GeoRE: A Relation Extraction Dataset for Chinese Geometry Problems

Wei Yu, Mengzhu Wang, Xiaodong Wang
College of Computer
National University of Defense Technology

Xun Zhou, Yongfu Zha, Yongjian Zhang
College of Computer and Imformation Science
Chongqing Normal University

Shuyu Miao
Ant Group
`miaoshuyu.msy@antgroup.com`

Jingdong Liu
School of Software Engineering
Chongqing University of Posts and Telecommunication

Abstract

Relation extraction is an important foundation for many natural language understanding applications, as well as geometry problem solving. In this paper, we present GeoRE, a relation extraction dataset for Chinese geometry problems. To the best of our knowledge, GeoRE is the first Chinese relation extraction dataset about geometry problems. It consists of 12,901 geometry problems on 43 shapes, covering 19 positional relations and 4 quantitative relations. We experiment with various state-of-the-art (SOTA) models and the best model achieves only 70.3% F1 value on GeoRE. This shows that GeoRE presents a challenge for future research.

1 Introduction

Geometry is a subject that studies the structure and properties of space. It is the basic content of mathematics. In recent years, geometry problem solving has been gaining more attention in NLP and AI-EDU community [17, 16, 9, 15]. Solving geometry problems is a compulsory course for cultivating students' abstract thinking and spatial perception in high school education. According to the natural language description of the given problem and the corresponding diagram, it needs to identify geometric relations, apply theorem knowledge, and perform algebraic calculations to derive the numerical value of the answer. The above process is extremely complicated and the step of identifying and extracting geometric relations is foundation. Once the geometric relations in the geometry problem cannot be accurately identified, the geometric problem cannot be solved correctly. Although some well-labeled datasets [12, 1, 17, 16] for solving geometric problems have been proposed, the corpus size is very small and these datasets are all in English. To this end, a new large-scale Chinese relation extraction dataset named GeoRE that aims to extract geometric relations from the given geometry problem text, is proposed in this paper.

As Figure 1 illustrates, given a geometry problem, our corpus annotates shapes, positions, values, and relations. A geometric relation triplet consists of two or three entities and a relation between them. These triplets act as an important role of human knowledge and explicitly or implicitly hidden in the plain text. Extracting these geometric relational facts could benefit downstream applications, e.g, geometry problem solving and geometry diagram drawing.

Different from the traditional general relation extraction (RE) dataset, GeoRE is a relation extraction corpus focusing on specific domains. It consists of 12,901 geometry problems on 43 shapes, covering 19 positional relations and 4 quantitative relations, all labeled by 6 college students spending a total of 850 man-hours. GeoRE has significant geometric domain characteristics:

35th Conference on Neural Information Processing Systems (NeurIPS 2021) Workshop on Math AI for Education (MATHAI4ED).

Annotators check geometric question：

在等腰直角三角形DBC中，角BDC等于90度，BF平分角DBC，与CD相交于点F，延长BD到A，使DA等于DF。求证：三角形FBD与三角形ACD全等。
In the isosceles right triangle DBC, the angle BDC is equal to 90 degrees. BF bisects the angle DBC and intersects the CD at point F. BD is extended to A making DA is equal to DF. Proof: Triangle FBD and triangle ACD are congruent.

Annotators create：

形状(shape)	等腰直角三角形 (Isosceles Right Triangle)	DBC
	角(Angle)	BDC, DBC
	线段(Line Segment)	BF, CD, BD, DA, DF
	点(Point)	F, A
位置(Position)	交(Intersect), 平分(Bisect)	
数值(Value)	90	
数量关系 (Quantitative Relation)	等于(Equal): (BDC, 90)，(DA, DF) 全等(Congruent): (FBD, ACD)	
位置关系 (Positional Relation)	交(Intersect): (BF, CD, F) 平分(Bisect): (BF, DBC)	

The quantitative and positional relations of shapes

Figure 1: Our corpus annotates shapes, positions, values, and relations according to the given geometry problems. The example contains multiple shapes and relations.

i. The shapes in the geometry problems are represented by a sequence of letters. It can easily cause ambiguity in reference. For example, the *DBC* in Figure 1 can refer to either an isosceles right triangle *DBC* or an angle *DBC*.

ii. A lot of geometric symbols in the geometry problems, such as $\perp, \|, \angle, \frown, \odot, \equiv, \cong, \triangle$, etc.

iii. There are many entities and relationships in each geometry problems, and the same entity has different relations with multiple entities.

To assess the task difficulty, we experiment with several SOTA models and the best model only achieves a 70.3% F1 value. This suggests that general RE models still have a large room for improvement on the geometric relation extraction dataset.

2 Corpus Construction

All shapes, positions, values and relations are labeled and reviewed by 6 Chinese college students. As shown in Figure 2, we develop our dataset in three steps, spending a total of 850 hours of human labor: section 2.1 Dataset Collection and Cleaning, section 2.2 Semi-automatic Generation of Entity Relation, section 2.3 Review and Checking.

2.1 Dataset Collection and Cleaning

The corpus about News and encyclopedias can be easily collected from the website of Wikipedia and People's Daily. However, it is difficult to collect enough data from a website for the geometry problems.

To tackle this issue, we firstly search on the Internet with the keyword-几何题库 (geometry problems bank). Then, we manually filter out the list of websites to be crawled from the returned results. Finally, we use the urllib library to request HTML and Beautiful Soup 4 [1] to parse the HTML node to get geometry problems.

The crawled geometry problems are messy and need to be cleaned up. We unify the format and letter case, delete redundant blank characters, and complete the incomplete question.

[1] https://www.crummy.com/software/BeautifulSoup/

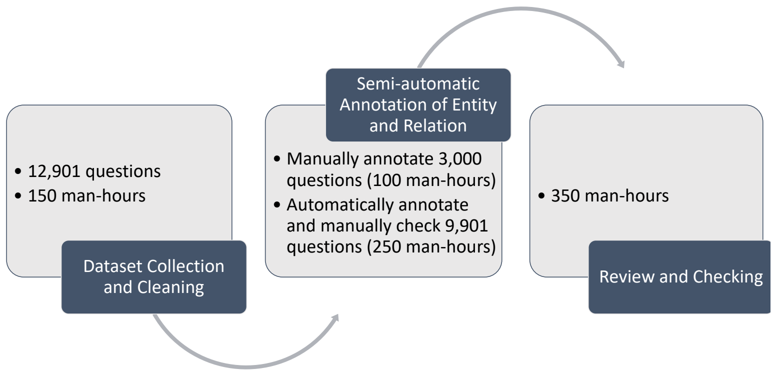

Figure 2: The annotation process of our GeoRE corpus.

2.2 Semi-automatic Annotation of Entity and Relation

For cleaned 12,901 geometry problems, we ask six college students to annotate shapes, positions, values and relations. Since the shapes and relations of geometry problems are limited, we use a semi-automatic generation method to annotate entity and relation. First, each student manually annotates 500 questions to find out what shapes, positions, values, and relations are roughly present in the dataset. Then, we write regular expressions according to the situation to roughly annotate the question automatically. Finally, the automatically annotated 9,901 questions are assigned to the annotator for correction.

2.3 Review and Checking

Once the geometry problems is labeled, we ask a different annotator to review and check if the shapes and relations are right. For a geometry problems with multiple possible annotations, the reviewers double check whether the annotation is correctly labeled under our protocol.

2.4 Dataset Statistics

Our corpus has 12,901 geometry problems. Each question has a maximum of 456 characters and a minimum of 10 characters. The average length of the question is 57. Different numbers of questions are distributed in the interval from 10 to 456 and a large number of questions with a length of 10-100. For the distribution of question length, see Appendix A. GeoRE contains realistic data extracted from the web and covers most of the shapes and relations in elementary geometry. Line Segment, Intersect, and Equal are the most frequent items, and the amount of shapes and relations is uneven. A large number of entity-relation and such uneven distribution make it more difficult to extract relations. For a list of shapes and relations, see Appendix B.

3 Experiments

3.1 Benchmark Approaches

In order to analyze the difficulty of our corpus, we experiment with several state-of-the-art relation extraction models.

- **NovelTagging** [26] proposes a novel tagging scheme that incorporates both relation roles and entity types and converts the joint extraction to a sequence tagging problems.
- **CopyMTL** [23] generates the triplets sequentially using copy mechanism, and applies multi-task learning to solve the problems of generating multi-token entities.
- **ETL-Span** [22] applies a novel decomposition strategy, which first distinguishes all head entities, and then identifies corresponding tail entities and relations.

Table 1: Main results of the relation extraction models on NYT24 and GeoRE.

Model	NYT24			GeoRE		
	P	R	F1	P	R	F1
NovelTagging[26]	0.621	0.375	0.422	0.568	0.197	0.292
CopyMTL[23]	0.757	0.687	0.720	0.593	0.556	0.568
ETL-Span[22]	0.841	0.746	0.791	0.701	0.706	0.703

3.2 Experimental Results

We take standard Precision (P), Recall (R) and F1 score as our evaluation metrics. A triplet is considered to be correctly extracted if and only if its relation type and entities are exactly matched. We evaluate the performance of several models on our test set. These models are SOTA on the New York Times (NYT24) dataset. For the final training dataset, we randomly split the examples into 8,000 train, 2000 dev, 2901 test.

Table 1 presents the experimental results. A closer inspection of the table shows the performance of models that performed well on the NYT24 dataset significantly are reduced on our test set. What stands out is that the F1 value of CopyMTL drops by nearly 16%. It indicates that our corpus is challenging. For the task of extracting geometric relations, the SOTA models still have a large room for improvement.

4 Related Work

Table 2: Comparison of relation extraction datasets.

Dataset Name	Level	Relations	Train	Test	Manual Annotation
SemEval 2010 Task 8 [6]	sentence	18	8,000	2,717	Yes
NYT10 [13]	sentence	52	455,412	172,415	No
NYT11 [8]	sentence	24	335,843	1,450	Test
NYT24 [24]	sentence	24	56,196	5,000	No
NYT29 [18]	sentence	29	63,306	4,006	No
WebNLG [24]	sentence	216	5,519	703	Yes
ACE05 [19]	sentence	7	9,038	1,535	Yes
CoNLL04 [14]	sentence	5	1,153	288	Yes
GDS [11]	sentence	4	13,161	5,663	Yes
TACRED [25]	sentence	41	90,755	15,509	Yes
FewRel 2.0 [4]	sentence	100	56,000	14,000	Yes
WikiReading [7]	document	884	14.85M	3.73M	No
DocRED [21]	document	96	4,053	1,000	Yes
GeoRE	sentence	23	10,000	2,901	Yes

Table 3: Compare with the geometric problem solving dataset.

Dataset	question	word	shape	grade	operator type
GeoShader [1]	102	/	4	6-10	$\{+, -, \times, \div, \square^2, \sqrt{\square}\}$
GEOS [17]	186	4343	4	6-10	$\{+, -, \times, \div, \square^2, \sqrt{\square}\}$
Geometry 3K [12]	3002	36736	6	6-12	$\{+, -, \times, \div, \square^2, \sqrt{\square}, sin, cos, tan\}$
GeoRE(ours)	12901	735357	43	6-12	/

Relation extraction is a basic task of natural language processing that attracts much attention. In recent years, relation extraction datasets have been released continuously. We present the basic information of the currently widely used relation extraction datasets in Table 2. Compared with those datasets, our GeoRE is a medium-scale sentence-level manually annotated dataset. Different from the traditional Wikipedia and News corpus, our corpus is domain-specific (geometry problems) and aims to extract positional and quantitative relations of geometric shapes in the text. In addition, we compare with the dataset for solving geometric problems in Table 3. It can be seen that our corpus has the largest number of questions and geometric shapes.

5 Conclusion

In this paper, we introduce GeoRE, a large complex and domain-specific relation extraction dataset, which directly benefits both NLP and AI-EDU communities. Experimental results on several state-of-the-art models on GeoRE suggest plenty of space for improvement.

References

[1] Chris Alvin, Sumit Gulwani, Rupak Majumdar, and Supratik Mukhopadhyay. Synthesis of solutions for shaded area geometry problems. In Vasile Rus and Zdravko Markov, editors, *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2017, Marco Island, Florida, USA, May 22-24, 2017*, pages 14–19. AAAI Press, 2017.

[2] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795, 2013.

[3] Li Dong, Furu Wei, Ming Zhou, and Ke Xu. Question answering over freebase with multi-column convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 260–269. The Association for Computer Linguistics, 2015.

[4] Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. FewRel 2.0: Towards more challenging few-shot relation classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6250–6255, Hong Kong, China, November 2019. Association for Computational Linguistics.

[5] Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. Creating training corpora for NLG micro-planners. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 179–188. Association for Computational Linguistics, 2017.

[6] Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval@ACL 2010, Uppsala University, Uppsala, Sweden, July 15-16, 2010*, pages 33–38. The Association for Computer Linguistics, 2010.

[7] Daniel Hewlett, Alexandre Lacoste, Llion Jones, Illia Polosukhin, Andrew Fandrianto, Jay Han, Matthew Kelcey, and David Berthelot. Wikireading: A novel large-scale language understanding task over wikipedia. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics, 2016.

[8] Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke S. Zettlemoyer, and Daniel S. Weld. Knowledge-based weak supervision for information extraction of overlapping relations. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pages 541–550. The Association for Computer Linguistics, 2011.

[9] Mark Hopkins, Ronan Le Bras, Cristian Petrescu-Prahova, Gabriel Stanovsky, Hannaneh Hajishirzi, and Rik Koncel-Kedziorski. Semeval-2019 task 10: Math question answering. In Jonathan May, Ekaterina Shutova, Aurélie Herbelot, Xiaodan Zhu, Marianna Apidianaki, and Saif M. Mohammad, editors, *Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2019, Minneapolis, MN, USA, June 6-7, 2019*, pages 893–899. Association for Computational Linguistics, 2019.

[10] Luyang Huang, Lingfei Wu, and Lu Wang. Knowledge graph-augmented abstractive summarization with semantic-driven cloze reward. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5094–5107, Online, July 2020. Association for Computational Linguistics.

[11] Sharmistha Jat, Siddhesh Khandelwal, and Partha P. Talukdar. Improving distantly supervised relation extraction using word and entity based attention. In *6th Workshop on Automated Knowledge Base Construction, AKBC@NIPS 2017, Long Beach, California, USA, December 8, 2017*. OpenReview.net, 2017.

[12] Pan Lu, Ran Gong, Shibiao Jiang, Liang Qiu, Siyuan Huang, Xiaodan Liang, and Song-Chun Zhu. Inter-gps: Interpretable geometry problem solving with formal language and symbolic reasoning. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 6774–6786. Association for Computational Linguistics, 2021.

[13] Sebastian Riedel, Limin Yao, and Andrew McCallum. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part III*, volume 6323 of *Lecture Notes in Computer Science*, pages 148–163. Springer, 2010.

[14] Dan Roth and Wen-tau Yih. A linear programming formulation for global inference in natural language tasks. In *Proceedings of the Eighth Conference on Computational Natural Language Learning, CoNLL 2004, Held in cooperation with HLT-NAACL 2004, Boston, Massachusetts, USA, May 6-7, 2004*, pages 1–8. ACL, 2004.

[15] Mrinmaya Sachan, Avinava Dubey, Eduard H. Hovy, Tom M. Mitchell, Dan Roth, and Eric P. Xing. Discourse in multimedia: A case study in extracting geometry knowledge from textbooks. *Comput. Linguistics*, 45(4):627–665, 2019.

[16] Mrinmaya Sachan and Eric P. Xing. Learning to solve geometry problems from natural language demonstrations in textbooks. In Nancy Ide, Aurélie Herbelot, and Lluís Màrquez, editors, *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics, *SEM @ACM 2017, Vancouver, Canada, August 3-4, 2017*, pages 251–261. Association for Computational Linguistics, 2017.

[17] Min Joon Seo, Hannaneh Hajishirzi, Ali Farhadi, Oren Etzioni, and Clint Malcolm. Solving geometry problems: Combining text and diagram interpretation. In Lluís Màrquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 1466–1476. The Association for Computational Linguistics, 2015.

[18] Ryuichi Takanobu, Tianyang Zhang, Jiexi Liu, and Minlie Huang. A hierarchical framework for relation extraction with reinforcement learning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 7072–7079. AAAI Press, 2019.

[19] Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. Ace 2005 multilingual training corpus. 2006.

[20] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 1112–1119. AAAI Press, 2014.

[21] Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. Docred: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 764–777. Association for Computational Linguistics, 2019.

[22] Bowen Yu, Zhenyu Zhang, Xiaobo Shu, Tingwen Liu, Yubin Wang, Bin Wang, and Sujian Li. Joint extraction of entities and relations based on a novel decomposition strategy. In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 2282–2289. IOS Press, 2020.

[23] Daojian Zeng, Haoran Zhang, and Qianying Liu. Copymtl: Copy mechanism for joint extraction of entities and relations with multi-task learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9507–9514. AAAI Press, 2020.

[24] Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 506–514. Association for Computational Linguistics, 2018.

[25] Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 35–45. Association for Computational Linguistics, 2017.

[26] Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1227–1236. Association for Computational Linguistics, 2017.

A Distribution of question length

Figure 3 is the distribution diagram of the question length in GeoRE. Part B is the complete illustration and Part A is an enlarged view of B from 0 to 25. In general, the question lengths in the GeoRE dataset range from 10 (minimum) to 456 (maximum). Among them, the number of questions in the range of 10 to 100 accounted for 96% of the total, and the number of samples with a length of 59 are the most.

B Distribution of shapes and relations

Table 4 is a list of shapes and relations in the GeoRE dataset, including 43 shapes, 19 positional relations, and 4 quantitative relations. In order to better analyze the distribution, we make statistics on all shapes and relations. The distributions of shapes, positional and quantitative relations, respectively shown in Figure 4-6.

Figure 4 shows the distribution of shapes in the GeoRE dataset. It can be found that Line Segment, Point, and Triangle are the most numerous shapes. There are also a certain number of common shapes such as Angle, Parallelogram, and Rectangle. Figure 5 is the distribution of positional relations. It can be seen that Intersect, Perpendicular, and Parallel are the most numerous positional relations. A certain number of common positional relations such as On, Midpoint, and Bisect are also shown in the figure. The distribution of quantitative relations in the GeoRE dataset is shown in Figure 6. Equal is the most quantitative relations in the figure. The remaining three quantitative relationships are relatively small. Through the above statistics on shapes and relationships, we can find that the GeoRE dataset covers a variety of geometric shapes and relations. The proposal of GeoRE not only enriches the relation extraction community, but also proposes a new benchmark for the identification of geometric relations for solving geometry problems.

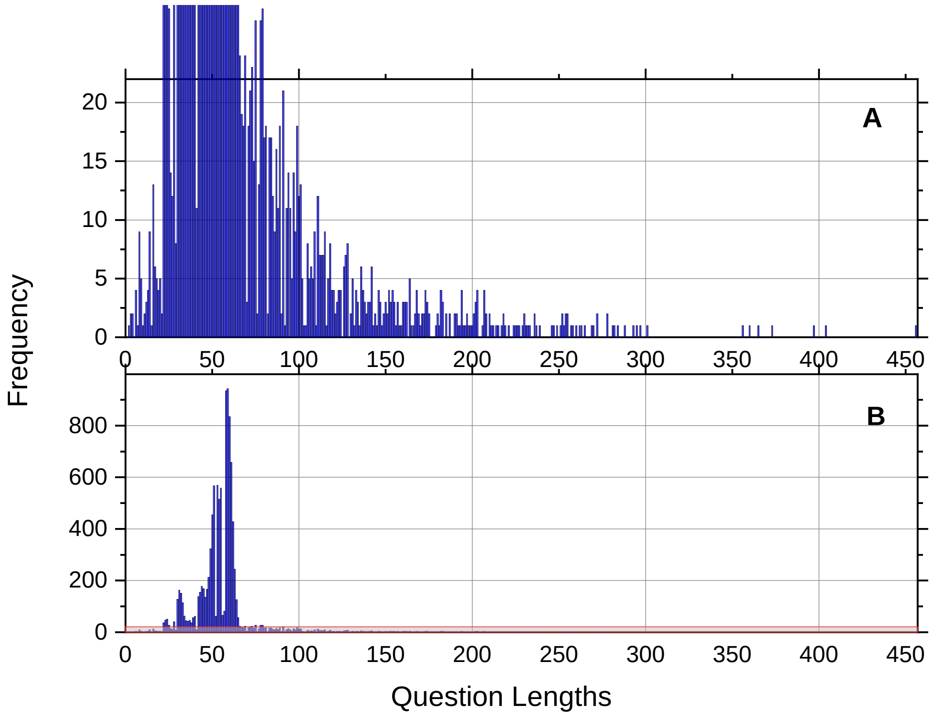

Figure 3: Distribution of question sizes in GeoRE. Part B is the complete illustration and Part A is an enlarged view of B from 0 to 25.

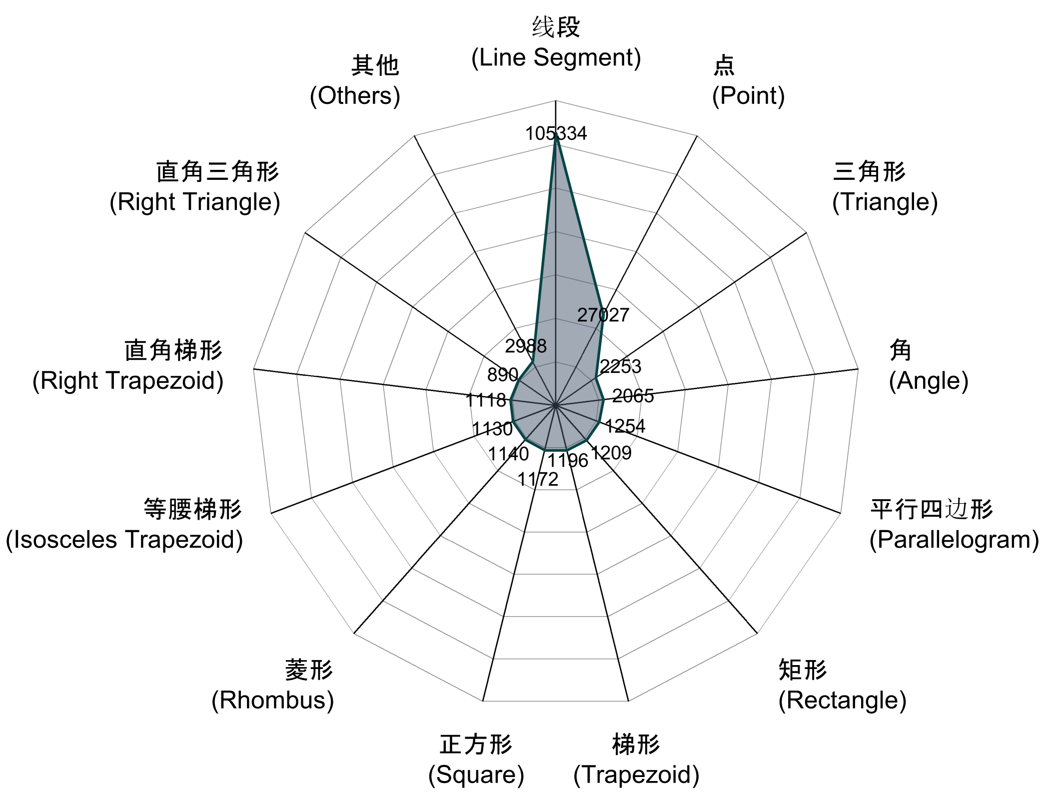

Figure 4: Distribution of shapes in the GeoRE.

Table 4: List of shapes and relations.

43 shapes			
中点 (midpoint)	等腰三角形 (isosceles triangle)	菱形 (rhombus)	内接三角形 (inscribed triangle)
内接正方形 (inscribed rectangle)	正方形 (square)	平行四边形 (parallelogram)	半圆 (semicircle)
钝角三角形 (obtuse angle)	切点 (tangent point)	直角边 (right-angled edge)	扇形 (sector)
三角形 (triangle)	等腰直角三角形 (isosceles right triangle)	垂线 (perpendicular)	点 (point)
对角线 (diagonal)	弦 (chord)	直线 (line)	长方形 (rectangle)
等腰梯形 (isosceles trapezoid)	直角三角形 (right triangle)	半径 (radius)	割线 (secant line)
锐角三角形 (acute triangle)	圆 (circle)	切线 (tangent line)	斜边 (oblique edge)
梯形 (trapezoid)	射线 (ray)	外接圆 (circumcircle)	中线 (midline)
弧 (arc)	角 (angle)	等边三角形 (equilateral triangle)	四边形 (quadrilateral)
垂足 (foot of perpendicular)	高 (alititude)	劣弧 (minor arc)	矩形 (rectangle)
线段 (line segment)	直角梯形 (right trapezoid)	直径 (diameter)	
19 positional relations			
垂直平分 (perpendicular bisect)	过 (through)	部分重叠 (partial overlap)	内接 (inscribed)
外接 (circumscribed)	对称 (symmetry)	延长 (production)	同侧 (ipsilateral)
交 (intersect)	平行 (parallel)	两侧 (both sides)	平分 (bisect)
上 (on)	共用 (share)	垂直 (perpendicular)	外切 (to touch externally)
切 (tangent)	外 (out)	落在 (on)	
4 quantitative relations			
等于 (equal)	全等 (congruent)	相似 (similar)	倍数 (time)

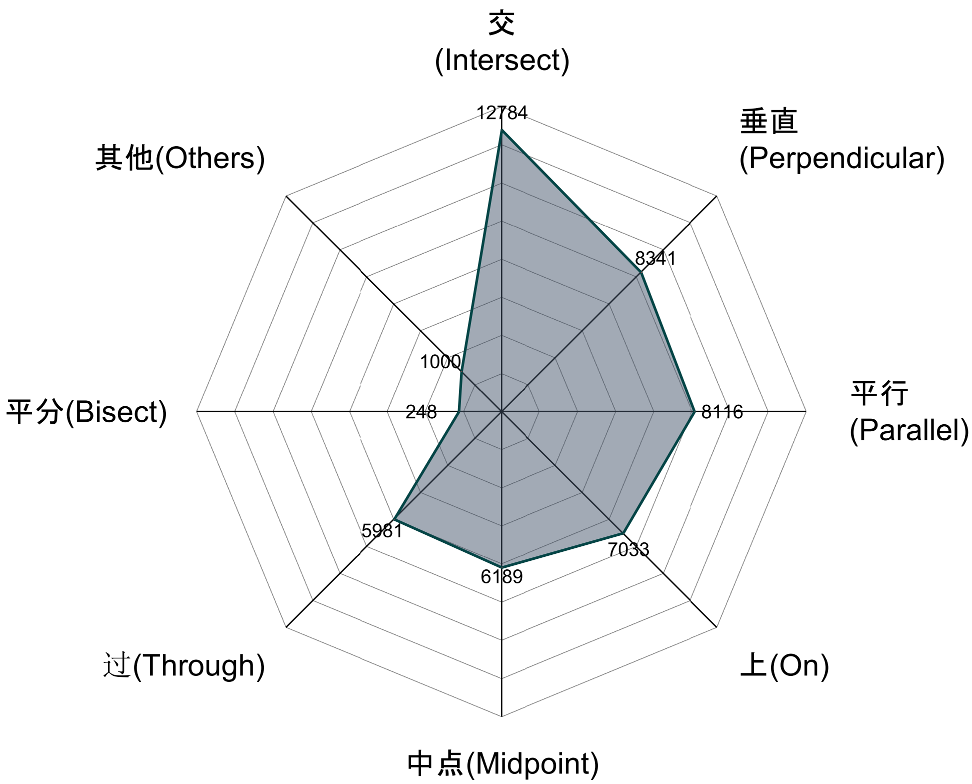

Figure 5: Distribution of positional relations in the GeoRE.

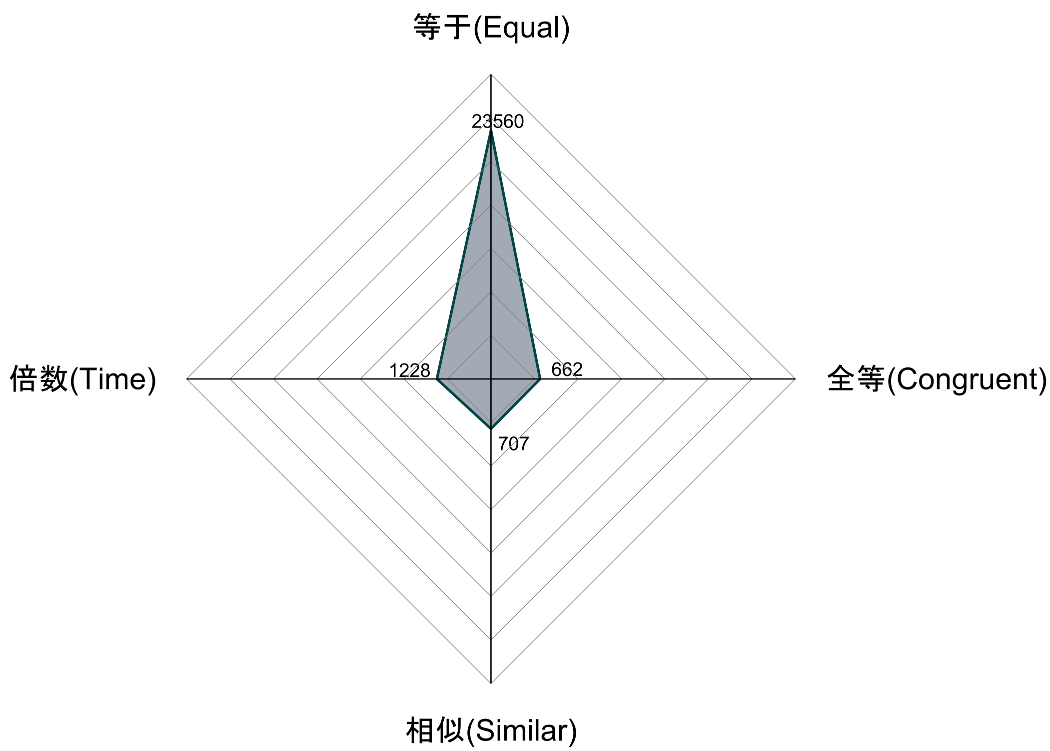

Figure 6: Distribution of quantitative relations in theGeoRE.