Generating Informative Conversational Response using Recurrent Knowledge-Interaction and Knowledge-Copy

Xiexiong Lin Weiyu Jian Jianshan He Taifeng Wang Wei Chu
Ant Financial Services Group
{xiexiong.lxx,weiyu.jwy,yebai.hjs}@antfin.com
{taifeng.wang,weichu.cw}@alibaba-inc.com

Abstract

Knowledge-driven conversation approaches have achieved remarkable research attention recently. However, generating an informative response with multiple relevant knowledge without losing fluency and coherence is still one of the main challenges. To address this issue, this paper proposes a method that uses recurrent knowledge interaction among response decoding steps to incorporate appropriate knowledge. Furthermore, we introduce a knowledge copy mechanism using a knowledge-aware pointer network to copy words from external knowledge according to knowledge attention distribution. Our joint neural conversation model which integrates recurrent Knowledge-Interaction and knowledge Copy (KIC) performs well on generating informative responses. Experiments demonstrate that our model with fewer parameters yields significant improvements over competitive baselines on two datasets Wizard-of-Wikipedia(average Bleu +87%; abs.:0.034) and DuConv(average Bleu +20%; abs.:0.047) with different knowledge formats (textual & structured) and different languages (English & Chinese).

1 Introduction

Dialogue systems have attracted much research attention in recent years. Various end-to-end neural generative models based on the sequence-to-sequence framework (Sutskever et al., 2014) have been applied to the open-domain conversation and achieved impressive success in generating fluent dialog responses (Shang et al., 2015; Vinyals and Le, 2015; Serban et al., 2016). However, many neural generative approaches from the last few years confined within utterances and responses, suffering from generating uninformative and inappropriate responses. To make responses more meaningful and expressive, several works on the dialogue system exploiting external knowledge. Knowledge-driven methods focus on generating more informative and meaningful responses via incorporating structured knowledge consists of triplets (Zhu et al., 2017; Zhou et al., 2018; Young et al., 2018; Liu et al., 2018) or unstructured knowledge like documents (Long et al., 2017; Parthasarathi and Pineau, 2018; Ghazvininejad et al., 2018; Ye et al., 2019). Knowledge-based dialogue generation mainly has two methods: a pipeline way that deals with knowledge selection and generation successively (Lian et al., 2019), and a joint way that integrates knowledge selection into the generation process, for example, several works use Memory Network architectures (Sukhbaatar et al., 2015) to integrate the knowledge selection and generation jointly (Dinan et al., 2018; Dodge et al., 2015; Parthasarathi and Pineau, 2018; Madotto et al., 2018; Ghazvininejad et al., 2018). The pipeline approaches separate knowledge selection from generation, resulting in an insufficient fusion between knowledge and generator. When integrating various knowledge, pipeline approaches lack flexibility. The joint method with the memory module usually uses knowledge information statically. The confidence of knowledge attention decreasing at decoding steps, which has the potential to produce inappropriate collocation of knowledge words. To generate informative dialogue response that integrates various relevant knowledge without losing fluency and coherence, this paper presents an effective knowledge-based neural conversation model that enhances the incorporation between knowledge selection and generation to produce more informative and meaningful responses. Our model integrates the knowledge into the generator by using a recurrent knowledge interaction that dynamically updates the attentions of knowledge selection via decoder state and the updated knowledge attention assists in decoding the next state, which

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 41–52
July 5 - 10, 2020. ©2020 Association for Computational Linguistics

maintains the confidence of knowledge attention during the decoding process, it helps the decoder to fetch the latest knowledge information into the current decoding state. The generated words ameliorate the knowledge selection that refines the next word generation, and such repeated interaction between knowledge and generator is verified to be an effective way to integrate multiple knowledge coherently that to generate an informative and meaningful response when knowledge is fully taken account of.

Although recurrent knowledge interaction better solves the problem of selecting appropriate knowledge for generating the informative response, the preferable integration of knowledge into conversation generation still confronts an issue, i.e., it is more likely that the description words from external knowledge generated for the dialog response have a high probability of being an oov(out-of-vocabulary), which is a common challenge in natural language processing. A neural generative model with pointer networks has been shown to have the ability to handle oov problems (Vinyals et al., 2015; Gu et al., 2016). Very few researches on copyable generative models pay attention to handle external knowledge, while in knowledge-driven conversation, the description words from knowledge are usually an important component of dialog response. Thus, we leverage a knowledge-aware pointer network upon recurrent knowledge interactive decoder, which integrates the Seq2seq model and pointer networks containing two pointers that refer to utterance attention distribution and knowledge attention distribution. We show that generating responses using the knowledge copy resolves the oov and the knowledge incompleteness problems.

In summary, our main contributions are: (i) We propose a recurrent knowledge interaction, which chooses knowledge dynamically among decoding steps, integrating multiple knowledge into the response coherently. (ii) We use a knowledge-aware pointer network to do knowledge copy, which solves oov problem and keeps knowledge integrity, especially for long-text knowledge. (iii) The integration of recurrent knowledge interaction and knowledge copy results in more informative, coherent and fluent responses. (iv) Our comprehensive experiments show that our model is general for different knowledge formats (textual & structured) and different languages (English & Chinese). Furthermore, the results significantly outperform competitive baselines with fewer model parameters.

2 Model Description

Given a dataset $D = \{(X_i, Y_i, K_i)\}_{i=1}^{N}$, where N is the size of the dataset, a dialog response $Y = \{y_1, y_2, \ldots, y_n\}$ is produced by the conversation history utterance $X = \{x_1, x_2, \ldots, x_m\}$, using also the relative knowledge set $K = \{k_1, k_2, \ldots, k_s\}$. Here, m and n are the numbers of tokens in the conversation history X and response Y respectively, and s denotes the size of relevant knowledge candidates collection K. The relevant knowledge candidates collection K is assumed to be already provided and the size of candidates set is limited. Each relevant knowledge element in candidate collection could be a passage or a triplet, denoted as $k = \{\kappa_1, \kappa_2, \ldots, \kappa_l\}$, where l is the number of the tokens in the knowledge element. As illustrated in Figure 1, the model KIC proposed in this work is based on an architecture involving an encoder-decoder framework (Sutskever et al., 2014) and a pointer network (Vinyals et al., 2015; See et al., 2017). Our model is comprised of four major components: (i) an LSTM based utterance encoder; (ii) a general knowledge encoder suitable for both structural and documental knowledge; (iii) a recurrent knowledge interactive decoder; (iv) a knowledge-aware pointer network.

2.1 Utterance Encoder

The utterance encoder uses a bi-directional LSTM (Schuster and Paliwal, 1997) to encode the utterance inputs by concatenating all tokens in the dialogue history X and obtain the bi-directional hidden state of each x_i in utterance, denoted as $H = \{h_1, h_2, \ldots, h_m\}$. Combining two-directional hidden states, we have the hidden state h_t^* as

$$h_t^* = [\overrightarrow{LSTM}(x_t, h_{t-1}); \overleftarrow{LSTM}(x_t, h_{t+1})]. \quad (1)$$

2.2 Knowledge Encoder

As illustrated in Model Description, the knowledge input is a collection of multiple knowledge candidates K. The relevant knowledge k_i can be a passage or a triplet. This paper provides a universal encoding method for both textual and structured knowledge. The relevant knowledge is represented as a sequence of tokens, which are encoded by a transformer encoder (Vaswani et al., 2017), i.e., $z_t = Transformer(\kappa_t)$. Static attention a_i^k is

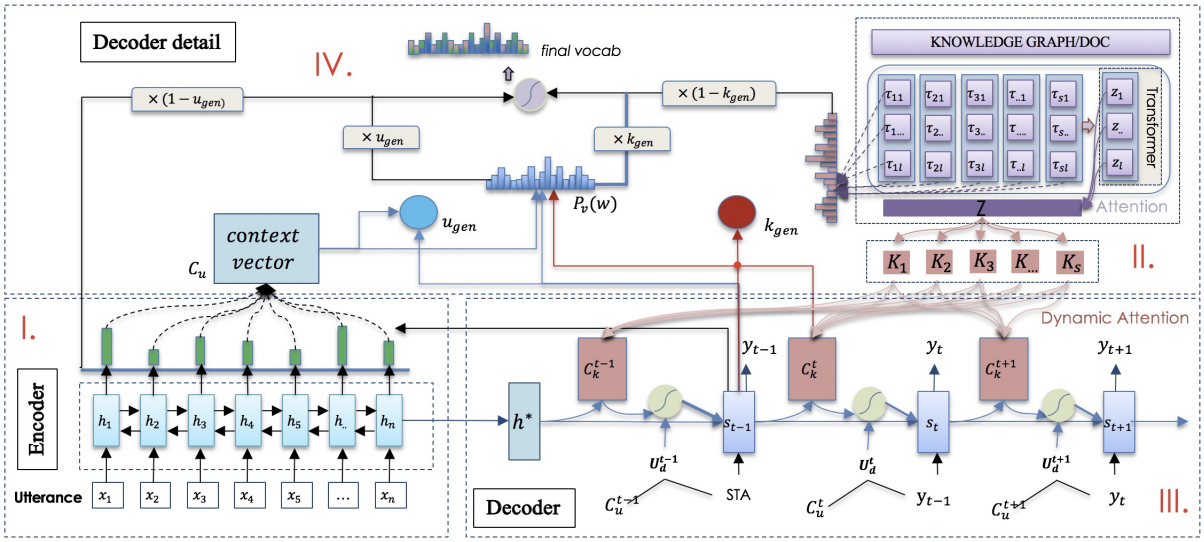

Figure 1: The architecture of KIC. Here, U_d^t is calculated by decode-input and utterance context vector C_u^t at current step , C_k^t represents the knowledge context vector resulted from dynamic knowledge attention. u_{gen} and k_{gen} are two soft switches that control the copy pointer to utterance attention distribution and knowledge attention distribution, respectively.

used to encode knowledge $Z = \{z_1, z_2, \ldots, z_l\}$ to obtain the overall representation K^{rep} for the relevant knowledge as

$$a_i^k = softmax(V_z^T \tanh(W_z z_i)) \quad (2)$$

$$K^{rep} = \sum_{i=1}^{l} a_i^k z_i, \quad (3)$$

where V_z^T and W_z are learnable parameters. So far we have the knowledge representations for the knowledge candidate collection C_k^{rep}.

2.3 Recurrent Knowledge Interactive Decoder

The decoder is mainly comprised of a single layer LSTM (Hochreiter and Schmidhuber, 1997) to generate dialogue response incorporating the knowledge representations in collection C_k^{rep}. As shown in Figure 1, in each step t, the decoder updates its state s_{t+1} by utilizing the last decode state s_t, current decode-input U_d^t and knowledge context C_k^t. The current decode-input is computed by the embeddings of the previous word $e(y_t)$ and utterance context vector C_u^t. We provide the procedure as

$$e_i^t = v_e^T \tanh(W_h h_i + W_s^u s_t + b_{ua}) \quad (4)$$

$$u^t = softmax(e^t) \quad (5)$$

$$C_u^t = \sum_{i=1}^{m} u_i^t h_i \quad (6)$$

$$U_d^t = V_u[e(y_t), C_u^t] + b_u, \quad (7)$$

where $V_u, b_u, v_e, W_h, W_s^u, b_{ua}$ are learnable parameters.

Instead of modeling knowledge selection independently, or statically incorporating the representation of knowledge into the generator, this paper proposes an interactive method to exploit knowledge in response generation recurrently. The knowledge attention d^t updates as the decoding proceeds to consistently retrieve the information of the knowledge related to the current decoding step so that it helps decode the next state correctly, which writes as

$$\theta_i^t = v_k^T \tanh(W_k K_i^{rep} + W_s^k s_t + b_{ak}) \quad (8)$$

$$d^t = softmax(\theta^t) \quad (9)$$

$$C_k^t = \sum_{i}^{s} d_i^t K_i^{rep}, \quad (10)$$

where v_k, W_k, W_s^k, b_{ak} are learnable parameters. A knowledge gate g^t is employed to determine how much knowledge and decode-input is used in the generation, which is defined as

$$g^t = sigmoid(V_g[U_d^t, C_k^t] + b_g), \quad (11)$$

where V_g and b_g are learnable parameters. As the steps proceed recurrently, the knowledge gate can dynamically update itself as well. Hence, the decoder updates its state as:

$$s_{t+1} = LSTM(s_t, (g_t U_d^t + (1 - g^t) C_k^t)) \quad (12)$$

2.4 Knowledge-Aware Pointer Networks

Pointer networks using a copy mechanism are widely used in generative models to deal with oov problem. This paper employs a novel knowledge-aware pointer network. Specifically, we expand the scope of the original pointer networks by exploiting the attention distribution of knowledge representation. Besides, the proposed knowledge-aware pointer network shares extended vocabulary between utterance and knowledge that is beneficial to decode oov words. As two pointers respectively refer to the attention distributions of utterance and knowledge, each word generation is determined by the soft switch of utterance u_{gen} and the soft switch of knowledge k_{gen}, which are defined as

$$u_{gen} = \sigma(w_{uc}^T C_u^t + w_{us}^T s_t + w_u^T U_d^t + b_{up}) \quad (13)$$

$$k_{gen} = \sigma(w_{kc}^T C_k^t + w_{ks}^T s_t + w_g^T U_g^t + b_{kp}), \quad (14)$$

where $w_{uc}^T, w_{us}^T, w_u^T, b_{up}, w_{kc}^T, w_{ks}^T, w_g^T, b_{kp}$ are learnable parameters. The U_g^t here is defined as

$$U_g^t = V_g[e(y_t), C_k^t] + b_g, \quad (15)$$

where V_g, b_g are learnable parameters. Therefore, the final probability of the vocabulary w is

$$P_{final}(w) = (\lambda u_{gen} + \mu k_{gen}) P_v(w) + \lambda(1 - u_{gen}) \sum_i u_i^t + \mu(1 - k_{gen}) \sum_i d_i^t, \quad (16)$$

$$P_v(w) = softmax(V_2(V_1[s_t, C_u^t, C_k^t] + b_1) + b_2), \quad (17)$$

where V_1, V_2, b_1, b_2, λ and μ are learnable parameters under constrain $\lambda + \mu = 1$. Note that if the word is an oov word and does not appear in utterance, $P_v(w)$ is zero and we copy words from knowledge instead of dialogue history.

3 Experiments

3.1 Datasets

We use two recently released datasets Wizard-of-Wikipedia and DuConv, whose knowledge formats are sentences and triplets respectively.

Wizard-of-Wikipedia (Dinan et al., 2018): an open-domain chit-chat dataset between agent wizard and apprentice. Wizard is a knowledge expert who can access any information retrieval system recalling paragraphs from Wikipedia relevant to the dialogue, which unobserved by the agent apprentice who plays a role as a curious learner. The dataset contains 22311 dialogues with 201999 turns, 166787/17715/17497 used for train/valid/test, and the test set is split into two subsets, Test Seen(8715) and Test Unseen(8782). Test Seen has 533 overlapping topics with the training set; Test Unseen contains 58 topics never seen before in train or validation. We do not use the ground-truth knowledge information provided in this dataset because the ability of knowledge selection during generation is a crucial part of our model.

DuConv (Wu et al., 2019b): a proactive conversation dataset with 29858 dialogs and 270399 utterances. The model mainly plays the role of a leading player assigned with an explicit goal, a knowledge path comprised of two topics, and is provided with knowledge related to these two topics. The knowledge in this dataset is a format of the triplet(subject, property, object), which totally contains about 144k entities and 45 properties.

3.2 Comparison Approaches

We implement our model both on datasets Wizard-of-Wikipedia and DuConv, and compare our approach with a variety of recently competitive baselines in these datasets, respectively. In Wizard-of-Wikipedia, we compare the approaches as follows:

- **Seq2Seq:** an attention-based Seq2Seq without access to external knowledge which is widely used in open-domain dialogue. (Vinyals and Le, 2015)
- **MemNet(hard/soft):** a knowledge grounded generation model, where knowledge candidates are selected with semantic similarity(hard); / knowledge candidates are stored into the memory units for generation (soft). (Ghazvininejad et al., 2018)
- **PostKS(concat/fusion):** a hard knowledge grounded model with a GRU decoder where knowledge is concatenated (concat); / a soft model use HGFU to incorporated knowledges with a GRU decoder.(Lian et al., 2019)
- **KIC:** Our joint neural conversation model named knowledge-aware pointer networks and recurrent knowledge interaction hybrid generator.

While in dataset DuConv, a Chinese dialogue dataset with structured knowledge, we compare to the baselines referred in (Wu et al., 2019b)

that consists of **retrieval-based** models as well as **generation-based** models.

3.3 Metric

We adopt an automatic evaluation with several common metrics proposed by (Wu et al., 2019b; Lian et al., 2019) and use their available automatic evaluation tool to calculate the experimental results to keep the same standards. Metrics include Bleu1/2/3, F1, DISTINCT1/2 automatically measure the fluency, coherence, relevance, diversity, etc. Metric F1 evaluates the performance at the character level, which mainly uses in Chinese dataset DuConv. Our method incorporates generation with knowledge via soft fusion that does not select knowledge explicitly, therefore we just measure the results of the whole dialog while not evaluate performances of knowledge selection independently. Besides, we provide 3 annotators to evaluate the results on a human level. The annotators evaluate the quality of dialog response generated on fluency, informativeness, and coherence. The score ranges from 0 to 2 to reflect the fluency, informativeness, and coherence of results from bad to good. For example, of coherence , score 2 means the response with good coherence without illogical expression and continues the dialogue history reasonably; score 1 means the result is acceptable but with a slight flaw; score 0 means the statement of result illogically or the result improper to the dialog context.

3.4 Implement Detail

We implement our model over Tensorflow framework(Abadi et al., 2016). And our implementation of point networks is inspired by the public code provided by (See et al., 2017). The utterance sequence concats the tokens of dialog history and separated knowledge. And the utterance encoder has a single-layer bidirectional LSTM structure with 256 hidden states while the response decoder has a single-layer unidirectional LSTM structure with the same dimensional hidden states. And the knowledge encoder has a 2-layer transformer structure. We use a vocabulary of 50k words with 128 dimensional random initialized embeddings instead of using pre-trained word embeddings. We train our model using Adagrad (Duchi et al., 2011) optimizer with a mini-batch size of 128 and learning rate 0.1 at most 130k iterations(70k iterations on Wizard-of-Wikipedia) on a GPU-P100 machine. The overall parameters are about 44 million and the model size is about 175MB, which decreases about 38% against the overall best baseline PostKS(parameters:71 million, model size: 285M)

3.5 Results and Analysis

3.5.1 Automatic Evaluation

As the experimental results on Wizard-of-Wikipedia with automatic evaluation summarized in Table 1, our approach outperforms all competitive baseline referred to recently working (Lian et al., 2019), and achieves significant improvements over most of the automatic metrics both on Seen and Unseen Test sets. The Bleu-1 enhances slightly in Test Seen while improving obviously in Test Unseen. Bleu-2 and Bleu-3 both yield considerable increments not only in Test Seen but in Test Unseen as well, for example, the Bleu-3 improves about 126% (absolute improvement: 0.043) in Test Seen and about 234%(absolute improvement: 0.047) in Test Unseen. The superior performance on metrics Bleu means the dialog response generated by model KIC is closer to the ground-truth response and with preferable fluency. As all

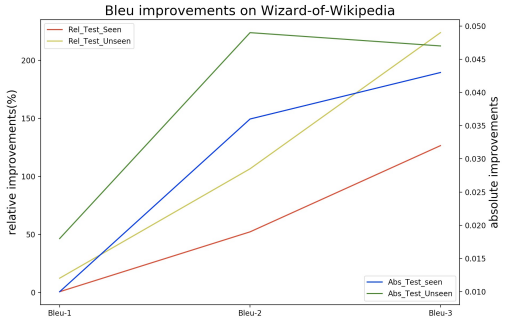

Figure 2: Bleu improvements on Wizard-of-Wikipedia.

Bleu metrics are shown in Figure 2, we can find that the improvement of result increasing with the augment of Bleu's grams, which means the dialog response produced via model KIC is more in line with the real distribution of ground-truth response in the phrase level, and the better improvement on higher gram's Bleu reflects the model have preferable readability and fluency. Generally, the ground-truth responses in datasets make up with the expressions from knowledge which conduces to the informativeness of response. As the recurrent knowledge interaction module in model KIC provides a mechanism to interact with the knowledge when decoding words of dialog response step by step. Moreover, the knowledge-aware pointer

Models	Test Seen		Test Unseen	
	Bleu-1/2/3	DISTINCT-1/2	Bleu-1/2/3	DISTINCT-1/2
Seq2Seq	0.169/0.066/0.032	0.036/0.112	0.150/0.054/0.026	0.020/0.063
MemNet(hard)	0.159/0.062/0.029	0.043/0.138	0.142/0.042/0.015	0.029/0.088
MemNet(soft)	0.168/0.067/0.034	0.037/0.115	0.148/0.048/0.023	0.026/0.081
PostKS(concat)	0.167/0.066/0.032	0.056/0.209	0.144/0.043/0.016	0.040/0.151
PostKS(fusion)	0.172/0.069/0.034	0.056/0.213	0.147/0.046/0.021	0.040/0.156
KIC(ours)	**0.173/0.105/0.077**	**0.138/0.363**	**0.165/0.095/0.068**	**0.072/0.174**

Table 1: Automatic Evaluation on Wizard-of-Wikipedia. The results of baselines are taken from (Lian et al., 2019).

Models	F1	Bleu-1	Bleu-2	DISTINCT-1	DISTINCT-2	ppl
norm retrieval	34.73	0.291	0.156	0.118	**0.373**	-
norm Seq2Seq	39.94	0.283	0.186	0.093	0.222	10.96
generation w/o klg.	28.52	0.29	0.154	0.032	0.075	20.3
generation w/ klg.	36.21	0.32	0.169	0.049	0.144	27.3
norm generation	41.84	0.347	0.198	0.057	0.155	24.3
KIC(ours)	**44.61**	**0.377**	**0.262**	**0.123**	0.308	**10.36**

Table 2: Automatic Evaluation on DuConv. Here, klg. denotes knowledge and norm stands for normalization on entities with entity types, norm generation is the PostKS in Table1. The results of baselines are taken from (Wu et al., 2019b).

network in KIC allows copying words from the expression of knowledge while decoding. Therefore, the dialog response generated by KIC contains relatively complete phrases of knowledge that as knowledge-informativeness as the ground-truth response. In addition, the improvements of metrics Bleu increase from Test Seen to Test Unseen, that is to say, the KIC with an advantage in case of unseen knowledge guided dialogue, which shows that our model is superior to address the dialogues with topics never seen before in train or validation. Besides, the metrics DISTINCT also achieves impressive results and prior than most of the baselines, about average 77% over the most competitive method PostKS. The metrics DISTINCT mainly reflects the diversity of generated words, whose improvements indicating that the dialogue response produced by KIC could present more information. In addition to experiments on Wizard-of-Wikipedia, we also conduct experiments on DuConv to further verify the effectiveness of our model on structured knowledge incorporated conversation. As the dataset DuConv released most recently that we compare our model to the baselines mentioned in the (Wu et al., 2019b) which are first applied to the DuConv including both retrieval-based and generation-based methods. The results presented in Table 2 show that our model obtains the highest results in most of the metrics with obvious improvement over retrieval and generation methods. Concretely, the F1, average Bleu, average DISTINCT, and ppl are over the best results of baseline norm generation about 6.6%, 20.5%, 115.8%, and 5.5%. Similar to Wizard-of-Wikipedia, the impressive augments of metrics demonstrate that the model has the capacity of producing appropriate responses with fluency, coherence, and diversity.

Metrics	Wizard-of-Wikipedia	DuConv
Fluency	1.90	1.97
Coherence	1.50	1.64
Informativeness	1.12	1.62

Table 3: Human Evaluation for the results of KIC.

3.5.2 Human Evaluation

In human evaluation, according to the dialogue history and the related knowledge, the annotators evaluate the quality dialog responses in terms of fluency and coherence. The score ranges from 0 to 2; the score is as higher as the responses are more fluent, informative, and coherent to the dialog context and integrate more knowledge. Manual evaluation results are summarized in Table 3, the model achieves high scores both in Wizard-of-Wikipedia and DuConv, meaning that the responses generated by KIC also with good fluency, informativeness,

Models	F1	Bleu-1	Bleu-2	DISTINCT1	DISTINCT2	Parameters
Part1: seq2seq w/o klg.	26.43	0.187	0.100	0.032	0.088	**43.47M**
Part2: Part1 + w/ klg.	36.59	0.313	0.194	0.071	0.153	43.50M
Part3: Part2 + klg. copy	43.35	0.365	0.249	0.122	0.301	43.59M
KIC: Part3 + dyn. attn.	**44.61**	**0.377**	**0.262**	**0.123**	**0.308**	43.63M

Table 4: Automatic Evaluation on progressive components of model KIC over DuConv. Here, klg. and dyn.attn. denote knowledge and dynamic attention, klg.copy stands for knowledge-aware pointer networks. Metrics remain consistent with Table 2.

Models	Test Seen		Test Unseen	
	Bleu-1/2/3	DISTINCT-1/2	Bleu-1/2/3	DISTINCT-1/2
Part1	0.122/0.049/0.024	0.026/0.07	0.113/0.037/0.014	0.013/0.033
Part2	0.154/0.086/0.060	0.117/0.305	0.140/0.071/0.048	0.038/0.089
Part3	0.165/0.097/0.071	0.129/0.341	0.155/0.088/0.062	0.070/0.168
KIC	**0.173/0.105/0.077**	**0.138/0.363**	**0.165/0.095/0.068**	**0.072/0.174**

Table 5: Automatic Evaluation on progressive components of model KIC over Wizard-of-Wikipedia. Here, Part1,Part2 and Part3 are the same with Table 4. Metrics remain consistent with Table 1.

and coherence in human view, close to the superior performance of automatic evaluation.

3.6 Ablation Study

We conduct further ablation experiments to dissect our model. Based on the Seq2Seq framework, we aggrandize it with each key component of model KIC progressively and the results are summarized in Table 4 and Table 5. We first incorporate knowledge into Seq2Seq architecture with dot attention of knowledge and use a gate to control the utilization of knowledge during generation, and the results achieve considerable improvement with the help of knowledge. And then, we apply knowledge-aware pointer networks over the model illustrated in last step to introduce a copy mechanism, which increases effect significantly demonstrates the facilitation of knowledge-aware copy mechanism to produce dialogue response with important words adopted from utterance and knowledge. In the end, we replace the knowledge dot attention by dynamic attention updated with decode state recurrently, which is the whole KIC model proposed in this paper, and the experimental results show that such amelioration also achieves an impressive enhancement. The dynamic update of knowledge attention during decoding effectively integrates multiple knowledge into the response that improves the informativeness. The performances of the model are gradually improved with the addition of components, meaning that each key component of the model KIC plays a crucial role. Additionally, with the considerable improvement at each progressive step, the model size and the parameters just increase slightly, which means the model KIC has a good cost performance.

3.7 Case Study

As shown in Figure 3, we present the responses generated by our proposed model KIC and the model PostKS(fusion), which achieves overall best performance among competitive baselines. Given utterance and knowledge candidates, our model is better than PostKS(fusion) to produce context-coherence responses incorporating appropriate multiple knowledge with complete descriptions. The model KIC prefers to integrate more knowledge into dialogue response, riching the informative without losing fluency. Furthermore, our model has an additional capability of handling oov problem, which can generate responses with infrequent but important words (which are oov words most of the time) from the knowledge context, like the "Alfred Hitchcock Presents" in Figure 3. We also compare to the result of the model with static knowledge attention, whose result mismatches between the "award" and the representative work "Alfred Hitchcock Presents". The static knowledge attention calculated before decoding, the information and confidence losing with the decoding step by step, leading to mispairing the expression of multiple knowledge. While the recurrent knowledge interaction helps the decoder to fetch the closest knowledge information into the current decoding

<table>
<tr><td rowspan="8">Hist.</td><td colspan="2" rowspan="8">A: 给你推荐一部口碑不错的爱情电影吧。 (Let me recommend you a nice romantic movie.)
B: 好啊，你推荐吧。(Okay, recommend please.)
A: 叫迷恋荷尔蒙，有时间可以带对象一起看看。(Soldier's Girl, you can watch it with your girlfriend when you have time.)
B: 不好意思，我比较喜欢自己看。(Well, I prefer to watch it myself.)
A: 其实你是单身吧。(Actually, I guess you are single.)
B: 哈哈，这都让你猜到了，我都不好意思了。(Aha, you got it, so embarrassed.)
A: 嗯嗯。迷恋荷尔蒙这部电影的导演弗兰克·皮尔森可真棒，把这部电影拍的出神入化。(Uh-huh. The director of this movie Frank Pierson is really awesome, what an amazing film he makes.)
B: 这个导演的名字好耳熟啊！(The director's name sounds so familiar!)</td><td rowspan="8">Knowledge</td><td rowspan="5">弗兰克·皮尔森
Frank Pierson</td><td>祖籍
Hometown</td><td>美国
America</td></tr>
<tr><td>性别
Gender</td><td>男
Male</td></tr>
<tr><td>代表作
Masterpiece</td><td>希区柯克悬念故事集
Alfred Hitchcock Presents</td></tr>
<tr><td>职业
Occupation</td><td>导演
Director</td></tr>
<tr><td>出生日期
Birth Date</td><td>1925 - 5 - 12
1925 - 5 - 12</td></tr>
<tr><td rowspan="3">迷恋荷尔蒙
Soldier's Girl</td><td>导演
Director</td><td>弗兰克 · 皮尔森
Frank Pierson</td></tr>
<tr><td>口碑
Comment</td><td>口碑不错的爱情电影
A nice romantic movie</td></tr>
<tr><td>获奖
Award</td><td>美国金球奖(2004；第 61 届)提名-电视类-最佳迷你剧
The Golden Globe Awards (2004; 61st) nominated - TV categories - Best Mini Series</td></tr>
<tr><td rowspan="4">Res.</td><td>Target</td><td colspan="4">哈哈。弗兰克·皮尔森导演出生 1925 年 5 月 12 号，是位美国人哟。
Aha. Director Frank Pierson was born on May 12, 1925, and is an American.</td></tr>
<tr><td>PostKS (fusion)</td><td colspan="4">他的代表作是<unk>。
His masterpiece is <unk>.</td></tr>
<tr><td>KIC (static)</td><td colspan="4">他是美国人，还获得过希区柯克悬念故事集。
He is an American, and he was awarded Alfred Hitchcock Presents.</td></tr>
<tr><td>KIC (dynamic)</td><td colspan="4">他是 1925 年 5 月 12 日出生的，他还有一部代表作叫希区柯克悬念故事集。
He was born on May 12, 1925. He also has a masterpiece called Alfred Hitchcock Presents.</td></tr>
</table>

Figure 3: Case study of DuConv. The <unk> means the out-of-vocabulary. KIC(static) denotes the model using static knowledge attention instead of recurrent knowledge interaction. Knowledge used in responses are in bold letters. Inappropriate words are highlighted with red color.

state, which superior to learn the coherent collocation of multiple knowledge. Some more cases of Wizard-of-Wikipedia and DuConv will present in the appendix section.

4 Related Work

Conversation with knowledge incorporation has received considerable interest recently and is demonstrated to be an effective way to enhance performance. There are two main methods in knowledge-based conversation, retrieval-based approches(Wu et al., 2016; Tian et al., 2019) and generation-based approaches. The generation-based method which achieves more research attention focuses on generating more informative and meaningful responses via incorporate generation with structured knowledge (Zhu et al., 2017; Liu et al., 2018; Young et al., 2018; Zhou et al., 2018) or documental knowledge(Ghazvininejad et al., 2018; Long et al., 2017). Several works integrate knowledge and generation in the pipeline way, which deal with knowledge selection and generation separately. Pipeline approaches pay more attention to knowledge selection, such as using posterior knowledge distribution to facilitate knowledge selection (Lian et al., 2019; Wu et al., 2019b) or used context-aware knowledge pre-selection to guide select knowledge (Zhang et al., 2019). While various works entirety integration the knowledge with generation in an end-to-end way, which usually manage knowledge via external memory module. (Parthasarathi and Pineau, 2018) introduced a bag-of-words memory network and (Dodge et al., 2015) performed dialogue discussion with long-term memory. (Dinan et al., 2018) used a memory network to retrieve knowledge and combined with transformer architectures to generate responses. The pipeline approaches lack of flexibility as constricted by the separated knowledge selection, and the generation could not exploit knowledge sufficiently. The end-to-end approaches with memory module attention to knowledge statically, when integrating multiple knowledge into a response are easier to be confused. Whereas we provide a recurrent knowledge interactive generator that sufficiently fusing the knowledge into generation to produce more informative dialogue responses.

Our work is also inspired by several works of text generation using copy mechanisms. (Vinyals et al., 2015) used attention as a pointer to generate words from the input resource by index-based copy. (Gu et al., 2016) incorporated copying into seq2seq learning to handle unknown words. (See et al., 2017) introduced a hybrid pointer-generator that can copy words from the source text while retaining the ability to produce novel words. In task-oriented dialogue, the pointer networks were also used to improve copy accuracy and mitigate

the common out-of-vocabulary problem (Madotto et al., 2018; Wu et al., 2019a). Different from these works, we extend a pointer network referring to attention distribution of knowledge candidates that can copy words from knowledge resources and generate dialogue responses under the guidance of more complete description from knowledge.

5 Conclusion

We propose a knowledge grounded conversational model with a recurrent knowledge interactive generator that effectively exploits multiple relevant knowledge to produce appropriate responses. Meanwhile, the knowledge-aware pointer networks we designed allow copying important words, usually oov words, from knowledge. Experimental results demonstrate that our model is powerful to generate much more informative and coherent responses than the competitive baseline models. In future work, we plan to analyze each turn of dialogue with reinforcement learning architecture, and to enhance the diversity of the whole dialogue by avoiding knowledge reuse.

References

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283.

Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. 2018. Wizard of wikipedia: Knowledge-powered conversational agents. *arXiv preprint arXiv:1811.01241*.

Jesse Dodge, Andreea Gane, Xiang Zhang, Antoine Bordes, Sumit Chopra, Alexander Miller, Arthur Szlam, and Jason Weston. 2015. Evaluating prerequisite qualities for learning end-to-end dialog systems. *arXiv preprint arXiv:1511.06931*.

John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159.

Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. 2018. A knowledge-grounded neural conversation model. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Jiatao Gu, Zhengdong Lu, Hang Li, and Victor OK Li. 2016. Incorporating copying mechanism in sequence-to-sequence learning. *arXiv preprint arXiv:1603.06393*.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Rongzhong Lian, Min Xie, Fan Wang, Jinhua Peng, and Hua Wu. 2019. Learning to select knowledge for response generation in dialog systems. *arXiv preprint arXiv:1902.04911*.

Shuman Liu, Hongshen Chen, Zhaochun Ren, Yang Feng, Qun Liu, and Dawei Yin. 2018. Knowledge diffusion for neural dialogue generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1498.

Yinong Long, Jianan Wang, Zhen Xu, Zongsheng Wang, Baoxun Wang, and Zhuoran Wang. 2017. A knowledge enhanced generative conversational service agent. In *DSTC6 Workshop*.

Andrea Madotto, Chien-Sheng Wu, and Pascale Fung. 2018. Mem2seq: Effectively incorporating knowledge bases into end-to-end task-oriented dialog systems. *arXiv preprint arXiv:1804.08217*.

Prasanna Parthasarathi and Joelle Pineau. 2018. Extending neural generative conversational model using external knowledge sources. *arXiv preprint arXiv:1809.05524*.

Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*.

Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.

Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. 2015. End-to-end memory networks. In *Advances in neural information processing systems*, pages 2440–2448.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Zhiliang Tian, Wei Bi, Xiaopeng Li, and Nevin L Zhang. 2019. Learning to abstract for memory-augmented conversational response generation. In

Proceedings of the 57th Conference of the Association for Computational Linguistics, pages 3816–3825.

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.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In *Advances in Neural Information Processing Systems*, pages 2692–2700.

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Chien-Sheng Wu, Richard Socher, and Caiming Xiong. 2019a. Global-to-local memory pointer networks for task-oriented dialogue. *arXiv preprint arXiv:1901.04713*.

Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. 2019b. Proactive human-machine conversation with explicit conversation goals. *arXiv preprint arXiv:1906.05572*.

Yu Wu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li. 2016. Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. *arXiv preprint arXiv:1612.01627*.

Hao-Tong Ye, Kai-Ling Lo, Shang-Yu Su, and Yun-Nung Chen. 2019. Knowledge-grounded response generation with deep attentional latent-variable model. *arXiv preprint arXiv:1903.09813*.

Tom Young, Erik Cambria, Iti Chaturvedi, Hao Zhou, Subham Biswas, and Minlie Huang. 2018. Augmenting end-to-end dialogue systems with commonsense knowledge. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Yangjun Zhang, Pengjie Ren, and Maarten de Rijke. 2019. Improving background based conversation with context-aware knowledge pre-selection. *arXiv preprint arXiv:1906.06685*.

Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Commonsense knowledge aware conversation generation with graph attention. In *IJCAI*, pages 4623–4629.

Wenya Zhu, Kaixiang Mo, Yu Zhang, Zhangbin Zhu, Xuezheng Peng, and Qiang Yang. 2017. Flexible end-to-end dialogue system for knowledge grounded conversation. *arXiv preprint arXiv:1709.04264*.

A Additional Comparison

In dataset Wizard-of-Wikipedia, (Lian et al., 2019) used the metrics Bleu1/2/3, distinct1/2 to evaluate their work, which different from the origin metrics (PPL, F1) used in (Dinan et al., 2018). In main body, we adopted metrics from (Lian et al., 2019) and compared the baselines presented in their work. We also implements a comparison using PPL&F1 metrics and compare to the methods listed in their paper. The results are summerized in Table 6 and Table 7. The Two-Stage Transformer Memory Networks with knowledge dropout(artificially prevent the model from attending to knowledge a fraction of the time during training) performs best in Test-Seen situation, while our KIC model achieves the best performance at Test-Unseen situation.

Models	Test Seen	
	PPL	F1
E2E MemNet (no auxiliary loss)	66.5	15.9
E2E MemNet (w/ auxiliary loss)	63.5	16.9
Two-Stage MemNet	54.8	18.6
Two-Stage MemNet (w/ K.D.)	**46.5**	**18.9**
KIC	51.9	18.4

Table 6: Comparisons with metrics from (Dinan et al., 2018) over Test-Seen. K.D. denotes knowledge dropout which involves artificial effort.

Models	Test Unseen	
	PPL	F1
E2E MemNet (no auxiliary loss)	103.6	14.3
E2E MemNet (w/ auxiliary loss)	97.3	14.4
Two-Stage MemNet	88.5	17.4
Two-Stage MemNet (w/ K.D.)	84.8	17.3
KIC	**65.8**	**17.3**

Table 7: Comparisons with metrics from (Dinan et al., 2018) over Test-Unseen. K.D. denotes knowledge dropout which involves artificial effort.

B Additional Cases

We have analyzed many cases both on Wizard-of-Wikipedia and DuConv, some of them are presented from Figure 4 to Figure 9. Our model KIC performs well in generating a fluent response coherent to the dialogue history as well as integrating multiple knowledge. Even in no history context situation (the model first to say), the KIC also has the capability of incorporating knowledge to start a knowledge relevant topic.

Hist.	[] Wizard first to say	
Knowledge	**K1**	The domestic cat ("Felis silvestris catus" or "Felis catus") is a small, typically furry, carnivorous mammal.
	K2	They are often valued by humans for companionship and for their ability to hunt vermin.
	K3	There are more than seventy cat breeds recognized by various cat registries.
	K4	Cat senses fit a crepuscular and predatory ecological niche.
Res.	**Target**	I love cats, I have one names Sam is a small, typically furry, carnivorous mammal.
	KIC (ours)	I have a domestic cat that is a small, typically furry, carnivorous mammal

Figure 4: Case of wizard-of-wikipedia with no dialog history.

Hist.	A: "I love cats, I have one names Sam is a small, typically furry, carnivorous mammal." B: "Cats are my favorite animal, I have one too, I think he is a tom since he is a boy", A: "that is neat. They are often called house cats when kept as indoor pets.", B: "Ahh, that totally makes since, I never let my kitty outside, I am terrified of the gross things they may bring back inside the house!"	
Knowledge	**K1**	They are often called house cats when kept as indoor pets or simply cats when there is no need to distinguish them from other felids and felines.
	K2	They are often valued by humans for companionship and for their ability to hunt vermin.
	K3	There are more than seventy cat breeds recognized by various cat registries.
	K4	Cat senses fit a crepuscular and predatory ecological niche.
Res.	**Target**	Cats are great pets. They are often valued by humans for companionship
	KIC (ours)	They are often valued by humans for companionship and for their ability to hunt vermin.

Figure 5: Case of wizard-of-wikipedia with long knowledge copy.

Hist.	A: "One of my favorite foods is a nice steak with wine" B: "Which type of steak are you talking about as we have fish steaks, ground meat steaks, pork steak and many more varieties of steaks." A: "Typically I would go with a meat or pork type steak. I love bacon wrapped pork chops"	
Knowledge	**K1**	Pork is one of the most commonly consumed meats in the world.
	K2	In the United States, pork chops are the most commonly consumed meat cut from the pork loin.
	K3	Chops are commonly served as an individual portion.
	K4	Pork chops are not processed and are leaner than other cuts of pork.
Res.	**Target**	Definitely. Everyone like pork. It is one of the most commonly consumed meats in the world.
	KIC (ours)	I love pork chops which is the most commonly consumed meats cut in the world.

Figure 6: Case of wizard-of-wikipedia with multiple knowledge integration.

Hist.	No History. First to say.	**Knowledge**	告白 Confessions	类型 Genre	惊悚 Horror
				国家 Country	日本 Japan
				是否 上映 If Released	已上映 Released
				领域 Field	电影 Movie
				发布日期信息 Release Date	5 年前上映 Released 5 years ago
			哈纳莱伊湾 Hanalei Bay	国家 Country	日本 Japan
				发布日期信息 Release Date	今年将会上映 Will be released this year
				类型 Genre	剧情 Drama
Res.	**Target**	喜欢看日本的电影么？ Do you like watching Japanese movies?			
	KIC (ours)	你喜欢看日本电影吗？ Do you like watching Japanese movies?			

Figure 7: Case of DuConv with no dialog history.

<table>
<tr><td rowspan="10">Hist.</td><td rowspan="10">A: 女性视角是什么样的呢？(What is the female perspective?)
B: 这个不清楚。(I don't know.)
A: 总要在伤害过后才更明白爱。(Always understand love after the injury.)
B: 哎！是啊。(Ah! Yeah.)
A: 有部叫蓝白红三部曲之白，你看一下这个电影，描述的就是女性视角。(A movie called Trois couleurs: Blanc, which describes a female perspective, you can watch it)
B: 好的，我会认真观看一下。(Okay, I will watch it carefully.)
A: 这个电影是朱丽叶·比诺什的代表作，她跟你一样是双鱼座的哦！(This movie is the masterpiece of Juliet Binoche. She is a Pisces like you!)
B: 还挺巧，不过不是很了解。(What a coincidence, but I don't know her very well.)</td><td rowspan="10">Knowledge</td><td rowspan="6">朱丽叶 · 比诺什
Juliette Binoche</td><td>祖籍
Hometown</td><td>法国巴黎
Paris, France</td></tr>
<tr><td>性别
Gender</td><td>女
Female</td></tr>
<tr><td>星座
Constellation</td><td>双鱼座
Pisces</td></tr>
<tr><td>评论
Comment</td><td>很有母性韵味的女人。。。角色也都收放自如 A woman with a very maternal charm. . . Her roles are all acted freely.</td></tr>
<tr><td>标签
Tag</td><td>女性 视角
Female perspective</td></tr>
<tr><td rowspan="4">蓝白红三部曲之白
Trois couleurs: Blanc</td><td>标签
Tag</td><td>女性 视角
Female perspective</td></tr>
<tr><td>口碑
Comment</td><td>口碑不错
Good Reputation</td></tr>
<tr><td>时光网短评
Mtime
Short Comment</td><td>总 要 在 伤害 过后 才 更 明白 爱 。 Always understand love after the injury.</td></tr>
<tr><td rowspan="2">Res.</td><td>Target</td><td colspan="4">她的祖籍是在法国巴黎，网友觉得她很有母性韵味，角色的饰演也都很到位。
Her hometown is Paris, France. Netizens think she has a maternal charm and the roles were all played well.</td></tr>
<tr><td>KIC (ours)</td><td colspan="4">她祖籍是法国巴黎的，网友评论说她是很有母性的女人，角色也都收放自如。
Her hometown is Paris, France. Netizens commented that she was a very maternal woman, and the roles were all acted freely.</td></tr>
</table>

Figure 8: Case of DuConv with long knowledge copy.

<table>
<tr><td rowspan="9">Hist.</td><td rowspan="9">A: 2009 获得英国帝国奖最佳女演员提名的知道是谁吗？ (Do you know who was nominated for best actress in the British Empire Awards in 2009?)
B: 不知道，你知道的话告诉我吧！(I don't know, tell me if you know it!)
A: 欧嘉·柯瑞兰寇，她很迷人，容貌，演技还棒。(Olga Kurylenko, she is charming, her looks and acting are also good.)
B: 听你说应该很美。(She should be beautiful then.)
A: 必须的啊，而且演技还好，主演的一部恐怖片可是吓得我不轻(It must be, and her acting is also good. She starred in a horror movie, which scared me a lot.)
B: 是哪部电影啊，都可以吓得到你(Which movie that can scare you?)</td><td rowspan="9">Knowledge</td><td rowspan="4">欧嘉 · 柯瑞兰寇
Olga Kurylenko</td><td>祖籍
Hometown</td><td>美国
America</td></tr>
<tr><td>性别
Gender</td><td>女
Female</td></tr>
<tr><td>评论
Comment</td><td>她很迷人，容貌，演技
She is charming, looks, acting</td></tr>
<tr><td>职业
Occupation</td><td>演员
Actor</td></tr>
<tr><td>主要成就
Major Achievements</td><td>2009 获得英国帝国奖最佳女演员提名
Nominated for Best Actress in the British Empire Awards in 2009</td></tr>
<tr><td rowspan="4">玛拉
Mara</td><td>导演
Director</td><td>Clive Tonge
Clive Tonge</td></tr>
<tr><td>上映时间
Release Time</td><td>2018 年
2018</td></tr>
<tr><td>类型
Genre</td><td>恐怖
Horror</td></tr>
<tr><td rowspan="2">Res.</td><td>Target</td><td colspan="4">是那部 2018 年上映的玛拉，有时间你可以看一看
It's the Mara that was released in 2018. You can take a look at it sometime.</td></tr>
<tr><td>KIC (ours)</td><td colspan="4">名字叫做玛拉，2018 年上映的，导演是 Clive Tonge
The name is Mara and it was released in 2018. The director is Clive Tonge.</td></tr>
</table>

Figure 9: Case of DuConv with multiple knowledge integration.