2017 10th International Symposium on Computational Intelligence and Design

Semantic Reasoning of Question Answering over Heroes of the Marshes Based on Concept Knowledge Tree

Ziqi Lin[1,2], Wancheng Ni[1], Haidong Zhang[1], Meijing Zhao[1], Yu Liu[1]
1. Integrated Information System Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China
2. University of Chinese Academy of Science, Beijing, China
e-mail: {linziqi2013, wancheng.ni, haidong.zhang, meijing.zhao, yu.liu}@ia.ac.cn

***Abstract*—To make computers obtain a deeper thinking, we propose a question semantic representation model based on concept knowledge tree, which integrates concepts and knowledge to express questions' semantics. Combined the question semantic representation model and related knowledge, we introduce a knowledge-driven semantic reasoning algorithm to recursively obtain new knowledge for deepening the comprehending of questions. We evaluate the performance of our system on a collected dataset, and the top-1 accuracy reaches to 62.1%. Experimental results demonstrate our proposed methods improve the depth of question comprehending.**

Keywords-knowledge-based question answering; semantic reasoning; concept knowledge tree;

I. INTRODUCTION

Question answering systems aim to automatically answer questions proposed in natural language, which can help people to find corresponding information more efficiently [1]. Nowadays, many semantic parsing models, information retrieval methods and neural network are introduced to analyze questions and answers, which achieves great success. And many open-domain knowledge bases have been built, which contain many entities and relations. Compared to various questions, these knowledge is still incomplete. Knowledge is never finite for both people and computers. The biggest difference between computers and human is that there are no intelligence. In knowledge-based question answering (KB-QA) systems, how to obtain answers with incomplete knowledge is a challenging problem. If we make computers have reasoning ability, it would be possible to make computers comprehend knowledge. Most researchers focus on designing knowledge and question representation models to analyze natural language questions and extract corresponding answers, neglecting the answers reasoning. Moreover, the potential relationships between searching process and extracting answers also benefit the reasoning, increasing the comprehending ability for complex questions.

Considering these problems, we integrate concept knowledge tree (CKT) that contains general concepts, domain knowledge, and semantic relationships into question answering systems to increase questions' semantic connections and enlarge questions' semantic representations.

The main contributions of this paper are two parts. First, we introduce a Chinese question semantic representation model based on the CKT to enhance the ability for questions' semantic representations. Second, a semantic reasoning algorithm is proposed to comprehend questions with related concepts or knowledge. It recursively evolves semantic representation for questions and generates new knowledge from concepts base and domain knowledge base, which gradually approaches the real answers.

II. AN OVERVIEW OF OUR SYSTEM

We transform the CKT into question answering systems and introduce a KB-QA system to further mine questions' semantics. It mainly includes: preprocessing, question representation generation, concepts and knowledge nodes mapping, semantic reasoning, and answers merging and filtering, as shown in Figure 1.

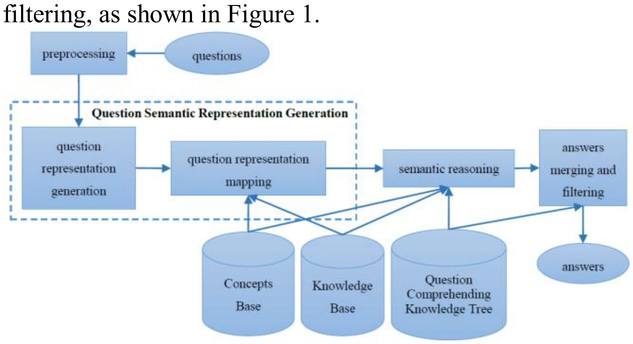

Figure 1. Architecture of our KB-QA system.

A. *Question Preprocessing*

Each question can be divided into main clause and one or more dependent clauses. The main clause refers to the sentence containing interrogatives, while dependent clauses refer to adverbial sentences, attribute sentences or others. To obtain question components, we perform word segmentation, part of speech tagging and syntax analysis with Language Technology Platform [2].

B. *Question Semantic Representation Generation*

It consists of question representation generation that uses Chinese basic phrases and defined transformational rules to develop its structured expressions, and question representation mapping that connects structured representations with concepts or knowledge nodes in the bases. Formalized representations can obtain the relations

2473-3547/17 $31.00 © 2017 IEEE
DOI 10.1109/ISCID.2017.36

among questions' elements and extract the core of questions, and the mapping process embeds more semantics from the bases, enhancing machines to understand questions.

C. Question Semantic Reasoning

We propose a knowledge-driven semantic reasoning algorithm that integrates concepts base, knowledge base and question comprehending knowledge tree to further understand question semantics and reason its real intentions. The question comprehending knowledge tree (QCKT) uses to bridge the gap between questions and domain knowledge, which leverages the question representations, concepts relations and the connection between interrogative pronouns and knowledge to build this knowledge tree.

D. Answer Merging and Filtering

To avoid overlap and highlight the main idea of answers, we leverage some defined rules to merge answers generated from the main clause and dependent clauses of questions. Moreover, we filter out incorrect answers by obtaining the knowledge of interrogative pronouns and knowledge types.

III. CONCEPT KNOWLEDGE TREE

Concept knowledge tree [3-6] as a knowledge representation model contains concepts, knowledge and semantic relations, which can encode human knowledge into computer systems.

CKT defines formal semantic representations to describe language concepts. It divides concepts into independent concepts and compound concepts. The independent concepts are basic semantic representation units and can construct the compound concepts. The CKT defines three types formal representations to express compound concepts, including semantic constraint, semantic logic, and semantic state.

The semantic constraint is mainly used to denote compound concepts which contain concepts modified and attributes constraint relationships, and is represented a two-tuples (1):

$$\begin{aligned} &Semantic\ Constraint = \\ &< Constraint\ Concept : Core\ Concept > \end{aligned} \tag{1}$$

where core concept is the essence of the semantic constraint; constraint concept uses to modify the core concept.

The semantic logic describes the logic combination relation, which conjunctions connect multiple independent concepts or compound concepts. It includes two components: logical relations and concepts list. And the logical relations have five types: "and", "or", "not", "entailment", and "list"; "entailment" denotes the progressive relationship and causality of concepts; "list" indicates a simple arrangement of several concepts. The semantic logic is expressed as a two-tuples (2):

$$Semantic\ Logic = \left(Logic Type, Concepts\ List \right) \tag{2}$$

The semantic state is used to describe the compound structure of events, which is made up of predicate concepts, subject concepts, object concepts and state concepts, as shown in (3):

$$\begin{aligned} Semantic\ State = [< Subject\ Concepts > \{State\ Concepts\} \\ Predicate Concepts < Object\ Concepts >] \end{aligned} \tag{3}$$

Notably, the semantic constraint, the semantic logic, and the semantic state can nest with each other, for example, "big red flowers" is expressed as "<big:<red:flowers>>".

IV. QUESTION SEMANTIC REPRESENTATION MODEL

To further understand question semantics, we propose a question semantic representation model that integrates concepts, knowledge and semantic relationships coming from the CKT. The related concepts and knowledge defined in CKT can extend question semantics. The mapping process from the relations of question components to the relations in CKT benefits semantic reasoning, enhancing machines to recognize the real intentions.

We define the basic semantic representation unit for the model, which is semantic nodes. The semantic node includes six components: question representation, part of speech, matching knowledge nodes, matching concepts, matching attributes, matching relations, as shown in (4):

$$\begin{aligned} &Question = \{question\ representation, \\ &\ part\ of\ speech, matching\ knowledge\ nodes, \\ &\ matching\ concepts, matching\ attributes, \\ &matching\ relations\} \end{aligned} \tag{4}$$

where question representation is a formal representation of the question; matching knowledge nodes, matching concepts, matching attributes, and matching relations are the matching knowledge and concepts of question's elements. And the matching knowledge and concepts extend question semantics.

Semantic nodes make up phrases, phrases and semantic nodes are nested with each other to create questions. So questions can be formally expressed as semantic constraints, semantic logics or semantic states.

For example, "《水浒传》中燕青的外号是?" (means "What is the nickname of Yan Qing in Heroes of the Marshes?"), the question representation is expressed as (5):

[<<<<水浒传:中>:燕青>:外号>>是]   (5)

where the question representation is a semantic state, and the subject of question is a nested semantic constraint.

TABLE I. THE COMPONENT OF QUESTION SEMANTIC REPRESENTATION.

Component	The Content of Component
Part of speech	[<<<<nz:nd>:nh>:n>>v]
Matching Concepts	[<<<<:中_9>: >:外号_1 >>是_12]
Matching Nodes	[<<<<[]:[]>:[燕青]>:[]>>[]]
Matching Attributes	(水浒事件,燕青秋林渡射雁,水浒人物,***燕青***),(水浒事件,燕青劈牌定对,水浒人物,***燕青***)
Matching Relations	(卢俊义,主仆,***燕青***),(***燕青***,救助,卢俊义),(***燕青***,主仆,李固),(***燕青***,结拜,李师师),(***燕青***,朋友,许贯忠),(***燕青***,交战,任原)

Besides, the question semantic representation has several components, as shown in Table 1. The bold italics represent the matching parts of words with a domain knowledge base.

A word is likely to represent different concepts, for example, “棒子” can represent two different concepts, that are “棍子” (stick) and “玉米” (maize); so the symbol “_number” is used to distinguish the different concepts of the same word. Figure 2 gives the partial attributes of the knowledge node “燕青”.

TreeName	NodeName	AttributeName	AttributeValue
水浒人物	燕青	排名	第36位
水浒人物	燕青	民族	汉
水浒人物	燕青	朝代	宋朝
水浒人物	燕青	性别	男
水浒人物	燕青	籍贯	土居（今北京）
水浒人物	燕青	绰号	浪子
水浒人物	燕青	星号	天巧星
水浒人物	燕青	上山前身份或职业	卢俊义奴仆

Figure 2. The attributes of node “燕青”.

V. QUESTION SEMANTIC REASONING

To further deduce question semantics, we combine knowledge search and answer extraction to construct a knowledge-driven reasoning algorithm. It recursively parses question components to generate new question semantic representation containing related concepts or knowledge based on question comprehending knowledge tree, and leverage the structured representation to extract answers from concept knowledge tree. The reasoning process uses concepts and knowledge to connect question representation with CKT, which gradually deepens machines’ understanding. Therefore, the semantic reasoning is a dynamic and iteration comprehending process. To mine the critical information of questions, we use question semantic representations to extract the center of question.

The question semantic reasoning makes the semantic reasoning for each clause and the main clause, and it calls the clause semantic reasoning. Compared to clauses, the main clause needs extracting the core part from its object, which is the center of question; we replace the object of the main clause with the center of question to build a new main clause for the semantic reasoning. According to different representations, we obtain the center of question from a certain part of clause representation. For example, we get the center of question from the core element when the object of the main clause is a semantic constraint.

A. The clause semantic reasoning

The clause semantic reasoning uses a recursive method to implement the clause semantic reasoning based on the matching knowledge and rules, which is the foundation of our method. In fact, the question representation is a nested structure, that semantic nodes, semantic constraints, semantic logics, and the semantic states can nest with each other to express a question. Therefore, the clause semantic reasoning starts to make a semantic reasoning from the innermost layer of the clause representation by using a recursive method. With the continuous reasoning, a nested representation gradually evolves into a semantic node. Then we use the semantic node to replace the previous representation and generate new queries to obtain new knowledge. Here, we specify the semantic reasoning of semantic constraints as the next paragraph, and the algorithm also adapts to other representations.

First, the algorithm makes a recursive call to the clause semantic reasoning if the element of semantic constraint isn’t a semantic node. Then we construct new clauses by the elements’ semantic reasoning results and generate new queries to obtain related knowledge; through the previous semantic reasoning, these elements of semantic constraint may match new nodes, concepts, attributes or relations, so we use these new matching knowledge to replace these elements for building new clauses. Finally, the algorithm makes semantic reasoning for new clauses. If the elements of semantic representation are semantic nodes, the algorithm makes semantic reasoning for the original clause directly.

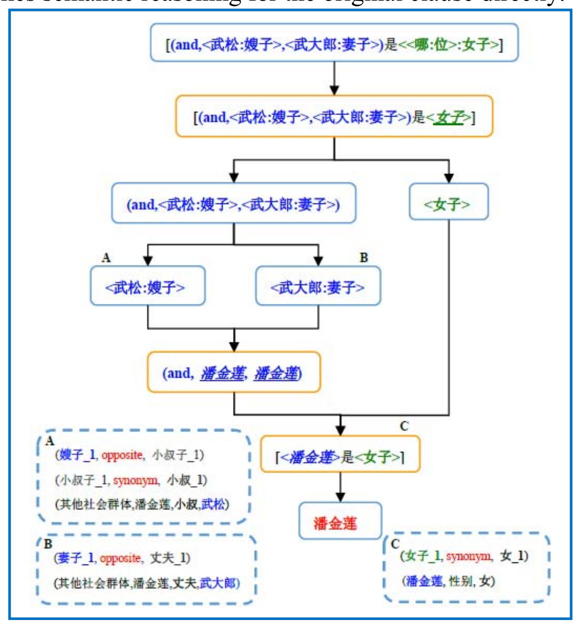

Figure 3. An Instance of Question Answering.

Figure 3 specifies the semantic reasoning process of the question “《水浒传》中，武松的嫂子、武大郎的妻子是哪位女子？” (In Heroes of the Marshes, who is Wu Song’s sister-in-law and Wu Dalang’s wife?). And the dotted boxes give the used concepts and knowledge.

Firstly, the algorithm makes a semantic reasoning for the dependent clauses (“《水浒传》中”). Without matching knowledge, we cannot get any results.

Secondly, through the semantic constraints “<<哪:位>:女子>”, we get the question center is “女子” (female) and create a new main clause; then we make a semantic reasoning for the new main clause.

Thirdly, we use a recursive call to make semantic reasoning for each part of the new main clause; so we make a reasoning for “<武松:嫂子>” (<Wu Song:sister-in-law>), and the algorithm uses the concept “嫂子_1” to obtain the opposite relation concept “小叔子_1” and a synonymous concept “小叔_1”, so we get “潘金莲” as the reasoning result; similarly, we make a reasoning for “<武大郎:妻子>” (<Wu Dalang:wife>), and use the opposite relation concept “丈夫_1” to get “潘金莲” as the reasoning result; the two parts are “and” relations, so the reasoning result is “潘金莲”.

Fourthly, we update the main clause representation with the reasoning result of the subject and generate queries to search the related knowledge about “潘金莲”; then we get a new main clause “[<潘金莲>是<女子>]” and a knowledge node “[潘金莲]”.

Finally, we make a semantic reasoning for the new main clause; depending on the sex attribute of knowledge node and a synonymous concept “女_1”, we get the final answer “潘金莲”.

Compared to other KB-QA systems, we make the semantic reasoning, and it can improve the depth of comprehending and the confidence of answers. Although some KB-QA systems also can get the correct answer “潘金莲”, their systems may think the answer is a person whom “武松” and “武大郎” both have a relation with. When they have two or more mutual acquaintances, the final answers may be incorrect.

VI. Experiments

A. Dataset

Our dataset comes from a game show “一站到底”. We collect question-answer pairs from 2012 to 2013 in a website (http://www.yzddtk.com/). In this paper, we deal with the “SHI” of wh-questions, which the predicate is “SHI”. We extract the “SHI” of wh-questions about Heroes of Marshes from these data, which has 95 questions. The statistics of dataset is shown in Table 2.

TABLE II. Distribution of the data.

Dataset	Number of “SHI” of Wh-questions	Number of Questions	Average Length of Questions	Average Words of Questions
Extracted Dataset	95	95	24.16	17.63
2012 Dataset	6239	9436	25.18	17.53
2013 Dataset	4422	6766	26.53	18.51

We construct a domain knowledge base about Heroes of the Marshes based on the CKT. Besides, we integrate the semantic knowledge of questions and domain knowledge to construct a question comprehending knowledge tree that supports the semantic reasoning process, which contains various lexical level semantic knowledge and reasoning rules.

B. Results and Analysis

We integrate domain knowledge base, concepts base and question comprehending knowledge tree to construct the KB-QA system. We perform the KB-QA system on the Heros of the Marshes dataset and evaluate its performance at top-1 level accuracy.

TABLE III. Experimental results.

Dataset	Method	Accuracy
Heroes of the Marshes Dataset	IR	28.42%
	IR+Reason	54.74%
	IR+Reason+Filter	62.11%
	IR+Revise	31.58%
	IR+Reason+Revise	57.89%
	IR+Reason+Filter+Revise	65.26%

Table 3 shows our method performance, “IR” denotes the information retrieval method. From the first and second methods or fourth and fifth methods, we observe that the semantic reasoning algorithm obviously improve the performance of the system. From the third and sixth methods, we observe that revising the preprocessing results can achieve a better result. Compared with the second and third methods, the filtering answers method apparently improve the performance of the system. Experiment results indicate that the knowledge in QCKT is beneficial for computers to understand user intentions.

VII. Conclusions

In our research, we introduce concept knowledge tree into question answering systems and develop a Chinese question semantic representation model. It uses relative concepts, knowledge and semantic relationships, included in concepts and knowledge bases, to strengthen questions’ semantics. At last, the question semantic reasoning algorithm iteratively develops question semantic representation and generates new knowledge from the bases to deduce question-answer pairs, which gradually obtain the final answers. In the future, we will explore more complex questions and increase much more knowledge of questions and concepts relations to enhance semantic reasoning.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61379099.

References

[1] A. Bouziane, D. Bouchiha, N. Doumi, and M. Malki, "Question answering systems: Survey and trends," Procedia Computer Science, vol.73, 2015, pp. 366-375, doi:10.1016/j.procs.2015.12.005.

[2] W. Che, Z. Li, and T. Liu, "Ltp: A chinese language technology platform," in Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, ACL Press, 2010, pp. 13-16.

[3] Y. Gao, "A concept-based knowledge representation system," Microelectronics & Computer, vol.21, 2004, pp. 71-74, doi:10.3969/j.issn.1000-7180.2004.09.019.

[4] P. Lu, "Research on semantic understanding of modern chinese preposition," Phd. Thesis, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 2008.

[5] H. Yu, Y. Gao, and Y. Yang, "Construction and application of domain knowledge based on knowledge tree," Application Research of Computers, vol.25, 2008, pp. 3246-3248, doi:10.3969/j.issn.1001-3695.2008.11.012.

[6] M. Zhao, "Research on semantic understanding of the meaning of chinese dictionary (MCD)," Phd. Thesis, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 2015.