Improving Dialogue Generation with Commonsense Knowledge Fusion and Selection

Dongjun Fu[1], Chunhong Zhang[2](✉), Jibin Yu[1], Qi Sun[2], and Zhiqiang Zhan[1]

[1] State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing 100876, China
{fudongjun,yujibin,zqzhan}@bupt.edu.cn

[2] Key Laboratory of Universal Wireless Communications, Ministry of Education,
Beijing University of Posts and Telecommunications, Beijing 100876, China
{zhangch,qisun}@bupt.edu.cn

Abstract. Knowledge-aware dialogue generation aims to generate informative and meaningful responses with external knowledge. Existing works are still insufficient to encode retrieved knowledge regardless of the dialogue context, which probably leads to the introduction of irrelevant information. In this paper, we propose a dialogue generation model named CKFS-DG, which filters out context-irrelevant and off-topic knowledge to reduce the influence of redundant knowledge. Specifically, we design a knowledge-enriched encoder and a topic fact predictor to improve the quality of fusion knowledge. For achieve the knowledge-enriched encoder, we put forward a context-knowledge attention mechanism to dynamically filter out irrelevant knowledge conditioned on context. For the topic fact predictor, we utilize the probability distribution on retrieved facts to retain on-topic knowledge. The experimental results on English Reddit and Chinese Weibo dataset demonstrate that CKFS-DG outperforms the state-of-the-art neural generative methods in knowledge utilization, and CKFS-DG could reduce the influence of irrelevant knowledge to generate reasonable responses.

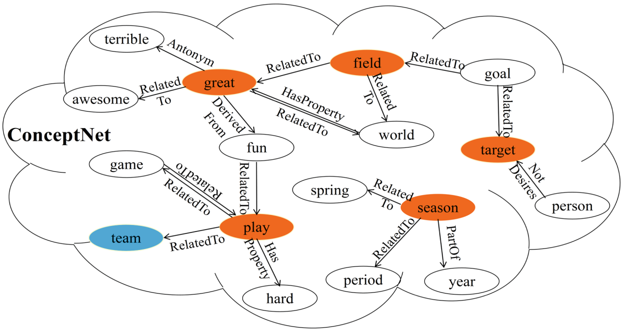
Keywords: Dialogue generation · Knowledge aware · Knowledge selection · Attention mechanism · Topic prediction

1 Introduction

The open-domain dialogue generation task is designed to generate a reasonable response for a given post. However, different from a human, a machine can merely extract limited information from the post and cannot associate the dialogue with background knowledge [1]. Consequently, it is difficult for a machine to completely comprehend the post and thus generate diverse and informative response. To address this problem, some prior studies begin to enhance the performance of dialogue generation by external knowledge [2]. In recent years,

G. Memmi et al. (Eds.): KSEM 2022, LNAI 13368, pp. 93–108, 2022.
https://doi.org/10.1007/978-3-031-10983-6_8

many researches integrate commonsense knowledge graphs as additional representations, and generate responses conditioned on both the post and the extra knowledge.



Post: that would be great if they had to play at target field the first season .
Ground-truth Response: its about time a mn team won there .
Generated Response 1 : well that would be a great idea . and that 's pretty awesome
Generated Response 2 : well , if the hawks had to play at target , they 'd be a great team .

Fig. 1. An example of dialogue generation, including a post-response pair and retrieved facts from knowledge graph based on the entity words (orange words) in the post. The fact whose tail entity (blue words) appears in the truth response is the golden fact. The two generated responses are given without/with considering the topic of the dialogue. (Color figure online)

To fully leverage the retrieved facts, Seq2Seq [3] framework with knowledge-aware mechanism is proposed to integrate the retrieved facts in encoder and generator module [1,2]. In encoder module, retrieved knowledge facts are encoded into additional semantic representation, which facilitates the understanding of the post. In generator module, retrieved facts are read as one of the word sources for response generation. However, most frameworks to retrieve knowledge facts do not consider specific dialogue context, which probably results in introducing noise in knowledge integration. On the one hand, some candidate facts retrieved by off-topic entity words in the post may be redundant. As illustrated in Fig. 1, generated response 1 is generic because the model focuses on the entity “great”, and generated response 2 is relatively reasonable because of noting “play” that is the topic of this dialogue. Obviously, topic entities are essential and retrieved facts through them may be more meaningful for developing conversation [4]. On the other hand, an entity may have multiple meanings, but only one specific meaning is involved in a particular context. Some retrieved knowledge facts based on the multi-meaning entity can be irrelevant to the current dialogue [5]. If irrelevant facts are encoded in knowledge integration, it might introduce redundant information for response generation. Therefore, we argue that it will be necessary to fuse relevant knowledge facts in post representation and select appropriate topic facts in response generation.

To address the aforementioned challenges, in this paper, we propose CKFS-DG (a model for dialogue generation with **C**ommonsense **K**nowledge **F**usion and **S**election) based on the Seq2Seq framework with two major knowledge-aware components, (1) a *knowledge-enriched encoder* that fuses filtered knowledge as additional semantic representation of the post, and (2) a *topic fact predictor* that predicts topic entities and facts, which facilitates the selection of appropriate knowledge facts for response generation. The motivation to design the knowledge-enriched encoder is to enhance the semantic of the post by combining relevant knowledge representations. Contextual knowledge attention mechanism is designed as a filter to dynamically filter out irrelevant knowledge based on the contextual vector of the post. For selecting appropriate topic facts, topic fact predictor is introduced to generate topic fact probability distribution over the retrieved facts, whose probability is adopted to guide word selection in response generation. We evaluate CKFS-DG on English Reddit and Chinese Weibo dataset [5] to demonstrate its effectiveness over the state-of-the-art dialogue generation methods. Our contributions are summarized as follows:

- We propose a knowledge-enriched encoder with context-knowledge attention mechanism to dynamically filter out irrelevant knowledge and enhance the semantic of the post by combining external knowledge representations;
- We design a topic fact predictor to generate topic fact distributions over the retrieved facts, which facilitates accurate knowledge selection;
- Experiments on Reddit and Weibo demonstrate the effectiveness of the proposed method on benchmarks of knowledge-aware dialogue generation.

2 Related Works

Knowledge-aware dialogue generation aims to generate informative and meaningful responses with external knowledge, such as additional texts [6] or knowledge graphs [4,7]. CCM [2] first applies a large-scale commonsense knowledge graph to facilitate the generation of a response with one-hop graph attention mechanisms. Some studies consider that the multi-hop graph is likely to contain more informative knowledge [1,4]. However, these models encode all retrieved knowledge to a representation, ignoring that the retrieved knowledge may contain irrelevant information that are useless for dialogue generation.

Hence, knowledge selection module that could select the appropriate knowledge gains much attention in knowledge-aware dialogue generation [8]. Generally, existing methods for combining knowledge selection and response generation can be grouped into two categories: a joint way and a pipeline way [9]. The joint approaches integrate knowledge selection into the generation process, that consistently select knowledge related to the current decoding step [9,10]. The joint ways result in the decoder being designed more complexly. The pipeline approaches separate knowledge selection from generation [11,12]. Some studies adopt attention mechanism [13,14] to filler out irrelevant knowledge, but could not utilize the actual knowledge as supervised training. ConKADI [5] utilizes the posterior knowledge distribution over the retrieved facts to select felicitous facts for generation. However, these works hardly consider the role of topic entities for generating on-topic responses. Different from some works attempting to capture the topic in the dialogue [4,15], we predict topic words and facts based on posterior information in responses, that is proven useful for knowledge selection.

In this work, we adopt a pipeline approach to focus on the knowledge selection easily. Different from previous works, we design context-knowledge attention mechanism to filter out irrelevant knowledge based on context, and a topic fact predictor as posterior knowledge selection module for reducing the influence of irrelevant knowledge in dialogue generation.

3 Methodology

3.1 Task Formulation and Model Overview

Knowledge-aware dialogue generation is defined as given a post $X = (x_1, ..., x_n)$ and a set of candidate knowledge facts $F = \{f_1, f_2, .., f_N\}$ to generate a response $Y = (y_1, ..., y_m)$. The words in the post X can be divided into entity words and common words. Candidate facts are retrieved from knowledge graph based on the entity words in the post [2]. A candidate fact f_i is formally a triple $< h_i, r_i, t_i >$, including head entity, relation and tail entity. In particular, knowledge aware dialogue generation targets to generate a response which can not only express consistent semantics as the post, but also embody the entity words contained in the candidate knowledge facts explicitly. The goal of training is to maximize the posterior probability of generating the truth response $\sum_{(X,Y,F)\in\mathcal{D}} \frac{1}{|\mathcal{D}|} p(Y|X, F)$.

The overview of the proposed model is shown in Fig. 2, which is consists of four components. First, *Context Encoder* encodes an utterance into contextual representation. Second, *Knowledge-enriched Encoder* encodes the post by fusing the filtered knowledge. Context-knowledge attention mechanism is proposed to dynamically filter out irrelevant facts in word-level based on the context. Third, *Topic Fact Predictor* calculates the topic fact probability distributions over retrieved facts to guide the generation. Finally, *Response Generator* generates a response by selecting from vocabulary, entity words, and copied words.

3.2 Context Encoder

The context encoder extracts information from the utterance by encoding the sequence into contextual representations, with bi-directional GRU network [16]:

$$\begin{aligned} \mathbf{h}_{\mathrm{t}}^{\mathrm{f}} &= GRU^{f}\left(\mathbf{h}_{\mathrm{t}-1}^{\mathrm{f}}, \mathbf{x}_{\mathrm{t}}, \mathbf{e}_{\mathbf{x}_{\mathrm{t}}}\right) \\ \mathbf{h}_{\mathrm{t}}^{\mathrm{b}} &= GRU^{b}\left(\mathbf{h}_{\mathrm{t}+1}^{\mathrm{b}}, \mathbf{x}_{\mathrm{t}}, \mathbf{e}_{\mathbf{x}_{\mathrm{t}}}\right) \end{aligned} \tag{1}$$

where $\mathbf{x}_{\mathrm{t}} \in \mathbb{R}^{K}$ is the word embedding corresponding to x_t. To enhance semantics, we add the matched entity embedding vector $\mathbf{e}_{\mathbf{x}_{\mathrm{t}}} \in \mathbb{R}^{d_e}$ of x_t, which will be a d_e-dimensional zero vector if x_t is a common word. The contextual state of the post is denoted as $H^x = (\mathbf{h}_1^{\mathrm{x}}, ..., \mathbf{h}_{\mathrm{n}}^{\mathrm{x}})$, and $\mathbf{h}_{\mathrm{t}}^{\mathrm{x}} = \left[\mathbf{h}_{\mathrm{t}}^{\mathrm{f}}; \mathbf{h}_{\mathrm{t}}^{\mathrm{b}}\right]$.

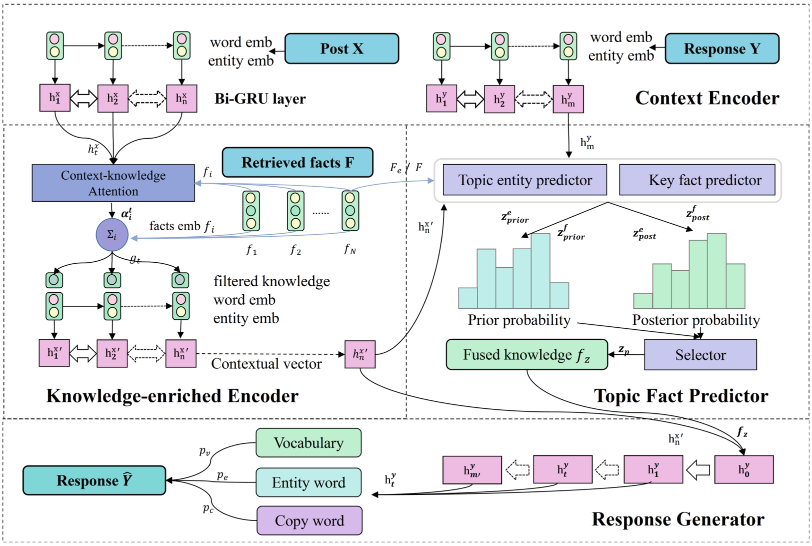


Fig. 2. Overall architecture of the proposed CKFS-DG. It contains four parts: Context Encoder, Knowledge-enriched Encoder, Topic Fact Predictor and Response Generator.

3.3 Knowledge-Enriched Encoder

The knowledge-enriched encoder is designed to encode the post by combining retrieved knowledge as additional semantic representations, which facilitates the understanding of a post. Apparently, as shown in Fig. 1, some preliminary retrieved knowledge facts may be irrelevant to the dialogue. Therefore, when using external knowledge to enhance context representation, it is necessary to filter out irrelevant knowledge based on the dialogue context.

Context-knowledge Attention Mechanism. We design the context-knowledge attention mechanism to dynamically filter out irrelevant knowledge in the word-level, inspired by previous works [14]. The major difference lies in that the context-knowledge attention mechanism generates filtered fact vector $\boldsymbol{g}_{\mathrm{t}}^{x}$ for t-th word x_t in post, because each word may focus on different parts of the candidate facts. Formally, filtered fact vector is calculated:

$$\boldsymbol{g}_{\mathrm{t}}^{x}=\sum_{i=1}^{N}\alpha_i^t\boldsymbol{f}_i, \alpha_i^t=softmax(\beta_i^t) \tag{2}$$

$$\beta_i^t=\left(\mathbf{W_h}\mathbf{h}_{\mathrm{t}}^{\mathrm{x}}\right)^{\top}\tanh\left(\mathbf{W_f}\boldsymbol{f}_i\right) \tag{3}$$

where $\boldsymbol{f}_i=[\boldsymbol{h}_i;\boldsymbol{r}_i;\boldsymbol{t}_i]\in\mathbb{R}^{d_e+d_r+d_e}$ is the embedding vector of i-th fact. $\mathbf{W_h}$ and $\mathbf{W_k}$ are trainable. α_i^t is the attention weight that measures the relevance of the contextual state $\mathbf{h}_{\mathrm{t}}^{\mathrm{x}}$ and the fact $\boldsymbol{f}_i$. Given contextual state $\mathbf{h}_{\mathrm{t}}^{\mathrm{x}}$, context-knowledge attention mechanism tries to discard the redundant parts of candidate facts, and remain relevant parts to form the filtered knowledge vector $\boldsymbol{g}_t^x$.

Knowledge-Enriched Encoder. To better fuse the filtered knowledge, knowledge-enriched encoder encodes the post into knowledge-aware representation. Referring to Eq. 1, bi-directional GRU network reads post and filtered knowledge vector, then outputs a contextual state sequence that contains relevant external knowledge information:

$$\begin{aligned}\mathbf{h}_{\mathrm{t}}^{\mathrm{f}'} &= GRU^{f'}\left(\mathbf{h}_{\mathrm{t}-1}^{\mathrm{f}'}, \mathbf{x}_{\mathrm{t}}, \mathbf{e}_{\mathbf{x}_{\mathrm{t}}}, \boldsymbol{g}_{\mathrm{t}}^{x}\right)\\ \mathbf{h}_{\mathrm{t}}^{\mathrm{b}'} &= GRU^{b'}\left(\mathbf{h}_{\mathrm{t}+1}^{\mathrm{b}'}, \mathbf{x}_{\mathrm{t}}, \mathbf{e}_{\mathbf{x}_{\mathrm{t}}}, \boldsymbol{g}_{\mathrm{t}}^{x}\right)\end{aligned} \tag{4}$$

where $\mathbf{e}_{\mathbf{x}_{\mathrm{t}}}$ is matched entity embedding vector and $\boldsymbol{g}_{\mathrm{t}}^{x}$ is filtered knowledge vector. In context encoding, the encoder encodes information from three aspects: word vectors, entity vectors, filtered knowledge vectors. After the knowledge-enriched encoder, the contextual state of the post is denoted as $H^{x'} = (\mathbf{h}_1^{\mathrm{x}'}, ..., \mathbf{h}_{\mathrm{n}}^{\mathrm{x}'})$, and the hidden state representation $\mathbf{h}_{\mathrm{t}}^{\mathrm{x}'} = \left[\mathbf{h}_{\mathrm{t}}^{\mathrm{f}'}; \mathbf{h}_{\mathrm{t}}^{\mathrm{b}'}\right]$.

3.4 Topic Fact Predictor

Considering that some retrieved facts may be redundant for current dialogue, we select appropriate facts for generating responses via the topic fact predictor. The topic fact predictor mainly consists of topic entity predictor and key fact predictor. The topic entity predictor predicts distribution over entity words in the post, denoting the probability of an entity word becoming discussion topic. The key fact predictor is designed to predict distribution over candidate facts, meaning that the probability of a fact selected in response generation.

Topic Entity Predictor. The topic entity predictor infers the topic probability over entity words in the post. As seen in Fig. 1, candidate facts are retrieved by entity words such as "great" and "play". But "play" looks more like a discussion topic. The candidate facts retrieved through topic entity may be more useful for response generation. Inspired by teacher-student network [14], posterior knowledge selection takes the context and response as input and generates the posterior distribution $\mathbf{z}_{\mathrm{post}}^{e}$ over entities as soft label. The prior distribution $\mathbf{z}_{\mathrm{prior}}^{e}$ is trained to be close to the posterior, and generates without response as input:

$$\begin{aligned}\mathbf{z}_{\mathrm{post}}^{e} &= softmax\left(tanh\left(\mathbf{F}_{e}\mathbf{W}_{\mathrm{post}}^{\mathrm{e}}\right) \cdot tanh\left(\left[\mathbf{h}_{\mathrm{n}}^{\mathrm{x}'}; \mathbf{h}_{\mathrm{m}}^{\mathrm{y}}\right]\mathbf{W}_{\mathrm{post}}^{\mathrm{h,e}}\right)^{\top}\right)\\ \mathbf{z}_{\mathrm{prior}}^{e} &= softmax\left(tanh\left(\mathbf{F}_{e}\mathbf{W}_{\mathrm{prior}}^{\mathrm{e}}\right) \cdot tanh\left(\mathbf{h}_{\mathrm{n}}^{\mathrm{x}'}\mathbf{W}_{\mathrm{prior}}^{\mathrm{h,e}}\right)^{\top}\right)\end{aligned} \tag{5}$$

where $\mathbf{F}_{e} \in \mathbb{R}^{|\mathcal{T}| \times d_e}$ is the embedding matrix of entity words in the post, and $\mathcal{T}$ is the set of entities. $\mathbf{W}_{\mathrm{post}}^{\mathrm{e}}, \mathbf{W}_{\mathrm{prior}}^{\mathrm{e}}, \mathbf{W}_{\mathrm{post}}^{\mathrm{h,e}}$ and $\mathbf{W}_{\mathrm{prior}}^{\mathrm{h,e}}$ are trainable parameters. *softmax* and *tanh* are activation functions. $\mathbf{h}_{\mathrm{n}}^{\mathrm{x}'}$ is the contextual representation of

the post X obtained by the knowledge-enriched encoder and $\mathbf{h}^{\mathrm{y}}_{\mathrm{m}}$ is the contextual representation of the response Y obtained by the context encoder.
Key Fact Predictor. The key fact predictor selects the facts that highly coincide with the dialogue. Given the context and fact representation, the predictor outputs a probability distribution $\mathbf{z}^f$ over the F by feedforward neural networks:

$$\begin{aligned}\mathbf{z}^f_{\text{post}} &= softmax\left(tanh\left(\mathbf{F}\mathbf{W}^{\mathrm{f}}_{\text{post}}\right)\cdot tanh\left(\left[\mathbf{h}^{\mathrm{x}'}_{\mathrm{n}};\mathbf{h}^{\mathrm{y}}_{\mathrm{m}}\right]\mathbf{W}^{\mathrm{h,f}}_{\text{post}}\right)^{\top}\right)\\ \mathbf{z}^f_{\text{prior}} &= softmax\left(tanh\left(\mathbf{F}\mathbf{W}^{\mathrm{f}}_{\text{prior}}\right)\cdot tanh\left(\mathbf{h}^{\mathrm{x}'}_{\mathrm{n}}\mathbf{W}^{\mathrm{h,f}}_{\text{prior}}\right)^{\top}\right)\end{aligned} \tag{6}$$

where $\mathbf{F}\in\mathbb{R}^{N\times(d_e+d_r+d_e)}$ is the embedding matrix of candidate facts F. We calculate the probability distribution $\mathbf{z}^p$ of candidate facts:

$$\mathbf{z}^p = \begin{cases}\lambda\mathbf{z}^e_{\text{post}} + (1-\lambda)\,\mathbf{z}^f_{\text{post}} & \text{if train}\\ \lambda\mathbf{z}^e_{\text{prior}} + (1-\lambda)\,\mathbf{z}^f_{\text{prior}} & else\end{cases} \tag{7}$$

where λ is a hyperparameter to control the contribution of distribution. $\mathbf{z}^p$ is used to calculate the topic fact representation: $\mathbf{f_z} = \mathbf{z}^p\cdot\mathbf{F}$.

The loss function of topic fact predictor consists of three parts: the Bag-of-Words (BoW) [17], Cross Entropy (CE) and Kullback-Leibler divergence (KLD) loss [18]. The purpose of the BoW loss is to measure the accuracy of contextual and topic fact vector to response generation. Meanwhile, the label $\mathbf{I}^e, \mathbf{I}^f$ are 0–1 indicator vectors to supervise the training of $\mathbf{z}^e_{\text{post}}$ and $\mathbf{z}^f_{\text{post}}$. $\mathbf{I}^e_i$ is either 1 or 0, denoting whether the i-th entity word in post is one of topic entities or not. $\mathbf{I}^f_i$ indicates whether the target entity in i-th facts is in the truth response or not. The labels are applied in CE loss. KLD loss is used to force prior distribution and posterior distribution to become as close as possible. Thus, the training objective of the key fact predictor module is to minimize a loss:

$$\begin{aligned}\mathcal{L}_p = \mathcal{L}_{BoW} &+ \mathcal{L}_{CE}(\mathbf{z}^e_{\text{post}}, \mathbf{I_e}) + \mathcal{L}_{CE}(\mathbf{z}^f_{\text{post}}, \mathbf{I_f})\\ &+ \mathcal{L}_{KLD}(\mathbf{z}^e_{\text{post}}, \mathbf{z}^e_{\text{prior}}) + \mathcal{L}_{KLD}(\mathbf{z}^f_{\text{post}}, \mathbf{z}^f_{\text{prior}})\end{aligned} \tag{8}$$

3.5 Response Generator

The response generator is used to generate the sentence conditioned on context and topic fact vector. Formally, the hidden states of decoder are computed:

$$\mathbf{h^y_t} = GRU\left(\mathbf{h^y_{t-1}}, [\mathbf{u_{t-1}}; \mathbf{c_{t-1}}]\right) \tag{9}$$

where $\mathbf{u_{t-1}} = [\mathbf{y_{t-1}}; \mathbf{e}_{\mathbf{y}_{t-1}}]$ connects word embedding and entity embedding of the last predicted token y_{t-1}; $\mathbf{c_{t-1}}$ is attentive context vector [19]; the initialization state of the decoder is $\mathbf{h}^{\mathbf{y}}_0 = \tanh([\mathbf{h}^{\mathbf{x}'}_{\mathbf{n}}; \mathbf{f_z}]\mathbf{W_{init}})$.

The current word y_t is generated by choosing from the vocabulary, entity words or copy words. Therefore, P_v, P_e, P_c are the probability distributions over vocabulary, entity words and copy words, respectively, calculated as follows:

$$\begin{aligned} &P_v\left(y_t\right)=softmax\left(elu([\mathbf{h_t^y};\mathbf{u_{t-1}};\mathbf{c_t}]\mathbf{W}_{v_1})\mathbf{W}_{v_2}\right) \\ &P_e\left(y_t\right)=\gamma_t\mathbf{z}^p+(1-\gamma_t)softmax(elu(\mathbf{F}\mathbf{W}_{ef})\cdot elu([\mathbf{h_t^y};\mathbf{u_{t-1}}]\mathbf{W}_{et})^\top) \\ &P_c\left(y_t\right)=softmax(elu(H^{x'}\mathbf{W}_{cx})\cdot elu([\mathbf{h_t^y};\mathbf{u_{t-1}};\mathbf{c_t}]\mathbf{W}_{ct})^\top) \end{aligned} \tag{10}$$

where $\mathbf{z}^p$ is topic fact distribution, calculated in topic fact predictor. $\gamma_t = \text{sigmoid}([\mathbf{h_t^y};\mathbf{u_t};\mathbf{c_t}]\mathbf{W}_\gamma)$ is a gate to control the contribution of topic fact distribution. Next, we employ three selection gates to dynamically generate different kinds of words:

$$p\left(y_t\right)=\nu_t^v p_v\left(y_t\right)+\nu_t^e p_e\left(y_t\right)+\nu_t^c p_c\left(y_t\right) \tag{11}$$

$$\nu_t=\left[\nu_t^v,\nu_t^e,\nu_t^c\right]=softmax\left([\mathbf{h_t^y};\mathbf{u_{t-1}};\mathbf{c_t}]\mathbf{W}_p\right)\in\mathbb{R}^3 \tag{12}$$

where ν is the gate to control the contribution of three types of words. The loss function $\mathcal{L}_n$ of the generator module consists of two parts: the first part is the log-likelihood of the generated response; the second part is cross-entropy loss, which aims to supervise the gated prediction distribution:

$$\mathcal{L}_n=-\sum\log P\left(y_t\mid y_{t-1:1},X,F\right)-\sum\mathbf{I}_t\cdot\log\left(\nu_t\right) \tag{13}$$

where $\mathbf{I}_t\in\mathbb{R}^3$ is a 0–1 indicator vector. For example, if t-th word in truth response is an entity word, $\mathbf{I}_t=[0,1,0]$. Finally, the overall loss to train CKFS-DG is computed: $\mathcal{L}=\mathcal{L}_n+\mathcal{L}_p$.

4 Experiments

4.1 Datasets

We conduct experiments on two large-scale dialogue datasets: English Reddit [2] and Chinese Weibo [5], which are open-domain single-round dialogue datasets. Both datasets are aligned with the commonsense knowledge graph ConcetNet[1]. The statistics of Reddit and Weibo dataset are provided in Table 1. The statistics table includes the number of dialogue pairs in training and test sets. Additionally, the number of candidate facts and the average number of facts per dialogue pair are included. Golden facts are the facts whose target entity appears in the response Y. According to statistics, each dialogue pair involves a large number of candidate facts and few golden facts, that makes it difficult to select appropriate facts for response generation.

[1] Available at: https://conceptnet.io.

Table 1. The statistics of knowledge-aware dialogue dataset Reddit and Weibo.

Datasets	Train	Test (valid)	Facts	Avg facts	Avg golden facts
Reddit [2]	1,352,961	40,000	149,803	85.01	1.009
Weibo [5]	1,019,908	56,661	696,466	77.66	1.293

4.2 Baselines

We compare the performance of CKFS-DG with seven neural generative methods. These models are divided into three categories: without knowledge as input, with one-hop knowledge graph and with two-hop knowledge graph. Firstly, **Seq2Seq** [3] takes the post as input. **Copy** [20] can reproduce words from posts. And then, **GenDS** [7] utilizes entity words in one-hop knowledge graph. **CCM** [2] exploits knowledge graph with graph attention mechanisms to capture the semantics of the knowledge facts. **ConKADI** [5] designs felicitous fact recognizer to detect the facts that highly coincide with the dialogue context. Finally, **ConceptFlow** [1] simulates the dialogue flow in the two-hop knowledge graph space. **TSGADG** [4] employs two-hop based static graph attention to deepen the understanding of context. Considering that our model only utilizes one-hop knowledge graph, we mainly compare with the first two categories.

4.3 Experimental Setup

In the experiment, our model is implemented with Tensorflow[2]. Most hyper-parameters are consistent with ConKADI [5]. In detail, we use a fixed English vocabulary of 30,000 words, and a Chinese vocabulary of 50,000 words. The word embeddings adopt the Glove word embeddings with the dimension of 300. TransE embedding [21] is used for the entity and relations representations in the facts, whose dimension is 100. The state size of GRU network, used in encoder and decoder, is 512. The adam optimizer is used for training, and the initial learning rate is 0.0001. We halve the learning rate when perplexity [22] increases on validation data, and stop training if the perplexity improves for two successive iterations. The batch size is 100. The maximum number of epochs is 25.

4.4 Evaluation Metrics

To measure the quality of the generated responses, we introduce the common evaluation metrics [5,10] from five aspects:

Knowledge Utilization: E_{match} [2] is the average number of overlapping entities in the generated responses and candidate facts, that measures the model's ability to use the tail entities in the candidate facts. E_{use} [5] further considers the number of head entities to evaluate the utilization of entities. E_{recall} [5] is the ratio of overlapping entities between the generated responses and the ground-truth responses, which evaluates the accuracy of knowledge selection.

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

Embedding-Based Relevance: Emb_{avg} [23] evaluates the similarity between the generated responses and the ground-truth responses by using the averaged word embedding. Emb_{ex} [23] uses each dimension's extreme value of word embedding.

Overlapping-Based Relevance: BLEU-2 [24] measures n-gram overlap rate between the generated response and the ground-truth response.

Diversity: Distinct-2 [25] is a ratio of distinct bigrams in the generated responses, assessing the diversity of generated responses.

Informativeness: Entropy [26] is computed by averaging word-level entropy in the generated responses, measuring informativeness of generated responses.

4.5 Results and Analysis

Experimental Results. We compare CKFS-DG with the baselines on Reddit and Weibo dataset. Experiment results are shown in Tables 2 and 3.

In knowledge utilization evaluation, CKFS-DG outperforms the baseline methods in selecting appropriate knowledge, which can be proved by E_{recall} on Reddit dateset. The E_{recall} of our model is improved by 21.4% compared to ConKADI. The proposed topic fact predictor has the potential to select the accurate facts from candidate knowledge facts. Although E_{match} drops by 3.2%, the overall E_{use} increases by 12.6%. The advantages of our model lie in utilizing the copy mechanism, that exploits entity words in posts. The experiments on the Weibo dataset also basically achieve the effectiveness of ConKADI. The results demonstrate that our model selects entity words in candidate facts and has high utilization of the knowledge for response generation.

In embedding-based relevance evaluation, Emb_{ex} of our model improves by 15% compared to ConKADI on Reddit, indicating that the semantics between the responses generated by our model and the ground-truth responses are more similar. In overlapping-based relevance and diversity evaluation, BLUE-2 and dist-2 value are moderate compared to other models. Analysis of these two

Table 2. Experimental Results on Reddit. † means that the result is borrowed from ConKADI [5] and § means that the result is borrowed from TSGADG [4].

Metrics	Entity Score			Embedding		Overlap (%) BLEU-2	Diversity (%) dist-2	Informativeness Entropy
	E_{match}	E_{use}	E_{recall}	Emb_{avg}	Emb_{ex}			
Seq2Seq† [3]	0.41	0.52	0.04	0.868	0.837	4.81	1.77	7.59
Copy† [20]	0.14	0.67	0.09	0.868	0.841	**5.43**	8.33	7.87
GenDS† [7]	1.13	1.26	0.13	0.876	0.851	4.68	3.97	7.73
CCM† [2]	1.08	1.33	0.11	0.871	0.841	5.18	5.29	7.73
ConKADI† [5]	1.24	1.98	0.14	0.867	0.852	3.53	18.78	8.50
ConceptFlow§ [1]	1.26	-	-	0.82	0.81	5.14	12.28	8.14
TSGADG§ [4]	**1.57**	-	-	0.88	0.63	5.15	**27.25**	**8.53**
CKFS-DG(Ours)	1.20	**2.23**	**0.17**	**0.88**	**0.865**	4.93	14.39	8.42

Table 3. Experimental Results on Weibo. † means that the result is borrowed from ConKADI [5].

Metrics	Entity Score			Embedding		Overlap (%) BLEU-2	Diversity (%) dist-2	Informativeness Entropy
	E_{match}	E_{use}	E_{recall}	Emb_{avg}	Emb_{ex}			
Seq2Seq† [3]	0.33	0.58	0.13	0.770	0.500	2.24	1.04	6.09
Copy† [20]	0.33	0.68	0.13	0.786	0.501	2.28	2.18	6.13
GenDS† [7]	0.75	0.84	0.26	0.789	0.524	2.09	1.66	5.89
CCM† [2]	0.99	1.09	0.28	0.786	0.544	3.26	2.59	6.16
ConKADI† [5]	**1.48**	**2.08**	**0.38**	**0.846**	0.577	**5.06**	23.93	9.04
CKFS-DG(Ours)	1.44	2.00	**0.38**	0.816	**0.592**	4.32	**24.88**	**9.57**

indicators by CCM and ConKADI, the high BLUE-2 based on word character similarity losts the diversity of generated responses to a certain extent [5]. In informativeness evaluation, entropy of our model improves by 5.8% compared to ConKADI on Weibo, that further confirms the advantages of our model in integrating knowledge and generating informative responses. We also found that the entropy is slightly lower than ConKADI on Reddit. The reason may be that the collected sources of the two datasets are different, and there are inherent differences of datasets, as mentioned in [5]. The experimental results prove that our method has the ability to selecting appropriate knowledge facts to generate informative responses.

Ablation Study. We conduct further ablation experiments to dissect our model. The experimental results are shown in Table 4. We analyze the influence and role of each module by comparing the experimental results of the model with and without this module. Specifically, (1) **w/o TEP** is the model without topic entity predictor, whose role is to predicte dialogue topic entity in the post. (2) **w/o KE** is the model removing knowledge-enriched encoder, which is based on the context-knowledge attention mechanism and generates knowledge-enriched contextual representation of post. (3) **w/o TFP** is the model without topic fact predictor, including topic entity and fact prediction of the current dialogue.

Table 4. Results of ablation study on Reddit and Weibo. Comparative experiments include: (1) "-w/o TEP" without Topic Entity Predictor, (2) "-w/o KE" without Knowledge-enriched Encoder and (3) "-w/o TFP" without Topic Fact Predictor.

	Reddit					Weibo				
Metrics	E_{recall}	Emb_{ex}	BLEU-2	dist-2	Entropy	E_{recall}	Emb_{ex}	BLEU-2	dist-2	Entropy
CAFS-DG	**16.4**	**0.865**	4.938	14.392	8.426	**37.45**	**0.592**	**4.321**	24.877	**9.566**
- w/o TEP	15.93	0.571	5.154	**24.059**	**8.971**	36.12	0.588	4.085	**27.587**	9.541
- w/o KE	13.59	0.849	3.966	15.764	8.263	36.46	0.567	4.127	26.611	9.462
- w/o TFP	12.6	0.855	**5.614**	7.374	7.857	32.34	0.542	3.671	4.590	6.579

The results show that (1) when our model removes the topic entity predictor, E_{recall} drops by 2.8% and Emb_{ex} drops by 33.9% on Reddit. It shows that predicting the topic entities helps the model generate responses related to the topic, and there is a high semantic similarity between the generated responses and the target responses. Intuitively, the topic entity predictor will make the model pay more attention to candidate facts related to topic entities in response generation. Without topic entity predictor, dict-2 decreases by 10.9% on Weibo, indicating that focusing on the topic of the dialogue may lose the diversity of generated responses. (2) Without knowledge-enriched encoder, E_{recall} and BLEU-2 decrease by 17.1% and 19.7% on Reddit respectively. The module uses the context-knowledge attention mechanism to construct filtered knowledge representation, and generates knowledge-enriched contextual vector of post. The performance of the model is improved with this module, showing that the knowledge-enriched encoder module is helpful to strengthen the semantic representation of the post. (3) Removing the topic fact predictor, E_{recall} drops by 13.6% and entropy drops by 31.3% on Weibo. It demonstrates the importance of topic fact predictor to generate informative and diverse responses with appropriate entities in retrieved facts.

Case Study. As shown in Table 5, we discuss two typical cases, including generated responses by our model and baselines CopyNet, CCM and ConKADI.

In case 1, candidate facts were retrieved based on the entity words in the post such as “great” and “play”. The retrieved facts whose head entity is “play” should receive more attention. Comparing with “great”, “play” is more likely to be the topic point of the current conversation. CCM and ConKADI do not select golden facts and generate generic responses. Our model focuses on the topic entity “play” and selects facts to generate a reasonable response. From the perspective of response quality, our model generates a natural response according to dialogue topic.

In case 2, it is obvious that “paper” and “work” in the post are important topic entities, but “zombies” are not. ConKADI paid attention to the “paper” while also paying attention to the “zombies” when generating responses. Our model utilizes candidate facts to infer “graduation” from “work” and generates relatively fluent responses.

We further visualize the probability distribution over candidate facts for the above cases in Fig. 3. The distribution plots on the left are generated by ConKADI, and the plots on the right are generated by CKFS-DG. In each plot, the words on the left are the entity words in the post, and no more than 25 knowledge facts are retrieved based on each entity word. The probability distribution value for each fact is distinguished by the color of the plots.

Obviously, in case 1, the focus of ConKADI is the candidate facts whose head entity word is “great”. Different from ConKADI, our model pays attention to the facts “⟨play, RelatedTo, team⟩” which is used for generate responses. In case 2 our model focuses on knowledge facts related to “work”. Our model further utilizes candidate facts to infer the word “graduation” based on the dialogue

Table 5. Two cases of knowledge-aware dialogue generation. The comparison models include CopyNet, CCM, ConKADI, and our model CKFS-DG. The orange words are the entity words in the post, the blue words are the entity words in the true response, and the green words are entity words in the candidate facts.

case 1 (Reddit)	Post	that would be great if they had to play at target field the first season
	Knowledge	⟨play, RelatedTo, team⟩ ⟨great, RelatedTo, awesome⟩ ⟨great, RelatedTo, money⟩ ⟨great, HasContext, music⟩ ...
	Golden Fact	⟨play, RelatedTo, team⟩
	True Response	its about time a mn team won there.
	GenDS [7]	i would n't be surprised if they did n't have to play at target field field
	CCM [2]	i 'd be happy with that .
	ConKADI [5]	well that would be a great idea. and that 's pretty awesome.
	CKFS-DG(Ours)	well, if the hawks had to play at target, they 'd be a great team.
case 2 (Weibo)	Post	被论文和工作弄疯了，现在暂时缓了一下下，但是一大波僵尸还在前方啊 Be crazy for the papers and work, now it is temporarily slowing down, but a large wave of zombies is still ahead.
	Knowledge	⟨毕业,Causes,工作⟩ ⟨加油,MotivatedByGoal,工作⟩ ⟨graduation,Causes,work⟩ ⟨Come on,MotivatedByGoal,work⟩ ⟨工作,HasProperty,苦⟩ ⟨麻烦,NotDesires,工作⟩ ... ⟨Work,HasProperty,bitter⟩ ⟨Trouble, NotDesires, work⟩ ...
	Golden Fact	⟨毕业,Causes,工作⟩ ⟨graduation,Causes,work⟩
	True Response	劳逸结合，熬过毕业季 Combine work and rest to survive the graduation season
	ConKADI [5]	是啊，加油，我把论文写完了，前方的僵尸 Yes, come on, I have finished writing the thesis, the zombie in front
	CKFS-DG(Ours)	弄的挺麻烦，你说，毕业的时候，加油 It was very troublesome , say that I will work hard when I graduate

context. Compared to ConKADI, our model introduces topic entity predictor to focus on the relevant topic entities, which helps to select appropriate knowledge.

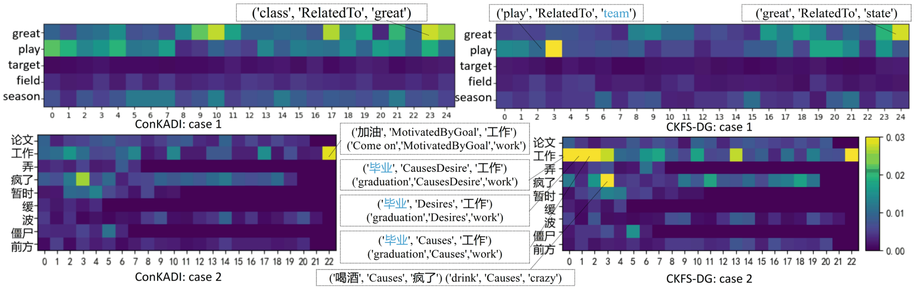


Fig. 3. Visualization of probability distributions over candidate facts for the two cases. We only show the 25 candidate facts retrieved for each entity word in posts.

5 Conclusion and Future Work

We propose a knowledge-aware dialogue generation model CKFS-DG to integrate external knowledge into response generation. In particular, we design context-knowledge attention mechanism to filter out redundant knowledge for knowledge enhanced context representation, and topic fact prediction mechanism to select appropriate knowledge facts for response generation. Experimental results on Reddit and Weibo datasets demonstrate the effectiveness of CKFS-DG in selecting appropriate knowledge and generating informative responses. In the future, we intend to introduce different forms of knowledge into the generation model effectively and naturally.

Acknowledgements. This work was supported by the National Key R&D Program of China under Grant No. 2019YFF0302601.

References

1. Zhang, H., Liu, Z., Xiong, C., Liu, Z.: Grounded conversation generation as guided traverses in commonsense knowledge graphs. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 2031–2043 (2020)
2. Zhou, H., Young, T., Huang, M., Zhao, H., Xu, J., Zhu, X.: Commonsense knowledge aware conversation generation with graph attention. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 4623–4629 (2018)
3. Sutskever, I., Vinyals, O., Le, Q. V.: Sequence to Sequence Learning with Neural Networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)
4. Zhou, S., Rong, W., Zhang, J., Wang, Y., Shi, L., Xiong, Z.: Topic-aware dialogue generation with two-hop based graph attention. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 7428–7432 (2021)
5. Wu, S., Li, Y., Zhang, D., Zhou, Y., Wu, Z.: Diverse and informative dialogue generation with context-specific commonsense knowledge awareness. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5811–5820 (2020)
6. Ghazvininejad, M., et al.: A knowledge-grounded neura conversation model. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 5110–5117 (2018)
7. Zhu, W., Mo, K., Zhang, Y., Zhu, Z., Peng, X., Yang, Q.: Flexible end-to-end dialogue system for knowledge grounded conversation. CoRR, abs/1709.04264 (2017)
8. Moon, S., Shah, P., Kumar, A., Subba, R.: Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 845–854 (2019)
9. Lin, X., Jian, W., He, J., Wang, T., Chu, W.: Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 41–52 (2020)

10. Wang, W., Gao, W., Feng, S., Chen, L., Wang, D.: Adaptive posterior knowledge selection for improving knowledge-grounded dialogue generation. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management, pp. 1989–1998 (2021)
11. Liu, Z., Niu, Z. Y., Wu, H., Wang, H.: Knowledge aware conversation generation with explainable reasoning over augmented graphs. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 1782–1792 (2019)
12. Wang, J., Liu, J., Bi, W., Liu, X., He, K., Xu, R., Yang, M.: Improving Knowledge-Aware Dialogue Generation via Knowledge Base Question Answering. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, pp. 9169–9176 (2020)
13. Dinan, E., Roller, S., Shuster, K., Fan, A., Auli, M., Weston, J.: Wizard of Wikipedia: Knowledge-Powered Conversational Agents. In: Proceedings of the 7th International Conference on Learning Representations, (2019)
14. Wang, Y., Wang, Y., Lou, X., Rong, W., Hao, Z., Wang, S.: Improving Dialogue Response Generation Via Knowledge Graph Filter. In: 2021 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), pp. 7423–7427 (2021)
15. Zhong, P., Liu, Y., Wang, H., Miao, C.: Keyword-guided neural conversational Model. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 14568–14576 (2021)
16. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1724–1734 (2014)
17. Zhao, T., Zhao, R., Eskenazi, M.: Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 654–664 (2017)
18. Kullback, S., Leibler, R.A.: On information and sufficiency. The annals of mathematical statistics, pp. 79–86 (1951)
19. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. CoRR, abs/1508.04025 (2015)
20. Gu, J., Lu, Z., Li, H., Li, V.O.: Incorporating copying mechanism in sequence-to-sequence learning. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 199–208 (2016)
21. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Advances in Neural Information Processing Systems, pp. 2787–2795 (2013)
22. Serban, I. V., Sordoni, A., Bengio, Y., Courville, A., Pineau, J.: Hierarchical neural network generative models for movie dialogues. CoRR, abs/1507.04808, pp. 434–441 (2015)
23. Liu, C. W., Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., Pineau, J.: How not to evaluate your dialogue system: an empirical study of unsupervised evaluation metrics for dialogue response generation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2122–2132 (2016)
24. Tian, Z., Yan, R., Mou, L., Song, Y., Feng, Y., Zhao, D.: How to make context more useful an empirical study on context-aware neural conversational models. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 231–236 (2017)

25. Li, J., Galley, M., Brockett, C., Gao, J., Dolan, 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 (2015)
26. Mou, L., Song, Y., Yan, R., Li, G., Zhang, L., Jin, Z.: Sequence to backward and forward sequences: a content-introducing approach to generative short-text conversation. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 3349–3358 (2016)