

# MRE : Multi Relationship Extractor for Persona based Empathetic Conversational Model

Bharatram Natarajan  
SRIB\*

bharatram.nitt@gmail.com

Abhijit A Nargund  
SRIB\*

nargundabhi@gmail.com

## Abstract

Artificial intelligence(AI) has come a long way in aiding the user requirements in many fields and domains. However, the current AI systems do not generate human-like response for user query. Research in these areas have started gaining traction recently with explorations on persona or empathy based response selection. But the combination of both the parameters in an open domain haven't been explored in detail by the research community. The current work highlights the effect of persona on empathetic response. This research paper concentrates on improving the response selection model for PEC dataset, containing both persona information and empathetic response. This is achieved using an enhanced **multi relationship extractor** and **phrase based information** for the selection of response.

## 1 Introduction

Empathetic response generation refers to the ability of a system to understand the people mentality or the feelings of a user and provide an appropriate response. In the area of NLP, empathetic conversational models have shown a positive impact on the user compared to normal responses [Liu and Picard (2005); Zhou et al. (2020); Lin et al. (2020); Li et al. (2021)]. Empathetic response creates a personal connect with the user and achieves significantly higher engagement by providing seamless conversational experience as shown by this example 1. You can see how empathetic response aids in different kind of engagement with the user and also provide suggestions.

Recent years have seen an increase in exploration on empathetic response generations using neural conversational models. Lin et al. (2020) have developed CAIRE to generate empathetic response using user emotions for better connections with the

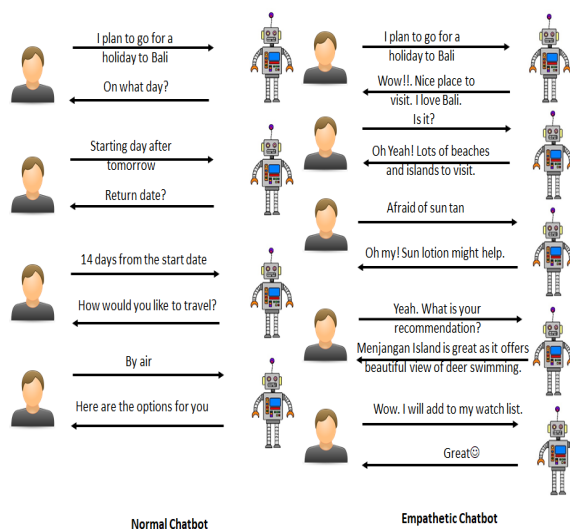


Figure 1: Normal Chatbot Vs Empathetic Chatbot

user. Li et al. (2021) have used causal emotional information to understand a particular user emotional state and correspondingly learn response for each emotion class. Zhou et al. (2020) have developed an AI model named XiaoIce for better human understanding and communication. (Rashkin et al., 2018) has created custom datasets for the said problems named EMPATHETIC DIALOGUES (ED). These dataset has proven that pre-trained retrieval models like BERT and its variants are able to reply with more empathy when trained with such dataset.

The above explorations doesn't take into account the persona of the user. As persona aides in better conversational response, this area started gaining traction by the research community. Demasi et al. (2020) developed Crisisbot where it uses the persona of the user in the conversation to provide complex response to train hotline counselors for suicide prevention task. Song et al. (2019) explored the way to generate sustainable and coherent response, using persona in a conversation, where each request will have many possible responses. Wang et al.

Samsung R&D Institute India - Bangalore

(2021) developed an emotion-affective open domain chatbot where they use knowledge graph to extract personal information and incorporate into the system for consistent personality.

The presence of Persona have shown impact in the conversational response. Hence, exploration on the impact of persona on empathetic response started gaining traction. Persona have shown to impact empathetic natural language generation capability. We have noticed that empathetic response of system is different from two different users for the same input user utterance or query. Roller et al. (2020) incorporated persona with empathy on response generation for several dataset but the impact of persona on empathy is not explored. Zhong et al. (2020) has presented a novel large-scale multi-turn Persona-based Empathetic Conversation (PEC) using two contrasting sentimental domains from social media Reddit. They have proposed novel Cobert architecture (combination of BERT with basic Co-Attention mechanism) to select the appropriate response for the post and penalise opposite response pairs. This is the first approach where Persona information is used to influence the empathetic response selection. The current architecture takes the full user context, personas and response for gathering bi-attention and selecting appropriate response. Co-attention won't be able to capture phrase level impact on the response selection. Moreover the influence of the different phrases from different positions might not be captured well.

To address this issue, we are proposing multi relationship extraction using BERT for influencing the selection of appropriate response and penalising the un-important ones.

In summary, the contributions of the papers is summarized as below

- We propose phrase importance planner to extract n-gram impact of the phrase in unigram.
- We propose multi relationship extraction using BERT where we use phrase importance planner along with the entire context to impact response.

## 2 Related Work

### 2.1 Retrieval based Conversational Model

Lots of work have happened in neural conversational model for response selection task. The task

is approached in 3 stages which are as shown below.

- **Encoding** The encoding module encodes the input tokens into contextual vector representation using encoders like BERT, ELMO or non contextual vector representations like Glove embedding.
- **Matching** The matching module measures the co-relation between user context and response using persona details with different attention techniques.
- **Aggregation** The aggregation module aides in summarising the matching module information along the sequence axis to get the final representation.

Humeau et al. (2019) introduced polyencoders for handling pairwise comparison between 2 sequences using the combination of cross encoders and bi-encoders. Cross-encoders calculates attention over all the target labels and hence will be slow in calculation. Bi-encoders calculate individual pair attention and will be faster.

## 3 PEC Dataset

This section explains the high level information on PEC dataset gathered by Zhong et al. (2020). For full detailed analysis, please refer to the above paper. PEC dataset is available in huggingface as <sup>1</sup>.

**Data Source** The data has been collected by the author Zhong et al. (2020) from Reddit, a discussion forum where users can discuss any topics on sub-reddits. The data is made from two contrasting sentiments related sub-reddits namely **Happy** and **Offmychest**. Here the comments are more empathetic than casual conversations.

**Conversation Collection** Reddit has threads where there will be a single post containing direct and indirect comments. These threads are organized in a tree where the root represent post and comments are represented as nodes connecting to parent comment node or root post node. If there are n nodes then author extracts n-1 conversations where each conversation starts from root node and ends at n-1 non root nodes. The author has split the data into 80:10:10 for training, validation and testing.

**Persona Collection** Persona sentences are collected from all the posts and comments that the

<sup>1</sup><https://huggingface.co/datasets/pec>

user has written. There are strict rules applied to fetch the persona sentences from the provided posts which are listed below

- Presence of the word "i" in the post.
- Presence of atleast one verb
- Presence of atleast one noun or adjective.
- Presence of atleast one content word.

**Data Processing** We follow the data processing steps followed by the author [Zhong et al. \(2020\)](#). These steps are listed below.

- Each conversation has at most 6 most recent turns.
- Each post is between 2 and 90 words.
- Each comment is between 2 and 30 words.
- Each speaker has atleast one persona sentence.
- The last speaker is different from the first speaker in each conversation. The reason is last comment is considered as empathetic response rather than reply to the user.
- Remove all special symbols, URLs and image captions.
- Lowercase all the utterances.

Sample conversation of PEC dataset for happy and offmychest are present in Table 1

## 4 Multi Relationship Extractor using BERT Embedding

This section introduces the task of response selection and briefly explains novel architecture on addressing the task at hand as shown in Figure 2.

### 4.1 Task Definition

We denote the training conversational dataset as  $D_X$ .  $D_X$  is a set of  $n$  conversations. Each conversation is in the format of  $(U_X, P_X, Y_X)$  where  $U_X = U_{X1}, U_{X2}, U_{X3}, \dots, U_{Xn}$  indicates the  $n$  user context utterances,  $P_X = P_{X1}, P_{X2}, P_{X3}, \dots, P_{Xp}$  denotes the  $p$  persona sentences for the respondent and  $Y$  denotes the target response for the user context. We can formulate the response selection problem as  $f(U_X, P_X, Y_X)$  where we assign highest probability to true candidate  $Y_X$  and lowers the score of negative candidates given  $X$  and  $P$ . When we infer the model, the model will select the best candidate from the candidate list by selecting highest probability.

Conversations	
OffMyChest	Happy
why is it ok for women to wear skirts in a business casual environment , but men can't wear shorts ?	got the best t - shirt ever today !
skirts generally are n't comfortable . you can't do much in them other than walk , unless they 're long and even then ...	prepare your inbox for pussy puns
my issue is n't comfort , it 's sweating my balls off . ladies also get to wear sleeveless shirts !	i do n't get it ..
... not sweating your balls off is comfort .	she 's a girl wearing a shirt with a cat on it . you know what , no ... this has got ta be a troll attempt
Persona of respondent	
i was going to say " welcome to being a woman "	i have 2 characters .
i just wanted to share .	i respond the same way to gal gadot tickling me
i make lots of male friends easy .	i know that feel , unfortunately

Table 1: Sample data for **OffMyChest** and **Happy** classes in the PEC dataset. First 4 rows represent Conversation and last 3 rows represent Persona details of the respondent.

### 4.2 BERT Embedding

This module handles the first stage of retrieval model namely Embedding step. In this module, we encode the user context utterances, persona utterances and response utterances using BERT pre-trained model ([Devlin et al., 2018](#)). User context utterances is obtained by concatenating the list of sentences uttered by the user in order. Persona utterances are obtained by random ordered concatenation of list of persona utterances for the respondent. Response utterances are obtained by concatenating the list of response sentences. When we encode context utterances, persona utterances and response utterances using BERT, we get vector representation of context, persona and response. Context

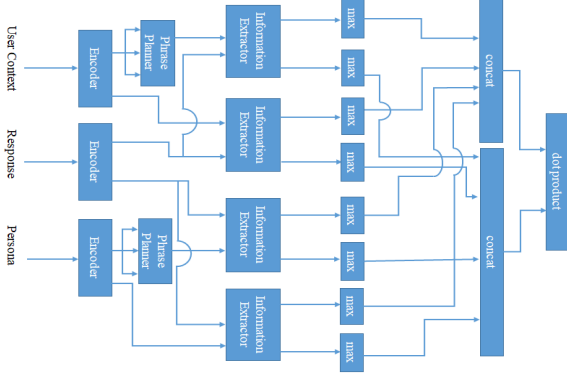


Figure 2: Multi Relationship Extractor using BERT. Information Extractor module address the importance of phrase or context level learning over response and vice versa. Phrase Planner handles phrase importance of user context or persona inputs by incorporating the importance of bigram and trigram over unigram.

vector representation will be  $C \in \mathbb{R}^{c \times d}$  where  $c$  is the maximum sequence length of the user context utterances. Persona vector representation will be  $P \in \mathbb{R}^{p \times d}$  where  $p$  is the maximum sequence length of the persona utterances. Response vector representation will be  $R \in \mathbb{R}^{r \times d}$  where  $r$  is the maximum sequence length of the response utterances. One important information is that we use different segment ids for user context utterances and response utterances. Now Context vector representation and Persona vector representation will pass through Phrase Planner Module.

### 4.3 Phrase Planner Module

This submodule aides in capturing the phrase importance for both Context vector and Persona vector separately as shown in Figure 3. For Context vector, this is done by capturing unigram, bigram and trigram information using 3 different 2d convolution module with sliding window size of 1, 2 and 3 tokens respectively keeping padding same. The importance of bigram and trigram on unigram is captured using Multi Head Attention(MHA) module developed by Vaswani et al. (2017) as shown in Figure 4. We use separate MHA module for calculating different positional bigram importance on unigram and different positional trigram importance on unigram. MHA achieves this by dividing the Query(Q), Key(K), Value (V) input into equal chunks and process each chunk in parallel. Each chunk calculates the weights for V using scaled dot product followed by softmax as shown in Equation 1. The scalar dot product calculates the affinity or

importance of Query on Key. Now we concatenate each chunk output at last dimension layer to get final matrix.

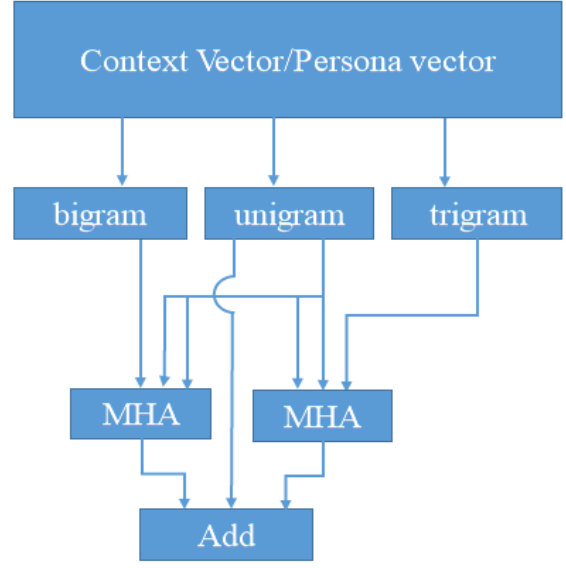


Figure 3: Phrase Planner Module.

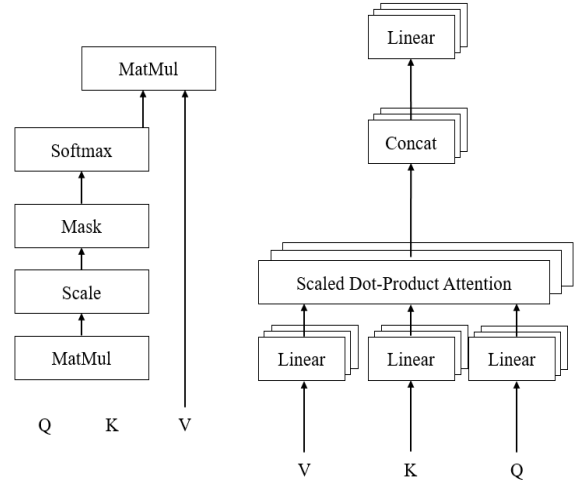


Figure 4: Multi Head Attention.

$$Attention(Q, K, V) = softmax(QK^T / \sqrt{d})V \quad (1)$$

For bigram, we pass the Query as bigram, Key as unigram and Value as unigram in this module. Similarly we calculate for Trigram importance in unigram. Finally we add unigram, bigram importance and trigram importance modules to get final matrix named Context phrase Planner. This module maintains the dimension of the input vector. Hence the output for Context phrase Planner is  $C^{PP} \in \mathbb{R}^{c \times d}$ . Similarly we apply different Phrase Planner Mod-

ule for Persona vector and generate Persona phrase planner as  $P^{PP} \in \mathbb{R}^{p \times d}$ . Now we pass the Context phrase Planner, Persona phrase planner, Context vector, Persona vector and response vector will be passed through Information Extractor module.

#### 4.4 Information Extractor

This module is responsible for mutual importance relations projection between the following learnings.

- Context vector C and Response vector R.
- Context Phrase Planner  $C^{PP}$  and Response vector R.
- Persona vector P and Response vector R.
- Persona Phrase Planner  $P^{PP}$  and Response vector R.

This module handles the second stage of retrieval model namely Matching step. Sample flow of Information Extractor module is shown in Figure 5 For a given  $C^{PP}$  and R, we are calculating

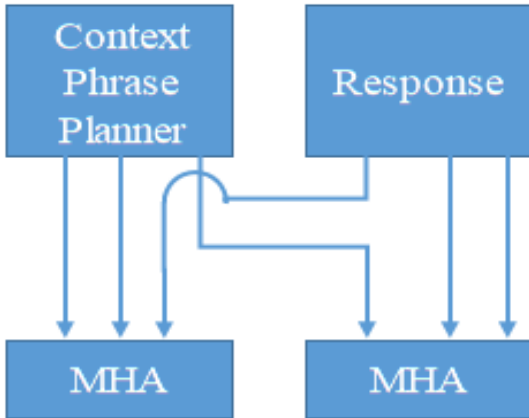


Figure 5: Mutual Information Projector contains 2 inter relations MHA where 1 MHA calculates the importance of Context phrase Planner to response and other calculates the importance of Response to Context phrase Planner.

the importance of  $C^{PP}$  that should be projected to R and the importance of R that should be incorporated on  $C^{PP}$ . This is done with the help of 2 Multi Head Attention modules. The first Multi Head attention module calculates the importance of  $C^{PP}$  on R by passing Query as  $C^{PP}$ , Key as R and Value as R. The scale dot product attention, followed by softmax calculates the affinity of  $C^{PP}$  and R to create matrix  $\epsilon \in \mathbb{R}^{p \times r}$ . The weighted

affinity matrix will be multiplied with Value matrix to get impact of  $C^{PP}$  on R for  $R_{PP} \in \mathbb{R}^{p \times d}$ . Similarly we use another Multi head Attention to calculate the importance of R on  $C^{PP}$ . Here we use Query as R, Key as  $C^{PP}$  and Value as  $C^{PP}$ . The scale dot product attention, followed by softmax calculates the affinity of  $C^{PP}$  and R to create matrix  $\epsilon \in \mathbb{R}^{r \times p}$ . The weighted affinity matrix will be multiplied with Value matrix to get impact of R on  $C^{PP}$  as  $C_R^{PP} \in \mathbb{R}^{r \times d}$ . The same is applied for Context vector C and Response vector R, Persona vector P and Response vector R and Persona Phrase Planner and Response vector R to get  $C_R \in \mathbb{R}^{r \times d}$ ,  $R_C \in \mathbb{R}^{c \times d}$ ,  $P_R \in \mathbb{R}^{r \times d}$ ,  $R_P \in \mathbb{R}^{p \times d}$ ,  $R_{PPP} \in \mathbb{R}^{p \times d}$  and  $P_R^{PP} \in \mathbb{R}^{r \times d}$ . All these learnings are passed through max pooling layer which takes the maximum along the sequence dimension to generate  $R_{C^{PP}max} \in \mathbb{R}^d$ ,  $C_{Rmax}^{PP} \in \mathbb{R}^d$ ,  $C_{Rmax} \in \mathbb{R}^d$ ,  $R_{Cmax} \in \mathbb{R}^d$ ,  $R_{PPPmax} \in \mathbb{R}^d$ ,  $P_{Rmax}^{PP} \in \mathbb{R}^d$ ,  $P_{Rmax} \in \mathbb{R}^d$ ,  $R_{Pmax} \in \mathbb{R}^d$ .

#### 4.5 Dot Product

We concatenate the response importance learnings as  $R_f = R_{C^{PP}max}; R_{Cmax}; R_{PPPmax}; R_{Pmax} \in \mathbb{R}^{4d}$  and  $U_{xf} = C_{Rmax}^{PP}; C_{Rmax}; P_{Rmax}^{PP}; P_{Rmax} \in \mathbb{R}^{4d}$ . Then we use dot product to calculate final matching score as shown in equation 2.

$$f(U_X, P_X, Y_X) = dot(R_f, U_{xf}) \quad (2)$$

Model is optimized by reducing the cross entropy loss between target true candidate and final matching score.

## 5 Experiments

This section explains the baseline models, experimentation and model comparisons.

### 5.1 Baseline Models

We compare our models with BoW, HLSTM, Bi-encoder, Co-BERT for PEC dataset.

- **BoW**: tri-encoder architecture with average word embedding for context, response and persona.
- **HLSTM**: Makes use of utterance level Bi-LSTM and context level Bi-LSTM. Also all encoders share same utterance level Bi-LSTM.

Models	Happy				OffMyChest				All			
	R@1	R@10	R@50	MRR	R@1	R@10	R@50	MRR	R@1	R@10	R@50	MRR
BoW	10.2	45.6	85.2	21.8	13.9	51.6	87.1	26.2	15.4	52.9	86.7	27.4
HLSTM	15.7	53.6	91.6	28.1	17.6	55.7	91.8	30.2	22.2	63.0	94.8	35.2
DIM	31.3	67.0	95.5	43.0	40.6	72.6	96.4	51.2	39.3	74.6	97.3	50.5
Bi-encoder	32.4	71.3	96.5	45.1	42.4	78.4	97.6	54.5	42.4	78.4	97.6	54.5
Poly-encoder	33.7	72.1	96.7	46.4	43.4	79.3	97.7	55.3	42.3	79.2	98.1	54.4
Co-Bert	36.2	73.0	96.9	48.4	47.0	79.7	97.8	58.0	45.1	80.5	98.3	56.7
Our Model	<b>37.8</b>	<b>75.2</b>	<b>97.5</b>	<b>49.6</b>	<b>48.1</b>	<b>81.2</b>	<b>98.3</b>	<b>59.4</b>	<b>46.2</b>	<b>81.2</b>	<b>98.7</b>	<b>57.4</b>

Table 2: Comparison of state of the art models for Happy, OffMyChest and All

- **DIM**: Makes use of fine grained matching and hierarchical aggregation to learn rich matching information.
- **Bi-encoder**: State of the art BERT based model for empathetic response selection.
- **PolyEncoder**: gains an understanding of latent attention codes for finer grained matching.

## 5.2 Evaluation Metrics

We follow the same evaluation metrics proposed by Zhong et al. (2020). We evaluate the models on Recall@k where k candidates needs to be selected from C samples. We abbreviate it as R@k. We use k as 1, 20 and 50. We use C as 100. We also measure Mean reciprocal Rank (MRR). MRR calculates the mean of reciprocal of the rank of the correct response. The rank of the correct response is calculated by finding the position of the correct response id inside the list of predicted response ids sorted in decreasing order by probabilities.

## 5.3 Baseline Comparison

Table 2 shows the experimentation results of the models for PEC dataset namely Happy, Offmychest and All.

From the above table, we are able to observe that sentence representation is one of the most important critical factor for response selection. Another important factor that is noticeable is the fine grained matching logic which aides in better response selection. Sentence representation information importance is visible between BoW, HLSTM, DIM and Bi-encoder where Bi-encoder model has outperformed other models. CoBert has performed best amongst all the other models (except our model) mainly because of first-order

and second-order multi-hop co-attention calculation which aided in better response, user context pair with the help of persona. Our model is able to defeat Cobert model because of additional phrase level projection of user context and persona on response. In addition, the mutual information extractor module aided in better relationship between persona, user context to response which enhanced the response selection.

## 6 Conclusion

We are able to observe that additional phrase level information flow, both for user context as well as persona, aided in better relationship building between response and context as well response and persona which in turn aided in better response selection. In addition the Multi head attention aided in multi phrase positional information capture which resulted in better learning representation.

## References

- Orianna Demasi, Yu Li, and Zhou Yu. 2020. A multi-persona chatbot for hotline counselor training. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 3623–3636.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Samuel Humeau, Kurt Shuster, Marie-Anne Lachaux, and Jason Weston. 2019. Poly-encoders: Transformer architectures and pre-training strategies for fast and accurate multi-sentence scoring. *arXiv preprint arXiv:1905.01969*.
- Yanran Li, Ke Li, Hongke Ning, Xiaoqiang Xia, Yalong Guo, Chen Wei, Jianwei Cui, and Bin Wang. 2021. Towards an online empathetic chatbot with

- emotion causes. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2041–2045.
- Zhaojiang Lin, Peng Xu, Genta Indra Winata, Farhad Bin Siddique, Zihan Liu, Jamin Shin, and Pascale Fung. 2020. Caire: An end-to-end empathetic chatbot. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13622–13623.
- K Liu and Rosalind W Picard. 2005. Embedded empathy in continuous, interactive health assessment. In *CHI Workshop on HCI Challenges in Health Assessment*, volume 1, page 3.
- Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv preprint arXiv:1811.00207*.
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M Smith, et al. 2020. Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637*.
- Haoyu Song, Wei-Nan Zhang, Yiming Cui, Dong Wang, and Ting Liu. 2019. Exploiting persona information for diverse generation of conversational responses. *arXiv preprint arXiv:1905.12188*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Weixuan Wang, Xiaoling Cai, Chong Hsuan Huang, Haoran Wang, Haonan Lu, Ximing Liu, and Wei Peng. 2021. Emily: Developing an emotion-affective open-domain chatbot with knowledge graph-based persona. *arXiv preprint arXiv:2109.08875*.
- Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. 2020. Towards persona-based empathetic conversational models. *arXiv preprint arXiv:2004.12316*.
- Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, 46(1):53–93.