Knowledge-aware Dialogue Generation with Hybrid Attention (Student Abstract)

Yaru Zhao, Bo Cheng, Yingying Zhang
State Key Laboratory of networking and switching technology,
Beijing University of Posts and Telecommunications
{zhaoyaru,chengbo,YingyingZhang}@bupt.edu.cn

Abstract

Using commonsense knowledge to assist dialogue generation is a big step forward for dialogue generation task. However, how to fully utilize commonsense information is always a challenge. Furthermore, the entities generated in the response do not match the information in post most often. In this paper, we propose a dialogue generation model which uses hybrid attention to better generate rational entities. When a user post is given, the model encodes relevant knowledge graphs from a knowledge base with a graph attention mechanism. Then it will encode the user post and graphs with a co-attention mechanism, which effectively encodes complex related data. Through the above mechanism, we can get a better mutual understanding of post and knowledge. The experimental results show that our model is more effective than the current state-of-the-art model (CCM).

Introduction

The dialogue generation task is designed to generate reasonable responses based on the posts. Currently, dialogue generation tasks are either fully data-driven or based on external knowledge. In this paper, we focus on the latter, we try to utilize knowledge graph as external knowledge to generate responses.

Conversational system is one of the ultimate goals of artificial intelligence development. Fully data-driven neural models seem to generate dialogues which are smooth but lack actual content and practical significance. As a result, many studies have turned to dialogue generation task involving external knowledge.

In recent years, many researches have integrated various forms of external knowledge (such as text and knowledge graph) into the task of dialogue generation(Ghazvininejad et al.,2018; Zhou et al., 2018), and the performance of dialogue system has been significantly improved. However, how to use external knowledge to the extreme is a problem. There is a lack of mutual understanding between post and knowledge, so the entities in the generated response cannot match the meanings in the post.

The co-attention mechanism has been widely used in natural language processing tasks such as reading comprehension and question-answering (Zhong et al., 2019). It can better capture text paragraphs and information related to the question. However, it has not yet been mentioned in this task.

In this paper, we propose a knowledge-aware dialogue generation model with graph attention and co-attention.The model first retrieves the knowledge associated with post, and then uses the knowledge to generate a reasonable response. In order to realize mutual understanding between post and knowledge, we use the hybrid attention mechanism in encoding, that is to say, the graph attention mechanism encodes knowledge, followed by the co-attention mechanism encodes post and knowledge. Finally, the dynamic graph attention mechanism (Zhou et al., 2018) is used to select knowledge and generate response.

Model

Task Definition. Given a post X = $x_1, ..., x_n$, where n denotes the number of words in the post, and some commonsense knowledge graphs G = $g_1, g_2, , g_t$ which are corresponded to the post. The task is to generate a proper response Y = $y_1, ..., y_m$, where m is the length of response. The overall framework of the model is shown in Figure 1.

Representation of the post and KG. We use Glove embedding to initialize the representation of words in the post as X_p. And we obtain representation of external knowledge with the TransE embedding as X_g. **Encoder with Hybrid Attention.** First we apply graph attention to compute a more comprehensive representation of related graphs, which takes the joint information representation of the entities and relations into account, so we get $X_{g'}$. Then we use co-attention mechanism to learn the representation of post and knowledge facts that are relevant to it. To some extent, information fusion is realized. In the end, we get the input of encoder:

$$A_{gp} = X_{g'}(X_p^T) \tag{1}$$

$$C_p = softmax(A_{gp}^T)X_{g'} \tag{2}$$

$$D_p = [X_p^T; C_p] \tag{3}$$

$$C_g = D_p(softmax(A_{gp})) \tag{4}$$

$$C_{coa} = [X_{g'}; C_g^T] \tag{5}$$

where the letter T represents matrix transposition, softmax() function is column-wise normalization, and [;] is column-wise concatenation of two vectors. At this point, we have

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

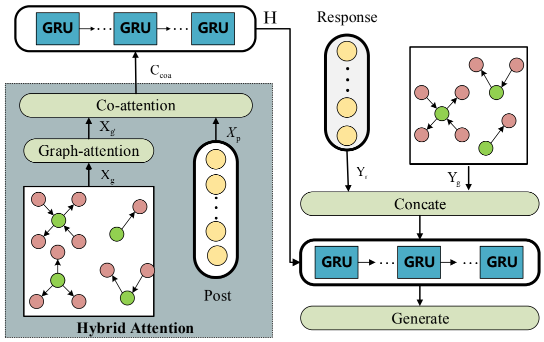

Figure 1: The overview of the model.

mutual information representations of post and knowledge. Then we feed the above result to the encoder which is made up of n-layer GRUs.

$$h_t = GRU(h_{t-1}, C_{coa}(x_t)) \tag{6}$$

Knowledge-aware decoder. The knowledge-aware decoder is to generate proper responses by fully considering external knowledge. We followed the previous work (Zhou et al., 2018) using dynamic graph attention mechanism to get the distribution over generic words and entity words respectively. y_t is a concatenation of the embedding of response y_r and the representation of the related knowledge y_g.

$$s_t = GRU(s_{t-1}, [c_{t-1}; y_{t-1}]) \tag{7}$$

Finally, we utilize the distribution to select a word or an entity for generation. This distribution is a coincidence distribution that is weighted by common words and entities, and the weight is calculated by s_t and c_t.

$$output = P(y_t|y_{<t}, c_t) \tag{8}$$

c_t is complex context vector, which is a combination of the context vector of encoder, graph information, information of triples in the graph and response.

Experiment

Dataset. We use Reddit (Zhou et al.2018) single-round dialogue dataset, which includes 3,384,185 training pairs, 10,000 validation pairs and 20,000 test pairs. The external knowledge is from ConceptNet.

Baselines. We compare our model with the following models: Seq2Seq (Sutskever, Vinyals, and Le 2014), CopyNet (Zhu et al.2017), MemNet (Ghazvininejad et al. 2018), and CCM (Zhou et al. 2018).

Training Details. We applied TransE embeddings for entities and relations, and 300 dimensional Glove word vectors for post words. The embedding size of entities and relations is set to 100. The vocabulary size was set to 30,000. The encoder and decoder layers were set to 2-layer GRU with a hidden size of 512. We used the Adam optimizer. The learning rate of first five epochs was set to 0.0005, which aimed to converge faster to save training time, then we adjust it to 0.00005 to get a good learning effect. The batch size was set to 100.

Model	Overall	High	Medium	Low	OOV
Seq2Seq	47.02	42.41	47.25	48.61	49.96
MemNet	46.85	41.93	47.32	48.86	49.52
CopyNet	40.27	36.26	40.99	42.09	42.24
CCM	39.18	35.36	39.64	40.67	40.87
HADG (ours)	**38.69**	**35.01**	**39.23**	**40.11**	**40.67**

Table 1: Automatic evaluation with perplexity.

Model	Overall	High	Medium	Low	OOV
Seq2Seq	0.717	0.713	0.740	0.721	0.669
MemNet	0.761	0.764	0.788	0.760	0.706
CopyNet	0.960	0.910	0.970	0.960	0.960
CCM	1.180	1.156	**1.191**	1.196	1.162
HADG (ours)	**1.185**	**1.173**	**1.191**	**1.2**	**1.177**

Table 2: Automatic evaluation with entity score.

Results and Discussion. Perplexity and entity score (Zhou et al. 2018) are adopted to evaluate the performance of our model. As shown in Table1, perplexity of our model is lower than all the baselines, which indicates that the sentences generated by our model are more consistent with the grammatical structure. As shown in Table2, entity score of our model exceeds all the baselines, which proves our model can capture entities more efficiently. The above experimental results show that our method can indeed make post and knowledge establish a closer connection, which is very beneficial for both encoding and decoding.

Acknowledgments

This work is supported by the National Key Research and Development Program of China under grant 2018YFB1003804, in part by the National Natural Science Foundation of China under grant 61921003, 61972043. Bo Cheng is the corresponding author.

References

Ghazvininejad,M.; Brockett,C.; Chang,M.-W.; Dolan,B.; Gao,J.; Yih,W.-t.; and Galley, M. 2018. A knowledge-grounded neural conversation model. In *AAAI*.

Zhou, H.; Young, T.; Huang, M.; Zhao, H.; Xu, J.; and Zhu, X. 2018. Commonsense knowledge aware conversation generation with graph attention. In *IJCAI*.

Victor Zhong.; Caiming Xiong.; Nitish Shirish Keskar.; and Richard Socher. 2019. Coarse-grain fine-grain coattention network for multi-evidence question answering. In *ICLR*.

Sutskever, I.; Vinyals, O.; and Le,Q.V.2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*.

Zhu,W.;Mo,K.; Zhang,Y.; Zhu,Z.; Peng, X.; and Yang, Q. 2017. Flexible end-to-end dialogue system for knowledge grounded conversation. In *arXiv preprint*, arXiv:1709.04264.