See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/331418056

Domain specific automatic Chinese multiple-type question generation

Conference Paper · December 2018

DOI: 10.1109/BIBM.2018.8621162

CITATION
1

READS
56

3 authors, including:

Tianlin Zhang
Chinese Academy of Sciences
12 PUBLICATIONS 13 CITATIONS

SEE PROFILE

Pei Quan
Chinese Academy of Sciences
12 PUBLICATIONS 13 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

UCAS_HPDM View project

intelligent transportation View project

All content following this page was uploaded by Tianlin Zhang on 12 November 2019.

The user has requested enhancement of the downloaded file.

Domain specific automatic Chinese multiple-type question generation

Tianlin Zhang[1,2,3], Ying liu[1,2,3], Pei Quan[1,2,3]

School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 101400, China[1]

Data Mining and High Performance Computing Lab., University of Chinese Academy of Sciences, Beijing 101400, China[2]

Key Lab. of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China[3]

zhangtianlin172@mails.ucas.ac.cn, yingliu@ucas.edu.cn, quanpei17@mails.ucas.ac.cn

***Abstract*—Question Generation is a rising research field of artificial intelligence in education and staff training. It is an effective strategy to conduct knowledge evaluation and performance appraisal. How to generate a variety of good quality questions has been an essential issue. In this paper, we propose an automatic method that generates Chinese multiple-type question by combing diverse knowledge bases and similar language features. Our research is mainly focused on Tradition Chinese Medicine (TCM) domain because it contains many Chinese-based resources and complex relationships. But the method is flexible and can be easily applied to other domains. Experimental results demonstrate that the proposed method obtain a good result.**

***Keywords*—Automatic Question Generation, Multiple-type question generation, Natural language processing, Knowledge evaluation, TCM domain**

I. Introduction

Knowledge evaluation and performance appraisal are essential for educational institution and enterprises. Researches manifest that questioning is an effective assessment strategy. But traditional generating questions manually require an intense amount of human labor and time [1].Motivated by these challenges, Automatic Question Generation(AQG) has emerged.

AQG is an important and challenging problem in natural language processing. In a specific domain, such as Traditional Chinese Medicine(TCM), AQG can efficiently evaluate medical students or provide continuing education for doctors and nurses. In order to better meet the assessment requirements, the diversity and quality of the questions become very important.

II. Ralated Work

Many automatic question generation systems have been proposed with different methods, such as embedding-based [2], WordNet-based [3] and ontology-based [4] question generation. Meanwhile, these methods just designed in English. The research of Chinese question generation is still immature. The major approaches take advantage of statistical, semantic and syntactical information [5] [6] [7]. But all of them do not make full use of the various knowledge information in the domain and merely aim at a single type of question. So in this paper, we propose an automatic method for generating Chinese multiple-type questions in a specific domain. The major contributions of this paper can be summarized as follows:

- Constructing a dataset which combines multiple knowledge bases in a specific domain, and using it to generate Chinese questions automatically.
- Utilizing some rules to generate multiple types of questions, including Factual question, Yes/No question and Multiple-Choice question.
- For Multiple-Choice question generation, the distractors are generated by combining various similar language features.

Our research is mainly focused on automatic question generation from TCM domain but the presented strategy is flexible and can be easily applied to other domains.

III. Methodology

Fig. 1 shows the whole framework of our QG system. As preparation, we build up a Knowledge base related to TCM. Then, we use some rules and NLP technologies to generate multiple-type question.

A. Knowledge base Construction

We build a dataset as the knowledge base for question generation. The contents of dataset come from unstructured and semi-structured texts, weblogs, domain term base, rule base and ontology base, etc. associated with TCM.

These data can be obtained from web crawling, Chinese pharmacopoeia and Dictionary of prescription. Especially, for unstructured texts like article and textbook, in order to get the core contents and generate high-quality questions, we should select some representative sentences which contain the core concepts. We use a graphical algorithm called TextRank [8]to extract core sentences.

B. Factual Questions Generation

Factual questions(a.k.a. wh-questions, FQ) are a common type of question whose answers are specific facts. We mainly generate factual questions from the core sentences in the knowledge base. The generation mainly involves three parts [9] : 1) Finding the appropriate question word in the sentence, 2) Deciding on whether to use What or Why, 3) Transforming question.

In order to find the appropriate question word, we must preprocess the sentences. Due to the lack of clear boundaries between words in a Chinese sentence, Chinese is unique to the Western language. So we need to segment word firstly. Besides, the preprocessing step includes part of speech tagging, named entity recognition, dependency parsing and semantic role label parsing(SRL). This grammatical and semantical information is essential for question generation. In our research, we use the LTP platform [10] to do these preprocessing work. Fig.2 presents an example of processing result in LTP.

In our work, we can find the location, number, a person name and other information that can be used to generate questions by using named entity recognition. In addition, we define some rules based on dependency parsing and SRL to get the relations between subjects and objects in the sentence, such as causality. And we also identify other roles like some adverbial.

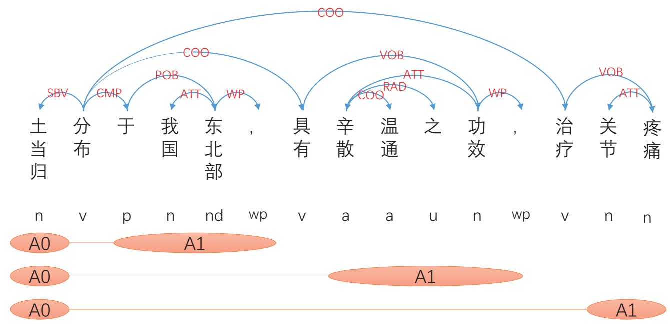

Fig. 2: Example of processing result in LTP.

We can decide what type of the factual question to use after finding the above information. Meanwhile, we would utilize our sentence pattern and question generation rules defined to transform the sentence into a question. The rules and examples are shown in Table I.

C. Yes/No Questions Generation

Yes/No questions (YNQ) are a type of question that can be used to understand the difference between different concepts. knowledge graphs store information about lots of entities and relationships between them, in addition, they include different concepts which span different fields of TCM textbooks. Thus we use our ontology base and knowledge graphs to generate YNQ.

The relationships between different entities that captured by knowledge graphs are usually expressed in the form of triples. For example, 上位词(牛蒡子，活血祛瘀药) Hypernym (Arctium, Drugs for activating blood circulation to dissipate blood stasis). 治疗(人工牛黄，癫痫) Treatment (artificial bezoar, epilepsy).These triples are referred to as certainty which can be transformed to judgement sentences. For instance, The above triples are changed into牛蒡子是活血化瘀药(Arctium belongs to drugs for activating blood circulation to dissipate blood stasis) and人工牛黄能治疗癫痫(Artificial bezoar can treat epilepsy).

Because there have many triples in ontology base and knowledge graphs, we use an algorithm related concept semantic similarity in triples [11] to find a confusing description to replace the original attributes. 上位词(菟丝子，利水消肿药) Hypernym(Semen Cuscutae, diuresis-promoting herbal medicine)has another kind of medicine which will add interfering words.

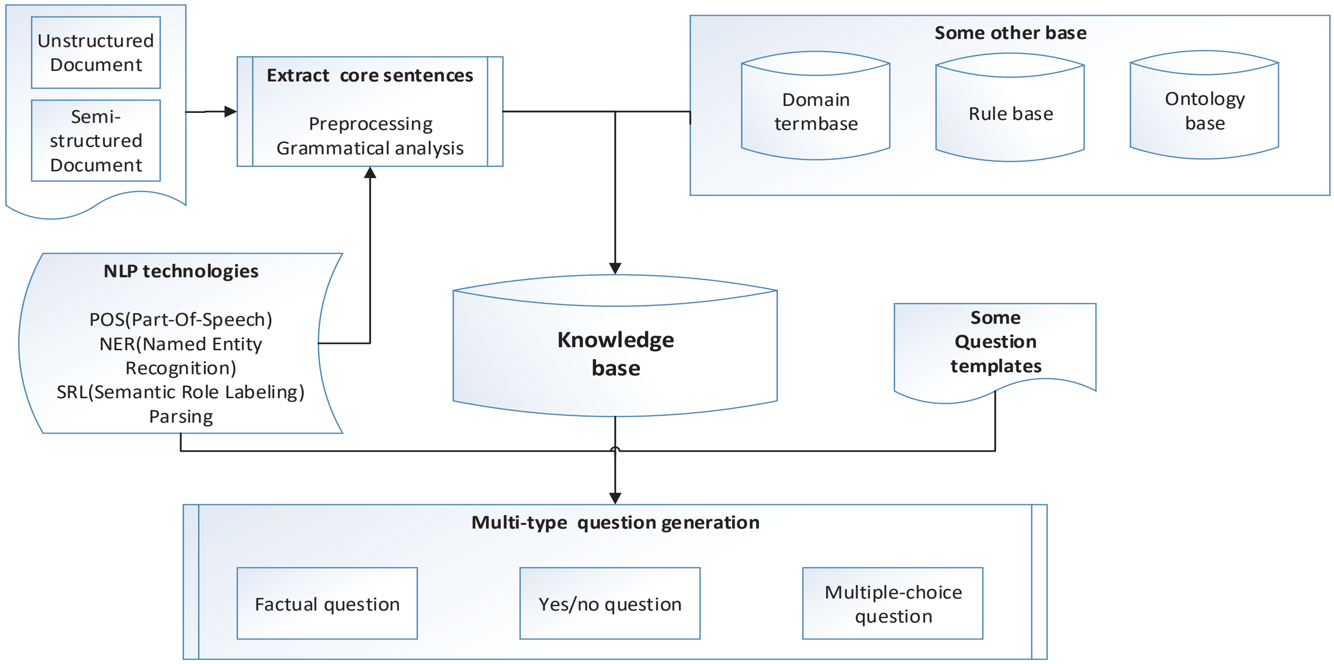

Fig. 1: The framework of QG system

According to such judgement sentences and interfering words in the previous examples, we can generate the following questions by utilizing some rules.

牛蒡子是利水消肿药么？

(Question: Does Arctium belong to diuresis-promoting herbal medicine?)

人工牛黄能治疗癫痫么？

(Question: Can Artificial bezoar treat epilepsy?)

D. Multiple-Choice Questions Generation

Multiple-choice questions(MCQ) are efficacious and popular way of assessment in which users need to select the correct answer from some distractors(a set of alternative choices which can distract users).

MCQ consist of three parts: 1) the suitable sentence as question 2) the correct answer 3) distractors. The way of selecting appropriate sentences is the same as factual questions generation. And the words used for the question are mainly terminology or some significant concept.

What is the most important is good quality distractors generation, because good quality distractors can better test the user’s mastery of knowledge. Good quality distractors refer to alternative answers which are semantically related to the correct answer and better confusing [12].

Different from the traditional merely ontology-based or embedding-based approaches, in our work, we combine many different similar language features to generate distractors.

- POS similarity. Ensuring the POS tags are same. Distractors and correct answer are all noun or verb.
- Affix similarity. Affixes include prefixes and suffixes. The distractors have common affix with correct answer. For example,红枣(Jujube) and 黑枣(Diospyros Iotus Linn) 虎膏(tiger fat) and 虎骨(tiger bone).
- Embedding similarity. Embedding similarity indicates the similar semantic information, By using an efficient word embedding strategy word2vec [13] to find similar semantic representation word. For example, 散瘀消肿(Dissipating blood stasis and swelling)，生肌止痛(producing muscles and relieving pain)，清热利咽(clearing away heat and relieving pharynx) have semantically similarity.
- Frequency similarity. In domain textbooks, similar word frequency indicates a similar degree of importance.
- General knowledge similarity. It refers to the most confusing terms and concepts which come from web blog and ontology base etc. For instance, 半枝莲(Scutellaria barbata D.Don)和半支莲(Large flower Purslane).In Chinese, only one word is different, but the efficacy of medicine

TABLE I: Factual question rules and examples

Factual question type	Rule	Pattern	Question template	Example
什么(What)	contains the explanation of the rule words* of the terminology	Contains(term's rw*, A0)	Replace(A1, 是什么)	川乌的功效是祛除风湿。 The effect of the Aconitum carmichaeli is to remove rheumatism.
				川乌的功效是什么？ What's the effect of the Aconitum carmichaeli?
什么人(Who)	contains person name, personal pronouns, and types of people	Contains(nr or prp, A0)	Replace(A0, 什么人)	阴虚火旺的人忌用川穹。 People with Yin-deficiency and Fire-hyperactivity are not allowed to use Ligusticum Wallichii Franch.
				什么人忌用川穹？ Who should not use Ligusticum Wallichii Franch?
哪种(Which)	subject contains terminology	Contains(term, A0)	Replace(A0, 哪种term's parent)	郁金的主要成分包括樟脑和姜黄素。 The main components of Curcuma include d-camphor and curcumin.
				哪种中药成分中包括樟脑和姜黄素？ Which TCM includes d-camphor and curcumin?
为什么(Why)	contains Causal vocabulary	Contains(cv.., A1)	Replace(cv, 为什么)	川楝子不可过量或持续使用，因为川楝素为强积累物质。 Melia toosendanin can not be used excessively or continuously,because toosendanin is a strong accumulation of matter.
				为什么川楝子不可过量或持续使用？ Why Melia toosendanin can not be used excessively or continuously?
哪里(Where)	contains Place adverbial	Contains(ns or pa, A1)	Replace(A1, 在哪里)	冬虫夏草主要产自云贵高原地区。 Cordyceps sinensis is mainly produced in Yunnan-Guizhou Plateau.
				冬虫夏草主要产自哪里？ Where is Cordyceps sinensis produced?
多少(How many)	contains a number following quantity	Contains(m+q，A0)	Replace(A0, 需要多少q)	银柴胡需3-10g煎汤内服用于治疗盗汗。 Take radices stellariae dichotomae 3 to 10 grams of decoction for oral administration to treat night sweat.
				需要多少g银柴胡煎汤内服治疗盗汗？ How many grams of radices stellariae dichotomae for oral administration to treat night sweat?

Notes:rule words* indicate some special question words, such as "功效(contraindication)","禁忌(pesticide effect)","成分(component)".A0 is the agent, A1 is the object.
m is number, nr or pre is personal pronouns, cv is causal vocabulary, ns or pa is place adverbial, q is quantity. Contains(A, B) indicates that A is in B, Replace(X, Y) indicates that X is replaced by B in question generation.

is completely different, so it is very confusing.

By counting the number of times the distractors are selected in the tester's actual answer results, and based on the difficult requirement of the question, the selection of similar features can be freely matched.

The Multiple-choice question about TCM as the instance shown in the Fig.3.

以下哪种中药具有驱虫解毒的功效？（Which of the following Chinese medicine has the efficacy of expelling parasite and detoxifying?）
A.大叶地耳根(root of Combritum latifolium)
B.大叶凤仙花(Impatiens apalophylla)
C.大叶千斤拔(Flemingia macrophylla)
D.大叶骨牌草(Microsorium fortunei(Moore)Ching)

麝香具有治疗以下哪种病症的功能？(What kind of symptoms can musk treat?)
A.吐血(hematemesis)
B.须发早白(premature graying hair)
C.疮疡肿毒(swelling sores)
D.疟疾(malaria)

Fig. 3: Examples of Multiple-choice question.

IV. Experiment and Analysis

A. Experimental data

In this study, we use TCM data from various sources. There are more than 30000 semi-structured data crawling from the network. Unstructured data come from Chinese textbooks including *Chinese Medicine (7th edition)*, *Chinese Medicine Prescription*, *Modern Research on 500 Kinds of Chinese Medicine*, *surgery of traditional Chinese medicine*, etc. There are 10715 Chinese medicinal materials and introduction from Chinese medicine information inquiry Platform. In addition, we utilize a knowledge map based on TCM language system, which contains 1905 entities and 31 semantic relationships.

B. Experimental results

In the domain of TCM, more than five thousands questions were generated by using our knowledge base. As far as we know, there is no benchmark dataset for Chinese question generation, which makes it difficult to directly compare.

We invited some students to evaluate the questions generated by our automatic method. We randomly selected 200 questions for each question type to evaluate. The evaluation is shown in Table II:

TABLE II: qualitative feedback for questions.

Type of comment	**coveraged count**		
	FQ	YNQ	MCQ
grammatically well-formed	156.6	187	174
mostly well-formed.with slight problems	27.6	11	16
has grammatical problems	15.8	2	10
semantically adequate	185	181.4	188
mostly adequate,with slight problems	11.2	15	10
has semantically problems	3.8	3.6	2

It can be seen from the evaluation that the majority of our generation questions are standard in grammar and semantics. And there are some unsatisfactory aspects, such as some keywords chosen are not important. But all in all, our method is feasible.

V. Conclusion

Automatic question generation is a challenging task. It is a very effective assessment method as well, whether in education or staff training. Thus automatic question generation is in real demand. In this paper, we present an automatic Chinese multiple-type question generation method in a specific domain by combing with many strategies. The questions generated are of high quality and variety. In the future, we will design an efficient generation method and integrating the question generation tool into a learning management system.

Acknowledgment

This project was partially supported by Grants from Natural Science Foundation of China ♯ 71671178 /♯ 91546201 /♯ 61202321, and the open project of the Key Lab of Big Data Mining and Knowledge Management. It was also supported by Hainan Provincial Department of Science and Technology under Grant No. ZDKJ2016021, and by Guangdong Provincial Science and Technology Project 2016B010127004.

References

[1] Soleymanzadeh, Katira. "Domain Specific Automatic Question Generation from Text." ACL 2017, Student Research Workshop 2017:82-88.

[2] Kumar, Girish, R. Banchs, and L. F. D'Haro. "RevUP: Automatic Gap-Fill Question Generation from Educational Texts." Tenth Workshop on Innovative Use of Nlp for Building Educational Applications 2015:154-161.

[3] Mitkov, R. "Computer-aided generation of multiple-choice tests."International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings IEEE, 2003:15.

[4] Stasaski, Katherine, and M. A. Hearst. "Multiple Choice Question Generation Utilizing An Ontology." The Workshop on Innovative Use of Nlp for Building Educational Applications 2017:303-312.

[5] Liu, Ming, V. Rus, and L. Liu. "Automatic Chinese Factual Question Generation." IEEE Transactions on Learning Technologies 10.2(2017):194-204.

[6] Ding, Xiang Min, and G. U. Hong-Bin. "Automatic generation technology of Chinese multiple-choice items based on ontology." Computer Engineering & Design 31.6(2010):1397-1400.

[7] Liu, Ming, V. Rus, and L. Liu. "Automatic Chinese Multiple Choice Question Generation Using Mixed Similarity Strategy." IEEE Transactions on Learning Technologies PP.99(1939):1-1.

[8] Niu, Jianwei, et al. "OnSeS: A Novel Online Short Text Summarization Based on BM25 and Neural Network." Global Communications Conference IEEE, 2017:1-6.

[9] Flor,Michael and Riordan,Brian. "A Semantic Role-based Approach to Open-Domain Automatic Question Generation." Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications, 2018:254-263.

[10] Che, Wanxiang, Z. Li, and T. Liu. "LTP: A Chinese Language Technology Platform. " Journal of Chinese Information Processing 2.6(2010):13-16.

[11] Xiao, Min, L. Zhong, and Q. Xiong. "Semantic Similarity between Concepts Based on OWL Ontologies." International Workshop on Knowledge Discovery & Data Mining 2009:749-752.

[12] Goodrich, Hubbard C. "Distractor Efficiency in Foreign Language Testing." Tesol Quarterly 11.1(1977):69-78.

[13] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." International Conference on Neural Information Processing Systems Curran Associates Inc. 2013:3111-3119.

View publication stats