Research on Chinese Cuisine QA Service System

Ling Xia   Fuji Ren

(The University of Tokushima)

Abstract

This paper presents our research on Chinese cuisine question answering service system. We first established question taxonomy in cooking domain. Then an integrated approach on question classification was presented. To deal with the problem of answer generating, we proposed an answer planning methodology. We collected cooking documents from web and built an off-line knowledge sources. After putting further annotation on domain documents, a two-stage tactic was exploit, namely, answer content planning and answer surface realization. We carried out the experiments based on the proposed methods and the result shows that the system can satisfy user's demand

Key words: question answering (QA) system, domain knowledge, question classification , answer generation

1. Introduction

Automatic question answering system involves the research areas including computer linguistics, information science and artificial intelligence, etc., and has been one of hot topics of the computer applications. Recent research on QA system has mostly concentrated on open domain. But up to date, the domain-independent QA system has not reached the level of real world application, while the restricted-domain question answering (RDQA), which can exploit the domain knowledge to reduce the complexity of language understanding, have been successfully developed.

Diet is not only the primary necessity of the people, it is also good for health and mood. Although people's daily foods increasingly rely on market supplements, the warmly home made dinners are always the best for families, which are un-replaceable by lunch boxes or fast foods, because having dinner at home will make people feeling the happiness of life and encouraging people. For enriching computer aided cooking service, we aim to build a cooking QA service system which serves as an intelligent agent to guide meals (e.g. dish cooking, cooking techniques) in our daily life. Our main challenge is: 1) to establish question taxonomy in cooking domain and find the question classification rules; 2) to put effective annotation on domain document and find the suitable answer generating strategies for different question type. The rest of this paper is organized as follows: Section 2 gives a description of our question taxonomy and question classification. Section 3 presents the annotation scheme of cooking document and proposed the answer planning methodology for answer generating. Section 4 is evaluation and discussion. Finally we give conclusion and future work in section 5.

2. Question Analysis

2.1   Question Taxonomy and Domain Word

The service objects of our cooking QA system are the people who are interested in cooking and want to cook dishes by themselves. They would like to know the cooking methods of some cuisine and some basic cooking techniques. People can use many ways to ask questions for a same thing, for example, "请问麻婆豆腐的做法？(Please tell me the way to cook Mapo bean curd?) " and "麻婆豆腐应该怎样做？(How to cook Mapo bean curd?)", the answer of two questions is all the recipe of "Mapo bean curd ". In this work, we categorized questions according to their expected answer type. After a survey on the frequently asked cooking questions from Internet, we extracted the most common and largest of four types question and set up our own question taxonomy. Table2.1 shows question types and examples.

Table 2.1 Question categories and examples.

question type	examples
raw material	做熊掌豆腐需要哪些原材料? (What's the raw material of the bear's paw-like bean curd?)
cooking recipe	干煸牛肉丝怎么做? (How to cook the Saute of jerked beef with chili?)
dish recommendation	川菜中的经典菜是哪几道? (Please recommend me some classic Sichuan dishes!)
cooking skill	炒菜时怎样勾芡? (How to put in starch when stir-frying dish?)

In RDQA, domain words is an important way to express the domain knowledge. Using the domain words can distinctively describe the subject of the question. In this work, we mainly summarized three kinds of domain words, namely the words used for expressing cuisine name, cooking raw material, and the words used for

describing cooking techniques. Table 2.2 shows some examples of our domain words.

Table 2.2 Types and examples of domain words.

Type	Examples
dish name	麻婆豆腐 (Mopo bean curd)
material	鸭舌、茄子 (duck's tongue, eggplant)
cooking skill	焯水、勾芡 (heat in water, put in starch)

2.2. Question classification

Question classification is an important component of QA system. It has two major functions. One is to reduce the set of candidate answer and thus increase the accuracy of returned answer, the other is to decide the strategy of answer extraction. Questions in a restricted domain may vary from general open-domain factoid and confine questions to very special kinds of questions that depend on the selected domain [1]. For instance, medical questions can be classified based on a hierarchical evidence taxonomy created by physicians [2]. Therefore it is difficult to use a general classification methodology for the development of a RDQA system.

In restricted domain, the service and the question types are relatively fixed, therefore, by the means of the domain knowledge, the question type can be recognized effectively based on rules. We extract classification features by virtue of the domain attributes, then according to matching filtering strategy, we classify the questions using the production rules. When the user's question can not be matched with filtering rules, we use the machine learning method to carry on the secondary classification to strengthen system's robustness [3]. Figure 2.1 shows the flow chart of our integrated classification approach.

where the used abbreviations are as follows :

    FW: focus word

    DN: dish name

    MI: main ingredient

    CF: cooking form

    CS: cooking skill

    TT: taste type

3. Answer Generation Based on Answer Planning Method

General question answering systems adhere to the pipeline architecture which mainly includes three parts:question analysis, information retrieval, and answer
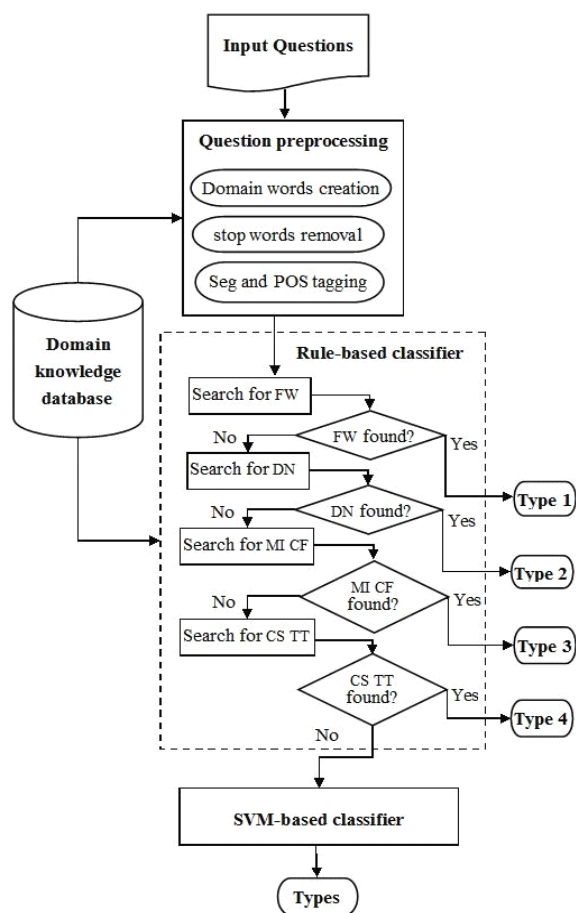

Figure 2.1 Overview of integrated classification

extraction. For example, recent research in restricted domain, such as sightseeing QA system, Analects of Confucius QA system, are worked in this way. They return answers by various similarity calculation and answer ranking [4]. However, how to generate a well-formed answer is a difficult task. In this work, we proposed the answer planning method to deal with the problem of answer generation.

Our QA system is aiming at real application of the cooking domain. We can obtain rich structural cooking documents from web and build an off-line knowledge sources. After annotating the corpus with function blocks, the answer information may extract directly from the knowledge database. We exploit a two-stage strategy for answer generating. i.e answer content planning and answer surface realization. The architecture of planning based answer generation is shown in Figure 3.1.

3.1 Answer content planning

Content planning decide what information must be communicated, namely, answer content selection and ordering. As refered to in section 2.1, we categorized the cooking questions into four types. The answers which correspond with the first three types of questions could

be generate from dish cooking documents, the answer components are function blocks which have been annotated. However, as the difference in domain characteristic, the answers corresponding the fourth type will be generated from the resources of cooking technique documents, and the answer component is an explanation paragraph.
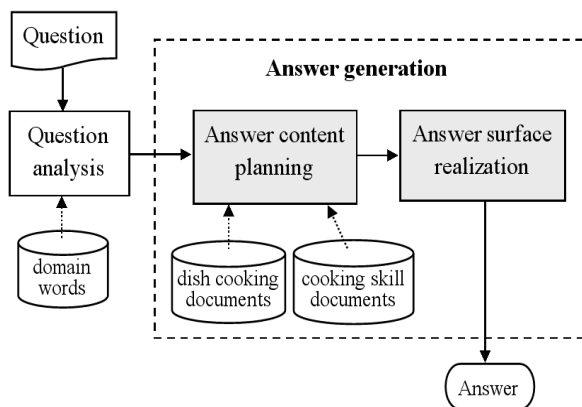

Figure.3.1 The architecture of answer generation

3.2 Answer surface realization

According to content planning, we can decide the answer content component from the related database. By answer surface realization, we can convert content components into actual answer. (e.g. Insert function words, order components, add punctuation, etc.). As our support system aimed at answering frequently asked cooking questions, so we argue that small number of predefined patterns can cover all the superficial variations in answer construction. Therefore, the strategy we proposed to answer surface realization is to use domain-specific language, and summarize the generic templates for answer sentences according to four question categories. Figure 3.2 shows the workflow of the answer surface realization.
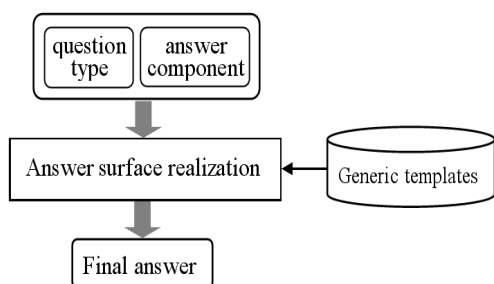

Figure 3.2   The flowchart of answer surface realization

4. Evaluation

We evaluate our approach by applying our answer generating method to 53 questions which have been classified by our classifier, and we distinguish three cases: correct answer, incorrect answer, and no answer. The performance of answer generation task is evaluated by means of coverage and accuracy. Here we define that coverage is the ratio of the number of answered questions to the total number of questions. Accuracy is the ratio between the number of the correct answers and the total number of answers. The result of evaluation is shown in Table 4.1.

Table 4.1 Evaluation of the answer generation task

QT	#Q	#CA	#ICA	#NA	Cov(%)	Acc(%)
1	10	9	0	1	90	100
2	17	15	0	2	88.24	100
3	12	11	0	1	91.67	100
4	14	13	1	0	100	92.86
total	53	48	1	4	92.42	97.96

5. Conclusion

This work research on Chinese cuisine QA service system. Based on question classification, we capture various aspects of the question-answer relationship by answer planning, match question subject with answer components which is extracted from related domain database, and generate succinct answer in natural language. The evaluation shows that a small number of linguistically motivated domain features can efficiently classify questions, and natural language generation techniques can be flexibly used to generate answer for restricted domain question answering system.

Acknowledgments

References

［1］ D. Molla and J. Gonzalez. Question answering in restricted domains: An overview. In Proceedings of Computational Linguistics, pages 41–61, 2007.

［2］ H. Yu, C. Sable, and H. Zhu. Classifying medical questions based on an evidence taxonomy. In Proc. AAAI 2005 workshop on question answering in restricted domains, 2005.

［3］ L. Xia, Z. Teng, and F. Ren. Question classification for chinese cuisine question answering system. IEEJ Trans on Electrical and Electronic Engineering, Vol. 4 , No. 6, 2009.

［4］ H. Q. Hu, P. L. Jang, and F. J. Ren. A new question answering system for chinese restricted domain. IEICE Transactions on Information and Systems, E89-D(6):1848－1859, 2006.