MDKB-Bot: A Practical Framework for Multi-Domain Task-Oriented Dialogue System (Postprint)

Authors: Lao, Yadi, Liu, Weijie, Gao, Sheng, Li, Si

Date: 2022-11-27T00:00:00+00:00

Abstract

One of the major challenges to build a task-oriented dialogue system is that dialogue state transition frequently happens between multiple domains such as booking hotels or restaurants. Recently, the encoder-decoder model based on the end-to-end neural network has become an attractive approach to meet this challenge. However, it usually requires a sufficiently large amount of training data and it is not flexible to handle dialogue state transition. This paper addresses these problems by proposing a simple but practical framework called Multi-Domain KB-BOT (MDKB-BOT), which leverages both neural networks and rule-based strategy in natural language understanding (NLU) and dialogue management (DM). Experiments on the data set of the Chinese Human-Computer Dialogue Technology Evaluation Campaign show that MDKB-BOT achieves competitive performance on several evaluation metrics, including task completion rate and user satisfaction.

Full Text

Preamble

DATA PAPER

MDKB-Bot: A Practical Framework for Multi-Domain Task-Oriented Dialogue System

Yadi Lao, Weijie Liu, Sheng Gao† & Si Li
Pattern Recognition and Intelligence System Laboratory, Beijing University of Posts and Telecommunications, Beijing 100876, China

Keywords: Dialogue system; Knowledge base; Natural language understanding; Slot filling; Natural language generation

Citation: Y. Lao, W. Liu, S. Gao & S. Li. MDKB-Bot: A practical framework for multi-domain task-oriented dialogue system. *Data Intelligence* 1(2019), 176-186. doi: 10.1162/dint_a_{00010}

Received: August 27, 2018; **Revised:** November 30, 2018; **Accepted:** December 6, 2018

Abstract

One of the major challenges in building a task-oriented dialogue system is that dialogue state transitions frequently occur between multiple domains, such as booking hotels or restaurants. Recently, encoder-decoder models based on end-to-end neural networks have become an attractive approach to address this challenge. However, these models typically require large amounts of training data and lack flexibility in handling dialogue state transitions. This paper proposes a simple yet practical framework called Multi-Domain KB-BOT (MDKB-BOT), which leverages both neural networks and rule-based strategies in natural language understanding (NLU) and dialogue management (DM). Experiments on the dataset from the Chinese Human-Computer Dialogue Technology Evaluation Campaign demonstrate that MDKB-BOT achieves competitive performance across several evaluation metrics, including task completion rate and user satisfaction.

1. Introduction

Over the past decade, dialogue systems have emerged as an attractive research topic and can be classified into open-domain and task-oriented systems. A general approach to dialogue system design treats it as a retrieval problem by learning relevance matching scores between user queries and system responses. Inspired by recent advances in deep learning, building an end-to-end dialogue system has become a popular approach due to its flexibility and extensibility. For example, encoder-decoder models based on recurrent neural networks (RNNs) directly maximize the likelihood of desired responses when previous dialogue history is available. However, two major drawbacks of these systems are the requirement for multiple training corpora and their tendency to generate generic responses such as "I do not know," which limit generalization ability—especially for task-oriented systems that require knowledge from multiple domains to understand users' underlying intents.

Compared to the end-to-end approach, designing a task-oriented dialogue system as a modularized pipeline is more feasible, where each essential component is trained individually: (1) Natural Language Understanding (NLU) to specify task domain and user intent and extract slot-value pairs, (2) Dialogue Manager (DM) to track dialogue state and guide users toward desired goals, and (3) Natural Language Generation (NLG) to produce responses. A key challenge for task-oriented dialogue systems is that dialogue state transitions frequently occur between multiple domains. If early components make mistakes in slot value extraction, errors accumulate and severely impair the entire system' s functionality.

To address this complex dialogue state transition problem, we adopt a modularized pipeline architecture and propose Multi-Domain KB-BOT (MDKB-BOT), which leverages both rule extraction and neural networks. Evaluation experiments on the Chinese Human-Computer Dialogue Technology Evaluation Campaign dataset show that MDKB-BOT can robustly handle frequent changes in user intent across three domains (flight, train, and hotel) and achieves competitive scores based on human evaluation metrics.

2. Related Work

As mentioned previously, significant research efforts have applied deep learning to task-oriented dialogue systems. One of the most effective approaches builds a modularized pipeline system by connecting NLU, DM, and NLG components. Traditional NLU approaches model domain classification and intent detection as sentence classification tasks while treating slot-value pair extraction as a sequence labeling problem. A desirable NLU system should be robust to intent errors and slot errors, particularly for slot filling.

For example, Xu and Sarikaya [1] applied an RNN for contextual domain classification and used a triangular conditional random field (CRF) based on a convolutional neural network for intent detection and slot filling. Jaech, Heck, and Ostendorf [2] applied multi-task learning to leverage knowledge from source domains with abundant data to improve performance on target domains with limited data. Bapna et al. [3] explored the role of context information in NLU by injecting previous dialogues into an RNN-based encoder and a memory network.

Meanwhile, many attempts have been made to improve DM architecture. Recent research indicates that reinforcement learning (RL) holds promise for planning dialogue policies based on current dialogue state. Williams, Asadi, and Zweig [4] proposed Hybrid Code Networks (HCNs), which combine supervised learning and reinforcement learning. HCNs select dialogue actions at each step by optimizing task completion rewards using policy gradient [5]. Faced with the sparse nature of reward signals in RL, Peng et al. [6] designed an end-to-end hierarchical RL framework where a MANAGER selects current goals (such as specific domain tasks) and a WORKER takes actions to help users complete subtasks. Inspired by recent RL advances, Mrkšić et al. [7] introduced a belief tracker that overcomes the drawback of requiring large hand-crafted lexicons to capture linguistic variation in user language. Their Neural Belief Tracking (NBT) models can reason over pre-trained word embeddings of system output, user utterance, and candidate database pairs.

For NLG, most current work applies information retrieval techniques to large query-response databases or uses template-based methods with rule sets to map frames to natural language. Dusek and Jurcicek [8] encoded frames based on syntax trees and used seq2seq models for generation.

3. Proposed Framework

The proposed framework is illustrated in Figure 1 [Figure 1: see original paper], comprising NLU, DM, and NLG components. Implementation details for these components are described in Sections 3.1 through 3.3.

Figure 1. The overall framework consists of three components: (1) NLU module, which predicts intent domain and extracts slot values from user utterances, (2) DM module, which outputs dialogue actions to the NLG module, and (3) NLG module, which generates the final response.

3.1 Natural Language Understanding (NLU)

The main NLU tasks involve domain classification, intent detection, and slot filling, as illustrated in Figure 1.

3.1.1 Domain Classification We adopt the convolutional neural network proposed by Kim [9] for domain classification. Let $W \in \mathbb{R}^{v \times d}$ be the d-dimensional word embedding table, where v is vocabulary size. Sentence semantic representation of user query $X \in \mathbb{R}^{n \times d}$ is obtained by looking up each word in W, where n is the number of words in the query. A 1-D convolutional layer then extracts n-gram features from X. However, convolutional neural networks (CNNs) may produce errors in cases containing descriptions from multiple domains, such as: "The train is cheaper, but to save time, give me airline flight schedules and flight timetables." Therefore, our online model employs rule-based strategies to address misclassification by constructing keyword lists from both corpora and databases (e.g., city name lists).

3.1.2 Slot Filling Slot filling is treated as a named entity recognition task using the popular begin-in-out (BIO) format to represent tags for each word in a query. Long Short-Term Memory (LSTM) scans the words and outputs representations: $\tanh(Wx + Uh + \tanh(c))$. To enhance slot-value pair extraction capability, a CRF network connects to the LSTM or bidirectional LSTM (BLSTM) output. The score of a sentence X along a tag path Y is calculated as the sum of transition score A and LSTM network score f:

$$S(X,Y) = \sum_i (A_{y_{i-1},y_i} + f_i(y_i, X))$$

where h represents the trainable parameters of the LSTM network.

Figure 2 [Figure 2: see original paper] shows a bidirectional LSTM network enhanced with a CRF layer on top. For our online model, we apply both keyword matching and BLSTM-CRF to handle diverse or non-standard expressions.

Figure 2. BLSTM-CRF model for slot filling.

3.1.3 Intent Detection Based on slots extracted from BLSTM-CRF, we update the maintained dialogue state template. User intent is then inferred by comparing the predefined dialogue template with the new state template.

3.2 Dialogue Management (DM)

After the NLU module, we obtain output including the user intent domain and slot values for the current turn.

To avoid unnecessary dialogue turns over insignificant information, we divide slots into two categories: required slots and extra slots. Required slots such as `<departCity>` are necessary for task completion, while extra slots like `<trainValue>` and `<countRate>` may make dialogue tedious for users who don' t require them. We therefore only complete required slots for booking, but consider extra slots if users mention them during information retrieval.

To regulate dialogue flow during system-user interaction, the DM module updates conversation state and determines the next dialogue action. We divide this module into three states with detailed procedures as follows.

Initial State: At conversation beginning, utterances without explicit intention are considered casual talk. After identifying user intent, the system transitions to slot filling state. Note that the system stores slot information even before domain prediction, distributing it to corresponding slots afterward.

Slot Filling State: The main task at this state is to interactively obtain required slot information from users for response generation.

Recommendation State: Our bot lists results matching user demands through database retrieval and extraction. In case of failure, we implement strategies for similar recommendations: (1) remove extra slot limitations, (2) adjust departure time appropriately, (3) change cabin or train type, and (4) expand price range. Additionally, users can modify requests and return to slot filling or recommendation states.

Throughout the dialogue, when the system determines an intent cannot be completed, it promptly notifies users to avoid wasting time. For example, when a user wants to book a flight to a city without air service, continuing dialogue is unwise, so the system recommends alternative travel means. Users can also change intent anytime, and our system automatically stores common slot information during transitions.

3.3 Natural Language Generation (NLG)

At this point, we have obtained both the user query' s intention category and the next dialogue action for each turn, which guide the NLG module to generate natural language responses. Given the user' s slot list, we convert it into an SQL statement to retrieve information from the database storing train, flight, and accommodation data, checking for eligible items matching the user' s goal and selecting appropriate response templates.

Due to the shortage of large-scale dialogue corpora in these domains, we employ template-based NLG, requiring us to capture every possible slot state combination to presuppose dialogue templates. Once user dialogue actions match predefined sentence templates, we replace slot values with user history information. This approach ensures response controllability.

4. Experiments

In the Chinese Human-Computer Dialogue Technology Evaluation Campaign, we developed a task-oriented dialogue system to help users book flights, trains, and hotels.

4.1 Data Sets

Since only three databases were provided, we extended the Task 1 dataset for domain classification and rule extraction. We also annotated a 300-dialogue corpus (approximately 1,500 training sentences) with slot labels for evaluating LSTM-CRF and BLSTM-CRF models. Table 1 and Table 2 show dataset details.

Table 1. Statistics of training corpus for domain classification.

Table 2. Statistics of training corpus for slot filling.

4.2 Evaluation

For slot filling in NLU, we adopt the entity-level prediction F1 score commonly used in named entity recognition.

However, dialogue evaluation remains challenging. We use evaluation metrics from the Chinese Human-Computer Dialogue Technology Evaluation Campaign, including task completion rate, user satisfaction score, dialogue naturalness, number of turns, and robustness for uncovered cases.

5. Discussion

Table 3 shows slot filling results. We compare LSTM-CRF and BLSTM-CRF performance using unigram and unigram plus bigram features. As illustrated, accuracy improves by 1.7% when considering word sequence order with BLSTM. Using bigram features helps LSTM-CRF, though it slightly harms BLSTM performance due to short average bigram sequence length. A possible improvement is using character embeddings instead of word embeddings.

Table 3. Comparison of labeling performance on NLU.

Table 4 illustrates our system' s performance according to the aforementioned evaluation metrics. Most metrics are manually annotated except for average dialogue turns. Our system achieved the best scores in user satisfaction, dialogue naturalness, and boot ability due to our reasonable predefined dialogue intent transition templates. However, this led to decreased task performance, particularly when user intent was not identified or important slot values were incorrectly extracted.

Table 4. Dialogue quality evaluation results for top 4 teams in the competition.

6. Conclusion

This paper proposed a simple yet practical framework for multi-domain task-oriented dialogue systems. Our model leverages both neural networks and rule-based strategies to handle domain transition problems, achieving competitive results in the Chinese Human-Computer Dialogue Technology Evaluation Campaign, particularly for user-friendliness and utterance guidance metrics. Future work will apply end-to-end neural networks to NLG based on information extracted and maintained in NLU and DM to improve system performance.

Author Contributions

Y. Lao (laoyadi@bupt.edu.cn) and W. Liu (liuweijie@bupt.edu.cn) led the MDKB-Bot system development and designed the overall framework. S. Gao (gaosheng@bupt.edu.cn) and S. Li (lisi@bupt.edu.cn) summarized applications and drafted the paper. All authors participated in manuscript revision.

Acknowledgements

This work was supported by the Beijing Natural Science Foundation (No. 4174098), National Natural Science Foundation of China (No. 61702047), and the Fundamental Research Funds for the Central Universities (No. 2017RC02).

References

[1] P. Xu, & R. Sarikaya. Contextual domain classification in spoken language understanding systems using recurrent neural network. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 136-140. doi: 10.1109/ICASSP.2014.6853573.

[2] A. Jaech, L. Heck, & M. Ostendorf. Domain adaptation of recurrent neural networks for natural language understanding. arXiv preprint. arXiv:1604.00117, 2016.

[3] A. Bapna, G. Tür, D. Hakkani-Tür, & L. Heck. Sequential dialogue context modeling for spoken language understanding. In: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, 2017, pp. 103–114. doi: 10.18653/v1/W17-5514.

[4] J.D. Williams, K. Asadi, & G. Zweig. Hybrid code networks: Practical and efficient end-to-end dialogue control with supervised and reinforcement learning. arXiv preprint. arXiv: 1702.03274, 2017.

[5] R.J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning* 8(3-4)(1992), 229–256. doi: 10.1007/BF00992696.

[6] B. Peng, X. Li, L. Li, J. Gao, A. Celikyilmaz, S. Lee, & K.-F. Wong. Composite task-completion dialogue policy learning via hierarchical deep reinforcement learning. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2221–2230. doi: 10.18653/v1/D17-1237.

[7] N. Mrkšić, D.O. Séaghdha, T.-H. Wen, B. Thomson, & S. Young. Neural belief tracker: Data-driven dialogue state tracking. arXiv preprint. arXiv:1606.03777, 2016.

[8] O. Dusek, & F. Jurcicek. Sequence-to-sequence generation for spoken dialogue via deep syntax trees and strings. arXiv preprint. arXiv:1606.05491, 2016.

[9] Y. Kim. Convolutional neural networks for sentence classification. arXiv preprint. arXiv:1408.5882, 2014.

Author Biography

Yadi Lao is currently a graduate student at the Pattern Recognition and Intelligent System Laboratory, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. She received her Bachelor’ s degree from Beijing University of Posts and Telecommunications in 2016. Her research interests include natural language processing and reinforcement learning. She is the system designer of MDKB-Bot.

Weijie Liu is currently a graduate student at the Pattern Recognition and Intelligent System Laboratory, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. He received his Bachelor’ s degree from Central South University of Forestry and Technology in 2016. His research interests include natural language processing and machine reading comprehension. He is the system designer of MDKB-Bot.

Sheng Gao is an Associate Professor at the Pattern Recognition and Intelligent System Laboratory, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. He received his PhD degree from the Sixth University of Paris, France in 2011. He is a member of the Big Data Professional Committee of the Chinese Computer Society (CCF) and an expert of the National Natural Science Foundation of China. He was selected for the “Young Talent” Program of Beijing and the “Star-casting Program” of Microsoft Asia Research Institute. He has published more than 30 papers in internationally renowned journals and conferences such as *IEEE Transactions on*

Emerging Topics in Computing, *Knowledge-Based Systems*, International Conference on Data Mining and Knowledge Discovery (DMKD), AAAI Conference on Artificial Intelligence, International World Wide Web Conference (WWW), Industrial Conference on Data Mining (ICDM), European Conference on Machine Learning (ECML), and European Conference on Information Retrieval (ECIR). His research directions include machine learning, natural language processing, information recommendation, and social network analysis.

Si Li is a Lecturer at the Pattern Recognition and Intelligent System Laboratory, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. She received her PhD degree in Signal Information Processing from Beijing University of Posts and Telecommunications in 2012. Her research interests are natural language processing and information extraction. She has published many papers in top-level conferences including Annual Meeting of the Association for Computational Linguistics (ACL), International Conference on Research on Development in Information Retrieval (SIGIR), and Conference on Empirical Methods in Natural Language Processing (EMNLP).

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.