Empathetic Dialogue Generation with Pre-trained RoBERTa-GPT2 and External Knowledge

Ye Liu, Wolfgang Maier, Wolfgang Minker and Stefan Ultes

Abstract One challenge for dialogue agents is to recognize feelings of the conversation partner and respond accordingly. In this work, RoBERTa-GPT2 is proposed for empathetic dialogue generation, where the pre-trained auto-encoding RoBERTa is utilised as encoder and the pre-trained auto-regressive GPT-2 as decoder. With the combination of the pre-trained RoBERTa and GPT-2, our model realizes a new state-of-the-art emotion accuracy. To enable the empathetic ability of RoBERTa-GPT2 model, we propose a commonsense knowledge and emotional concepts extractor, in which the commonsensible and emotional concepts of dialogue context are extracted for the GPT-2 decoder. The experiment results demonstrate that the empathetic dialogue generation benefits from both pre-trained encoder-decoder architecture and external knowledge.

1 Introduction

With the development of Spoken Dialogue Systems (SDSs), people are no longer satisfied with the task-oriented interaction, like book a train ticket or make a reservation; but are additionally interested in chit-chat communication. An expected trait of chit-chat agents is to be able to identify the user emotion and express their empathy. For instance, the psychology study in [41] shows that talking about an emotional experience to someone and sharing their emotions contributes to emotional recovery from the event. Hence, exactly identifying the user emotion and appropriately expressing their empathy will be a desired trait for SDSs.

Ye Liu, Wolfgang Maier, Stefan Ultes
Mercedes-Benz AG, Sindelfingen, Germany e-mail: ye.y.liu@daimler.com, wolfgang.mw.maier@daimler.com, stefan.ultes@daimler.com

Wolfgang Minker
Ulm University, Ulm, Germany e-mail: wolfgang.minker@uni-ulm.de

Table 1 shows an empathetic dialogue from the EmpatheticDialogues dataset [27]. A speaker tells a listener the lonely situation that they are facing, and the listener tries to understand the speaker's feelings and responds accordingly. Even though sharing emotional experiences is a general manifestation for humans, it is a great challenge to train a chit-chat agent capable to understand the user emotion and respond empathetically.

Table 1 One empathetic dialogue in EmpatheticDialogues dataset.

Emotion	Lonely
Situation	All my friends live in a different country
Speaker	Hi, I feel so lonely sometimes because all my friends live in a different country.
Listener	Oh, I'm sure you are lonely. Maybe you can join some kind of club that lets you meet new friends?
Speaker	I was thinking about it! I wanted to join a group for local moms.
Listener	That's a good idea! This way you can also meet friends for yourself, but also maybe meet new friend's for your children to hang out with while you do with their moms!

Several works with Transformer-based encoder-decoder architecture [36] have been presented for empathetic dialogue generation, such as the multi-task learning [26, 27, 37] or mixture of experts [16]. However, the combination of a pre-trained auto-encoding encoder and a pre-trained auto-regressive decoder have not been explored for empathetic dialogue generation. In this work, the pre-trained Robustly optimized BERT approach (RoBERTa) [18] as encoder and the pre-trained Generative Pre-trained Transformer (GPT-2) [25] as decoder: RoBERTa-GPT2 encoder-decoder architecture is presented for empathetic dialogue generation. The experiments with EmpatheticDialogues dataset show that the combination of RoBERTa and GPT-2 highly improves the emotion recognition ability and realizes a new state-of-the-art emotion accuracy.

In addition to the advanced neural network architecture, some external knowledge also contributes to the empathetic dialogue generation. Humans generally understand the world and express implicit emotions based on their experience and knowledge. Also, [39] demonstrates that commonsense knowledge is fundamental for chit-chat agents to understand conversations and generate appropriate responses. As shown in Fig. 1, the underlying commonsensible and emotional concepts of the speaker utterance can help the listener to better understand what the speaker is talking about. Hence, we propose an Commonsense Knowledge and Emotional Concepts Extractor (CKECE) for GPT-2 decoder in our work, to enable the commonsense and empathetic response generation. In the CKECE, we firstly utilize KeyBERT [7] to extract the keywords from the dialogue context; then elicit the commonsensible and emotional concepts of the keywords based on commonsense knowledge: ConceptNet [34] and emotion lexicon: NRC_VAD [20]; finally the extracted concepts are fed into GPT-2 decoder in a more plain text format to guide the empathetic generation.

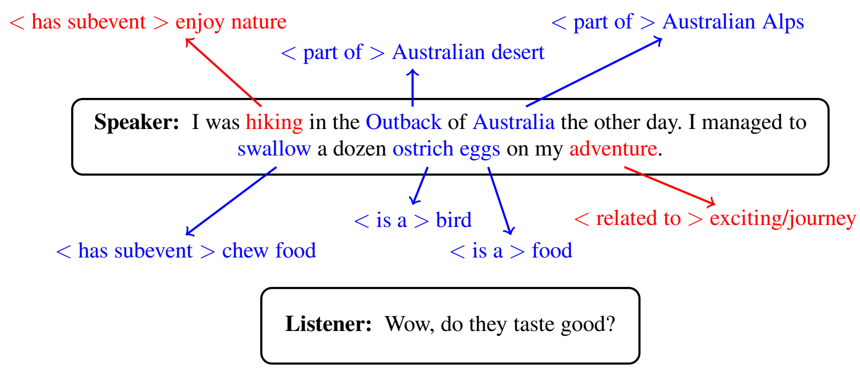

Fig. 1 An example of EmpatheticDialogues dataset with underlying commonsense knowledge (blue part) and emotional concepts (red part). (The special token in $<>$ represents the relation in commonsense knowledge: ConcepetNet [34].)

2 Related Work

Open-domain and chit-chat conversational models have been widely studied [30, 38]. With the rise of public accessible datasets [9, 15, 27] and data-driven learning approaches [35, 36], several works have attempted to make chit-chat dialogue more engaging. Some aim to improve the personalization of responses by conditioning the generation on a persona profile [11]. Then the PersonaChat dataset [42] was particularly introduced and the competition in ConvAI 2 challenge [5] demonstrated that the produced responses include more consistent personas by adding persona information into the model. However, the personalized dialogue models often can not take the feelings of their conversation partners into consideration. Besides the chit-chat research on displaying the consistent personality, some works focus on emotional and empathetic dialogue generation. The existing emotional dialogue models [3, 12, 32, 45, 46] generally generate the response depending on a predefined emotion, however, the empathetic dialogue models are capable of perceiving the emotion of the speaker and express their empathy without extra step to determine which emotion type to respond explicitly [33]. Hence, the empathetic dialogue model is more in line with the real world scenarios [14], because the listener is capable to infer the emotion of the speaker in human-human communication.

In recent years, several works have been presented for empathetic dialogue generation. [27] created a benchmark and dataset towards empathetic open-domain dialogue. [16] softly combined the possible emotional responses from several separate experts to generate the final empathetic response. [13] proposed an multi-resolution interactive empathetic dialogue model to evoke more emotional perception in dialogue generation. [14] proposed a multi-type knowledge aware empathetic dialogue generation framework to enhance the empathy of generations. The abovementioned approaches are all trained from scratch. [21] proposed BERT2BERT for Arabic empathetic response generation, while the encoder and decoder are both warm started using pre-trained auto-encoding AraBERT [1] parameters. [40] introduced EmpTransfo and [17] presented CAiRE, both are empathetic aware model adapted from GPT [24]. With the release of encoder-decoder model in Huggingface[1], where any pre-trained auto-encoding model as the encoder and any pretrained auto-regressive model as the decoder can be initialized as a sequence-to-sequence model, we are more interested in the performance of pre-trained auto-encoding encoder and auto-regressive decoder architecture for empathetic dialogue generation. Furthermore, [28] performed an extensive study on leveraging variable pre-trained models for sequence generation tasks and demonstrated that combining RoBERTa [18] and GPT-2 [25] achieves strong results. Hence, RoBERTa-GPT2 is proposed in this work for empathetic dialogue generation.

In addition, the corpora with emotion labelling play a significant role in empathetic dialogue generation. There are several interesting resources. [15] developed the DailyDialog dataset, with manually emotion labelling to each utterance. [9] collected the EmotionLines dataset from TV shows and human-to-human chats, where each utterance is further annotated with one of seven emotion-categorical labels. However, only 5% of the utterances in DailyDialog and 16.68% in EmotionLines have varied emotional labels and others are either "none" or "happy" labels. Hence, they are not suitable for empathetic dialogue generation because of the extremely unbalanced data distribution. [27] released an empathetic dialogue dataset: EmpatheticDialogues, which focuses explicitly on conversations about emotionally grounded personal situations and considers a richer, evenly distributed set of emotions. In our work, we conduct the experiment of empathetic dialogue generation with EmpatheticDialogues dataset.

3 The Proposed Method

In this work, we present the RoBERTa-GPT2 encoder-decoder architecture for empathetic dialogue generation, where the pre-trained auto-encoding RoBERTa [18] as encoder and pre-trained auto-regressive GPT-2 [25] as decoder. In addition, a Commonsense Knowledge and Emotional Concepts Extractor (CKECE), which is used to extract the relevant concepts from dialogue history, is proposed to enable the commonsensible and empathetic ability of GPT-2 decoder. In this section, the CKECE will be firstly introduced and then the RoBERTa-GPT2 architecture with extracted concepts will be shown.

[1] https://huggingface.co/transformers/model_doc/encoderdecoder.html

3.1 Commonsense Knowledge and Emotional Concepts Extractor: CKECE

For the CKECE, two knowledge sources: the commonsense knowledge ConceptNet [34] and the emotional lexicon NRC_VAD [20], and one keyword extraction tool, KeyBERT [7], are used. We firstly utilize the KeyBERT to extract the keywords of the dialogue context, and then filter out the most relevant commonsense knowledge and emotional concepts of the keywords with the confidence score of ConceptNet and emotional intensity of NRC_VAD.

3.1.1 The CKECE components

The three resources used in CKECE are introduced in the following:

KeyBERT[2] is a minimal and easy-to-use keyword extraction technique that leverages BERT embeddings and cosine similarity to find the keywords and keyphrases in a document that are the most similar to the document itself.

ConceptNet[3] is a large-scale and multilingual commonsense knowledge graph that describes general human knowledge in natural language. It comprises 5.9M assertions, 3.1M concepts and 38 relations. The nodes in ConceptNet are concepts and the edges are relations. Each (*concept1, relation, concept2*) triplet is an assertion. Each assertion is associated with a confidence score. The assertion confidence score are usually in the $[1, 10]$ interval. For example, (*loneliness, CausesDesire, socialize*) with confidence score of 3.464.

NRC_VAD[4] is a lexicon that includes a list of more than 20k English words and their Valence, Arousal, and Dominance (VAD) scores. For a given word and a dimension, the scores range from 0 (lowest) to 1 (highest). The interpretations of NRC_VAD dimensions are presented in Table 2. Such as, the VAD score vector $[V_a, A_r, D_o]$ of word "happiness" is [0.960, 0.732, 0.850].

Table 2 Interpretations of NRC_VAD dimensions.

Dimensions	Values	Interpretations
Valence (V_a)	[0, 1]	Negative-Positive
Arousal (A_r)	[0, 1]	Calm-Excited
Dominance (D_o)	[0, 1]	Weak-Powerful

[2] https://github.com/MaartenGr/KeyBERT

[3] https://conceptnet.io/

[4] https://saifmohammad.com/WebPages/nrc-vad.html

3.1.2 CKECE

To extract more relevant concepts, we firstly utilize the KeyBERT to extract the keywords from the dialogue context. In this step, the recommended KeyBERT model "distilbert-base-nli-mean-tokens" is used and only maximal top 10 keywords with score larger than 0 are retained. An example of extracted keywords is shown in Fig. 2.

Then, we pick out the commonsense concepts from ConceptNet based on the keywords and denote them in a tuple (keyword, relation, concept, scaled confidence score) as $\{\tau_k^i = (k_i, r_k^i, c_k^i, s_k^i)\}_{k=1,2,...,K}$ where the confidence score s is scaled by the following Eq. 1 *min-max* normalization.

$$min\text{-}max(s) = \frac{s - min_s}{max_s - min_s} , \tag{1}$$

where min_s is 1 and max_s is 10. The processed $s \in [0,1]$ and the *min-max* normalization is also used in [14, 44]. With *min-max* normalization, the example (*loneliness, CausesDesire, socialize*) with confidence score 3.464 in Section 3.1.1 is transformed into (loneliness, CausesDesire, socialize, 0.274) tuple with scaled confidence score 0.274. In order to pick out the most relevant concepts, the following tuples will be removed in this step:

- The keywords or concepts are stop words. (The union of stop words in NLTK [19] and SpaCy[5] are used.)
- The scaled confidence score is less than a pre-defined threshold α. We set α is 0.1 in this work, i.e. $s < 0.1$.
- The keywords and concepts are same or have the same stem. Like: (addition, Synonym, addition, 0.11); (actual, DerivedFrom, actually, 0.11).
- The relation is in an excluded relation list. i.e. $r \in [Antonym, ExternalURL, NotDesires, NotHasProperty, NotCapableOf, dbpedia, DistinctFrom, EtymologicallyDerivedFrom, EtymologicallyRelatedTo, SymbolOf, FormOf, AtLocation, DerivedFrom, SymbolOf]$

Furthermore, to enable the emotional concepts, we adopt NRC_VAD to compute emotion intensity values for the concepts c as the Eq. 2.

$$\eta(c) = min\text{-}max(\|V_a(c) - \frac{1}{2}, \frac{A_r(c)}{2}\|_2) , \tag{2}$$

where $\|\cdot\|_k$ denotes l_k norm. $V_a(c)$ and $A_r(c)$ represent valence and arousal score of concept c_i, respectively. When c not in NRC_VAD, we set $\eta(c)$ to the mid value of 0.5.

Lastly, the final score f in Eq. 3 is derived from three aspects: emotion intensity, semantic similarity and scaled confidence score. The semantic similarity $cos(k_i, c_k^i)$ is the cosine similarity between keyword and concept both embedded by the GloVe [23], which stands for global vectors for word representation and is an unsupervised learning algorithm for obtaining vector representations for words.

[5] https://github.com/explosion/spaCy

$$f(\tau_k^i) = \eta(c_k^i) + cos(k_i, c_k^i) + s_k^i \ , \tag{3}$$

We sort the candidate tuples in descending order of the final scores and select top 3 tuples for each keyword. Maximal 10 tuples are chosen for every dialogue context. Then the extracted concepts are arranged in a more plain textual form: "keyword <relation> concept", which is shown in Fig. 2, for GPT-2 decoder.

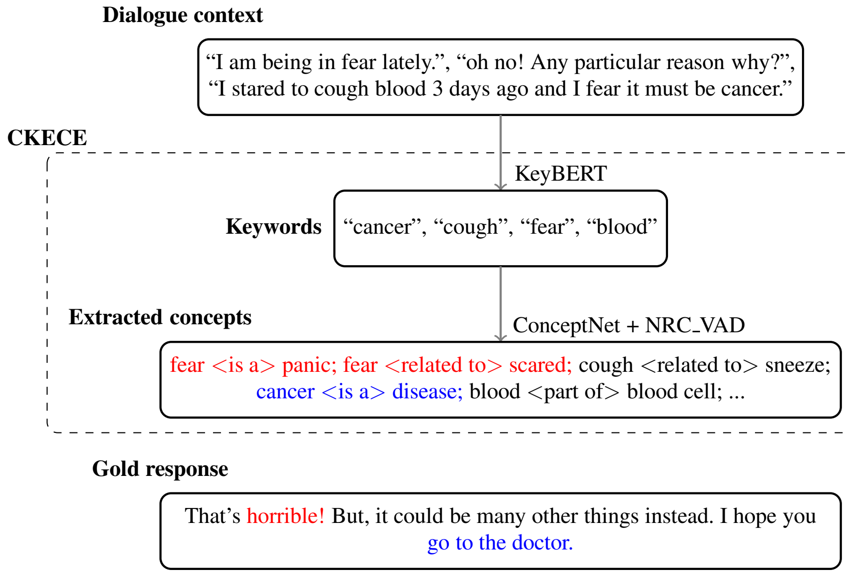

Fig. 2 An example for the process of CKECE for the dialogue context. The extracted emotional concepts and emotional word in gold response are marked in red. The blue part in extracted concepts and gold response share same commonsense knowledge.

3.2 Pre-trained RoBERTa-GPT2 encoder-decoder

The RoBERTa [18] and GPT-2 [25] are both large architectures pre-trained on large collections of texts. Then the pre-trained models are widely fine-tuned in downstream tasks. In this work, we explore the pre-trained RoBERTa-GPT2 as encoder-decoder architecture for empathetic dialogue generation.

3.2.1 The preliminaries of RoBERTa-GPT2

The pre-trained auto-encoding RoBERTa and pre-trained auto-regressive GPT-2 are introduced in the following:

RoBERTa[6] has the same architecture as BERT [4], but uses a byte-level Byte-Pair Encoding (BPE) [29] as a tokenizer (same as GPT-2) and improved the training procedure of BERT [4].

GPT-2[7] is a pre-trained large-scale unsupervised language model which generates coherent paragraphs of text. GPT-2 is also widely used in task-oriented dialogue generation [2, 22] and chit-chat dialogue generation [17, 43].

3.2.2 RoBERTa-GPT2

Fig. 3 shows our proposed RoBERTa-GPT2 encoder-decoder architecture for empathetic dialogue generation. The simplified input for RoBERTa encoder and GPT-2 decoder in Fig. 3 only shows the initial part of the sentences. And Fig. 2 and Fig. 3 share the same dialogue example.

The pre-trained RoBERTa as encoder process the dialogue context, where the $<$ CLS $>$ token is appended at the first place and $<$ SEP $>$ is for separating speaker utterance and listener utterance. The output of $<$ CLS $>$ token, pooled output, represents the entire meaning of the input. A linear layer with softmax activation is added on the top of pooled output for emotion classification. The encoder outputs will be fed to the GPT-2 decoder for cross-attention mechanism. As shown in Fig. 3, the input for GPT-2 decoder starts with extracted concepts. During the training, the gold response is also attached after concepts for faster convergence and separated by <SEP> token. It is noteworthy that only the response part without extracted concepts is the output of GPT-2 decoder for computing the generation loss during the training. That means, the response is generated conditioned on the contextual information of encoder outputs with cross-attention mechanism and emotional concepts of decoder inputs with self-attention mechanism by combining pre-trained RoBERTa and GPT-2. Lastly, all the parameters of RoBERTa-GPT2 are jointly trained end-to-end to optimize the emotion classification and response generation by minimising emotion cross entropy loss and maximum likelihood estimator (MLE) generation loss.

[6] https://github.com/pytorch/fairseq/tree/master/examples/roberta

[7] https://github.com/openai/gpt-2

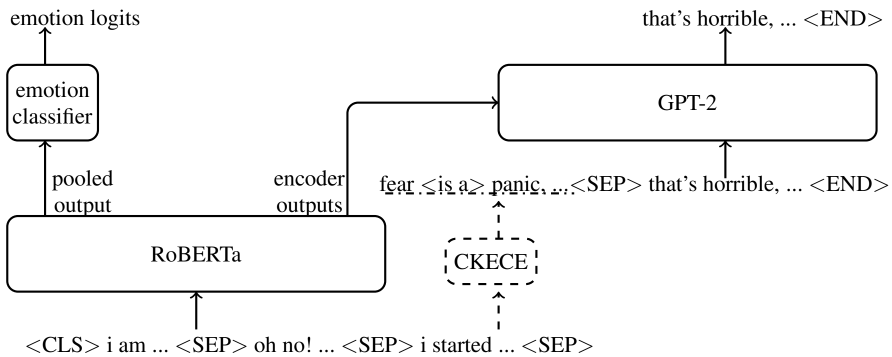

Fig. 3 Our proposed RoBERTa-GPT2 encoder-decoder architecture with CKECE guidance for empathetic dialogue generation.

4 Experimental Settings and Results Analysis

4.1 Dataset

We conduct our experiment on the large-scale multi-turn EmpatheticDialogues [27], which consists of 25k one-to-one open-domain conversation grounded in emotional situations. And the EmpatheticDialogues dataset provides 32 evenly distributed emotion labels.

4.2 Baselines

We compare our models with the following four baselines.

1) **Transformer** [36]: a Transformer-based encoder-decoder model trained with MLE generation loss.
2) **EmoPrepend-1** [27]: an extension of Transformer model with an additional supervised emotion classifier. The whole model is jointly trained by optimizing both the classification and generation loss.
3) **MoEL** [16]: another extension of Transformer model, which softly combines the outputs of the multiple listeners. Each listener is optimized to react to a certain emotion and generate an empathetic response.
4) **MK-EDG** [14]: a multi-type knowledge aware empathetic dialogue generation framework. Commonsense knowledge and emotional lexicon are used to enrich the dialogue utterance.

Additionally, to better analyse our proposed RoBERTa-GPT architecture for empathetic dialogue model, we also conducted **RoBERTa w/o GPT-2**: only RoBERTa encoder with emotion classifier trained with emotion loss; and **RoBERTa-GPT2 w/o CKECE**: RoBERTa-GPT2 without the guidance of external knowledge.

4.3 Training details

The RoBERTa-GPT2 is trained with batch size 16 and learning rate 1e−5. Early stopping is applied during the training for saving the best model. During decoding, we use the top-k [6] and nucleus sampling (top-p) [8] decoding algorithms with top-k equal to 5 and top-p equal to 0.9.

4.4 Automatic Evaluation Results

To evaluate the performance of RoBERTa-GPT2 model, we firstly adopt the Emotion Accuracy as the agreement between the ground truth emotion labels and the predicted emotion labels by the emotion classifier. In addition, Perplexity [31] values are utilized to measure the high-level general quality of the generation model. Furthermore, Distinct-1 and Distinct-2 [10] are used to measure the proportion of the distinct unigrams and bigrams in all the generated results to indicate diversity. Table 3 shows the evaluation results between our proposed methods and baselines. The results of MK-EDG in Table 3 are directly copied from [14], hence MK-EDG is absent from use cases in Table 4.

Table 3 Evaluation results between RoBERTa-GPT2 and baselines

Models	Emotion Accuracy	Perplexity	Distinct-1	Distinct-2
Transformer	-	35.56	0.41	1.49
EmoPrepend-1	0.3359	35.66	0.42	1.62
MoEL	0.3425	37.69	0.43	1.72
MK-EDG	0.3931	34.85	1.48	4.90
RoBERTa w/o GPT-2	0.3439	-	-	-
RoBERTa-GPT2 w/o CKECE	**0.5262**	14.97	1.62	10.47
RoBERTa-GPT2	0.5151	**13.57**	**2.04**	**11.68**

In Table 3, we observe our proposed RoBERTa-GPT2 outperforms the baselines by a large margin in terms of all automatic metrics and realizes a new state-of-the-art result. The emotion accuracy comparison between only RoBERTa encoder w/o GPT-2 and RoBERTa-GPT2 demonstrates that emotion classification benefits from the combination of RoBERTa encoder and GPT-2 decoder architecture. The RoBERTa-GPT2 has better results on Perplexity, Distinct-1 and Distinct-2 than RoBERTa-GPT2 w/o CKECE, indicating that injecting external knowledge contributes to the generation quality, though a little loss on emotion accuracy.

4.5 Use cases study

Table 4 Generated responses from Transformer, EmoPrepend-1, MoEL and RoBERTa-GPT2. The underline part shows the empathetic and contextual of RoBERTa-GPT2 response.

Emotion	excited
Dialogue context	**Speaker:** I'm taking a vacation and going on a cruise to Alaska next week!
Transformer	Where are you going?
EmoPrepend-1	That is awesome! Where are you going?
MoEL	I hope you have a great time!
RoBERTa-GPT2	I am sure you will enjoy the cruise!
Emotion	terrified
Dialogue context	**Speaker:** Do you even know how crazy it is to skydive? I must have been absolutely nuts when I did it. **Listener:** Oh I don't think I could do that. I have a fear of falling from high places. **Speaker:** It gave me the biggest rush that's for sure. But on the way down I was saying my prayers.
Transformer	That is a great idea. I have been there.
EmoPrepend-1	I would be too. I do not know what you mean.
MoEL	I am sorry to hear that. I hope everything works out for you.
RoBERTa-GPT2	That is a lot of stress! I am glad you did not get injured.
Emotion	excited
Dialogue context	**Speaker:** I couldn't wait to go to the concert.
Transformer	That sounds like a lot of fun!
EmoPrepend-1	That sounds like a lot of fun!
MoEL	That sounds like a lot of fun!
RoBERTa-GPT2	That is awesome! What kind of concert will it be?
Emotion	trusting
Dialogue context	**Speaker:** I let my daughter ride in the road by herself on her skateboard the other day.
Transformer	That is awesome! What was it?
EmoPrepend-1	That must have been a lot of fun!
MoEL	oh no! What happened?
RoBERTa-GPT2	Wow! I would be a little nervous if she did not skate at all.

Cases of generated responses from baselines Transformer, EmoPrepend-1, MoEL and our proposed method, RoBERTa-GPT2, are listed in Table 4. In the first case, Transformer and EmoPrepend-1 do not exactly understand what the speaker is saying. MoEL identifies the user emotion but the response is kind of generic. Besides the correct understanding user emotion, RoBERTa-GPT2 also knows that the speaker is talking about "cruise". The baselines in the second case do not correctly recognise the user emotion. Compared with generic response of the baselines in the third case, RoBERTa-GPT2 generates contextual response with a proper positive emotion by replying with "awesome". In the fourth case, the response of EmoPrepend-1 is generic and other two baselines do not understand the speaker, while RoBERTa-GPT2 generates coherent and informative response by showing the concern. All the cases in Table 4 show that our proposed RoBERTa-GPT2 can both handle with user emotion and dialogue content.

5 Conclusion and Outlook

In this work, we leverage pre-trained auto-encoding RoBERTa as encoder and pre-trained auto-regressive GPT-2 as decoder for empathetic dialogue generation. Meanwhile, the external knowledge: commonsense knowledge and emotional lexicon; are utilized to extract emotional and commonsensible concepts from dialogue context for GPT-2 decoder to enable the empathetic and contextual responses. Both automatic metrics and cases study show that our proposed RoBERTa-GPT2 outperforms the baselines and demonstrate that the empathetic dialogue generation benefits from pre-trained modelling and external knowledge.

In the future work, we will continually evaluate our proposed method for empathetic dialogue generation from human perspective. Meanwhile, we are also interested in other flexible methods for injecting external knowledge to empathetic dialogue system.

References

1. Antoun, W., Baly, F., Hajj, H.: Arabert: Transformer-based model for arabic language understanding. In: LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020, p. 9
2. Budzianowski, P., Vulić, I.: Hello, it's gpt-2-how can i help you? towards the use of pretrained language models for task-oriented dialogue systems. In: Proceedings of the 3rd Workshop on Neural Generation and Translation, pp. 15–22 (2019)
3. Colombo, P., Witon, W., Modi, A., Kennedy, J., Kapadia, M.: Affect-driven dialog generation. 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), pp. 3734–3743 (2019)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: 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), pp. 4171–4186 (2019)
5. Dinan, E., Logacheva, V., Malykh, V., Miller, A., Shuster, K., Urbanek, J., Kiela, D., Szlam, A., Serban, I., Lowe, R., et al.: The second conversational intelligence challenge (convai2). arXiv e-prints pp. arXiv–1902 (2019)
6. Fan, A., Lewis, M., Dauphin, Y.: Hierarchical neural story generation. In: ACL (1) (2018)
7. Grootendorst, M.: Keybert: Minimal keyword extraction with bert. (2020). DOI 10.5281/zenodo.4461265. URL https://doi.org/10.5281/zenodo.4461265

8. Holtzman, A., Buys, J., Du, L., Forbes, M., Choi, Y.: The curious case of neural text degeneration. In: International Conference on Learning Representations (2019)
9. Hsu, C.C., Chen, S.Y., Kuo, C.C., Huang, T.H., Ku, L.W.: Emotionlines: An emotion corpus of multi-party conversations. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
10. Li, J., Galley, M., Brockett, C., Gao, J., Dolan, W.B.: A diversity-promoting objective function for neural conversation models. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 110–119 (2016)
11. Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., Dolan, W.B.: A persona-based neural conversation model. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 994–1003 (2016)
12. Li, J., Sun, X.: A syntactically constrained bidirectional-asynchronous approach for emotional conversation generation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 678–683 (2018)
13. Li, Q., Chen, H., Ren, Z., Ren, P., Tu, Z., Chen, Z.: Empdg: Multi-resolution interactive empathetic dialogue generation. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 4454–4466 (2020)
14. Li, Q., Li, P., Chen, Z., Ren, Z.: Towards empathetic dialogue generation over multi-type knowledge (2020)
15. Li, Y., Su, H., Shen, X., Li, W., Cao, Z., Niu, S.: Dailydialog: A manually labelled multi-turn dialogue dataset. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 986–995 (2017)
16. Lin, Z., Madotto, A., Shin, J., Xu, P., Fung, P.: Moel: Mixture of empathetic listeners. 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), pp. 121–132 (2019)
17. Lin, Z., Xu, P., Winata, G.I., Siddique, F.B., Liu, Z., Shin, J., Fung, P.: Caire: An end-to-end empathetic chatbot. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13,622–13,623 (2020)
18. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized BERT pretraining approach. CoRR **abs/1907.11692** (2019). URL http://arxiv.org/abs/1907.11692
19. Loper, E., Bird, S.: Nltk: The natural language toolkit. In: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, pp. 63–70 (2002)
20. Mohammad, S.: 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), pp. 174–184 (2018)
21. Naous, T., Antoun, W., Mahmoud, R., Hajj, H.: Empathetic bert2bert conversational model: Learning arabic language generation with little data. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp. 164–172 (2021)
22. Peng, B., Zhu, C., Li, C., Li, X., Li, J., Zeng, M., Gao, J.: Few-shot natural language generation for task-oriented dialog. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, pp. 172–182 (2020)
23. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543 (2014)
24. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I.: Improving language understanding by generative pre-training
25. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019)
26. Rashkin, H., Smith, E.M., Li, M., Boureau, Y.L.: I know the feeling: Learning to converse with empathy (2018)

27. Rashkin, H., Smith, E.M., Li, M., Boureau, Y.L.: Towards empathetic open-domain conversation models: A new benchmark and dataset. In: ACL (1) (2019)
28. Rothe, S., Narayan, S., Severyn, A.: Leveraging pre-trained checkpoints for sequence generation tasks. Transactions of the Association for Computational Linguistics **8**, 264–280 (2020)
29. Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1715–1725 (2016)
30. Serban, I.V., Lowe, R., Charlin, L., Pineau, J.: Generative deep neural networks for dialogue: A short review (2016)
31. Serban, I.V., Sordoni, A., Bengio, Y., Courville, A., Pineau, J.: Hierarchical neural network generative models for movie dialogues. arXiv preprint arXiv:1507.04808 **7**(8), 434–441 (2015)
32. Shen, L., Feng, Y.: Cdl: Curriculum dual learning for emotion-controllable response generation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 556–566 (2020)
33. Skowron, M., Theunis, M., Rank, S., Kappas, A.: Affect and social processes in online communication–experiments with an affective dialog system. IEEE Transactions on Affective Computing **4**(3), 267–279 (2013)
34. Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
35. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Advances in Neural Information Processing Systems **27**, 3104–3112 (2014)
36. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: NIPS (2017)
37. Wei, W., Liu, J., Mao, X., Guo, G., Zhu, F., Zhou, P., Hu, Y.: Emotion-aware chat machine: Automatic emotional response generation for human-like emotional interaction. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 1401–1410 (2019)
38. Wolf, T., Sanh, V., Chaumond, J., Delangue, C.: Transfertransfo: A transfer learning approach for neural network based conversational agents (2019)
39. Young, T., Cambria, E., Chaturvedi, I., Zhou, H., Biswas, S., Huang, M.: Augmenting end-to-end dialogue systems with commonsense knowledge. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
40. Zandie, R., Mahoor, M.H.: Emptransfo: A multi-head transformer architecture for creating empathetic dialog systems. In: The Thirty-Third International Flairs Conference (2020)
41. Zech, E., Rimé, B.: Is talking about an emotional experience helpful? effects on emotional recovery and perceived benefits. Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice **12**(4), 270–287 (2005)
42. Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., Weston, J.: Personalizing dialogue agents: I have a dog, do you have pets too? In: ACL (1) (2018)
43. Zhang, Y., Sun, S., Galley, M., Chen, Y.C., Brockett, C., Gao, X., Gao, J., Liu, J., Dolan, W.B.: Dialogpt: Large-scale generative pre-training for conversational response generation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 270–278 (2020)
44. Zhong, P., Wang, D., Miao, C.: 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), pp. 165–176 (2019)
45. Zhou, H., Huang, M., Zhang, T., Zhu, X., Liu, B.: Emotional chatting machine: Emotional conversation generation with internal and external memory. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
46. Zhou, X., Wang, W.Y.: Mojitalk: Generating emotional responses at scale. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1128–1137 (2018)