Improving Dialogue Intent Classification with a Knowledge-Enhanced Multifactor Graph Model (Student Abstract)*

Huinan Xu[1], Jinhui Pang[1†], Shuangyong Song[2], Bo Zou[3]

[1] Beijing Institute of Technology, Beijing 100081, China
[2] Department of Big Data and AI, China Telecom, Beijing, China
[3] JD AI Research, Beijing 100176, China
{xuhuinan, pangjinhui}@bit.edu.cn, songshy@chinatelecom.cn, cdzoubo@jd.com

Abstract

Although current Graph Neural Network (GNN) based models achieved good performances in Dialogue Intent Classification (DIC), they leave the inherent domain-specific knowledge out of consideration, leading to the lack of ability of acquiring fine-grained semantic information. In this paper, we propose a Knowledge-Enhanced Multifactor Graph (KEMG) Model for DIC. We firstly present a knowledge-aware utterance encoder with the help of a domain-specific knowledge graph, fusing token-level and entity-level semantic information, then design a heterogeneous dialogue graph encoder by explicitly modeling several factors that matter to contextual modeling of dialogues. Experiment results show that our proposed method outperforms other GNN-based methods on a dataset collected from a real-world online customer service dialogue system on the e-commerce website, JD.

Introduction

DIC is an important task in dialogue understanding. Previous methods made use of RNN to generate contextual representation of dialogues. In view of the fact that RNN-based methods are unable to explore the abundant semantic relationships between non-continuous utterances, some researchers proposed GNN-based models to adopt speaker-specific and contextual modeling for dialogues (Ghosal et al. 2019). These works made use of edges to inject speaker dependency and semantic relations among utterances, whereas lots of fine-grained semantic information that can be extracted from domain knowledge have not been used yet.

In this paper, we propose a new DIC approach based on a multifactor dialogue graph enriched with domain-specific knowledge. Firstly, we propose a *knowledge-aware utterance encoder* to fuse the token-level and entity-level semantic information of utterances, where the entities are from a well designed Knowledge Graph (KG) of a specific domain. Then we present a *heterogeneous multifactor dialogue GNN* by explicitly modeling the interactions among utterances, speakers and local discourses. We conduct experiments on a dataset from JD[1] and achieve the best performance.

Approach

Knowledge-Aware Utterance Encoder

We combine the token-level and entity-level representations to get the knowledge-aware initial representations of utterances, where the entities come from a well-designed knowledge graph of a specific domain.

Let $U = [t_1, t_2, ..., t_n]$ denote the raw sequence of an utterance in dialogue D, and $W = [w_1, w_2, ..., w_n]$ denotes the token vectors of U. And the entity-level vector of U is $E = [g_1, g_2, ..., g_n]$, where $g_i = KGEmb(e_j)$ if t_i is in the span of entity e_j, else $g_i = 0$. Here $KGEmb$ denotes the KG embedding transformation, which can be freely chosen.

Then we align and stack the token-level and entity-level embedding matrices as $M = [[w_1F(g_1)], [w_2F(g_2)], ..., [w_nF(g_n)]]$, where F is either a linear or non-linear mapping function to eliminate the divergence between token and entity space. Then M is fed into a sequential encoding module consisting of CNN+LSTM to compute knowledge-aware utterance representations.

Heterogeneous Multifactor Dialogue GNN

We construct a multifactor graph for a dialogue by explicitly modeling several factors that matter to contextual modeling of dialogues. Each utterance is treated as a utterance node in the graph, and we define speaker nodes to represent the contextual information of speakers. Moreover, we add local discourse nodes to aggregate the consecutive utterances spoken by same speakers. Utterance nodes, local discourse nodes and speaker nodes are denoted by U_i, L_j, S_k in Figure 1, respectively.

The initial representations of utterance nodes are the outputs of *knowledge-aware utterance encoder*. In addition, speaker nodes and local discourse nodes are initialized randomly.

The built multifactor dialogue graph has five different types of edges: speaker edges, local edges, local-speaker edges, utterance-order edges and local-order edges to model the temporal and hierarchical dependency relations among three types of nodes. An example is shown in Figure 1.

*Supported by the National Key R&D Program of China under Grant No. 2020AAA0108600.

†corresponding author

[1]https://www.jd.com/

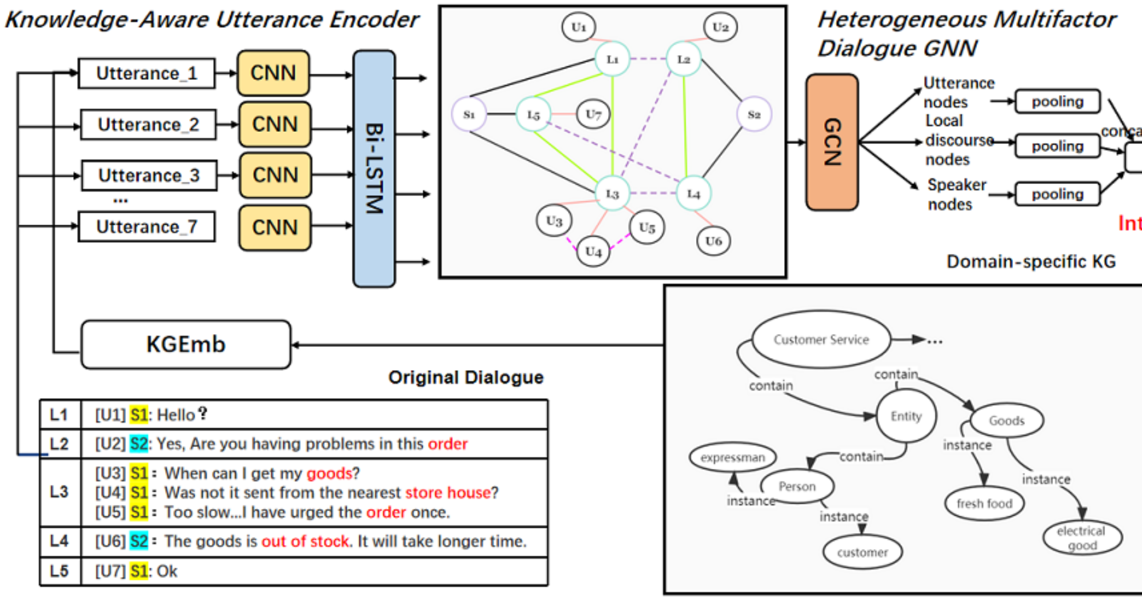

Figure 1: The framework of Knowledge-Enhanced Multifactor Graph Model.

Then we apply R-GCN to propagate contextual information among multi factors. The graph convolutional operation for node n_v at the $l+1$ layer can be defined as:

$$h_v^{(l+1)} = RELU\left(\sum_{r\in\mathcal{R}}\sum_{a\in N_r(v)} W_r^{(l)} h_a^{(l)} + b_r^{(l)}\right) \quad (1)$$

Finally, we concatenate the pooling results of output features of three types of nodes separately at each GCN layer as graph features for DIC.

Experiments

We evaluate our KEMG model on a dataset supplied by the customer service department of JD. Each dialogue consists of several utterances with speaker annotations. The dialogues are annotated with one of 50 intent labels. The dataset has 20,000 samples of dialogues, with a total of 437,060 utterances. We use 18,000 dialogues for training, 1,000 for validation, and the remaining for testing. Our experimental comparison mainly includes several GNN-based DIC models, DialogueGCN (Ghosal et al. 2019), RAT (Ishiwatari et al. 2020) and DAG (Shen et al. 2021).We adopt several evaluation indicators, which are Accuracy(Acc), Top-3 accuracy(Top-3), Top-5 accuracy(Top-5), Macro-F1(M-f1) and Weighted-F1(W-f1), to evaluate the performance of KEMG.

The comparison results are shown in Table 1. KEMG outperforms other GNN-based models, demonstrating that modeling the contextual information of the dialogue with multi factors and domain-specific knowledge contributes to a comprehensive understanding of the dialogue. In addition, we also test KEMG without *knowledge-aware utterance encoder* to verify the effect of domain-specific KG. The results show that KG promotes further improvement in the performance.

Model	Acc	Top-3	Top-5	M-F1	W-F1
DialogueGCN	61.30	83.90	90.80	52.48	58.70
RGAT	63.50	89.40	93.90	59.02	63.53
DAG	63.20	86.40	93.20	58.52	61.69
KEMG-w/o KG	66.50	89.30	94.40	60.64	65.30
KEMG	**67.70**	**90.40**	**95.20**	**61.06**	**66.48**

Table 1: Comparison with baseline methods on JD dataset.

Conclusion

In this paper, we propose KEMG to model the dialogue with a multifactor dialogue graph involved with domain-specific knowledge for dialogue intent classification. We believe KEMG can obtain a comprehensive understanding of the dialogue. Experimental results demonstrate KEMG outperforms current GNN-based methods. Future work will include extending KEMG to multi-label classification tasks.

References

Ghosal, D.; Majumder, N.; Poria, S.; Chhaya, N.; and Gelbukh, A. F. 2019. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In *EMNLP-IJCNLP 2019*, 154–164.

Ishiwatari, T.; Yasuda, Y.; Miyazaki, T.; and Goto, J. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proc. of EMNLP,2020*.

Shen, W.; Wu, S.; Yang, Y.; and Quan, X. 2021. Directed Acyclic Graph Network for Conversational Emotion Recognition. In *ACL/IJCNLP, 2021*, 1551–1560.