Integrating Background and General Knowledge for Dialogue Generation

Hongsong Wang, Haoxian Ye(✉), and Jiazhan Li

South China Normal University, Foshan 528225, China
850615324@qq.com

Abstract. The traditional sequence-to-sequence model generates responses that are smooth but empty in their content. Background-based dialogue is one solution that uses the context's unstructured knowledge to generate informative responses. The key point of background-based dialogue is knowledge extraction, but some conversations have poor performance in knowledge selection due to insufficient information. At the same time, to improve the satisfaction of the responses, this paper can enhance the amount of conversational knowledge while allowing the model to carry some emotional awareness by selecting external knowledge sources with emotional information. In this paper, we introduce the CEC model, which utilizes graph attention and a double-matching matrix for the selection of external and background knowledge. The generation process is conducted within each decoding step, considering the selected knowledge's content. We conduct experiments on the Holl-E dataset. According to the experimental results, our model CEC outperforms the previous model in terms of performance.

Keywords: Background Based Conversations · Dialogue Systems · Knowledge-Enhanced

1 Introduction

Conversational systems are currently a hot topic in NLP research. Studies [1] show that 80% of enterprises will be equipped with chatbots (conversational systems) by the end of 2021, and the market will grow to $9.4 billion by 2024.

A core definition of a text generation task is the capability to generate an expected output sequence using a provided input sequence, often known as a sequence-to-sequence task. Thanks to the development of deep learning [2], numerous deep learning networks have been suggested for application in dialogue systems, encompassing recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformers. These networks offer diverse approaches to handle the complexities of dialogue generation. Although there are already many models that already perform well, traditional sequence-to-sequence models do not understand discourse well and generate responses that tend to be general replies because the input text itself contains a smaller amount of knowledge in

M. Zhang et al. (Eds.): CCF NCCA 2023, CCIS 1959, pp. 127–138, 2024.
https://doi.org/10.1007/978-981-99-8764-1_10

the traditional sequence-to-sequence models, for example, the newly proposed sequence-to-sequence model [3] can dynamically capture the range of local contexts and can better extract semantic information but can only generate meaningless responses such as "I don't know" due to the lack of external knowledge. Some models introducing external knowledge [4,5] have emerged to address this problem. Some studies demonstrate that introducing external knowledge can enhance performance, e.g., Huang [6] introduced a knowledge graph that can answer 10% more responses than the original model; introducing a knowledge graph in story generation also helps to understand the storyline [7].

To address the aforementioned limitations [8], researchers have proposed background-based dialogue approaches. These methods aim to generate sensible and informative responses by utilizing a combination of background knowledge (unstructured information) and input dialogue. The objective is to produce responses that are contextually relevant and provide valuable information to enhance the conversation. Knowledge selection is one of the most critical modules in background-based dialogues, which requires identifying the appropriate knowledge from the background knowledge based on the conversation, which will directly affect the quality of the generated response.

Using appropriate external knowledge augmentation also enables model generated responses to be implicitly emotional because, like humans, machines need to rely on experience and external knowledge to express implicit emotions [9,10]. If a dialogue system has some empathy, it can generate more appropriate and fluent responses [11,12].

Background-based dialogue research is one of the classifications of external knowledge enhancement research, and the advantage of background-based dialogue over traditional non-knowledge enhancement methods is that unstructured external knowledge is used [13]. Recent studies have shown that the coverage of a single knowledge source is not sufficient [14], and the results of several studies have shown that using more knowledge sources can improve the performance of knowledge-enhanced dialogue models [14,15].

To tackle the challenges mentioned above, this paper introduces a common sense emotional context enhanced dialog model (CEC). To fully utilize all the information (session history, background knowledge, external knowledge), a double-matching approach is proposed to fuse the information for knowledge selection. First, the model encodes the conversation history and background knowledge separately and then uses double matching to obtain the relevance weights among conversation history, background, and sentiment. After knowledge selection gets the knowledge topic transformation vector and combines it with graph feature representation to generate naturally flowing and informative responses.

In this paper, we perform an experimental analysis of CEC on Holl-E [16]. The experimental results show that CEC significantly outperforms the baseline model in machine evaluation, with stronger performance in knowledge selection and the ability to generate more appropriate responses.

The summarized contributions of this paper are as follows:

(1) We propose a dialogue model for emotional knowledge enhancement (CEC). By introducing common-sense knowledge and emotional-emotional information, the information implicit in the session is taken into account when making knowledge selection, enhancing the accuracy of knowledge selection and generating more appropriate responses.
(2) We introduced external knowledge through composition and proposed a dual matching matrix to integrate conversations with knowledge from various sources to construct an affective topic guidance vector to guide response generation.

2 Model

This model aims to combine external knowledge based on background knowledge to improve the rationality of knowledge selection and generate responses that conform to backward and forward logic. Formally this paper gives the symbolic definition. Given a session $C = \{c_1, c_2, c_3, c_4, ..., c_{\|C\|}\}$, where c_n represents the n^{th} word, similarly for unstructured background knowledge $K = \{k_1, k_2, k_3, k_4, ..., k_{\|K\|}\}$, where k_n represents the n^{th} word. This model generates responses $= \{r_1, r_2, r_3, r_4, ..., r_{\|R\|}\}$ based on conversation and background knowledge. The overall model framework is shown in Fig. 1.

In this section, the four modules that make up the entire model are presented.

(1) Background Context Encoder
Using two independent encoders, a given history session and background knowledge are encoded, and then an aggregation operation is performed to obtain the history session vector H_C and background knowledge vector H_K.
(2) Emotional context graph and graph encoder
ConceptNet and NRC_VAD, two sentiment enhancement libraries, are used to form a sentiment context map G with session history C. Then it is input into the graph encoder to obtain the graph feature representation H_G.
(3) Knowledge Selection
Based on the double-matching matrix, the historical session H_C, graph feature representation H_G and background knowledge representation H_K are used for matching operations.
(4) Response decoder
The knowledge topic transformation vector $H^s_{GC\to k}$ and the graph feature representation H_G are stitched together to obtain the emotional topic guidance vector H^g_{GCK}, and the module performs vocabulary generation based on this vector.

The whole process can be summarized as putting the history session C and the background knowledge K into the context encoder. The session history is combined with the knowledge base to obtain the feature representation through the graph encoding layer. Then, the knowledge selection module chooses the relevant information, which then guides the response decoder in generating the final response.

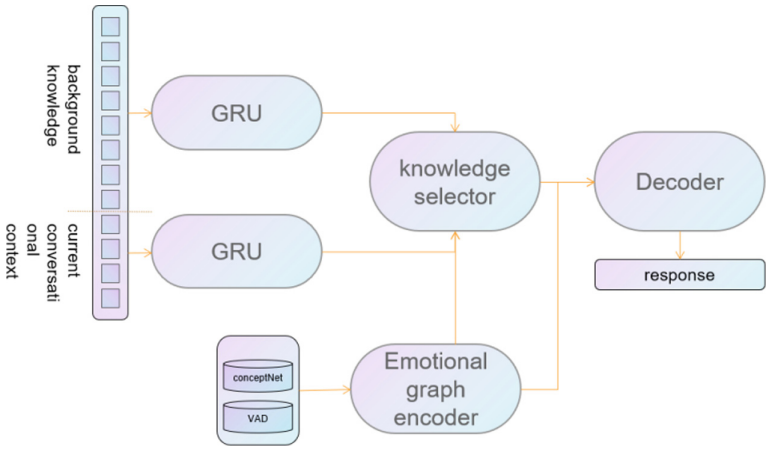

Fig. 1. The Overview of CEC

2.1 Background Context Encoder

We use two independent bidirectional GRUs to encode session history C and background knowledge K, respectively, to obtain $h_C = \{h_{c_1}, h_{c_2}, h_{c_3}, h_{c_4}, ..., h_{c_{\|C\|}}\}$ and $h_K = \{h_{k_1}, h_{k_2}, h_{k_3}, h_{k_4}, ..., h_{k_{\|K\|}}\}$.

$$h_{c_t} = BIGRU(e(c_t), h_{c_{t-1}}) \tag{1}$$

The parameters of these two GRUs are independent. We perform a highway transformation of these two vectors separately with the output of each layer of the bidirectional GRU to obtain a historical session H_C and background knowledge representation H_K for the next matching operation.

$$H_{k_t} = g_k(W_1[h_{k_t}, h_{X_{\|x\|}}] + b) + (1 - g_k)tanh(W_{n1}[h_{k_t}, h_{X_{\|x\|}}] + b) \tag{2}$$

$$g_k = \sigma(W_g[h_{k_t}, h_{X_{\|x\|}}] + b) \tag{3}$$

2.2 Emotional Context Graph and Graph Encoder

In this module, we use ConceptNet and NRC_VAD combined with Dialogue C to construct the sentiment graph G.Inspired by Li et al., we construct a series of candidate tuples $T_i = \{t_i^k = (c_i, r_k^i, x_k^i, s_k^i)\}_{k=1,2,3,...,K}$ for each non-deactivated word of the dialogue combined with the keywords in ConceptNet. The candidate tuples are filtered according to the following rules: (1) Only the tuples with confidence scores greater than 0.1 (s_k^i>0.1) are retained. (2) Use NRC_VAD to calculate the sentiment intensity value ($\mu(x_i^k)$) and select the k tuples with the highest scores. We compose the composition based on candidate tuples and dialogues, and the rules are as follows: (1) Two adjacent words will point to the next word in order. (2) The selected candidate sentiment words will point to his keywords (c_i).

For the graph encoder, we need to transform each vertex of the sentiment graph G. Similar to the transformer model, our proposed model utilizes both the position embedding layer and the word embedding layer. Additionally, we incorporate the vertex state embedding to further enhance the model's performance. Therefore, the vector representation of the entire vertex consists of three embeddings:

$$v_i = E_w(v_i) + E_p(v_i) + E_v(v_i) \tag{4}$$

Then go to the multiheaded graph attention mechanism to obtain a deeper representation of each vertex.

$$\hat{v_i} = v_i + \|_{n=1}^{H} \sum_{j \in A_i} a_{ij}^n W_v^n v_j \tag{5}$$

$$a_{ij}^n = a^n(v_i, v_j) \tag{6}$$

where H represents the number of multiheads. A_i is the adjacency matrix of G, and a^n is the self-attentive module of each head. To obtain a global contextual representation, after a multiheaded graph attention layer, we use the encoding layer of the transformer for global modelling to obtain a sentiment contextual graph representation $h_g = \{\bar{v_i}\}$.

$$h_i^l = LayerNorm(\hat{v}_i^{l-1} + MHA(\hat{v}_i^{l-1})) \tag{7}$$

$$\bar{v}_i^l = LayerNorm(h_i^l + FNN(h_i^l)) \tag{8}$$

where l represents the l^{th} layer of the coding layers, MHA represents the multiheaded attention module, and FNN represents a two-layer feedforward network with ReLU as the activation function.

2.3 Knowledge Selection

This module uses a double-matching matrix, the first of which is constructed using the potential representation of historical sessions H_C and background knowledge H_K derived in Sect. 3.1.

$$M_{kc}[i,j] = V_M tanh(W_{m_1} H_{k_i} + W_{m_2} H_{k_l}) \tag{9}$$

where V_M are the learnable vectors, and W_{m_1} and W_{m_2} are the learnable parameters. To match the sentiment map features with the background features, we first need to use a multilayer perceptron (MLP) to transform the h_g derived in Sect. 3.2 to obtain the H_G.

$$H_G = MLP(h_g) \tag{10}$$

We use a similar approach to obtain the second matching matrix M_{kg}:

$$M_{kg}[i,j] = V_{Mg} tanh(W_{mg_1} H_{k_i} + W_{mg_2} H_{g_l}) \tag{11}$$

For this dual matching matrix, we use the maximum pooling layer along the x-axis to obtain two perceptual background weight feature representations; each

element in the feature represents the weight of relevance to the context, with higher weights representing greater relevance:

$$W_{C\to K} = \max_x(M_{kc}) \tag{12}$$

$$W_{G\to K} = \max_x(M_{kg}) \tag{13}$$

Finally, we combine these two perceptual contextual weight feature representations to obtain the emotional contextual perceptual weight vector $W_{CG\to K}$. Although this vector captures the relationship between the context, sentiment map and background, it only considers the distribution of relationships in the word direction. It lacks a global perspective to derive the probability distribution of knowledge selection properly. Drawing inspiration from GLKS, we adopt sliding windows for the purpose of global knowledge selection. In the knowledge selection module, we employ the "m-size unfold and sum" and "m-size unfold and attention" operations. The former operation obtains the global semantic information, and the latter operation obtains the global attention weights.

In the first operation "m-size unfold and sum", we can obtain a sliding semantic representation by the following formula:

$$W'_{G\to K} = ([W'_{G\to K}]_{0:m}, ..., [W'_{G\to K}]_{N:N+m}, ...) \tag{14}$$

$$[W'_{G\to K}]_{N:N+m} = \sum_{i=N}^{N+m} W_{CG\to K}[i] \tag{15}$$

For the second operation, we use the "m-size unfold and attention" operation for the last layer of the background knowledge representation h_k to obtain the global attention H'_K:

$$H'_K = ([h'_K]_{0:m}, ..., [h'_K]_{N:N+m}, ...) \tag{16}$$

$$[h'_K]_{N:N+m} = \sum_{i=N}^{N+m} a_i h_K[i] \tag{17}$$

$$a_i = att(h_{c_{\|C\|}}, [h_{k_m} ... h_{k_{N+m}}]) \tag{18}$$

where a_i represents the attention weight of the session versus the background knowledge. Then we combine background knowledge K to generate knowledge topic transformation vectors $H^s_{CG\to k}$:

$$H^s_{CG\to k} = \sum_N P(K_N : K_{N+m}|C)[h'_K]_{N:N+m} \tag{19}$$

$$P(K_N : K_{N+m}|C) \propto \text{softmax}([W'_{CG\to K}]_{N:N+m}) \tag{20}$$

2.4 Response Decoder

During each decoding time step, the response decoder carries out a splicing operation utilizing the knowledge topic transformation vector $H^s_{GC\to k}$ and H_G in order to acquire the sentiment topic guidance vector H^g_{GCK}. Based on this vector, the response decoder obtains the probability of generating from the vocabulary and the probability of intercepting directly from the background and goes through a gate mechanism to finally decide how to generate.

First, we connect the decoded status code to $H^s_{GC\to k}$ and H_G:

$$H^g_{GCK} = [H^s_{GC\to k}, H_G, e(r_{t-1})] \tag{21}$$

where $e(r_{t-1})$ denotes the vector generated from the previous time step. Then we use the attention module to perform an attention operation on the knowledge-emotion topic vector with the background knowledge K, which will give us the background guidance vector $\bar{h}_{K_t}$. Similarly, we use the attention module to perform an attention operation with the session history C to obtain the session guidance vector $\bar{h}_{C_t}$:

$$\bar{h}_{K_t} = \sum_{i=1}^{\|K\|} a_{K_i} h_{K_i} \tag{22}$$

$$a_{K_i} = attention(H^g_{GCK_t}, h_K) \tag{23}$$

Then we join the two guidance vectors with the knowledge-emotion topic vector and use a readout layer to obtain an overall feature vector $\bar{h}_{r_t}$.

$$\bar{h}_{r_t} = \text{readout}(H^g_{GCK_t}, \bar{h}_{K_t}, \bar{h}_{C_t}) \tag{24}$$

Putting feature vectors $\bar{h}_{r_t}$ into linear layers with a softmax function to obtain the probability of generating words from the vocabulary:

$$P_v(r_t) = \text{softmax}(W_v \bar{h}_{r_t}) \tag{25}$$

For $P_k(r_t)$, we use an attention module for background knowledge to learn the intercept's start position pointer and end position pointer.

$$P_k(r_t) = \text{attention}(H^g_{GCK_t}, h_K) \tag{26}$$

Finally, we combine $P_v(r_t)$ and $P_k(r_t)$as follows:

$$P(r_t) = gP_v(r_t) + (1-g)P_k(r_t) \tag{27}$$

3 Experiments

3.1 Implementation Details

The word embedding size is configured as 300, while the hidden layer size is set to 256. The number of vocabulary words is limited to approximately 26,000, the length of the conversation history is limited to 65, and the length of the background knowledge is limited to 256. The optimizer uses Adam, and the batch size is set to 32. The entire model was trained for 20 rounds, and the highest scores were taken for comparison in the evaluation phase.

3.2 Datasets

To ensure a more accurate representation of the model's performance, we opted for Holl-E as the benchmark for our comparative experiments. The number of samples in the datasets is shown below.

Holl-E: This is a dataset with the correct labels containing background knowledge and the correct knowledge selection labels. The dataset focuses on the movie part, two people having a conversation about the movie plot, and each response will be a change or copy of the background knowledge to reply. The background knowledge consists of four parts: movie plot, reviews, professional commentary, and fact sheets related to the movie. The experiments in this paper use two versions of Holl-E: oracle background and mixed-short background. We partition the dataset into three according to its original partitioning method, with the training set containing 34486 samples, the validation set containing 4388 samples, and the test set containing 4318 samples (Table 1).

Table 1. Dataset sizes.

Datasets	train	validation	test
Holl-E	34486	4388	4318

3.3 Evaluation Metrics

In this paper, the evaluation metrics chosen for machine evaluation are ROUGE-1, ROUGE-2, and ROUGE-L. Since dialogue responses are generated using background knowledge, previous studies have shown that these metrics are consistent with BLEU. Therefore, employing these metrics would provide a comprehensive assessment of the model's performance.

3.4 Results

The experimental results are shown in the table. Table 2 and Table 3 subtables represent the results of the oracle background and oracle mixed-short background in Holl-E.

The experimental results demonstrate that CEC outperforms the baseline model across all metrics, providing evidence that CEC can enhance knowledge selection performance and generate more appropriate responses. Compared with BiDAF (extraction-based generation method), which benefits from combining extractive and generative approaches, CEC generates more reasonable and natural responses while using background knowledge well. RefNet uses span annotations, while CEC does not need additional annotation information and can better locate the correct background knowledge location. This is because we use guidance vectors and learn two pointers to locate background knowledge in the

generation process. Compared with AKGCM, which fuses knowledge graphs, and GLKS, which used to have the highest knowledge selection scores, CEC connects structured knowledge in a more rationalized way and, simultaneously, can significantly increase the performance of knowledge selection. Our advantage lies in the utilization of the double-matching matrix, which effectively fuses structured and unstructured knowledge to enhance knowledge information. This approach leads to a substantial improvement in knowledge selection performance while ensuring that empty responses are not generated. In both versions of the Holl-E dataset, we can observe that the same model in both tables (including CEC) performs better in the oracle mixed-short background version than in the oracle background. This is because the knowledge in the oracle background contains only one source, which has less information than in the oracle mixed-short background. Additionally, compared to the magnitude of the improvement of the baseline model in both datasets, we can observe that the improvement of CEC is not very significant. This may be because the knowledge richness in the dataset can already reach a standard level, and the added knowledge does not enhance it much. The above experimental analysis proves that it is essential to include additional knowledge in a session. Choosing the right way to integrate different knowledge types can improve response quality.

Table 2. Results on oracle background (256-word)

Methods	SR(%)			MR(%)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
no background Seq2Seq	27.15	9.56	21.48	30.91	11.85	24.81
oracle background						
GTTP [16]	29.82	17.33	25.09	35.08	22.05	30.06
BiDAF [17]	39.68	31.72	35.91	46.49	40.58	42.64
CaKe [18]	42.82	30.37	37.48	48.65	36.54	43.21
RefNet [19]	42.87	30.73	37.11	49.64	38.15	43.77
GLKS [20]	43.75	31.54	38.69	50.67	39.20	44.64
CEC(ours)	44.47	32.03	39.28	50.73	39.22	45.35

3.5 Ablation Study

Since the performance of CEC is consistent across datasets, the experiments in this section are conducted in the oracle background for ablation experiments only. We will analyze three aspects: (1) w/o emo_embedding+emo_match: No sentiment matching matrix and sentiment vector. (2) w/o emo_match:No sentiment matching matrix. (3) w/o emo_embedding:No emotion vector.

The experimental results are shown in Table 4. Both the sentiment matching matrix and the sentiment vector impact the final generation, and removing

Table 3. Results on mixed-short background (256-word)

Methods	SR(%)			MR(%)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
GTTP	30.77	18.72	25.67	36.06	23.7	
BiDAF	38.79	32.91	35.09	43.93	39.5	40.12
CaKe	41.26	29.43	36.01	45.81	34	40.79
AKGCM [21]		29.29	34.72			
PostKS [22]	27.52	9.21	21.23	31.57	12.55	25.15
SKT [23]	35.28	21.74	30.06	41.68	28.3	36.24
RefNet	41.33	31.08	36.17	47	36.5	41.72
DukeNet [24]	36.53	23.02	31.46	43.18	30.13	38.03
GLKS	44.52	33.05	39.63	50.06	38.87	45.12
MIKe [25]	37.78	25.31	32.82	44.06	31.92	38.91
CEC(ours)	44.58	33.22	39.7	50.69	39.33	45.29

either will degrade performance. Second, the performance degradation is most obvious if we remove the sentiment matching matrix (w/o emo_match) alone for knowledge selection. This demonstrates that adding additional sentiment-structured knowledge significantly improves the accuracy of knowledge selection and enhances model performance, possibly because the added knowledge is generated based on the current session and is, therefore, highly relevant and contains a more significant amount of valuable knowledge. Finally, to validate the effectiveness of the sentiment vector, we remove the sentiment vector (w/o emo_embedding) directly when combining the sentiment topic guidance vectors. The results demonstrate that adding sentiment vectors can improve the performance of the generation module, which means that sentiment vectors can provide additional sentiment information in addition to the session itself. It also improves the correctness of the selection knowledge when generating responses and making the responses more reasonable and justified.

Table 4. Ablation study

Methods	SR(%)			MR(%)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
w/o all	43.75	31.54	38.69	50.67	39.2	44.64
w/o match	43.91	31.52	38.73	50.65	39.21	45.17
w/o embedding	44.01	31.57	38.8	50.68	39.18	45.20
CEC(ours)	44.58	33.22	39.7	50.69	39.33	45.29

4 Conclusion

In this article, we introduce external knowledge by constructing a sentiment graph, generating a sentiment vector using graph attention, and then using a

matching matrix to combine the background knowledge with the sentiment vector to enhance both the precision of knowledge selection and the naturalness of response generation. The experimental results are better than all the baselines.

This paper introduces a sentiment knowledge base. Although it can improve the final response, the model does not explicitly model sentiment classification or recognition; therefore, this model can only restrict the generated responses to the session sentiment. To construct an empathic dialogue model, in the future, our work will focus on enhancing the model's capabilities in both emotion recognition and inference.

References

1. Abro, W.A., Aicher, A., Rach, N., Ultes, S., Minker, W., Qi, G.: Natural language understanding for argumentative dialogue systems in the opinion building domain. Knowl.-Based Syst. **242**, 108318 (2022)
2. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
3. Xu, M., Zeng, B., Yang, H., Chi, J., Chen, J., Liu, H.: Combining dynamic local context focus and dependency cluster attention for aspect-level sentiment classification. Neurocomputing **478**, 49–69 (2022)
4. Ghazvininejad, M., et al.: A knowledge-grounded neural conversation model. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
5. Zhong, P., Wang, D., Li, P., Zhang, C., Wang, H., Miao, C.: CARE: commonsense-aware emotional response generation with latent concepts. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 14577–14585 (2021)
6. Huang, L., Wu, L., Wang, L.: Knowledge graph-augmented abstractive summarization with semantic-driven cloze reward. arXiv preprint arXiv:2005.01159 (2020)
7. Guan, J., Wang, Y., Huang, M.: Story ending generation with incremental encoding and commonsense knowledge. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 6473–6480 (2019)
8. Zhou, K., Prabhumoye, S., Black, A.W.: A dataset for document grounded conversations. arXiv preprint arXiv:1809.07358 (2018)
9. Zhong, P., Wang, D., Miao, C.: Knowledge-enriched transformer for emotion detection in textual conversations. arXiv preprint arXiv:1909.10681 (2019)
10. 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)
11. Liu, S., et al.: Towards emotional support dialog systems. arXiv preprint arXiv:2106.01144 (2021)
12. Wang, L., et al.: CASS: towards building a social-support chatbot for online health community. Proc. ACM Hum. Comput. Interact. **5**(CSCW1), 1–31 (2021)
13. Wu, S., Wang, M., Li, Y., Zhang, D., Wu, Z.: Improving the applicability of knowledge-enhanced dialogue generation systems by using heterogeneous knowledge from multiple sources. In: Proceedings of the Fifteenth ACM International Conference on WEB Search and Data Mining, pp. 1149–1157 (2022)
14. Wu, S., Li, Y., Wang, M., Zhang, D., Zhou, Y., Wu, Z.: More is better: enhancing open-domain dialogue generation via multi-source heterogeneous knowledge. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 2286–2300 (2021)

15. Bai, J., Yang, Z., Liang, X., Wang, W., Li, Z.: Learning to copy coherent knowledge for response generation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 12535–12543 (2021)
16. Moghe, N., Arora, S., Banerjee, S., Khapra, M.M.: Towards exploiting background knowledge for building conversation systems. arXiv preprint arXiv:1809.08205 (2018)
17. Seo, M., Kembhavi, A., Farhadi, A., Hajishirzi, H.: Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603 (2016)
18. Zhang, Y., Ren, P., de Rijke, M.: Improving background based conversation with context-aware knowledge pre-selection. arXiv preprint arXiv:1906.06685 (2019)
19. Meng, C., Ren, P., Chen, Z., Monz, C., Ma, J., de Rijke, M.: RefNet: a reference-aware network for background based conversation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 8496–8503 (2020)
20. Ren, P., Chen, Z., Monz, C., Ma, J., de Rijke, M.: Thinking globally, acting locally: distantly supervised global-to-local knowledge selection for background based conversation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 8697–8704 (2020)
21. Liu, Z., Niu, Z.Y., Wu, H., Wang, H.: Knowledge aware conversation generation with explainable reasoning over augmented graphs. arXiv preprint arXiv:1903.10245 (2019)
22. Lian, R., Xie, M., Wang, F., Peng, J., Wu, H.: Learning to select knowledge for response generation in dialog systems. arXiv preprint arXiv:1902.04911 (2019)
23. Kim, B., Ahn, J., Kim, G.: Sequential latent knowledge selection for knowledge-grounded dialogue. arXiv preprint arXiv:2002.07510 (2020)
24. Meng, C., et al.: DukeNet: a dual knowledge interaction network for knowledge-grounded conversation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1151–1160 (2020)
25. Meng, C., Ren, P., Chen, Z., Ren, Z., Xi, T., Rijke, M.D.: Initiative-aware self-supervised learning for knowledge-grounded conversations. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 522–532 (2021)