

参考 Inter-GPS 设计数学几何图解析解读

Illusionna

2024 年 6 月 23 日

Table of Contents

1	Abstract	2
2	Introduction	2
3	Related Work	3
4	Geometry Formal Language	4
5	Geometry3K Dataset	5
5.1	Dataset Collection	5
5.2	Dataset Statistics	5
5.3	Comparisons with Existing Datasets	7
5.4	Human Performance	7
6	Geometry Problem Parser	7
6.1	Text Parser	7
6.2	Diagram Parser	8
7	Geometry Problem Solver	8
7.1	Symbolic Geometry Solver	9
7.2	Theorem Predictor (TP)	9
7.3	Low-first Search Strategy	10
8	Experiments	11
8.1	Experimental Settings	11
8.2	Comparisons with Baselines	12
8.3	Ablation Study and Discussion	13
9	Conclusion	15
10	Appendix	15

1 Abstract

几何问题推理近几年在 NLP 领域引起广泛关注，要求具备抽象问题的理解和公理化知识的符号推理。

现有数据集要么规模小要么不公开，所以作者贡献了一个最新的大规模基准 Geometry3K，由 3002 个几何问题和形式语言密集注解构成。

针对形式语言和符号推理，作者提出一个新颖的几何求解方法 Interpretable Geometry Problem Solver (Inter-GPS)。

通过规则文本解析和神经对象检测，Inter-GPS 首先将问题的文本和图像自动解析成形式语言。

与现有的内隐学习不同，Inter-GPS 融合定理知识作为条件规则，逐步进行符号化推理。

此外，作者设计了一个定理预判器推断定理应用序列，再反馈给符号求解器，以获得更加有效合理的搜索路径。

在 Geometry3K 和 GEOS 两类数据集上的大量试验，Inter-GPS 比现有方法都取得显著改进。

2 Introduction

解决几何问题是高中教育中培养学生抽象思维的重要课题。

几何问题需要解题者发掘几何关系、应用定理知识、代数计算、推导，最终得到正确答案。

形式语言是由一组特定规则的格式良好的字母和数字构成，通常用于语言学和数学领域。

作者提出的几何求解器在正式解决问题前，需要把问题转化成形式语言。

文本和图像转化形式语言在现有的方法中高度依赖人工注解，并且这些方法不能提供明确的推理过程。

有学者将几何问题求解任务简化成一个优化问题，从满足所有约束条件的备选方案中挑一个。

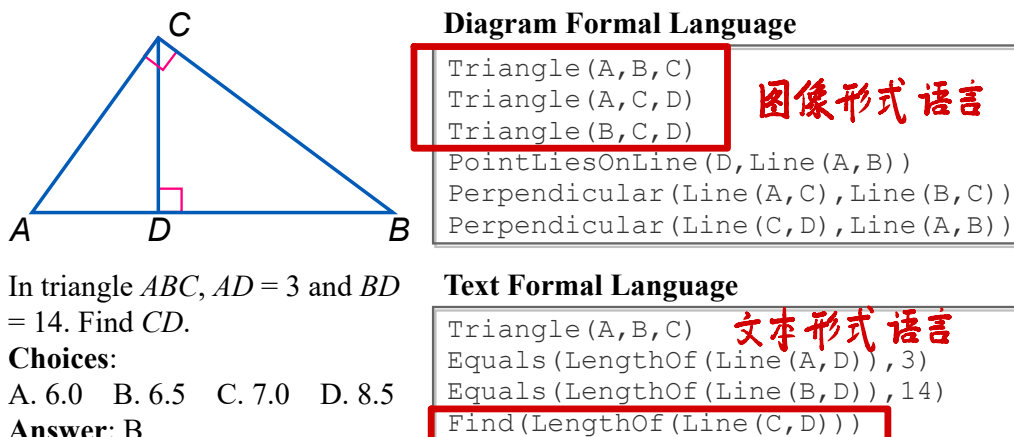


图 1: Geometry3K 的某个形式语言描述注解的几何题

作者构建的 Geometry3K 大规模基准数据集涵盖几何形状和文本问题，用形式语言对 3002 个问题进行统一的结构化描述注解，以评估算法在几何推理方面的性能。

Inter-GPS 有一个自动解析器，按照模板规则和解析图像准则，通过神经对象检测器将几何问题进行翻译。

Inter-GPS 不同于参数学习，它将几何问题任务表述为目标搜索，融合定理知识作为条件逐步进行符号推理。

Inter-GPS 具备可解释性。

作者设计的定理预判器推断 Inter-GPS 中可能的定理应用序列，以获得高效合理的搜索路径。

- 作者贡献一：引入公开的大规模基准数据集 Geometry3K；
- 作者贡献二：开发几何问题自动解析器，用于形式语言转化；
- 作者贡献三：提出新颖的 Inter-GPS 进行几何推理。

3 Related Work

Geometry3K 是从高中教科书中收集到的问题。

由于合适的数据稀疏，早期大多数自动求解几何系统的工作都集中于几何定理证明。

有学者尝试使用计算机视觉和自然语言处理求解几何问题。

然而，几何系统不能使用公理化知识执行精确的推理，之所以这样，是因为它将几何任务简化成优化问题，寻找出满足所有约束下的方案。

最近还有学者将几何原理写成霍尔语句规则和图文声明，解析成逻辑程序，输入到 Prolog 风格编程中求解问题。

然而这些学者的方法无法提供人类可读的解题步骤, 而且基于霍尔语句规则和内置求解器的参数学习, 会导致一个不可控的搜索过程。

- 实现数学问题可解释性解题步骤的第一个方向: 生成公式模板、操作程序、表达式树的中间结构结果;
- 实现数学问题可解释性解题步骤的第二个方向: 将数学问题翻译成符号语言并反复执行逻辑推理, 去预测最终结果, 这种方法更具可解释性。

作者受到学者语义解析工作启发, 文本和图像的结构化图解析和联合语义表示, 在可解释的几何问题求解中至关重要。

4 Geometry Formal Language

几何问题 P 定义: $P = (t, d, c)$, 其中 $t = text$, $d = diagram$, 正确答案:

$$c_i \in c = \{c_1, c_2, c_3, c_4\}$$

几何领域语言形式化问题定义 Ω 集合是由字面量语句的谓词和名词组成, 几何问题求解器的基本术语有如下三个定义:

1. 谓词是几何形状实体、几何关系、演算函数;
2. 字面量语句是一个谓词在一组像变量或常量的名字上的应用。一系列字面量语句构成形式化语言空间 Ω 中问题文本和图像的语义描述;
3. 原语是最小且不可分割的, 基础的几何元素像点、线段、图像中提取的弧段。

谓词 : 没着 、 相交于 、 是中点

字面量语句 : 找出三角形 ABC 的面积

Terms	Examples
<i>predicate</i>	Line, IntersectAt, IsMedianOf
<i>literal</i>	Find(AreaOf(Triangle(A, B, C)))

图 2: 几何形式化语言术语的例子

作者定义了 91 个谓词。

5 Geometry3K Dataset

5.1 Dataset Collection

Geometry3K 的样本来源自两本流行于 6–12 岁年龄的高中教科书所在的在线图书馆。

样本是由训练有素的本科学位者手动注解。

作者为了高精度地评估求解器性能，记录每个问题的目标和几何形状的数据。

Problem Text	Diagram	Choices	Text Literals	Diagram Literals
Find y . Round to the nearest tenth.		A. 18.8 B. 23.2 C. 25.9 D. 44.0 Answer: C	Find (y)	Equals (LengthOf (Line (A, B)), 32) Equals (LengthOf (Line (B, D)), y) Equals (MeasureOf (Angle (A, C, B)), 54) Equals (LengthOf (Line (A, D)), x) PointLiesOnLine (D, Line (A, C)) Perpendicular (Line (B, D), Line (C, D)) Equals (LengthOf (Line (A, B)), LengthOf (Line (B, C)))
Find the perimeter of $\square JKLM$.		A. 11.2 B. 22.4 C. 24 D. 44.8 Answer: B	Find (PerimeterOf (Parallelogram (J, K, L, M)))	Equals (LengthOf (Line (L, K)), 7.2) Equals (LengthOf (Line (M, L)), 4) Equals (LengthOf (Line (E, J)), 6) PointLiesOnLine (E, Line (M, L)) Perpendicular (Line (J, E), Line (E, L))
In $\odot K$, $MN = 16$ and $m\widehat{MN} = 98$. Find the measure of LN . Round to the nearest hundredth.		A. 6.93 B. 7.50 C. 8.94 D. 10.00 Answer: C	Circle (K) Equals (LengthOf (Line (M, N)), 16) Equals (MeasureOf (Arc (M, N)), 98) Find (LengthOf (Line (L, N)))	Equals (LengthOf (Line (J, K)), 10) Perpendicular (Line (P, K), Line (M, P)) PointLiesOnLine (P, Line (M, N)) PointLiesOnLine (P, Line (L, J)) PointLiesOnLine (P, Line (L, K)) PointLiesOnLine (K, Line (P, J)) PointLiesOnLine (K, Line (L, J)) PointLiesOnCircle (M, Circle (K)) PointLiesOnCircle (J, Circle (K)) PointLiesOnCircle (N, Circle (K)) PointLiesOnCircle (L, Circle (K))

图 3: Geometry3K 数据样例

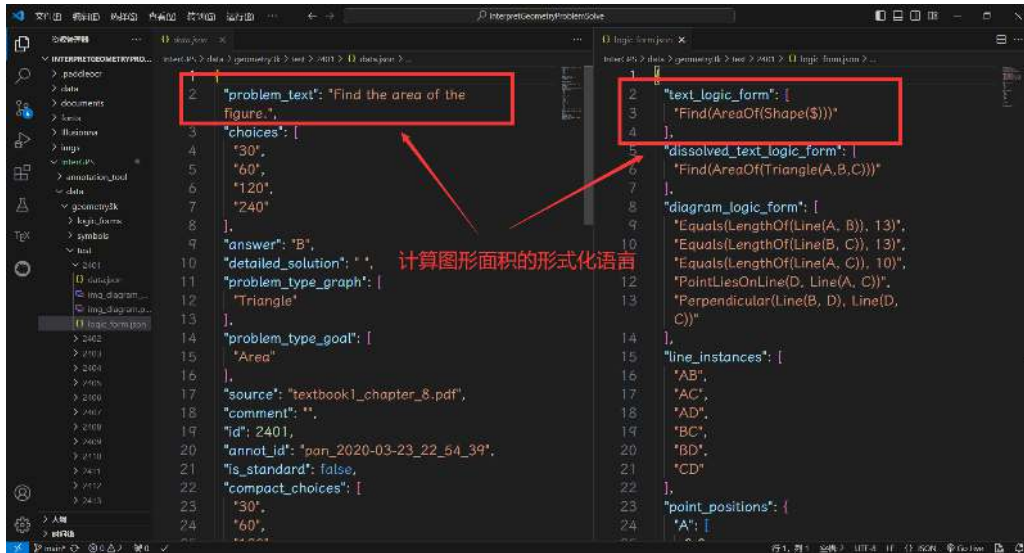


图 4: Geometry3K 数据 JSON 格式样例

5.2 Dataset Statistics

Geometry3K 由 3002 个几何问题构成。

训练集、验证集、测试集比例 0.7: 0.1: 0.2。

	Total	Train	Val	Test
Questions	3,002	2,101	300	601
Sentences	4,284	2,993	410	881
Words	30,146	20,882	2,995	6,269
Literals (Text)	6,293	4,357	624	1,312
Literals (Diagram)	27,213	19,843	2,377	4,993

图 5: Geometry3K 数据集的基本统计

长尾分布要求在文本内容下，求解器能够理解丰富的语义。

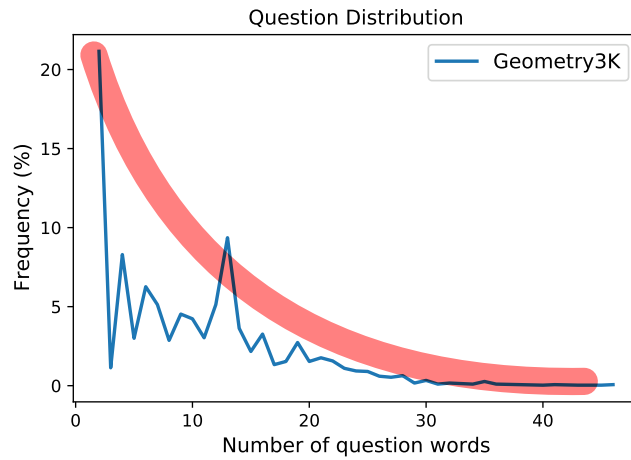


图 6: Geometry3K 数据集问题单词数量的长尾分布

Predicates (Text)	%	Predicates (Diagram)	%
Find	19.00	Line	30.89
Line	14.49	PointLiesOnLine	16.66
Equals	11.83	Equals	15.17
LengthOf	9.53	MeasureOf	10.46
MeasureOf	8.97	LengthOf	8.69
.....		
CircumscribedTo	0.05	Triangle	0.03
SumOf	0.04	Quadrilateral	0.02
HeightOf	0.04	Kite	0.01
BaseOf	0.04	HeightOf	0.01
IsHypotenuseOf	0.04	Square	0.01

图 7: Geometry3K 数据集形式描述谓词频率（频数至少大于五）

文本谓词分布比图像谓词更加均匀，主要原因在于几何问题的文本描述各式各样，像形状、属性、关系等，而图像只能显示点、线、弧的基本性质。

5.3 Comparisons with Existing Datasets

作者认为自己的数据集目前最大，加入了 GEOS 数据集线、三角形、正四边形和圆以外其他不规则的四边形和多边形。此外 Geometry3K 还引入更多未知变量和运算类型，可能要求解方程才能找到问题目标。

GEOS 数据集中 80.5% 的问题可以在没有关系图的情况下求解，而不提供问题图的情况下，Geometry3K 数据集中只有 1% 不到的问题可以求解。

5.4 Human Performance

作者将划分好的测试集推送到 Amazon Mechanical Turk 外包平台，每位注解者要求有高中或高中以上的学历。

25 分钟回答 10 个问题，所有问题至少 7 分钟，每个问题不得低于 10 秒钟。

作者还请了几十个理工科专业的毕业生回答这些问题，以评估人类专家表现。

对比于随机猜测的 25% 正确率，人类整体准确率达到 56.9%，专家准确率达到 90.9%。

Method	All	Angle	Length	Area	Ratio	Line	Triangle	Quad	Circle	Other
Random	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
Human	56.9	53.7	59.3	57.7	42.9	46.7	53.8	68.7	61.7	58.3
Human Expert	90.9	89.9	92.0	93.9	66.7	95.9	92.2	90.5	89.9	92.3
Q-only	25.3	29.5	21.5	28.3	33.3	21.0	26.0	25.9	25.2	22.2
I-only	27.0	26.2	28.4	24.5	16.7	24.7	26.7	30.1	30.1	25.9
Q+I	26.7	26.2	26.7	28.3	25.0	21.0	28.1	32.2	21.0	25.9
RelNet (Bansal et al., 2017)	29.6	26.2	34.0	20.8	41.7	29.6	33.7	25.2	28.0	25.9
FiLM (Perez et al., 2018)	31.7	28.7	32.7	39.6	33.3	33.3	29.2	33.6	30.8	29.6
FiLM-BERT (Devlin et al., 2018)	32.8	32.9	33.3	30.2	25.0	32.1	32.3	32.2	34.3	33.3
FiLM-BART (Lewis et al., 2020)	33.0	32.1	33.0	35.8	50.0	34.6	32.6	37.1	30.1	37.0
Inter-GPS (ours)	57.5	59.1	61.7	30.2	50.0	59.3	66.0	52.4	45.5	48.1
Inter-GPS (GT)	78.3	83.1	77.9	62.3	75.0	86.4	83.3	77.6	61.5	70.4

图 8: Inter-GPS 在 Geometry3K 上的推理结果与基准对比

6 Geometry Problem Parser

6.1 Text Parser

给定文本 t 的单词序列，文本解析器需要将其翻译成一组谓词变量构成的字面量 L_t 。

深度学习网络在 Seq2Seq 机器翻译任务中取得很好的表现。

- Seq2Seq 的语义解析器在 Geometry3K 无法令人满意原因一：Geometry3K 相对于 Seq2Seq 而言数据规模实在太小；
- Seq2Seq 的语义解析器在 Geometry3K 无法令人满意原因二：Seq2Seq 往往会给结果带来噪音，这对符号推理几何问题偏差敏感。

作者将规则解析方法、正则表达式和 BART 文本解析三者结合一体。

6.2 Diagram Parser

作者提出自动图解析器，无需人工注解干预，能够识别图中符号、平行、垂直等实体和关系。

第一步：霍尔变化提取几何图形中的基本物体，如点、线、弧、圆。

第二步：强目标检测器 RetinaNet 提取图像符号和文本区域。

第三步：使用 OCR 工具（MathPix 要付费）对区域中文本内容进行识别。

第四步：采用贪心算法，对获取到的原语集 P 和符号集 S 按照关系联系起来，每个符号分配最近的原语，忽略有效性。

第五步：将联系起来的任务视作为几何关系的优化问题。

$$\begin{cases} \min \sum_s dist(s_i, p_j) \times \mathbb{1}_{s_i \rightarrow p_j} ? 1 : 0 \\ s.t. (s_i, p_j) \in \text{可行解集 } F \end{cases} \quad (1)$$

其中，采用欧氏距离， F 定义符号链接原语的几何约束，如平行符号只能链接到两条斜率相同的直线，垂直符号只能链接到两条正交的直线。

7 Geometry Problem Solver

Inter-GPS 利用定理知识基础和人类可读搜索过程实现精确的显示符号推理。

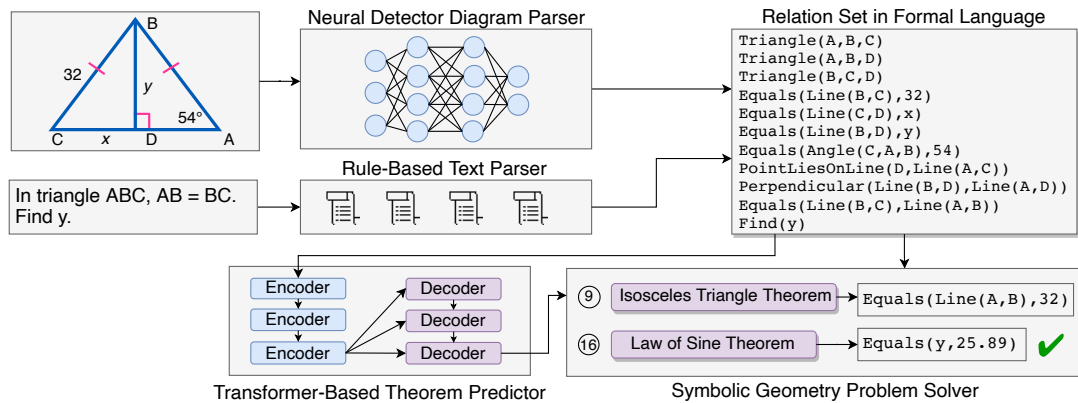


图 9: Inter-GPS 先将输入解析成关系集合的形式语言，然后定理序列预判器对关系集合进行符号推理，从而推导出答案，⑨ 和 ⑯ 表示定理顺序。

7.1 Symbolic Geometry Solver

Inter-GPS 总体思路是以**关系集合** \mathcal{R} 和**定理知识基础** \mathcal{KB} 作为输入，输出目标几何问题 g 的数值解 g^* 。

关系集合 R 定义给定问题中的几何属性和关系，并且用文本和图像的解析进行字面量的初始化。 R 进一步扩展了几何形状定义衍生的字面量。如，三角形被定义为三条相连的边，那么以下推导出来的信息会附加到 R 集合。

$$\triangle ABC \implies \text{Point}(A), \text{Point}(B), \text{Point}(C), \text{Line}(A, B), \text{Line}(B, C), \text{Line}(C, A)$$

定理知识基础 \mathcal{KB} 集合代表一个定理集合，每个定理 k_i 写成一个有前提 p 和结论 q 的条件规则。

对于搜索步骤 t 而言，如果 k_i 的前提 p 与当前关系集合 \mathcal{R}_{t-1} 匹配，则根据结论 q 更新关系集合。

$$\mathcal{R} \leftarrow k_i \wedge \mathcal{R}_{t-1}, \quad k_i \in \mathcal{KB}$$

应用几个定理后，建立已知值与未知问题目标 g 之间的方程，解出方程后，则 g 就能求解。

$$g^* \leftarrow \text{SOLVEEQUATION}(\mathcal{R}_t, g)$$

7.2 Theorem Predictor (TP)

一种可能的搜索策略是使用**暴力随机枚举定理集合中候选方案**，但随机搜索策略是低效的，并且可能因为早阶段有复杂定理而导致问题无法求解。

一个理想的几何求解器可以有**合理的定理应用序列**求解问题。

学术表现好的学生能通过大量解题训练获得求解问题的先验知识。

最近研究表明**神经引导搜索**可以加快搜索过程。

作者从定理集合中随机抽取多个样本生成应用序列，如果几何求解器 Inter-GPS 应用生成的序列求解了问题，则任务生成序列是正样本。具有最小长度的正序列样本视为拟合最优逼近。

作者收集 1501 个具有问题及其拟合最优逼近定理应用序列的训练样本。

给定问题的形式化描述 $L = \{l_1, l_2, \dots, l_m\}$ ，定理预判器目的是一个 token 一个 token 地重构拟合最优逼近定理序列 $T = \{t_1, t_2, \dots, t_n\}$ 。

作者将生成任务公式化为一个 Seq2Seq 问题，并使用 **Transformer-based** 模型来**生成定理序列 tokens**。

Transformer 解码器通过给定 $T = \{t_1, t_2, \dots, t_i\}$ 预测下一步定理的顺序。

训练 Seq2Seq 模型以优化负似然对数损失函数。

$$\mathcal{L}_{\text{TP}} = - \sum_{i=1}^n \log p_{\text{TP}}(t_i \mid t_1, t_2, \dots, t_{i-1})$$

其中, p_{rmTP} 是定理预测器模型参数化条件分布。

7.3 Low-first Search Strategy

在定理预判器预测应用定理序列之后, Inter-GPS 仍然有可能无法找到问题目标。人类在解决数学问题时倾向于优先使用简单定理, 以减少复杂计算。如果简单定理不可行, 才会转向更复杂的定理。

作者采用一种由学科知识驱动的高效启发式搜索策略。

作者把理论分为低阶定理集合 \mathcal{KB}_1 和高阶定理集合 \mathcal{KB}_2 , 如低阶的三角形内角和定理、全等三角形定理, 高阶的正弦定理。

在使用预测定理之后的每一步搜索中, 首先枚举低阶集合 \mathcal{KB}_1 来更新关系集合 \mathcal{R} 。

$$\mathcal{R} \leftarrow k_i \wedge \mathcal{R}_{t-1}, \quad k_i \in \mathcal{KB}_1$$

若低阶定理不能更新 \mathcal{R} , 再考虑高阶定理更新关系集合。

$$\mathcal{R}_t \leftarrow k_i \wedge \mathcal{R}_{t-1}, \quad k_i \in \mathcal{KB}_2$$

一旦找到问题目标 g 或搜索步骤达到最大值, 整个搜索过程就会停止。

Algorithm 1 Symbolic Geometry Solver

Input Literals \mathcal{L} , goal g , knowledge bases $\mathcal{KB}_1, \mathcal{KB}_2$
Output Numeric goal value g^* and theorem application \mathcal{S}

```
1: function SEARCH( $\mathcal{L}, g, \mathcal{KB}_1, \mathcal{KB}_2$ )
2:   Initialize relation set  $\mathcal{R}_0$  with  $\mathcal{L}, g^* = \emptyset, \mathcal{S} = \emptyset$ 
3:    $\mathcal{KB}_p \leftarrow \text{THEOPREDICTOR}(\mathcal{L})$   $\triangleright$  Predicted
4:   for  $k_i \in \mathcal{KB}_p$  do
5:      $\mathcal{R}_t \leftarrow k_i \wedge \mathcal{R}_{t-1}$ 
6:      $\mathcal{S}.\text{APPEND}(k_i)$ 
7:   end for
8:    $g^* \leftarrow \text{SOLVEEQUATION}(\mathcal{R}_t, g)$ 
9:   if  $g^* \neq \emptyset$  then
10:    return  $g^*$  and  $\mathcal{S}$ 
11:   end if
12:   while  $g^* = \emptyset$  and  $\mathcal{R}_t$  is updated do
13:     for  $k_i \in \mathcal{KB}_1$  do  $\triangleright$  Lower-order
14:        $\mathcal{R}_t \leftarrow k_i \wedge \mathcal{R}_{t-1}$ 
15:        $\mathcal{S}.\text{APPEND}(k_i)$ 
16:        $g^* \leftarrow \text{SOLVEEQUATION}(\mathcal{R}_t, g)$ 
17:       if  $g^* \neq \emptyset$  then
18:         return  $g^*$  and  $\mathcal{S}$ 
19:       end if
20:     end for
21:     for  $k_i \in \mathcal{KB}_2$  do  $\triangleright$  Higher-order
22:        $\mathcal{R}_t \leftarrow k_i \wedge \mathcal{R}_{t-1}$ 
23:        $\mathcal{S}.\text{APPEND}(k_i)$ 
24:        $g^* \leftarrow \text{SOLVEEQUATION}(\mathcal{R}_t, g)$ 
25:       if  $g^* \neq \emptyset$  then
26:         return  $g^*$  and  $\mathcal{S}$ 
27:       end if
28:     end for
29:   end while
30: end function
```

图 10: Inter-GPS 搜索算法

8 Experiments

8.1 Experimental Settings

Datasets and evaluation metrics

如果 Inter-GPS 未能在允许的步长内输出问题目标数值, 则从四个候选项中随机选择一个。就比较神经网络基准而言, 预测答案在四个选项中具有最大置信度得分。

Baselines

作者为几何求解器实现几个深度神经网络基准, 与 Inter-GPS 进行比较。

基准模型将几何问题求解任务形式化成分类问题，由序列编码器的文本嵌入和视觉编码器的图像表征提供。

Q-only 仅通过 Bi-GRU 编码器对自然语言中的问题文本进行编码。I-only 仅通过 ResNet-50 编码器作为输入对问题图像进行编码。

Q+I 分别使用 Bi-GRU 和 ResNet-50 对文字和图像进行编码。

RelNet 为了嵌入问题文本而实现的，它是实体和关系建模的强大方法。FiLM 实现了抽象图像问题的有效视觉推理。

FiLM-BERT 使用 BERT 编码器代替 GRU 编码器，FiLM-BART 使用 BART 编码器。

Implementation details

作者收集了一组 17 个几何定理来形成知识库。

为了生成定理序列正样本，每个问题尝试 100 次，最大序列长度为 20。

定理预判器使用了 Transformer 模型，有 6 层，12 个注意力头，隐藏层嵌入尺寸为 768。

Inter-GPS 的搜索步长设置为 100 步。

神经求解器选用 Adam 优化器，学习率设置 0.01，最大迭代次数 30 次。

为获得更精确结果，Inter-GPS 的每个试验重复三次。

8.2 Comparisons with Baselines

作者的 Inter-GPS 整体精度 57.5%，显著优于所有深度神经网络基准。

Inter-GPS 比人类精度更高。

带有真值（Ground truth）形式语言的 Inter-GPS 获得 20.8% 进一步提升。

Method	All	Angle	Length	Area	Ratio	Line	Triangle	Quad	Circle	Other
Random	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
Human	56.9	53.7	59.3	57.7	42.9	46.7	53.8	68.7	61.7	58.3
Human Expert	90.9	89.9	92.0	93.9	66.7	95.9	92.2	90.5	89.9	92.3
Q-only	25.3	29.5	21.5	28.3	33.3	21.0	26.0	25.9	25.2	22.2
I-only	27.0	26.2	28.4	24.5	16.7	24.7	26.7	30.1	30.1	25.9
Q+I	26.7	26.2	26.7	28.3	25.0	21.0	28.1	32.2	21.0	25.9
RelNet (Bansal et al., 2017)	29.6	26.2	34.0	20.8	41.7	29.6	33.7	25.2	28.0	25.9
FiLM (Perez et al., 2018)	31.7	28.7	32.7	39.6	33.3	33.3	29.2	33.6	30.8	29.6
FiLM-BERT (Devlin et al., 2018)	32.8	32.9	33.3	30.2	25.0	32.1	32.3	32.2	34.3	33.3
FiLM-BART (Lewis et al., 2020)	33.0	32.1	33.0	35.8	50.0	34.6	32.6	37.1	30.1	37.0
Inter-GPS (ours)	57.5	59.1	61.7	30.2	50.0	59.3	66.0	52.4	45.5	48.1
Inter-GPS (GT)	78.3	83.1	77.9	62.3	75.0	86.4	83.3	77.6	61.5	70.4

图 11: Geometry3K 数据集上 Inter-GPS 对比基准评估结果

Inter-GPS 还在 CEOS 数据集上获得最先进的几何求解器性能。

Method	Acc (%)
GEOS (Seo et al., 2015)	49
GEOS++ (Seo et al., 2015)	49
GEOS-OS (Sachan and Xing, 2017)	52
GEOS++AXIO (Sachan et al., 2017)	55
Inter-GPS (ours)	67

图 12: GEOS 数据集上评估结果

8.3 Ablation Study and Discussion

Search strategies

Inter-GPS 使用不同搜索策略解决问题所需的整体精度和平均步数。

Search strategies	Accuracy (%)	# Steps
Random	75.5 ± 0.2	13.2 ± 0.1
Low-first	77.3 ± 0.3	15.1 ± 0.2
Predict	77.5 ± 0.1	6.5 ± 0.1
Predict+Low-first (final)	78.3 ± 0.1	7.1 ± 0.1

图 13: Inter-GPS 在不同搜索策略下的性能

Predict 是指代使用定理预判器中的定理，然后再是随机定理序列，这么一种策略。

Predict 这种策略很大程度上将平均步数减少到 6.5 步。

Final 这种策略在剩下的搜索步骤中应用了预测定理一阶和低阶性质，获得整体最佳精度。

Problem parsers and literal sources

规则文本解析器达到 97% 准确率，而语义文本解析器只有 67% 准确率。

	Diagram w/o	Diagram	Diagram (GT)
Text w/o	25.0 ± 0.0	46.6 ± 0.7	58.7 ± 0.2
Text	25.4 ± 0.0	57.5 ± 0.2	75.0 ± 0.6
Text (GT)	25.4 ± 0.0	58.0 ± 1.7	78.3 ± 0.1

图 14: 不同字面量来源，预测值和真实值字面量对比性能

随着作者的问题求解器生成字面量，Inter-GPS 精度达到 57.5%。

Inter-GPS 与生成的文本和真值文本之间的差距很小，所以当前文本解析器执行得很好。

带有注解图像字面量的 Inter-GPS 改进 17.5% 的提升，意味图像解析器提升空间很大。

Searching step distribution

通过不同策略的平均搜索步骤数，比较正确求解问题的分布，作者的 Inter-GPS 采用 Predict+Low-first (final) 策略，两步就解决的问题有 65.97%，五步求解的问题有 70.06%。

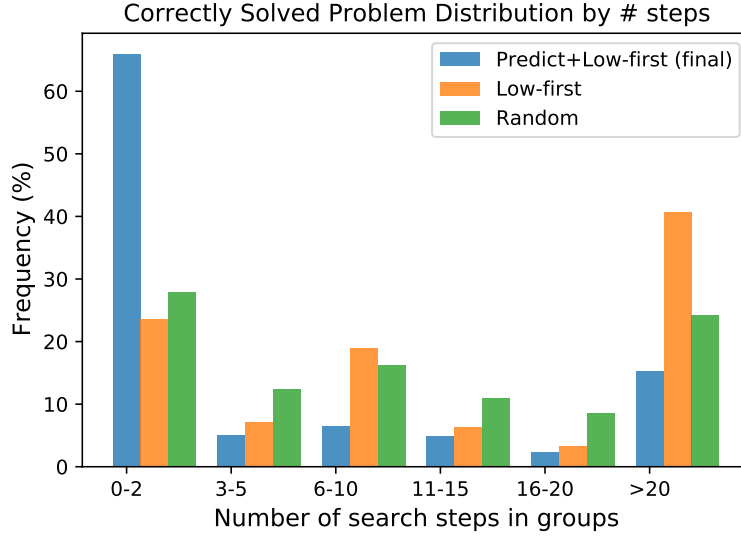


图 15: 正确求解问题的搜索步数分布

Neural geometry solvers

几何求解的神经网络基准在 Geometry3K 数据集上无法获得满意结果。这是因为神经网络方法输入的有意义的语义数据样本有限。

密集的隐式表征可能不合适像几何问题求解这样的逻辑推理。

作者将 Q+I 基准中的问题文本和图像的输入替换成真值文本和视觉形式化标注。

	Diagram (visual)	Diagram (formal)
Text (natural)	26.7	35.3
Text (formal)	34.6	35.9

图 16: 不同问题文本和图像表示下的神经网络求解器性能

9.2% 的提升改进表明，若学习到丰富的语义结构表征，神经网络模型在求解问题方面有很大潜力。

Failure cases

解析结构不准确和定理集不完整，可能会导致 Inter-GPS 无法找到解决方案。

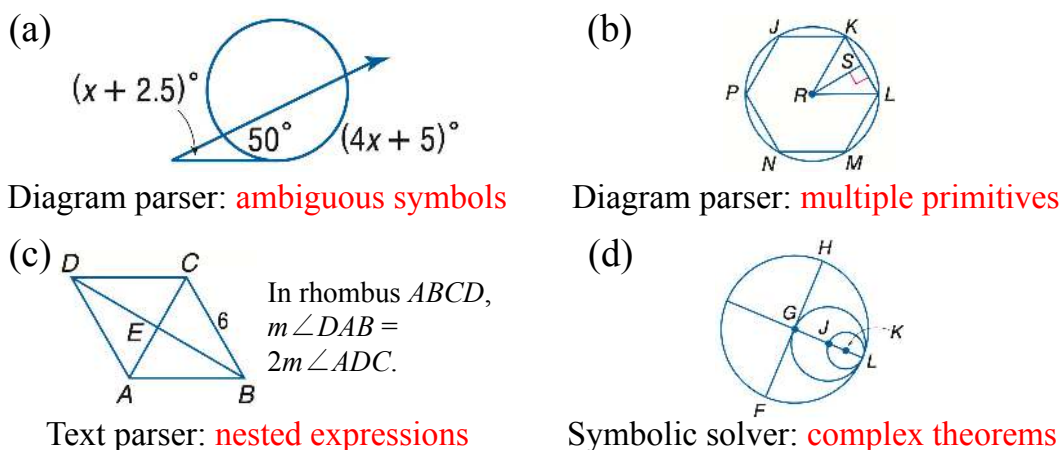


图 17: Inter-GPS 求解失败示例

譬如，图像中有模棱两可的注解或多个原语，文本中有嵌套表达式和不确定的指代，图像形状和阴影区域重叠复杂等等。

Interpretability in Inter-GPS

Inter-GPS 具有可解释性，首先问题内容解析成形式化语言的结构表示，然后显示地更新几何关系集进行符号推理，最后搜索过程依次应用合理的定理。

9 Conclusion

- 构建 Geometry3K 数据集；
- 提出 Inter-GPS 形式语言解析器，来对定理知识集进行符号推理；
- 提出 low-first 搜索策略的定理预测器，生成合理的定理应用序列。

10 Appendix

作者定义了 91 个谓词及其相应的字面量范式模板，分为六组，其中范式模板中的 \$ 表示未确定的形状。

- 几何形状；
- 一元几何属性；
- 通用几何属性；
- 二元几何关系；
- 形如 A-IsXOf-B 类型的几何关系；
- 数值属性和关系。

#	Predicates	Literal templates
1	Point	Point (A), Point (\$)
2	Line	Line (A, B), Line (m), Line (\$)
3	Angle	Angle (A, B, C), Angle (A), Angle (1), Angle (\$)
4	Triangle	Triangle (A, B, C), Triangle (\$), Triangle (\$1, \$2, \$3)
5	Quadrilateral	Quadrilateral (A, B, C, D), Quadrilateral (1), Quadrilateral (\$)
6	Parallelogram	Parallelogram (A, B, C, D), Parallelogram (1), Parallelogram (\$)
7	Square	Square (A, B, C, D), Square (1), Square (\$)
8	Rectangle	Rectangle (A, B, C, D), Rectangle (1), Rectangle (\$)
9	Rhombus	Rhombus (A, B, C, D), Rhombus (1), Rhombus (\$)
10	Trapezoid	Trapezoid (A, B, C, D), Trapezoid (1), Trapezoid (\$)
11	Kite	Kite (A, B, C, D), Kite (1), Kite (\$)
12	Polygon	Polygon (\$)
13	Pentagon	Pentagon (A, B, C, D, E), Pentagon (\$)
14	Hexagon	Hexagon (A, B, C, D, E, F), Hexagon (\$)
15	Heptagon	Heptagon (A, B, C, D, E, F, G), Heptagon (\$)
16	Octagon	Octagon (A, B, C, D, E, F, G, H), Octagon (\$)
17	Circle	Circle (A), Circle (1), Circle (\$)
18	Arc	Arc (A, B), Arc (A, B, C), Arc (\$)
19	Sector	Sector (O, A, B), Sector (\$)
20	Shape	Shape (\$)

Table 10: 20 predicates and corresponding literal templates for geometric shapes.

#	Predicates	Literal templates
1	RightAngle	RightAngle (Angle (\$))
2	Right	Right (Triangle (\$))
3	Isosceles	Isosceles (Polygon (\$))
4	Equilateral	Equilateral (Polygon (\$))
5	Regular	Regular (Polygon (\$))
6	Red	Red (Shape (\$))
7	Blue	Blue (Shape (\$))
8	Green	Green (Shape (\$))
9	Shaded	Shaded (Shape (\$))

Table 11: 9 predicates and corresponding literal templates for unary geometric attributes.

#	Predicates	Literal templates
1	AreaOf	AreaOf (A)
2	PerimeterOf	PerimeterOf (A)
3	RadiusOf	RadiusOf (A)
4	DiameterOf	DiameterOf (A)
5	CircumferenceOf	CircumferenceOf (A)
6	AltitudeOf	AltitudeOf (A)
7	HypotenuseOf	HypotenuseOf (A)
8	SideOf	SideOf (A)
9	WidthOf	WidthOf (A)
10	HeightOf	HeightOf (A)
11	LegOf	LegOf (A)
12	BaseOf	BaseOf (A)
13	MedianOf	MedianOf (A)
14	IntersectionOf	IntersectionOf (A, B)
15	MeasureOf	MeasureOf (A)
16	LengthOf	LengthOf (A)
17	ScaleFactorOf	ScaleFactorOf (A, B)

Table 12: 17 predicates and corresponding literal templates for general geometric attributes .

#	Predicates	Literal templates
1	PointLiesOnLine	PointLiesOnLine(Point(\$),Line(\$1,\$2))
2	PointLiesOnCircle	PointLiesOnCircle(Point(\$),Circle(\$))
3	Parallel	Parallel(Line(\$),Line(\$))
4	Perpendicular	Perpendicular(Line(\$),Line(\$))
5	IntersectAt	IntersectAt(Line(\$),Line(\$),Line(\$),Point(\$))
6	BisectsAngle	BisectsAngle(Line(\$),Angle(\$))
7	Congruent	Congruent(Polygon(\$),Polygon(\$))
8	Similar	Similar(Polygon(\$),Polygon(\$))
9	Tangent	Tangent(Line(\$),Circle(\$))
10	Secant	Secant(Line(\$),Circle(\$))
11	CircumscribedTo	CircumscribedTo(Shape(\$),Shape(\$))
12	InscribedIn	InscribedIn(Shape(\$),Shape(\$))

Table 13: 12 predicates and corresponding literal templates for binary geometric relations.

#	Predicates	Literal templates
1	IsMidpointOf	IsMidpointOf(Point(\$),Line(\$))
2	IsCentroidOf	IsCentroidOf(Point(\$),Shape(\$))
3	IsIncenterOf	IsIncenterOf(Point(\$),Shape(\$))
4	IsRadiusOf	IsRadiusOf(Line(\$),Circle(\$))
5	IsDiameterOf	IsDiameterOf(Line(\$),Circle(\$))
6	IsMidsegmentOf	IsMidsegmentOf(Line(\$),Triangle(\$))
7	IsChordOf	IsChordOf(Line(\$),Circle(\$))
8	IsSideOf	IsSideOf(Line(\$),Polygon(\$))
9	IsHypotenuseOf	IsHypotenuseOf(Line(\$),Triangle(\$))
10	IsPerpendicularBisectorOf	IsPerpendicularBisectorOf(Line(\$),Triangle(\$))
11	IsAltitudeOf	IsAltitudeOf(Line(\$),Triangle(\$))
12	IsMedianOf	IsMedianOf(Line(\$),Quadrilateral(\$))
13	IsBaseOf	IsBaseOf(Line(\$),Quadrilateral(\$))
14	IsDiagonalOf	IsDiagonalOf(Line(\$),Quadrilateral(\$))
15	IsLegOf	IsLegOf(Line(\$),Trapezoid(\$))

Table 14: 15 predicates and corresponding literal templates for A-IsXOf-B-type geometric relations.

#	Predicates	Literal templates
1	SinOf	SinOf(Var)
2	CosOf	CosOf(Var)
3	TanOf	TanOf(Var)
4	CotOf	CotOf(Var)
5	HalfOf	HalfOf(Var)
6	SquareOf	SquareOf(Var)
7	SqrtOf	SqrtOf(Var)
8	RatioOf	RatioOf(Var),RatioOf(Var1,Var2)
9	SumOf	SumOf(Var1,Var2,...)
10	AverageOf	AverageOf(Var1,Var2,...)
11	Add	Add(Var1,Var2,...)
12	Mul	Mul(Var1,Var2,...)
13	Sub	Sub(Var1,Var2,...)
14	Div	Div(Var1,Var2,...)
15	Pow	Pow(Var1,Var2)
16	Equals	Equals(Var1,Var2)
17	Find	Find(Var)
18	UseTheorem	UseTheorem(A.B.C)

Table 15: 18 predicates and corresponding literal templates for numerical attributes and relations.