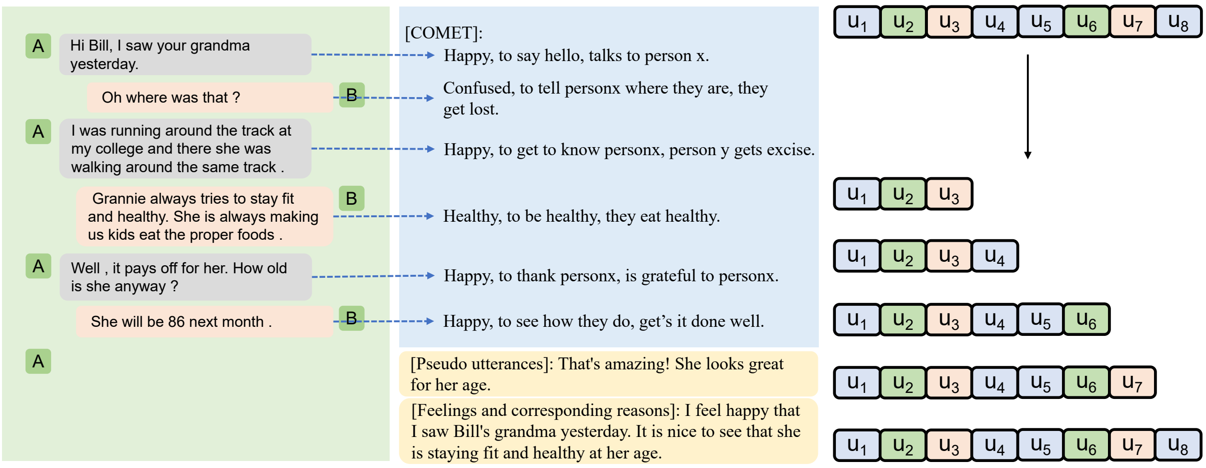
The datasets can not be directly used. We take each dialogue as an example and the emotion of the last utterance as the label. In training, we do not use the last utterance. DailyDialog is a dyadic dialogue dataset for emotion recognition. It contains more than 10,000 dialogues. We take each dialogue as a training example and take the last speaker of the dialogue as the addressee and the corresponding emotion as the label. MELD (Poria et al., 2019a) is a multimodal multiparty dialogue dataset designed for emotion recognition. However, it contains less than 2,000 dialogues. To get more training examples, we cut one dialogue into more dialogues. We keep a dialogue at least three utterances and cut it wherever the next speaker is different from the current speaker. Figure 3 shows how we process a dialogue with eight utterances. EmoryNLP (Zahiri and Choi, 2018) is a ERC dataset collected from *Friends*. We preprocess it the same way as MELD. Table 3 shows the statics after we preprocess the three datasets. We get each split of datasets from their original splits.

Dataset	train	dev	test
DailyDialog	11118	1000	1000
MELD	6125	685	1540
EmoryNLP	8345	1124	1140

Table 3: Statics of processed datasets.

C Adapting Codes to EIC

Since EIC is a new task, there are not many baselines for EIC. We adapt models that are original designed for ERC to do EIC. In our baselines, only DialogueInfer is designed for EIC. For other models, we mainly modify the code of their inputs and outputs.


A Hi Bill, I saw your grandma yesterday.

Oh where was that ? B

A I was running around the track at my college and there she was walking around the same track .

Grannie always tries to stay fit and healthy. She is always making us kids eat the proper foods . B

A Well , it pays off for her. How old is she anyway ?

She will be 86 next month . B

A

[COMET]:

Happy, to say hello, talks to person x.

Confused, to tell personx where they are, they get lost.

Happy, to get to know personx, person y gets excise.

Healthy, to be healthy, they eat healthy.

Happy, to thank personx, is grateful to personx.

Happy, to see how they do, get's it done well.

[Pseudo utterances]: That's amazing! She looks great for her age.

[Feelings and corresponding reasons]: I feel happy that I saw Bill's grandma yesterday. It is nice to see that she is staying fit and healthy at her age.

u_1 u_2 u_3 u_4 u_5 u_6 u_7 u_8
u_1 u_2 u_3
u_1 u_2 u_3 u_4
u_1 u_2 u_3 u_4 u_5 u_6
u_1 u_2 u_3 u_4 u_5 u_6 u_7
u_1 u_2 u_3 u_4 u_5 u_6 u_7 u_8

(a) (b)

Figure 3: (a) A dialogue example from DailyDialog and generated knowledge from COMET and our method. (b) An example of how we cut dialogues. This dialogue contains eight utterances. Colors denote speakers. We cut it into five dialogues.