Generating Natural Answers on Knowledge Bases and Text by Sequence-to-Sequence Learning

Zhihao Ye, Ruichu Cai, Zhaohui Liao, Zhifeng Hao, and Jinfen Li

School of Computer Science and Technology,
Guangdong University of Technology, Guangzhou, China
zhihaoye.chn@qq.com

Abstract. Generative question answering systems aim at generating more contentful responses and more natural answers. Existing generative question answering systems applied to knowledge grounded conversation generate natural answers either with a knowledge base or with raw text. Nevertheless, performance of their methods is often affected by the incompleteness of the KB or text facts. In this paper, we propose an end-to-end generative question answering model. We make use of unstructured text and structured KBs to establish an universal schema as a large external facts library. Each words of a natural answer are dynamically predicted from the common vocabulary and retrieved from the corresponding external facts. And our model can generate natural answer containing arbitrary number of knowledge entities through selecting from multiple relevant external facts by the dynamic knowledge enquirer. Finally,empirical study shows that our model is efficient and outperforms baseline methods significantly in terms of automatic evaluation and human evaluation.

Keywords: Natural Answers, Universal Schema, Sequence-to-Sequence Learning

1 introduction

Recent neural models of dialogue generation such as sequence to sequence model can be trained in an end-to-end and completely data-driven fashion. However, these fully data-driven models tend to generate safe responses that are boring and carry little information. In other words, these models can not have access to any external knowledge, which makes it difficult to respond substantively. From another perspective, we can consider generative question answering as a special case of knowledge grounded conversation. As the examples shown in table 1, daily conversations generally depends on individual's knowledge. Recently, some researchers proposed neural conversation model that can generate natural answers and knowledge-grounded responses either with knowledge base or with raw text.

Table 1. Examples of training instances for our model. The natural answer containing mutil-number of knowledge entities is generated based on both Knowledge Bases and Text.

KB fact	(Peking University, President,Yan Fu)
text fact	Yan Fu is the first President of Peking University.
UserA: Who was the first President of Peking University? **UserB:** The first President is Yan Fu	
KB fact	(The Journey to the West, author, Wu Chenen) (The Journey to the West, written time,Ming dynasty)
text fact	Wu Chenen who is the author of The Journey to the West was an outstanding novelist of the Ming dynasty.
UserA: Who is the author of "journey to the west". **UserB:** It was wu chengen in the Ming dynasty.	

In order to generate more contentful responses, more and more generative question answering systems and knowledge-grounded conversation model are proposed. On one hand, Ghazvininejad et al. [4] utilized external textual information as the unstructured knowledge. They found that unstructured knowledge can make a response more contentful. On the other hand, Yin et al.[18], He et al.[6] and Zhu et al.[20] have proposed generative question answering (QA) model that can generate natural answers by entities retrieved from the KB and seq2seq model[15]. But, the performance of model above are often affected by the incompleteness of the KB or text. How to generate more contentful resposes or natural answer by exploiting KB and text together is necessary to study.

In this paper, we propose our neural generative dialogue model, which can generate responses based on input message and external facts. For the first time, we propose our approach that combined text and KB library as our external facts by building the universal schema [10] to generate natural answer. In each time step of generating the natural answer, the possible word may come from common word vocabulary or knowledge entity vocabulary and the natural answer that contains the relevant arbitrary number of entities can be generated. Finally we conduct experiments on real-world datasets. Experimental results demonstrate that combining unstructured knowledge with structured knowledge is effective for generating natural answer, and our model is more efficient than the existing end-to-end QA/Dialogue model.

2 related work

Recently, sequence-to-sequence [7][15] learning, which can predict target sequence given source sequence, has been widely applied in dialogue systems. Shang et al. [14] first utilized the encoder and decoder framework to generate responses on micro-blogging websites. And after that, more and more dialogue system [16] [12] [13] on the basis of seq2seq framework were proposed. In our work, our model is also based on seq2seq framework and we try to combine the external facts composed of KB and text to generate more contentful responses.

Many researchers propose open domain dialogue system which can incorporate external knowledge to enhance reply generation. Han et al.[5] proposed a rule-based dialogue system by filling the response templates with retrieved KB. Ghazvininejad et al.[4] utilized external textual information as the unstructured knowledge. As demonstrated, the external textual information can convey more relevant information to responses. Some recent work used external structured knowledge graph to build end-to-end question answering systems. Yin et al. [18] proposed a seq2seq-based model where answers were generated in two ways, where one was based on a language model and the other was by some entities retrieved from the KB. He et al. [6]and [20] further studied the cases where questions require multiple facts and out-of-vocabulary entities.

In older to improve the performance of knowledge base QA model. Das et al. [3] extend universal schema to natural language question answering, employing memory networks to attend to the large amount of facts in the combination of text and KB. Inspired by them, we also have built the universal schema to combine KB and text and tried to employ a key-value MemNN model as our knowledge enquirer. But different from them, our model can generate more natural answer, rather than a single entity. Other work such as [19] [17], also put forward some models to exploit KB and the text together,but their formulations are totally different from ours.

3 Our Framework

3.1 Framework overview

In real-world environments, people prefer to reply one's question with a more natural way. Jsut like the example shown in table 1, When user A asks "Who is the first President of Peking University?", user B should answer: "The first President is Yan Fu" rather than only one entity or an answer that is not relevant to the question. For the above natural language question-answering scenario, in our workthe problem can be defined as:given an input message $Q = (x_1, x_2...x_L)$, the problem is to generate an appropriate response $Y = (y_1, y_2...y_L)$ based all possible facts form text and KB. And in order to try to solve the above problems, we propose an end-to-end generative question answering system, which is illustrated in Figure 1

3.2 Candidate Facts Retriever

The candidate facts retriever identifies facts that are related to the input message. In our work, the model retrieve the relevant text facts by firstly finding the relevant KB triples(subject-property-object) from the universal schema.

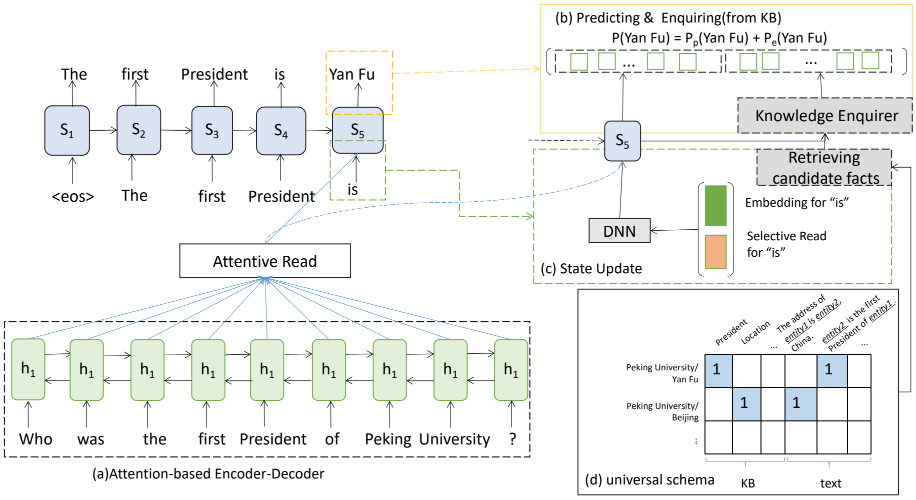

Fig. 1. The overview of our model. Our model consists of message encoder, candidate facts retriever, reply decoder, and universal schema containing external facts. When the user inputs a question, the knowledge retrieval module is firstly employed to retrieve related facts. And then message encoder encode the problem into hidden states. Finally, hidden state from message encoder are feed to reply decoder for generating natural answer.

Specifically We denote the entities of Q by $E = (e_1, e_2, .., e_m)$. E can be identified by keyword matching, or detected by more advanced methods such as entity linking or named entity recognition. Based on detected triples, we can retrieve the relevant facts from universal schema. Usually, question contains the information used to match the subject and property parts in a fact triple, and answer incorporates the object part information.

3.3 Question Encoder

In older to catch the user's intent and get hidden representations of input message. we employ a bidirectional GRU [2] [11] to transform the message $Q = (x_1, x_2...x_L)$ into a sequence of concatenated hidden states with two independent GRU. Once a message is encoded by message encoder, the forward and backward GRU respectively obtain the hidden state$\{\overrightarrow{h_1}, \overrightarrow{h_1}, ...\overrightarrow{h_L}\}$ and $\{\overleftarrow{h_L}, ...\overleftarrow{h_2}, \overleftarrow{h_1}\}$, where L is the maximum length of the message. The context memory of input message can be obtained by concatenated hidden state list $H_Q = \{h_1, ...ht, ..., h_L\}$, where h_t is equal to $[\overrightarrow{h}_t, \overleftarrow{h}_{(L-t+1)}]$. Besides, the last hidden state h_t is used to represent the entire message.

3.4 Reply Decoder

The reply decoder generates the final response Y based on the hidden representations of input message H_Q and candidate facts F_Q that come from the universal schema. There are two categories of possible words, the common words and knowledge words, in the generated response. Specifically, the probability of generating the answer :

$$p(y_1, y_2, ..., y_{L_Y}|H_Q, F_Q; \theta) = \\ p(y_1|H_Q, F_Q; \theta) \prod_{t=2}^{L_Y} p(y_t|y_1, y_2, ..., y_{t-1}, H_Q, F_Q; \theta) \quad (1)$$

where θ represents the parameters in the model. The generation probability of y_t is specified by

$$p(y_t|y_1, y_2, ..., y_{L_Y}, H_Q, F_Q; \theta) = p(y_t|y_{t-1}, z_t, s_t, H_Q, F_Q; \theta) \quad (2)$$

where s_t is the hidden state of the decoder model and $z_t \in \{0, 1\}$ is the value predicted by a binary classifier. In generating the t^{th} word y_t in the answer, the probability is given by the following mixture model.

$$p(y_t|y_1, y_2, ..., y_{L_Y}, H_Q, F_Q; \theta) = p_c(y_t|z = 0)p(z = 0|y_{t-1}, s_t, H_Q, F_Q; \theta) \\ + p_e(y_t|z = 1)p(z = 1|y_{t-1}, s_t, H_Q, F_Q; \theta) \quad (3)$$

Response words prediction classifier In order to generate the final response containing common words and knowledge words, we apply a MLP as a binary classifier and at each time step, feeding a time step s_{t-1}, y_{t-1}, the MLP classifier outputs a predicted value $z_t \in \{0, 1\}$. if $z_t = 0$, it means that the next generation word is from the entity vocabulary and in our work, the entity vocabulary contains all the "object" of the KB triples. And conversely, if $z_t = 1$, the next generation word is generated from common vocabulary. In summary, the y_t is generated as:

$$p(yt|y_{t-1}, z_t, s_t, H_Q, F_Q; \theta) = p_c(y_t)p(z = 0|y_{t-1}, s_t, H_Q, F_Q; \theta) \\ + p_e(y_t)p(z = 1|y_{t-1}, s_t, H_Q, F_Q; \theta) \quad (4)$$

Universal Schema To make full use of external facts from structured KBs and unstructured text , our external knowledge M comprise of both KB and text. And Inspired by Das et al. [3] we applied universal schema to integrate KB and text. Each cell of universal schema is in the form of key-value pair. Specifically, let $(s, r, o) \in K$ represent a KB triple, the key $\mathbf{k}$ is represented by concatenating the embeddings $\mathbf{s}$ and $\mathbf{r}$ and the object entity $\mathbf{o}$ is treated as it's value $\mathbf{v}$. For text, Let $(s, [w_1, ..., entity_1, ..., entity_2, w_n], o) \in T$ represent a textual fact, where

$entity_1$ and $entity_2$ correspond to the positions of the entities subject and object. We represent the key as the sequence formed by replacing $entity_1$ with subject and $entity_2$ with a special 'blank' token, i.e., $k = [w_1, ..., s, ..., blank, w_n]$, which is converd to a distributed representation using a bidirectional GRU and value as just the entity object $\mathbf{o}$.

Knowledge Enquirer We have chosen two implementations that have similar effect in our experiment as knowledge enquirer to calculate the matching scores between question and candidate facts. The first model is a two-layer MLP. The fact representation f is then defined as the concatenation of key and value. The list of all related facts' representations, $\{f\} = \{f_1, f_2, ..., f_{L_F}\}$ (L_F denotes the maximum of candidate facts), is considered to be a short-term memory of the large body external knowledge memory M. We define the matching scores function between question and facts as function is $S(q, s_t, f_j) = DNN1(q, s_t, f_j)$ where s_t is the hidden state of decoder at time t and DNN1 is the two-layer MLP. In addition, we also adopt the key-value MemNN proposeed by Miller et al.[8] where each memory slot consists of a key and value. It is worth noting that, excepting for question and related facts, We also need to use state s_t of decoding process as the input of the key-value MemNN because the matching results also depend on the state of decoding process at different times.

Common Word Generator To generate richer content and more matching answers to user questions, we applied a GRU model and attention mechanism to generate common words. Firstly, we calculate a message context vector c_t by using the attention mechanism [1] on the message hidden vectors H with the current generator hidden state s_{t-1}. And then, the word of the next time step s_t is obtained as $s_t = f(y_{t-1}, s_{t-1}, c_t)$. finally, the predicted target word y_t at time t is performed by a softmax classifier over a settled vocabulary (e.g. 40,000 words) through function g:$p(yt|y_{<t}, X) = g(y_{t-1}, s_t, c_t)$

State Update In the generic decoding process, each hidden state s_t is updated with the previous state s_{t-1}, the word embedding of previous predicted symbol y_{t-1}, and an optional context vector c_t (with attention mechanism). However, y_{t-1} may not come from entity vocabulary and not owns a word vector. Therefore, we modify the state update process. More specifically, y_{t-1} will be represented as $[e(y_{t-1}), \zeta_{k_{t-1}}]$, where $e(y_{t-1})$ is the word embedding associated with y_{t-1} and $\zeta_{k_{t-1}}$ are the weighted sum of hidden states in M_F corresponding to y_{t-1} .

$$\zeta_{kb_t} = \sum_{j=1}^{L_F} \delta_{tj} f_i \qquad \delta_{tj} = \begin{cases} \frac{1}{K} P_e(f_j|\cdot) & object(f_j) = y_t \\ 0 & otherwise \end{cases} \tag{5}$$

where $object(f)$ indicate the "object" part of fact f , and K are the normalization terms which equal $\sum_{j':object(f'_j)} P_e(f'_j|\cdot)$, which can consider the multiple positions matching y_t in external facts

4 Experiments

4.1 Dataset

For our experimental data, we used the data set provided by He et al.[6]. In addition, we have crawled the corresponding text facts from Baidu baike (a Chinese encyclopedia website). In our work, all "subject" entities and "object" entities of triples are used as encyclopedic items, and we crawl all article related to these encyclopedic items. The texts in Chinese in the data are converted into sequences of words using the Jieba Chinese word segmentor, then all related text facts were extracted through Keyword matching with KB triples. After extracting all the the relevant facts from the article, we used the facts from text and KB to establish the universal schema.

4.2 model

Firstly, we use seq2seq model with attention(seq2seq+atten) as one of our baselines, which is widely used in chit-chat dialogues system. And then, we also use generative QA model (GenQA [18]and COREQA [6]) as our baselines ,which can be applied in knowledge grounded conversation. Finally,We apply our model, and compared three types of external knowledge source which respectively comprise of only KB, only textual and universal schema containing both text and KB.

4.3 Evaluation metrics

We have compared our model with baselines by both automatic evaluation and human evaluation.

automatic evaluation Following the existing works, we employ the BLEU [9] automatic evaluation, which reflects the words occurrence between the ground truth and the generated response. And to measure the information correctness, we evaluate the performance of the models in terms of accuracy. Meanwhile,(same as COREQA [6]) we separately present the results according to the number of the facts which a question needs in knowledge base, including just one single fact (marked as Single), multiple facts (marked as Multi) and all (marked as Mixed). In our work, we randomly selected 5120 samples from data set as our test set, and the result is shown in table 2.

human evaluation We also recruit human annotators to judge the quality of the generated responses with aspects of Fluency, Correctness and grammar. All scores range from 1 to 5. Higher score represents better performance in terms of the above three metrics. In older to provide human evaluation, we randomly selected 300 samples from our test set, and the result is shown in table 3.

Table 2. The result of automatic evaluation on test data.

Models	BLEU	Single	Multi	Mixed
seq2seq+atten	0.39	20.1	3.5	19.4
GenQA	0.38	47.2	28.9	45.1
COREQA	-	58.4	42.7	56.6
our model$_{kb}$	0.42	56.2	45.9	54.7
our model$_{text}$	**0.45**	47.2	42.9	45.9
our model$_{text\&kb}$	0.43	**65.4**	**52.7**	**63.6**

4.4 Results

Table 2 shows the accuracies of the models on the test set. We can clearly observe that our model significantly outperforms all other baseline models and our model can generate correct answer that need single fact or multiple facts. This also proves that using KB and text as an external knowledge is helpful for generating more accurate natural answers and generating contentful responses.

Table 3. The result of human evaluation on test data.

Models	Fluency	Correctness	Grammar
seq2seq+atten	3.67	2.34	3.93
GenQA	3.56	3.39	3.73
our model$_{text\&kb}$	**4.12**	**4.42**	**4.19**

As illustrated in Table 3, the results show that our framework outperforms other baseline models. The most significant improvement is from correctness, indicating that our model can generate more accurate answer.

5 Conclusion and Future Work

In this paper, we propose an end-to-end generative question answering system to generate natural answers containing arbitrary number of knowledge entities.We establish an universal schema as large external fact library using unstructured text and structured KB. The experimental results show that our model can generate more natural and fluent answers and universal schema is a promising knowledge source for generating natural answer than using KB or text alone. However, after extracting related text facts from raw text through keyword matching with KB triples, a lot of useful text data also were discarded. In the future, we plan to explore ways to more effectively combine structured and unstructured knowledge with a fuller use of text.

References

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. Computer Science (2014)

2. Chung, J., Gulcehre, C., Cho, K.H., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. Eprint Arxiv (2014)
3. Das, R., Zaheer, M., Reddy, S., Mccallum, A.: Question answering on knowledge bases and text using universal schema and memory networks pp. 358–365 (2017)
4. Ghazvininejad, M., Brockett, C., Chang, M.W., Dolan, B., Gao, J., Yih, W.T., Galley, M.: A knowledge-grounded neural conversation model (2017)
5. Han, S., Bang, J., Ryu, S., Lee, G.G.: Exploiting knowledge base to generate responses for natural language dialog listening agents. In: Meeting of the Special Interest Group on Discourse and Dialogue. pp. 129–133 (2015)
6. He, S., Liu, C., Liu, K., Zhao, J., He, S., Liu, C., Liu, K., Zhao, J.: Generating natural answers by incorporating copying and retrieving mechanisms in sequence-to-sequence learning. In: Meeting of the Association for Computational Linguistics. pp. 199–208 (2017)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780 (1997)
8. Miller, A., Fisch, A., Dodge, J., Karimi, A.H., Bordes, A., Weston, J.: Key-value memory networks for directly reading documents pp. 1400–1409 (2016)
9. Papineni, S.: Blue ; a method for automatic evaluation of machine translation. In: Meeting of the Association for Computational Linguistics (2002)
10. Riedel, S., Yao, L., Mccallum, A., Marlin, B.M.: Relation extraction with matrix factorization and universal schemas. In: NAACL-HLT. p. xxixxii (2013)
11. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Press (1997)
12. Serban, I.V., Sordoni, A., Bengio, Y., Courville, A., Pineau, J.: Building end-to-end dialogue systems using generative hierarchical neural network models. In: Thirtieth AAAI Conference on Artificial Intelligence. pp. 3776–3783 (2016)
13. Serban, I.V., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A., Bengio, Y.: A hierarchical latent variable encoder-decoder model for generating dialogues (2016)
14. Shang, L., Lu, Z., Li, H.: Neural responding machine for short-text conversation pp. 52–58 (2015)
15. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks 4, 3104–3112 (2014)
16. Yao, K., Zweig, G., Peng, B.: Attention with intention for a neural network conversation model. Computer Science (2015)
17. Yao, L., Riedel, S., Mccallum, A.: Collective cross-document relation extraction without labelled data. University of Massachusetts - Amherst (2010)
18. Yin, J., Jiang, X., Lu, Z., Shang, L., Li, H., Li, X.: Neural generative question answering 27, 2972–2978 (2015)
19. Zeng, D., Liu, K., Chen, Y., Zhao, J.: Distant supervision for relation extraction via piecewise convolutional neural networks. In: Conference on Empirical Methods in Natural Language Processing. pp. 1753–1762 (2015)
20. Zhu, W., Mo, K., Zhang, Y., Zhu, Z., Peng, X., Yang, Q.: Flexible end-to-end dialogue system for knowledge grounded conversation (2017)