Graphs and Commonsense Knowledge improve the Dialogue Reasoning Ability

Minglei Gao[1], Sai Zhang[1], Xiaowang Zhang[1], Zhiyong Feng[1,2*], and Wenhuan Lu[1]

[1] College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[2] College of Intelligence and Computing, Shenzhen Research Institute of Tianjin University, Tianjin University, Tianjin, China
{minglei_gao3, zhang_sai, xiaowangzhang, zyfeng, wenhuan}@tju.edu.cn
* Corresponding Author

Abstract. Retrieving responses is a subtask in the dialogue system. The focus of the existing methods is semantic matching, but the reasoning ability is insufficient. The implicit feature association between the context and the candidate responses was not discovered. And this implicit feature association is precisely the key to realize reasoning. In this paper, we propose a new approach based on commonsense knowledge combined with graph features. We are using the advantages of graph structure in reasoning, putting the context and candidate responses in the same graph, and using commonsense knowledge to explicitly show the associated features, thereby improving the dialogue system's reasoning ability. Experiments show good performance through the effective combination of commonsense knowledge and graph structure.

Keywords: Retrieval Responses · Reasoning · GCN.

1 Introduction

Retrieving responses is an important approach in the dialogue system. Its goal is to choose the most suitable response based on the given context. The retrieved responses are often fluent and natural, with abundant information. The success of the retrieval can make the dialogue proceed more accurately and smoothly and can better enhance the user experience.

Previous work mainly concentrated on the matching relationship between context and candidate responses. It is well known that neural networks can learn multiple levels of rich information in semantics [1]. But, reasoning ability and the ability to capture commonsense information are insufficient [2]. Reasoning needs to learn key semantic features and perform effective reasoning based on the relationship between features. However, the feature information in the contextual semantics is not enough to support effective reasoning. It is very interesting to explore.

In this paper, we propose a graph reasoning model based on commonsense knowledge. Specifically, by extracting the key information in the context and candidate responses, and using commonsense knowledge to expand the key information explicitly, and using the grammatical information of the text to construct a graph, effective reasoning on the structured graph information can improve the reasoning ability of the model. By adding commonsense knowledge to the graph structure, an effective connection can be established between the context and the candidate response, which is more helpful for reasoning. The experimental results show that our model's reasoning ability has been dramatically improved with the help of graph structure and commonsense knowledge.

2   Our Approach

2.1   Problem Definition

Given a dataset $D = \{(U, C)\}$, where U represents dialogue context $U = \{u_1, u_2, \ldots, u_n\}$. And $C = \{c_1, c_2, c_3, c_4\}$, and c_i is a response candidate. The model is excepted to learn a function $f(U, c_i)$, which can evaluate the relevance between U and c_i.

2.2   Reasoning Graph

We construct the context and candidate responses into a graph. Specifically, the Grammar Parsing tool is used to analyze the context information and the candidate response. By merging the co-occurring word nodes in the context and the response, an effective connection is established on the graph. In addition, the ConceptNet knowledge base is used to find nodes related to the original node. ConceptNet is an extensive knowledge base of commonsense, containing a large number of nodes and relationships. We mark each context by part of speech, select verbs and nouns as key nodes, query the nodes around the nodes in ConceptNet, and add them to the graph. In particular, we delete the nodes whose credibility weight is less than 1 to get a complete analysis graph. With the help of ConceptNet, graph nodes have obtained richer commonsense information, which further helps improve reasoning ability.

2.3   Model Overview

The overall model is exhibited in Figure 1. Our model includes a semantic representation module, and a graph structure representation module. The semantic representation module uses the Xlnet [3], which is a pretrained model, and the graph structure representation module uses a Graph Convolutional Neural Network [4]. GCN can integrate and learn the feature information of nodes and connected nodes, making the feature representation of nodes more abundant. It is used to integrate expanded commonsense knowledge into our core nodes. Specifically, the feature representation of the node is obtained from Xl-
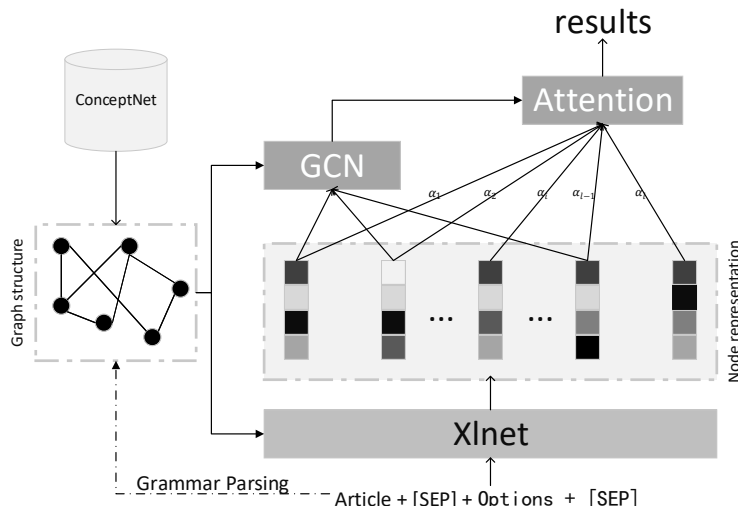
Fig. 1: The overall framework of our model.

net's context representation. The pre-training language model will fully consider the information of the token context for the representation of each token, so the representation is more accurate.

$$h_k = \sigma \left(W \sum_{n_i \in g_k} \frac{1}{|g_k|} h_{n_i} \right) \tag{1}$$

Where $g_k = \{n_0, \cdots, n_t\}$ represents the connected nodes of the k-th node, $|g_k|$ is the total number of the connected nodes, h_{n_i} is the representation of the token n_i, $W \in R^{d \times k}$ is the weight matrix.

To make the characteristics between nodes more obvious, the information of other nodes near the node is aggregated. Where l is the the layer of GCN. The z_i^l represents the aggregated results. N_i represents all connected nodes of the i-th node, and h_j^l is the j-th node representation. So far, we have obtained the neighbor information and updated node representation h_i^{l+1}.

$$z_i^l = \sum_{j \in N_i} \frac{1}{|N_i|} h_j^l \tag{2}$$

$$h_i^{l+1} = \sigma \left(W^l h_i^l + z_i^l \right) \tag{3}$$

In addition, we have added an attention mechanism to learn the importance of nodes. With the help of attention, the ability of model reasoning can be effectively improved.

$$\alpha_i = \frac{h^c \sigma \left(W_1 h_i^l \right)}{\sum_{j \in N} h^c \sigma \left(W_1 h_j^l \right)} \tag{4}$$

$$h^g = \sum_{i \in N} \alpha_i^l h_i^l \tag{5}$$

where h^c is the representation of the context, h^g is the graph representation.

3   Experiments

We test our proposed model on MuTual [5]. MuTual is constructed based on listening test data. We use these evaluation indicators R@1, R@2 and MRR. R@1 and R@2 are the recall at position 1 and 2 in 4 candidates, and MRR is the Mean Reciprocal Rank.

Our experimental comparison mainly includes the original Xlnet model, only commonsense knowledge node information without graph structure, and our model. The experimental results are as follows. We can see that when there is only commonsense knowledge without graph structure information, the model's performance is deteriorating compared with the original Xlnet model. Because the commonsense knowledge is isolated information without mutual connection, it is equivalent to introducing more noise. These introduced noises will increase the difficulty of model feature extraction, which leads to the deterioration of the model effect. When the graph structure is added, the isolated points are effectively combined, and the reasoning ability of the model is improved by learning the relationship between the nodes.
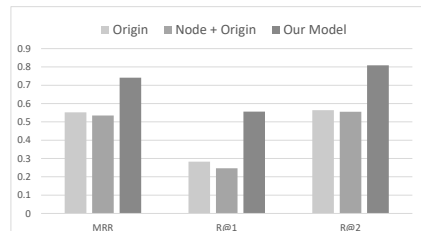
Model	MRR	R@1	R@2
Origin(Xlnet)	0.552	0.283	0.564
Node + Origin	0.535	0.247	0.555
Our Model	**0.741**	**0.556**	**0.809**

Fig. 2: The results of different models.

4   Conclusion and Future work

In this paper, we propose a new approach based on commonsense knowledge combined with graph reasoning to solve dialogue reasoning problems. The text is constructed into graphs through grammatical analysis, combined with the expansion of commonsense knowledge to make the relevance more obvious. In this way, the relationship between the context and the candidate response is more obvious, and the superiority of the graph in reasoning is fully utilized to enhance the reasoning ability of the model. In future work, we will try more variations model to test our approach.

5   Acknowledgments

References

1. Jawahar, G., Sagot, B., Seddah, D.: What does BERT learn about the structure of language? Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Volume 1: Long Papers. pp. 3651–3657. ACL, Italy, (2019)
2. Zhou, X., Li, L., Dong, D., Liu, Y., Chen, Y., Zhao, W.X., Yu, D., Wu, H.: Multi-turn response selection for chatbots with deep attention matching network. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers. pp. 1118–1127. ACL, Melbourne, Australia (2018)
3. Yang, Z., Dai, Z., Yang, Y., Carbonell, J.G., Salakhutdinov, R., Le, Q.V.: Xlnet: Generalized autoregressive pretraining for language understanding. Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, pp. 5754–5764, NeurIPS, Vancouver, BC, Canada (2019)
4. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, pp.3837–3845, Barcelona, Spain (2016)
5. Cui, L., Wu, Y., Liu, S., Zhang, Y., Zhou, M.: Mutual: A dataset for multi-turn dialogue reasoning. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 1406–1416. ACL, Online (2020)