



# Understanding Plane Geometry Problems by Integrating Relations Extracted from Text and Diagram

Wenbin Gan<sup>(✉)</sup>, Xinguo Yu, Chao Sun, Bin He, and Mingshu Wang

National Engineering Research Center for E-learning,  
Central China Normal University, Wuhan, China  
wenbingan@mails.ccnu.edu.cn

**Abstract.** Understanding problems is a critical and hard step for solving plane geometry problems. This paper presents a method for understanding plane geometry problems by integrating the information from text and diagram two modalities. Then high-confidence geometric relations are extracted for problem understanding through integrating the information separately extracted from text and diagram. And this paper also presents a function that can visually encode the extracted relations into the diagram to interactively present the visual effects of problem understanding results. The geometric relations are mined from the diagram after visual primitives are detected. The syntax-semantics ( $S^2$ ) model method is adopted to extract geometric relations from the text. And an integration process is used to couple the information from text and diagram to obtain the high-confidence geometric relations. The experimental results show that the proposed method can mine geometric relations in high accuracy and it can understand some problems that cannot be understood by using text only or by using diagram only.

**Keywords:** Plane geometry problem understanding  
Geometric relation extraction ·  $S^2$  model · Visually presentation  
Textual information extraction · Visual information extraction

## 1 Introduction

Automatically solving math problems is a long-standing research problem in AI [2, 6, 12] and it is a core technology in building intelligent educational systems to tutor learners. In this paper we focus on the understanding of plane geometry problems in which the question text is accompanied with a diagram (an example problem is shown in Fig. 1), which is a critical step of automatically solving geometry problems.

Generally, diagram and text are used complementary as effective means to state the problems clearly in geometry discipline. In some geometry problems, the diagram contains the necessary information to solve the problem which are omitted in the question text for avoiding repetition. On the other hand, the text

contains some decidable information that are ambiguous in a diagram because of imprecise scale. For the fully understanding of a geometry problem, it is necessary to propose the mechanism of integrating the information from both text and diagram.

This paper presents a novel method for understanding plane geometry problems by integrating the information from text and diagram. Then it uses LI-Geo, a learner-initiating geometry system, to interactively present the visual effects of the extracted relations to help learners understand the given geometry problems. The proposed problem understanding method identifies visual primitives from diagram and mines basic and derived geometric relations among the primitives, and uses an  $S^2$  model matching method to extract the textual entities and the geometric relations in the problem text. By integrating the visual information and the textual information, coreferences are established between the visual primitives and textual entities that refer to the same object, and some high-confidence geometric relations are found and visually encoded into the diagram to interactively present the visual effects of problem understanding results. Experimental results show that the proposed method has high accuracy in mining geometric relations from both diagram and text and it can understand some problems that cannot be understood by using text or diagram only. A user study also validate the usability of the proposed method in helping people understand geometry problems.

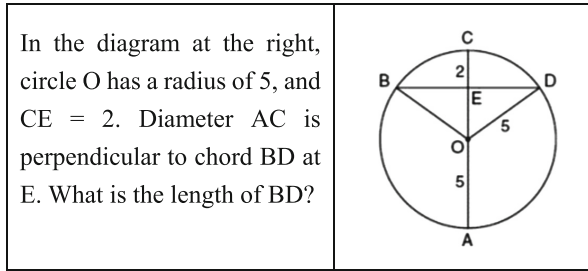


Fig. 1. A geometry problem where the question text is accompanied by a diagram.

## 2 Related Work

A majority of previous work in automatic problem understanding address two problems of diagram understanding and text understanding in isolation [10, 11]. For the understanding of the diagram in a geometry problem, a common approach is to detect the basic geometric primitives, mainly points, lines, circles, triangles, rectangles and so on. Hough Transform is a popular method for detecting lines and circles. Triangle and rectangle can be detected in a bottom-up manner where lines are linked together to form the big primitives [11]. Zhang and Fu [16] proposed a method using Hough transform and corner detection algorithm

to recognize and understand geometry diagram. The understanding is mainly focus on the finding of vertexes and lines, and their specific coordinate information. Seo et al. [11] proposed a method named G-ALIGNER for diagram understanding in geometry questions that discovers visual elements by maximizing agreement between textual and visual data. The use of textual information to assist in identifying the visual primitives improves the accuracy of primitive detection compared with hough-based method. In the higher-level understanding of the geometry diagram, some meaningful geometric information and properties implied in visual data are extracted from the geometry diagram. Chen et al. [1] proposed a method using geometric features retrieved from the diagram to find the underlying geometry theorem behind the diagram. It detects basic geometric primitives and mines basic geometric relations, then forms a undirected graph by representing the primitives as nodes and the relations as edges. A graph matching method is used to find the underlying theorems in the database. Liu et al. [7] propose a structure analysis method to better understand the spatial relationships of geometry diagram and describe a diagram in a series of features, such as local and global geometric attributes and spatial layout structure. These feature information can well represent the diagram. These works differ from our method in that they perform geometry understanding without considering the textual information. Moreover, these diagram analysis methods are insufficient for the geometry problems in which the diagrams label the values of line length or angle.

For the understanding of geometric textual information described in a problem, Guo et al. [4] proposed an algorithm to understand plane geometry proof problems in natural language (NL). This algorithm uses 196 sentence templates to transform the problems in NL into the problems described in the restricted geometric propositions. Regular expression matching is used to match a simple sentence with predefined relation patterns and get the contained relations. Mukherjee and Garain [9] developed another algorithm for formal representation of plane geometry proof problems. It used a knowledge base called GeometryNet to interpret the geometric meaning of an input text into diagram descriptions. Specifically, it decomposed the extracted entities into atomic entities by consulting the concepts in GeometryNet and used connector to link the entities to form a parse graph and then a translator is designed to translate the parse graph into structured summary of relation representation. Wong et al. [14] developed a system for understanding plane geometry proof problems and making conjectures. It represents each geometry relation as a relation frame consisting of several slots and predesigns a set of rules containing sentence templates for matching input sentences. It records the values of the attributes in the relation frame that are instantiated when a sentence is matched against a template. These works differ from our method because they perform geometry problem understanding by only considering the textual information, hence some information only present in the diagram cannot be obtained.

This paper is related to early work on understanding by integration of text and diagram [10–13]. Nakamura et al. [10] proposed a framework for semantic

understanding of a diagram by utilizing textual information. However, it assumes that the visual primitives in the diagrams are manually identified. Seo et al. [12] proposed a geometry solving system named GEOS, which understands geometry problems by combining text and diagram interpretation. This method uses the textual information to assist in identifying the visual primitives and extracts geometric relations by using statistical learning method. The combination of relations from two media improves the performance of problem solving, which also verifies the feasibility of proposed method by integrating textual and visual information in understanding plane geometry problems. However, the statistical learning method used in the text parsing is highly dependent on number of training examples, and since it is hard to obtain a large number of plane geometry problems and learning from a few examples makes it challenging for understanding a broader scope of plane geometry problems. In contrast to this method, we propose a method that also performs geometry problem understanding on the coordinated intake of information from both the text and the diagram but differs from it, the visual information and text information are obtained in isolation and a integration procedure is conducted subsequently to integrate both information. The primitives in the diagram are detected using a hierarchical detection algorithm and the relations (mainly quantity relations and spatial relations) are extracted and represented in the first order logic (FOL) like symbolic description. And an  $S^2$  model matching method is proposed to extract the relations in the text in high performance. Moreover, coreferences are built to align the visual primitives to their corresponding textual mentions, and high-confidence geometric relations are visually encoded into the diagram to interactively present the problem understanding results.

### 3 Overview of the Proposed Method

This section gives an overview of the proposed method of geometry problem understanding coupling textual and visual information. Before giving the problem formulation, some related concepts are first presented.

**Definition 1:** A geometric primitive is a visual element detected from a diagram. Four types of basic elements are used to form most of the diagrams in plane geometry, namely points, lines, circles and labels. All the primitives extracted from a diagram form a set  $P = \{P_1, P_2, \dots, P_m\}$ .

**Definition 2:** An entity mention is a word or phrase that indicates a primitive in the diagram. All the entity mentions extracted from the text form a set  $E = \{E_1, E_2, \dots, E_n\}$ .

**Definition 3:** An atomic proposition is a geometric relation by applying a predicate to a sequence of arguments (e.g.,  $isParallel(AB, CD)$ ). All the atomic propositions extracted from the text form a set  $R_T = \{R_T^1, R_T^2, \dots, R_T^i | R_T^i = predicate\langle E_1, \dots, E_a \rangle, a = 1, 2 \text{ or } 3\}$  and the atomic propositions extracted from the diagram form a set  $R_D = \{R_D^1, R_D^2, \dots, R_D^j | R_D^j = predicate\langle P_1, \dots, P_b \rangle, b = 1, 2 \text{ or } 3\}$ .

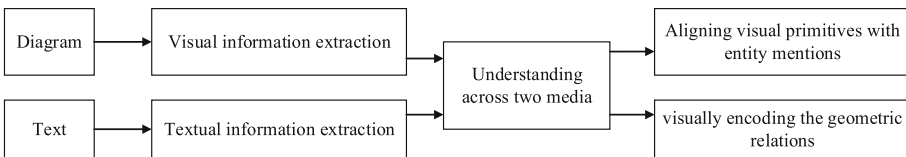
These atomic propositions are represented in the form of first order logic (FOL). Three kinds of atomic propositions exist in the geometry problems. They are unary, binary and ternary propositions, which contains different numbers of arguments (see Sect. 5.3). These propositions belong to two categories, namely position relation and quantity relation. For example, *parallel*( $AB$ ,  $CD$ ) is a position relation and *equalAngle* (*angle*( $ABC$ ),  $15^\circ$ ) is a quantity relation.

Given a geometry problem with text  $T$  and diagram  $D$ , the objective of understanding the problem is to extract the geometric propositions to represent the problem. It can be considered as two subtasks:

1. Extract a set of atomic propositions  $R_T = \{R_T^1, R_T^2, \dots, R_T^i\}$  from text  $T$ , and a set of atomic propositions  $R_D = \{R_D^1, R_D^2, \dots, R_D^j\}$  from diagram  $D$ .
2. Select a subset of atomic propositions from  $R_T$  and  $R_D$  to form a high-confidence relation set  $\hat{R} = \{R_1, R_2, \dots, R_l | R_l \in R_T \vee R_l \in R_D\}$  and ensure that the high-confidence relations in  $\hat{R}$  are compatible with both the text and the diagram.

To achieve these two subtasks, this paper presents a geometry problem understanding method taking the powerful paradigm of coupling the intake from both visual and textual information. Specifically, it consists of three steps, namely visual information extraction, textual information extraction, and the integration process to understand across two media, as shown in Fig. 2. To extract the visual information, geometric primitives are detected using computer vision technologies and geometric relations are mined by examining their corresponding algebraic relations using numeric verification. For extracting the textual information, a syntax-semantics ( $S^2$ ) model method is proposed to extract geometric relations from the text and form a set of atomic propositions. The integration process is used to fuse both visual and textual information and make mutual corroboration to obtain a set of high-confidence geometric relations which are both compatible with the text and the diagram.

To visually present the problem understanding result and provide the educational value for tutoring learners, we reactivate the visual primitives that have already been represented in the diagram and align them with the corresponding entity mentions in the text. In other words, to build the coreferences between the visual primitives and textual entities that refer to the same object. Moreover, the high-confidence geometric relations are also visually encoded into the diagram.



**Fig. 2.** The framework of the proposed method in understanding plane geometry problem.

## 4 Visual Information Extraction

This section is to present the extraction of visual information from the diagram. The visual information consists of geometric primitives and the various geometric relations among them. Extracting these information mainly consists of two procedures, namely geometric primitives detecting and geometric relation mining.

Given a diagram  $D$ , the geometric primitives detecting is to identify a set of primitives  $P = \{P_1, P_2, \dots, P_m\}$  from the diagram, and the procedure of geometric relation mining finds the geometric relations among the set  $P$  and obtain a set of atomic propositions  $R_D = \{R_D^1, R_D^2, \dots, R_D^j\}$ . The techniques of these two procedures are presented in the following subsections.

### 4.1 Geometric Primitives Detecting

Detecting geometric primitives from a diagram is a computer vision problem [8]. However, the extreme lack of both textural and color features in the diagram makes it different from the images typically studied in computer vision [7]. Hence the middle-level elements in the diagram such as the primitives turn out to play a significant role in extracting the visual information.

A geometric object recognition algorithm in [1] is adopted to detect the primitives. We promote the performance of this algorithm by using a hierarchical strategy, which first applies the connected component analysis method to segment the diagram into the body part and label part and then recognizes circles, recognizes lines, collects points of interest and recognizes labels from these two parts successively. The nearest neighbour principle is adopted to assign the recognized labels to the nearest geometric objects. Eventually a set  $C$  of circles, a set  $L$  of lines, a set  $I$  of points and a set  $B$  of labels contained in the diagram  $D$  can be obtained.

It is worth noting that the detected primitives are the basic primitives including points, lines, circles and labels, some geometric shapes such as triangles, parallelograms and trapezoids are not detected. The reason is that the complicated layout and the overlap of lines and circles in the diagram may produce many such geometric shapes and directly detecting them greatly deteriorates the diagram understanding efficiency. Moreover, some of them are not used in the problem. Hence we defer the detection of such shapes in the alignment process (see Sect. 7) by combining the entity mentions identified in the text and assembling basic primitives to form such geometric shapes.

The structure of a diagram is mainly depicted via the geometric relations among the primitives in the diagram. Hence the geometric relations indicated in the diagram should be mined based on the information of detected geometric primitives.

### 4.2 Geometric Relation Mining

Geometric relation mining plays an important role in understanding a diagram. By analyzing the geometry diagrams, eight basic geometric relations and four

**Table 1.** The basic geometric relations

Representation	Meaning	Representation	Meaning
pointOnLine(p, l)	point p lies on line l	equalDistance(AB,CD)	$\  AB \  = \  CD \ $
pointOnCircle(p, c)	point p lies on circle c	equalAngle(ABC,DEF)	$\angle ABC = \angle DEF$
parallel(l <sub>1</sub> , l <sub>2</sub> )	line l <sub>1</sub> is parallel to line l <sub>2</sub>	lcTangent(l, c)	line l is tangent to circle c
perpendicular(l <sub>1</sub> , l <sub>2</sub> )	line l <sub>1</sub> is perpendicular to line l <sub>2</sub>	ccTangent(c <sub>1</sub> , c <sub>2</sub> )	circle c <sub>1</sub> is tangent to circle c <sub>2</sub>

**Table 2.** The derived geometric relations

Relation representation	Meaning	Deriving rule
midPoint(C, AB)	point C is the midpoint of line AB	equalDistance(AC, BC) $\wedge$ pointOnLine(C,AB)
collinear(p <sub>1</sub> , p <sub>2</sub> , p <sub>3</sub> )	p <sub>1</sub> , p <sub>2</sub> , p <sub>3</sub> are colinear	pointOnLine(p <sub>1</sub> ,l) $\wedge$ pointOnLine(p <sub>2</sub> ,l) $\wedge$ pointOnLine(p <sub>3</sub> ,l)
intersect(p, l <sub>1</sub> , l <sub>2</sub> )	p is the intersection of line l <sub>1</sub> and line l <sub>2</sub>	pointOnLine(p, l <sub>1</sub> ) $\wedge$ pointOnLine(p, l <sub>2</sub> )
angleBisect(DB, ABC)	line DB bisect angle ABC	equalAngle(ABD,CBD)

derived geometric relations are proposed, as shown in Tables 1 and 2. The eight basic geometric relations can be used to describe most features of position and quantity of geometric primitives. The four derived geometric relations are derived from the basic geometric relations and they can describe higher level features of the diagram. All these geometric relations are represented as atomic propositions.

Given the set  $I$  of points, the set  $L$  of lines, and the set  $C$  of circles of a diagram with the set  $B$  of labels, geometric relation mining is to find a set  $R_D = \{R_D^1, R_D^2, \dots, R_D^j\}$  composed of basic geometric relations and the derived geometric relations. In general, a geometric relation can be certificated to be true if and only if its corresponding algebraic equality holds. Hence Numerical verification is used to examine the corresponding algebraic relations to obtain the basic geometric relations. For instance, to obtain the pointOnLine relations, we test each pair of point  $p \in I$  and line  $l \in L$  and calculate the *distance(p, l)*. If the value is less than a threshold<sup>1</sup>, then add the geometric relation *pointOnLine(p, l)* to the set  $R_D$ . By using this method, the eight basic geometric relations in Table 1 are obtained. After obtaining the basic geometric relations, we use the deriving rules shown in Table 2 to derive a series of relations and add them to the set  $R_D$ .

<sup>1</sup> Eight threshold parameters are used to specify tolerances to mine the eight basic relations. These values are acquired empirically by making experiments on a set of test diagrams with fixed size and then are automatically adjusted according to the size of given diagram.

## 5 Textual Information Extraction

This section is to present the extraction of textual information from the problem text. The textual information consists of entity mentions and the various geometric relations among them. The entity mentions are extracted using natural language analysis and then a syntax-semantics ( $S^2$ ) model method is proposed to extract geometric relations among these entity mentions from the text.

An algorithm is proposed using geometry  $S^2$  models to extract the geometric relations. It mainly consists of three steps, namely parsing and annotation, geometrical entity identification, and atomic proposition extraction, as depicted in Algorithm 1. The techniques of three steps of Algorithm 1 will be presented in the following subsections.

---

### Algorithm 1. Extracting textual information from problem text

---

**Input:** The text  $T$  of a plane geometry problem described in natural language.

**Output:** a set  $E$  of entity mentions and a set of atomic propositions  $R_T$ .

**Step 1:(Parsing and Annotation)** Uniform the problem text; Parse the problem text into phrases and annotate each phrases with POS;

**Step 2:(Element Identification)** Identify geometric relation words and entity mentions in each sentence, recognize the type of the extracted mentions and add the extracted entity mentions to  $E$ . All the results of extraction and recognition for a sentence form an annotation set  $F$ ;

**Step 3:(Relation extraction)** Use the  $S^2$  model to extract the relation of each sentence according to its annotation set  $F$  and form a set of atomic propositions  $R_T$ .

---

### 5.1 Parsing and Annotation

The goal of annotating the text of a plane geometry problem is to transform the problem text into a new form by doing parsing and annotation. ICTCLAS [15] is used to parse the text into phrases and to annotate these phrases with POS (part-of-speech) labels. A prepared geometric dictionary is used as the user dictionary of ICTCLAS to improve the accuracy of annotation.

### 5.2 Geometry Element Detection

Geometric relation words and entity mentions are important components of geometric relations. After annotation, these entity mentions are annotated with special categories of POS labels, which can be used to assist the extraction of entity mentions. A geometric entity representation is a duple  $e = (w, t)$  in which  $w$  is a phrase,  $t$  is the geometry type of  $w$ . Geometric relation words are extracted using keywords matching. A geometric relation representation is a duple  $J = (v, o)$  in which  $o$  is a representative relation word and  $v$  is the variant list of  $o$ . This paper have identified 48 kinds of geometric relations widely used in plane geometry problems.



### 5.3 Atomic Geometry Relation Extraction

Atomic geometry relation extraction is the key step of extracting textual information from problem text in Algorithm 1. To understand the techniques of this step, the preparation of list of atomic propositions and relation extraction procedure are presented, respectively.

**Preparation of List of atomic propositions:** The geometry relations in plane geometry can be divided into three types of unary, binary and ternary relations. Table 3 gives the examples of these three types of geometry relations. Each such relation corresponds to an atomic proposition so that there are 48 atomic propositions. Atomic propositions can be written in the form of first order predicate logic, abbreviated as FOL.

**Table 3.** Explanation of three types of geometry relations.

	Element representation	Semantics	FOL	#
Unary	(equilateral triangle, ABC, triangle)	ABC is an equilateral triangle	eqTriangle(ABC)	17
Binary	(parallel, AB, CD, line, line)	line AB is parallel to CD	parallel(AB, CD)	22
Ternary	(intersects, AB, CD, E, line, line, point)	line AB intersects CD at point E	Intersect(E,AB,CD)	9

**Definition 4:** An  $S^2$  model for plane geometry problems is defined as a triple  $N = (J, E, F)$ , where  $J$  represents geometric relation representation,  $E = \{e_1, e_2, e_3\}$  is the set of the involved elements, and  $F$  is the atomic propositions in FOL. Let  $\Pi = \{N_i = (J_i, E_i, F_i) | i = 1, 2, \dots, n\}$  denote all the prepared  $S^2$  models. It is also called as a pool of  $S^2$  models of plane geometry.

The pool of 48  $S^2$  models are used to extract all the atomic relations in the problem text as described in Procedure I.

**Procedure I: Extraction of geometry relations using the  $S^2$  models**

The input of this procedure is a set of simple sentences of text  $T$ . Each sentence  $S$  is annotated with its geometric entity mentions  $\mu$  and geometry relation representation  $\nu$ . The output is the contained atomic propositions in each sentence, denoted as  $R_T$ .

Load  $S^2$  models  $\Pi = \{N_i = (J_i, E_i, F_i) | i = 1, 2, \dots, n\}$ ;

Initialize  $R_T$  as empty;

While TRUE

    Pick a simple sentence from the sentence set

    For  $i$  from 1 to  $n$  do

        If matching  $J_i$  with  $\nu$  is FALSE continue;

        If matching  $E_i$  with  $\mu$  is FALSE continue;

        Put the instantiated  $F_i$  of  $N_i$  into  $\Delta$ ;

    If all sentences are processed break While loop;

The  $S^2$  model matching method can generate quite high-confidence geometric relations from the text. However, for some complex sentences containing many

geometric relations, the relations extracted may not fully reliable. Considering the sentence “AD and BC are produced to meet MN at E and F respectively”. Here, “AD, BC” and “E, F” are coordinate structures, and the intersect relation is indicated. It is difficult to directly use the Procedure I to obtain the right geometric relations because of the over-numbered geometric elements. Hence, for such cases, we over-generate the geometric relations to obtain all the possible ones from the sentence and defer the validation in the integration process.

## 6 Integration Process

This section presents the integration process of visual and textual information. Since the imprecision of diagram and the diverse statement of problem text, the intermediate results of visual and textual information are not fully reliable. Hence it is necessary to integrate both visual and textual information and make mutual corroboration to obtain a set of high-confidence geometric relations.

Given the textual relation set  $R_T$  and the visual relation set  $R_D$ , the integration process is to find a high-confidence relation set

$$\hat{R} = \{R_1, R_2, \dots, R_i\}, \text{ where } R_i \in R_D \bigvee R_i \in R_T. \tag{1}$$

Generally, the diagram often contains some important geometric relations that are not presented in the text. We call these relations as high-confidence visual relations. Hence, the visual relation set  $R_D$  is divided into two sets, namely the high-confidence visual relation set  $R_\Delta$  and the general visual relation set  $R_d$ .

$$R_D = R_\Delta \bigcup R_d \tag{2}$$

By analyzing the relations commonly appeared in the diagram,

$$R_\Delta = \{pointOnLine, pointOnCircle, collinear, intersect, equalDistance(line, number), equalAngle(angle, number)\}.$$

For example, the relation  $EqualDistance(OD, 5)$  in  $R_D$  in Fig. 1 represents the equal relation between a line and a number label. Such relations are confidently extracted from the diagram and form a high-confidence visual relation set  $R_\Delta$ . All the relations in  $R_\Delta$  should be added into  $\hat{R}$ .

Moreover, by using the visual information is not able to check the correctness of some geometric relations extracted from the text. Since the scale in the diagram are different from the text, the corresponding relations cannot be obtained from the diagram. We call these relations as high-confidence textual relations. Hence, the textual relation set  $R_T$  is divided into two sets, namely the high-confidence textual relation set  $R_\Omega$  and the general textual relation set  $R_t$ .

$$R_T = R_\Omega \bigcup R_t \tag{3}$$

By analyzing the relations commonly appeared in the text,

$$R_\Omega = \{congruentTriangle, similarTriangle, equalDistance(line, number), equalAngle(angle, number)\}.$$

These relations are correct to a large extent, therefore we directly add them into  $\hat{R}$ .

Therefore, the integration process in Eq. (1) is equal to find the set

$$\hat{R} = \{R_1, R_2, \dots, R_i\} \cup R_\Delta \cup R_\Omega, \text{ where } R_i \in R_d \cap R_t, R_\Delta \subset R_D, R_\Omega \subset R_T. \tag{4}$$

For each general textual relation  $R_j \in R_t$  from the text, we check whether it is also in the general visual relation set  $R_d$ . If it satisfies the text and the diagram simultaneously, we add it into  $\hat{R}$ , otherwise it is regarded as incorrect relation and is discarded.

Based on the above discussion, the procedure of integrating the relations extracted from diagram and text is described in Procedure II.

**Procedure II: Integration of relations extracted from diagram and text**

The input of this procedure is a set  $R_T$  of textual relations and a set  $R_D$  of visual relations. The output is a set  $\hat{R}$  of high-confidence relations.

- (i)  $R_\Omega, R_t \leftarrow R_T$ : divide the textual relation set  $R_T$  into two sets  $R_\Omega, R_t$ ;
- (ii)  $R_\Delta, R_d \leftarrow R_D$ : divide the visual relation set  $R_D$  into two sets  $R_\Delta, R_d$ ;
- (iii) Initialize  $\hat{R}$  as empty;

$$\hat{R} = \hat{R} \cup R_\Delta \cup R_\Omega;$$

For each  $R_j \in R_t$

If  $R_j \in R_d$  then  $\hat{R} = \hat{R} \cup R_j$ ;

return  $\hat{R}$ .

## 7 Alignment and Visually Presentation

To interactively present the problem understanding result, this section presents the alignment of geometric entities and geometric primitives and visually encoded the high-confidence geometric relations into the diagram.

The relation set  $\hat{R}$  (obtained in the Sect. 6) contains all the geometric entities occurred in the problem and their geometric relations. Hence we extract all the entities in  $\hat{R}$  and form an entity set E without repeated elements. The geometric primitives (detected in Sect. 4.1) form a set F<sup>2</sup>. A matrix  $W \in \{0, 1\}^{|E| \times |F|}$  is built to record the alignment.  $W_{i,j}$  identifies whether the  $i_{th}$  geometric entities  $E_i$  is aligned with the  $j_{th}$  geometric primitive  $F_j$ . This alignment is built by mapping the name of the geometric entities with the label of the corresponding primitives in the diagram. For example, the entity  $OB$  is mapped with a line  $l_3 := line(O, B)$  in Fig. 1.

<sup>2</sup> It is worth noting that the geometric primitives detected in Sect. 4.1 include the points, lines and circles and labels without some complex geometric shapes such as triangles, squares, parallelograms and trapezoids. To tackle this situation, we detect them only when they occurs in the entity set E, and use bottom-up method to combine the basic primitives into the geometric shapes. For example,  $\triangle ABC$  is detected by assembling three lines  $AB, AC$  and  $BC$ . All these detected shapes are also added into the set F.

For tutoring purpose in helping learners understand the problem, we also visually encode the geometric relations into the diagram. When the problem text goes on with mouse clicks, the related geometry elements and their relations in the diagram are highlighted and animated with various visually dynamic effects. This makes the understanding of diagram more vividly visualized and intuitive.

## 8 Experimental Results

This section is to evaluate the proposed method on understanding plane geometry problems. It first describes the setting for the experiments. Then it presents the results of the proposed method in mining geometric relations. To also better understand the performance on helping learners in geometry problem understanding, a user study is conducted.

### 8.1 Experimental Setup

**Dataset:** The datasets consist of the dataset used in [11] named as GeoE100, which contains the 100 plane geometry problems in English and the dataset prepared in this paper named as GeoC50, which contains 50 plane geometry problems in Chinese. These 50 problems are compiled from the test datasets used in [3, 5]. Every question has a textual description accompanied by a diagram. We manually annotate all the primitives in the diagram and the entity mentions in the text and build all the alignments between them. Moreover, for each problem we manually understand the problem and prepare a set of geometric relations as its groundtruth, which are required for finding the solutions. Table 4 gives the statistics of the problems and the groundtruth of GeoE100 and GeoC50. In GeoC50, the problem texts are much longer and the diagrams are more complicated than the problems in GeoE100.

**Table 4.** Statistics on the problems and the groundtruth of GeoE100 and GeoC50.

		Primitives <sup>a</sup>	Entity	V-relations	T-relations	VT-relations
GeoE100	Total	1015	482	821	436	648
	Average	10.2	4.8	8.2	4.4	6.5
GeoC50	Total	707	617	896	329	615
	Average	14.1	12.3	17.9	6.6	12.3

<sup>a</sup>The primitives here include the points, lines, circles and labels

**Evaluation measure:** Two tasks are evaluated in this experiment. Firstly, mining the geometric relations by integrating the textual and visual information. Secondly, testing the usability of the proposed method.

For the first task, we compare the mined relations with groundtruth relations by measuring them using precision, recall and  $F_1$ . For the second task, a user study is conducted to test the usability of the method in helping people understand geometry problems and to obtain feedbacks.

## 8.2 Results

**Mining geometric relations.** To study the performance of mining geometric relations by integrating textual and visual information, we compare the relation extracting results on all the test problems in GeoE100 and GeoC50 with ground truth relations. As shown in Table 5, precision is the number of correctly extracted relations divided by total number of extracted relations, recall is the number of correctly extracted relations divided by the number of relations in ground truth. The visual relations (V-relation) extracted from the diagram achieves 0.90 in  $F_1$  score. The entity mentions identified from the text obtain 100% in precision and recall, and the textual relation (T-relation) mined from the text achieves 0.92 in  $F_1$  score. This validates that the  $S^2$  models can extract geometric relations from problem texts both in Chinese and English. By integrating the textual and visual information (D-T integration), it achieves precision of 0.94 at the recall of 0.97 and 0.95 in  $F_1$  score in finding the high-confidence relations (HC-relation). These results show the effectiveness of integrating both textual and visual information to understand geometry problems. And they also show that the proposed method can understand problems that cannot be understood by using text or diagram only.

**Table 5.** The performance of mining geometric relations on the test problems.

		P	R	$F_1$
Diagram	V-relation	0.86	0.95	0.90
Text	Entity	1	1	1
	T-relation	0.89	0.96	0.92
D-T integration	HC-relation	0.94	0.97	0.95

**User study.** We built a learner-initiating interactive geometry system named LI-Geo (Fig. 3). A user study is conducted to test the usability of the system in helping people understand geometry problems. In LI-Geo, there are three separate areas for problem text presentation, diagram showing and geometric relation display. When click on the content in any of the three areas, the corresponding content in the other two will be activated and the dynamic visual effect will be presented in the diagram.

The test task was to understand the geometry problems provided by us and did not require any knowledge beyond senior school, so we recruited 12 graduate students who possessed the required geometry knowledge. we provided each subject 10 plane geometry problems with diagrams, and each subject was asked to select and understand 4 problems in LI-Geo. After trying the LI-Geo system, each subject was asked to answer a post-test questionnaire to grade the primitive detection accuracy, entity extraction accuracy, high-confidence relation extraction accuracy, the visual presentation, comfort of interaction, helpfulness in understanding problems, enjoyment of the tool, all in a 7-level Likert scale (1-very bad, 7-very good).

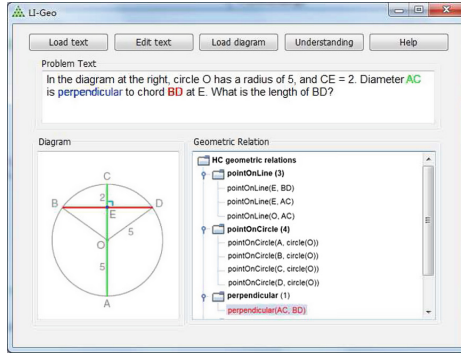


Fig. 3. The user interface of LI-Geo.

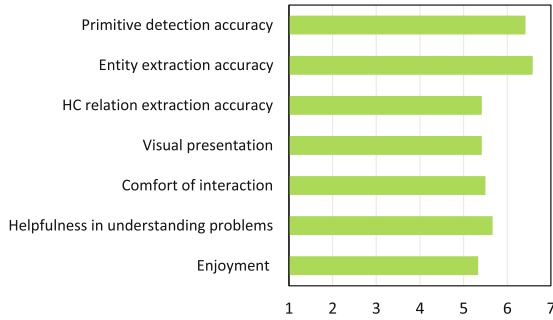


Fig. 4. Users' feedback of using LI-Geo.

Figure 4 exhibits the results of users' feedback. As shown, the primitive detection accuracy, entity extraction accuracy, high-confidence relation extraction accuracy and the visual presentation are all received good feedback from subjects. In addition, the subjects think it is comfortable and enjoyment to use the system and the system helps users in understanding geometry problems.

**Discussion.** The user study highlights the usefulness of the learner-initiating problem understanding tool. By using the interaction, the given and the goal of a problem and the geometric relations between the primitives (entities) will be better understood. By analyzing the geometric relations obtained by the system, an interesting finding is that some critical information for solving the problem but is not present in the text is obtained. For example, to solve the problem in Fig. 1, one has to know that  $pointOnLine(O, AC)$ ,  $collinear(O, E, C)$  and  $equalDistance(OC, 5)$ . Hence the proposed problem understanding method by integrating textual and visual information will facilitate the automated solving of problems. However, this research is ongoing, and how to use the problem understanding method in the task of geometry problem solving and how to use it to tutor learners are our future work.

## 9 Conclusion

This paper has presented a method for understanding plane geometry problem by integrating the information separately extracted from text and diagram. This paper has four contributions in techniques. First, it developed a method to extract relations from diagram. This method uses numerical verification to mine geometric relations after detecting the visual primitives. Second, it proposed an  $S^2$  model method to extract relations from the problem text. Third, it proposed a new method for understanding geometry problems by integrating textual and visual information. This method can understand a batch of plane geometry problems that cannot be fully understood from text only or from diagram only. Fourth, it developed a procedure to encode the extracted relations into the corresponding positions in the diagram. This procedure makes the understanding of plane geometry problems visualized and intuitive. The experimental results showed that the proposed method had a good performance. This work validates that coupling vision and NLP to process multi-model information helps improve textual or visual interpretations.

In the future, we want to extend the research in multiple directions. First, it is the good future job to develop the improved automatic solvers based on the method of problem understanding. Second, we plan to extend the method to understand geometry problems with hand-drawn input geometry diagrams. Third, we plan to apply the method that couple supplementary explanation extracted from multi-modality into understanding the problems from other subjects.

**Acknowledgments.** This work is supported by the Open Projects Program of National Laboratory of Pattern Recognition (No. 201600008) and the self-determined research funds of CCNU from the colleges' basic research and operation of MOE (No. ccnu16JYKX005).

## References

1. Chen, X., Song, D., Wang, D.: Automated generation of geometric theorems from images of diagrams. *Ann. Math. Artif. Intell.* **74**(3–4), 333–358 (2015)
2. Chou, S.C., Gao, X.S., Zhang, J.Z.: *Machine Proofs in Geometry: Automated Production of Readable Proofs for Geometry Theorems*. World Scientific, Singapore (1994)
3. Ge, Q.: Research of automated geometry reasoning and its application with constrained conditions. Ph.D. thesis, Central China Normal University (2011)
4. Guo, H.Y., Liu, Q.T., Chen, M., Huang, H., Ge, Q.: Research for facing the natural language of the geometry drawing. *Comput. Sci.* **39**(6A), 503–506 (2012)
5. Jiang, J.: iGeo: a theorem prover of the intelligent geometry software. Ph.D. thesis, Chengdu Computer Application Research Institute of Chinese Academy of Sciences (2006)
6. Jiang, J., Zhang, J.: A review and prospect of readable machine proofs for geometry theorems. *J. Syst. Sci. Complex.* **25**(4), 802–820 (2012)

7. Liu, L., Lu, X., Fu, S., Qu, J., Gao, L., Tang, Z.: Plane geometry figure retrieval based on bilayer geometric attributed graph matching. In: ICPR, pp. 309–314. IEEE (2014)
8. Montalvo, F.: Diagram understanding: the intersection of computer vision and graphics. Massachusetts Institute of Technology (1985)
9. Mukherjee, A., Garain, U.: Understanding of natural language text for diagram drawing. In: 13th International Conference on Artificial Intelligence and Soft Computing (2009)
10. Nakamura, Y., Furukawa, R., Nagao, M.: Diagram understanding utilizing natural language text. In: ICDAR, pp. 614–618. IEEE (1993)
11. Seo, M.J., Hajishirzi, H., Farhadi, A., Etzioni, O.: Diagram understanding in geometry questions. In: 28th AAAI Conference on Artificial Intelligence, pp. 2831–2838 (2014)
12. Seo, M.J., Hajishirzi, H., Farhadi, A., Etzioni, O., Malcolm, C.: Solving geometry problems: combining text and diagram interpretation. In: EMNLP, pp. 1466–1476 (2015)
13. Srihari, R.K.: Computational models for integrating linguistic and visual information: a survey. *Artif. Intell. Rev.* **8**(5-6), 349–369 (1994)
14. Wong, W.K., Huang, C.W., et al.: A computer-assisted environment for understanding geometry theorem proving problems and making conjectures. *Int. J. Intell. Inf. Database Syst.* **3**(3), 231–245 (2009)
15. Zhang, H.P., Liu, Q.: ICTCLAS. Institute of Computing Technology, Chinese Academy of Sciences (2002). [http://www.ict.ac.cn/freeware/003\\_ictclas.asp](http://www.ict.ac.cn/freeware/003_ictclas.asp)
16. Zhang, X., Fu, H.: Recognizing and understanding of plane geometry. *J. Comput. Appl.* **35**, 280–283, 341 (2015)