

# Research on Auto-solving Methods of You-sentence Word Problems

Ying zhou<sup>1</sup> Shengli Wang<sup>1</sup> Panfei Li<sup>1</sup> Yun Ren<sup>2</sup>

<sup>1</sup>Beijing Normal University, Beijing, China

<sup>2</sup>Beijing No12 High School, Beijing, China

## Abstract

Based on natural language processing technology, cognitive psychology theory and word problem teaching methods, word problem auto-solving and intelligent tutoring technologies have become hot issues of research in recent years. However, since word problems are described by natural language, the number of resolved problems is limited. You-sentence<sup>1</sup> is a common sentence in Chinese word problems and the method of understanding it is easily extended to other types of sentences. This paper proposes an algorithm for auto-understanding and auto-solving You-sentence word problems and develops a system based on it to see whether the algorithm works.

**Keywords:** auto-solving, auto-tutoring, word problem, intelligent tutoring

## 1. Introduction

Learning how to solve mathematical word problems is an important way to build and improve the pupils' problem-solving skills, but word problem teaching

---

1 'You' is a transliteration. It references '有' in Chinese and corresponds to 'there be', 'own', 'exist', 'possess', 'have', etc in English. You-sentence is that the predicates or predicate center of one sentence is the verb 'You' or You-structure in Chinese.

has always been a difficult and a weak link in mathematics teaching. The research in the area of word problem intelligent tutoring system has gained the attention of more and more researchers. These systems could adaptively present problems [1], generate word problems described by nature language within limited preconditions [2], provide useful information to help solving problems [3] or generate graphical clues to help understanding problem [4]. However, many problems also exist, for example, word problems are built-in and only answers are provided to students[1][3][4], which make them unaware of the errors in their resolution and unknown how to correct them.

In earlier research, some other researchers [5] had tried to solve these problems with word problem auto-solving. The tutoring system with the realization of word problem auto-solving not only solves word problems out of the word problem base, but it also identifies which part of the solving process goes wrong, points out other related problems to the students automatically, generate graphical clues to help students master solving methods, and offer teachers a detailed analysis report about students' wrong-procedures.

## 2. Related work

In recent years the research of the main systems and auto-solving methods throughout the world is as follows.

Bakman[6] developed a simulation program ROBUST that can understand the multi-step arithmetic word problems in English. ROBUST works by first parsing the problem text to split all sentences into simple sentences and then substituting the constant values in the simple sentences for the corresponding variables in a formulae. LIM-G[7] can comprehend one-step geometric word problem inputted in English. LIM-G uses InfoMap that is an ontology tool and framework template to comprehend the nature language. Considering that some semantic sentence models include sentence component such as subject, object and quantity etc. Ma[8] has solved the multi-step word problem of the first mathematical period of primary school with backward inference. SoMaPoTeS[9] uses Gate to extract the information from word problems to construct the semantic web. Then, the semantic web is transformed into equations that can be solved by Sage, an external algebraic equation system.

In conclusion, the system representation method represents an increasing number of types and number of sentences. But only a small number of problem types can be automatically solved at the present time, such as some problems related to Rate[2], Algebra, Arithmetic, Geometry, Probability and Proof, or some problems related to the context like problems about money[9]. We have a long way to go before word problem auto-solving can be used in a practical application.

### 3. Algorithm of Processing You-sentence

Confronting the existing auto-solving systems are relevant to problem types and context, the authors have analyzed a large number of word problems and found some sentences have nothing to do with problem types and context, especially in

certain assignment sentence which are used in all problems, for example ‘Lili has 3 apples’ and ‘Xiao hong is 15 years old.’ Our research team constructed a database with 5672 word problems, including 23662 sentences all together in which the You-sentence appears 3091 times and accounts for 15 percent. According to statistics, ‘You’ is the most frequent word and verb and next is ‘be’ close to 13%, meanwhile their characteristics are similar. Hence realizing the understanding of You-sentences is important for realizing auto-solving word problem and is the key to promote auto-solving accuracy significantly.

#### 3.1. Knowledge Representation Approach of You-sentences

According to the research results of You-sentence patterns in Chinese linguistics, You-sentence in word problems, the function of all elements in sentences containing Subject, Predicate, Object, Attribute, Adverbial and Complement and the semantic relationships among them, six sentence patterns of You-sentence were summarized as shown in Table 1, in which A represents ‘owner’, B1 represents ‘quantity and quantifier’, B2 represents ‘object’ and C represents the follow-up component.

Some You-sentences contain words such as ‘一共有 (have in total)’ (for example, Xiao wen has 46 photographs and pictures in total) and ‘原来有 (used to have)’ (for example, a school used to have 1200 students) etc. These words in You-sentences represent numeral relationship between objects or objects’ state, called variable to indicate that the number is a certain variable in a formula, for example, ‘3 hours’ in the sentence ‘a train arrived at its destination after 3 hours’ is the variable ‘time’ in the formula ‘time\*speed=distance’. The sentences including words like ‘一共有 (have in

total)’ are called complicated You-sentences and the words are called key words. On account of ‘key words-variable-value’ feature between the sentence components and the quantity, the paper employs the framework representation to represent these components, namely, You-sentences is

represented by 7 slots on the framework including owner, verb, object, quantity, quantifier, follow-up component and variable. The recognition, understanding and filling these components in framework are the key to success in solving word problems. An example is shown in Fig. 1.

Table 1: Sentence Patterns of You-sentence in Word Problems

Sentence Patterns	Example Sentences in Chinese	The English translation of Example
A+ You +B1	妮妮有 5 个	Nini has five.
A+ You +B1+B2+[C]	草地上有 17 只羊[在吃草]	There are 17 goats [grazing] on the grass.
A+ You +B2+B1	山坡上有松树 9 棵	There are 9 pines on the hillside.
You +B1+B2+[C]	有 16 只鹅[在河里游泳]	There are 16 geese [swimming in the river].
You +B2+B1	有柳树 6 棵	There are 6 willows.
B2+ You +B1	苹果有 8 个	There are 8 apples.

Framework name:	ID of the word problem
Sentence:	The school has 3 dance teams in rehearsals.(学校有三个舞蹈队在排练)
Owner:	the school
Verb:	has
Object:	dance team
Quantity:	3
Quantifier:	ge(the translation of ‘个’ that is a Chinese word)
Follow-up component:	in rehearsals
Variable	Null (To be confirmed)

Fig. 1: an Example of Framework Representation

### 3.2. Processing Algorithm of You-sentences

The understanding of You-sentence is defined as all its components, owner, verb, object, quantity, quantifier, follow-up component and variable are confirmed in this article. The algorithm that confirms these components is called a processing algorithm of You-sentences, and the algorithm flowchart is shown in Fig. 2 in which the crucial links of algorithm processing are as follows.

Make up sentences by labeling participles and part of speech on the base of ICTCLAS, a word segmentation system developed by Chinese Academy of Sciences, such as: The original sentence is: ‘草地上有公鸡 38 只(There are 38 cocks in the grass)’, and the result of Chinese

word segmentation and Part-Of-Speech Tagging is ‘草地/n 上/f 有/v 公鸡/n 38/m 只/q 。 /w ( the grass/n on/f have/v cocks/n 38/n zhi/q)’ in which ‘n’ represents noun, ‘f’ represents localizers, ‘v’ represents verb, ‘m’ represents number and ‘q’ represents quantifier.

Recognize quantity and quantifier. Put the content before ‘You’, ‘You’, the content between ‘You’ and the quantity, the quantity, the quantifier, the content after the quantifier and punctuation into cell 1, cell 2, cell 3 ... cell 8 respectively. If the content of this section does not exist, then the corresponding cell is empty. According to the part of speech m and q, the quantity and quantifier are confirmed.

Recognize the follow-up component. The sentence patterns of You-sentence

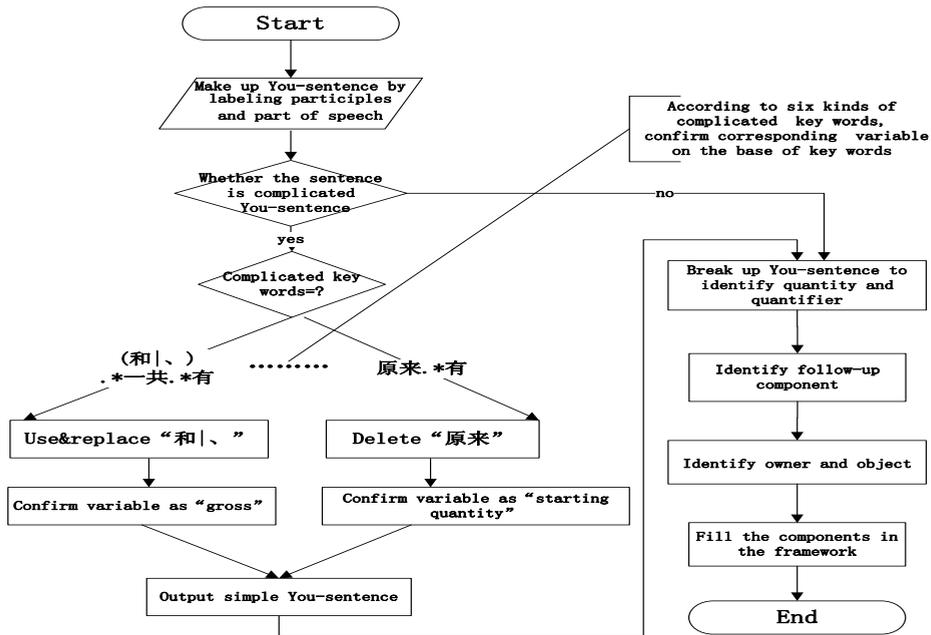


Fig. 2: The Algorithm of Processing You-sentence

indicate that the content after the quantifier is the ‘object’ and its follow-up component, and the study concludes that the follow-up component generally begins with the word which POS tagging is adverb (d), verb (v), vice-verb (vd), prepositions (p), vice-morpheme (dg), or conjunctions (c) and so on. So find the first position in which one of these words appear, and make the position as the dividing line, then the ‘object’ and the follow-up component will be separated.

Recognize the owner and object. After the above processing, the determined content are that ‘you’ is in the second cell, number is in the forth cell, quantifier is in the fifth cell, follow-up component is in the seventh cell and punctuation is in the eighth cell. Whether it is ‘owner’ or ‘object’ in the first, third and sixth cell need further analysis. The sentence patterns of You-sentence indicate that if the third cell and the sixth cell are not empty, only one of them has content and the content must be ‘object’. However,

there are two similar sentence patterns in You-sentence: B2 + You + B1 and A + You + B1. In both patterns, the corresponding slot of the content before ‘You’ is different and can be ‘owner’ slot or ‘object’ slot. How to determine the slot of this location is difficult to point out in this study. After analyzing word problems texts, ‘object’ is diverse, while ‘owner’ is the owner or the location of ‘object’ and often has features such as name and place name. Consequently, the following two laws can help to determine the ‘owner’. On the one hand, according to characteristics of ‘owner’, the value of ‘owner’ slot can be summarized as following: name (such as: 小红 / nr, Xiaohong), character (such as: 张阿姨 / n, aunt Zhang), location (such as: 文具店 / n, stationery shop), position + orientation (eg: 草地 /m 上 /f, on the grass), character/name + home (such as: 妮妮 /nr 家 /n, Nini’ s house), [school] + [grade] + [class] (e.g.: 新华 /nz 小学 /n 一 /m 年级 /n

二/m 班/q, Class 2, Grade 1, Xinhua primary school), sequence words (such as: the first, second), pronouns (such as: this, one, that.). In addition, it also includes the combination of several categories, such as: ‘二/m 年级/n 三/m 班/q 教室/n 里/f, in the classroom of Class 3, Grade 2’. These contents are built into a common knowledge base to match the first cell and judge whether it is ‘owner’. On the other hand, according to the value of ‘owner’ and ‘object’ slot of the sentences before or after this You-sentence, if the content is the ‘owner’ or ‘object’ in the previous or the next sentences, it is also ‘owner’ or ‘object’ in this sentence. Thus the above two laws can confirm the cells’ content are ‘owner’ or ‘object’.

If You-sentences included complicated key words, the processing will be different based on different key words. Variables are determined by these key words, but the Most of the variable slots in You-sentences are null and they are processed when solving problem.

#### 4. The word problem auto-solving system and examples of auto-solving method

The word problem auto-solving system is composed of seven modules: user interface module, preprocessing module, semantic understanding module, auto-solving module, problem answer presenting module, common knowledge base and mathematical formula base. The overall system structure is shown in Fig. 3.

From the user interface module users input the word problem in Chinese, then the Chinese word segmentation and Part-Of-Speech Tagging is accomplished and the stop-words are removed in the preprocessing module. These stop-words include: ‘问/v’, ‘问/v : /w’, ‘可/v 是/v’, ‘已/d 知/v’, ‘那么/d’, ‘则/d’, ‘若/c’, ‘如果/c’, ‘知道/v’ (‘inquire/v’, ‘inquire/v : /w’, ‘but/v’, ‘known/v’, ‘then/’, ‘so/d’,

‘for/c’, ‘if/c’, ‘know/v’) and so on. In the semantic understanding module, the word problem is split into several sentences by punctuation. The sentences components are filled into framework slots by different processing algorithm based on different sentence patterns. In auto-solving module, first, the variables in corresponding formula are determined by complicated key words. If there are no complicated key words related to all variables, it will infer the variable relevant to the sentence from the summarized rules. Then the formula for solving the problem is determined by one or two variables, the variables are replaced by quantities and the answer is deduced. If the calculation goes wrong, this type of error usually arises from the inconsistent ‘owners’ or the inconsistent ‘objects’ between two sentences. At this point, the system will call on a common knowledge base and continue to reason by the required common knowledge. The problem answer presenting module feeds the problem-solving steps and the final answer back to the users at last.

The system could automatically work out all word problems composed of You-sentences. The examples of word problem auto-solving are shown in Fig. 4. The figures present how to fill You-sentence in the framework, namely, the understanding results of You-sentence.

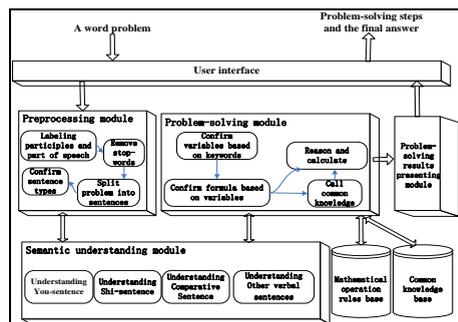


Fig. 3: The Architecture of Word Problem Auto-solving System



Fig. 4: The Word problem Composed of You-sentences And Comparative Sentences Understanding And Solving

## 5. Comparison and Prospects

Compared with the existing research, this paper changes the method of sentence comprehension to propose the new method using the algorithm to understand sentences and the research is more detailed. For instance, it can process the follow-up component, solving the conflict matching between one sentence and many sentence models and auto-solving all forms of You-sentences word problems.

Since the auto-solving technology enables the system to know every step of solving problems, the information could help teacher automatically acquire a detailed and real-time problem-solving report (correct or not). The system can also be applied to individualized counseling in homework and it will deliver the whole class assignment report to the teacher to assist in preparing lessons. Moreover, the system is suitable for foreign Chinese teaching and for keeping the elders' problem-solving skills.

## 6. References

[1] W. Schoppek and M. Tulis, "Enhancing arithmetic and word problem solving skills efficiently by individualized computer-assisted

practice". *The Journal of Educational Research*. vol. 103, no. 4, pp. 239-252, 2010.

[2] W. Sandra, *Generating Mathematical Word Problems: Predicting Item Difficulty Through Controlled Modifications*, Kansas:ProQuest, USA, 2006.

[3] A. Ahmad, H. B. Zaman, S. S. Salim and R. Zainuddin, Proc. "MINDA: A cognitive tool for solving mathematical word problems". *IEEE Symp. Information Technology (ITSim)*, IEEE Press, pp. 15-17, 2010.

[4] T. H. Huang, Y. C. Liu and H. C. Chang, "Learning achievement in solving word-based mathematical questions through a computer-assisted learning system in Taiwan". *Educational Technology & Society(SSCI)*, vol. 15, no. 1, pp. 148-259, 2012.

[5] D. G. Bobrow, "Natural Language Input for a Computer Problem Solving System". 1964.

[6] M. Buckley and D. Dietrich, "Verification of proof steps for tutoring mathematical proofs". *Proceedings of the 13th international conference on artificial intelligence in education*. pp. 560-562, 2007.

[7] W. K. Wong, S. C. Hsu and S. H. Wu, "Learner-initiating instruction model based on cognitive knowledge for geometry word problem comprehension". *Computers & Education*, vol. 48, no. 4, pp. 582-601, 2007.

[8] Y. H. Ma, "Representation the Semantic Frame and the Calculus Method for Automatic Solving Arithmetic Word Problem of Primary School". *Beijing Normal University, CHN*, 2010.

[9] C. Liguda and T. Pfeier, "Natural Language Processing and Information System". *Springer Berlin Heidelberg, GER*, 2012.