

Solving geometry problems: combining text and diagram interpretation

Minjoon Seo¹, Hannaneh Hajishirzi¹, Ali Farhadi^{1,2},
Oren Etzioni², Clint Malcolm¹

Sep. 20

EMNLP 2015

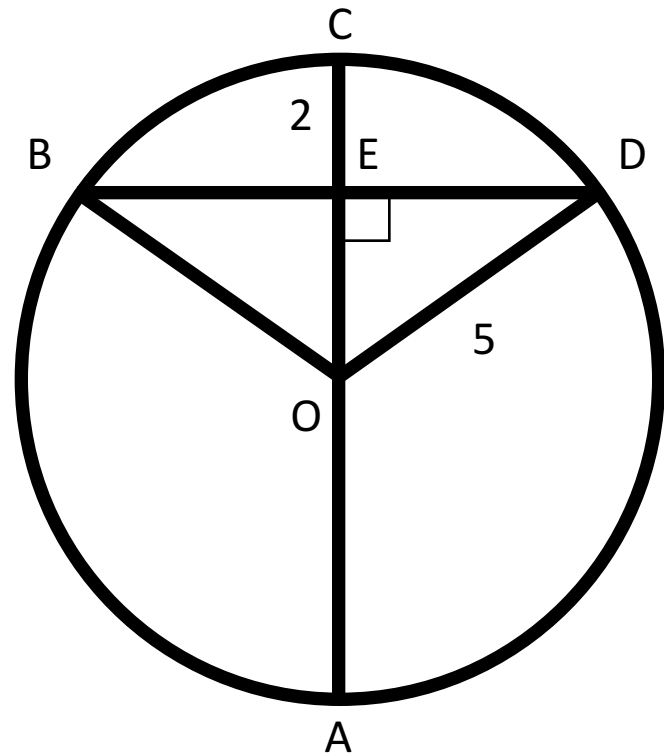
UNIVERSITY *of* 1
WASHINGTON

 ALLEN INSTITUTE 2
for ARTIFICIAL INTELLIGENCE

Geometry Word Problems

In the diagram at the right, circle O has a radius of 5, and $CE = 2$. Diameter AC is perpendicular to chord BD . What is the length of BD ?

- a) 2 b) 4 c) 6
d) 8 e) 10



Why geometry problems?

- Solving geometry word problems is challenging in AI
- Part of broader scope of solving math word problems
(Kushman et al., 2014; Hosseini et al., 2014; Roy et al., 2015; Dai et al., 2015; Shi et al., 2015)
- Interesting interplay between natural language and vision

Why geometry problems?

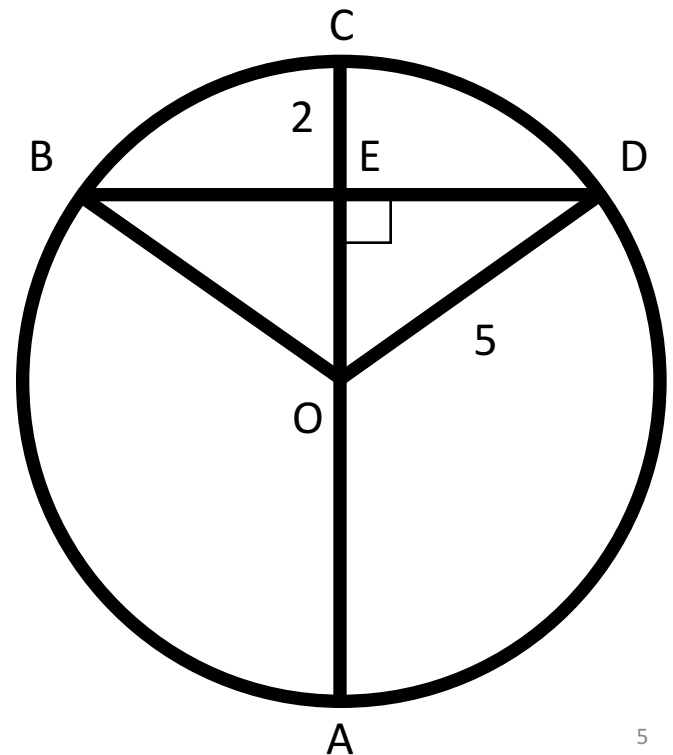
- Solving geometry word problems is challenging in AI
- Part of broader scope of solving math word problems
(Kushman et al., 2014; Hosseini et al., 2014; Roy et al., 2015; Dai et al., 2015; Shi et al., 2015)
- Interesting interplay between natural language and vision

In the diagram at the right, circle O has a radius of 5, and $CE = 2$. Diameter AC is perpendicular to chord BD . What is the length of BD ?

- a) 2 b) 4 c) 6
d) 8 e) 10

Why geometry problems?

- Solving geometry word problems is challenging in AI
- Part of broader scope of solving math word problems
(Kushman et al., 2014; Hosseini et al., 2014; Roy et al., 2015; Dai et al., 2015; Shi et al., 2015)
- Interesting interplay between natural language and vision



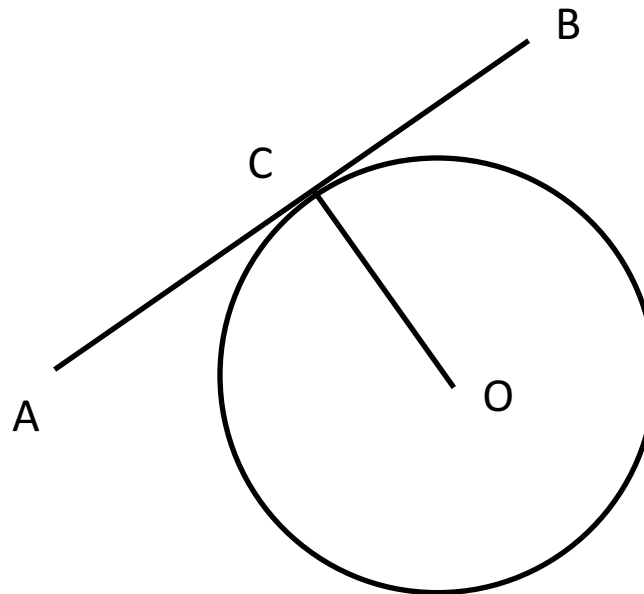
Why geometry problems?

- Solving geometry word problems is challenging in AI
- Part of broader scope of solving math word problems
(Kushman et al., 2014; Hosseini et al., 2014; Roy et al., 2015; Dai et al., 2015; Shi et al., 2015)
- Interesting interplay between natural language and vision
 - Closely related to language & vision and grounded language acquisition
 - Requires semantic understanding of each modality
- Has well-defined metric
- Interesting to NLP: unique characteristics of the geometry word problems.

Challenge #1

Interaction between Text and Diagram

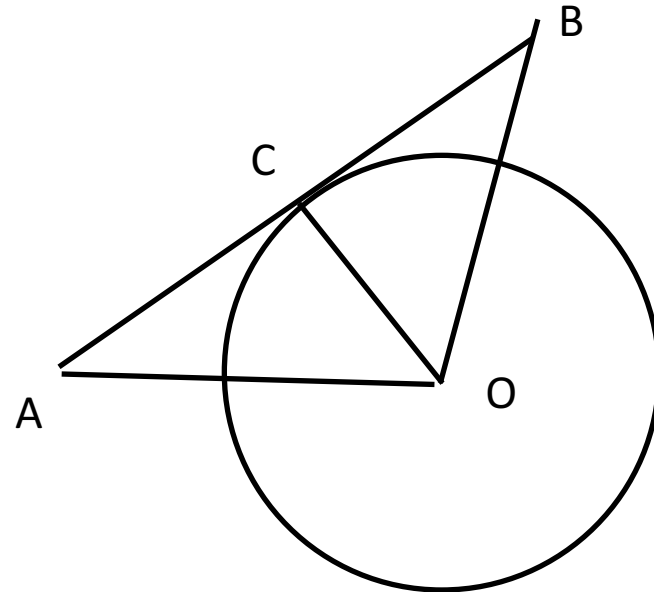
In the diagram at the right, the line is tangent to the circle.



- Previous work in semantic parsing and relation extraction does not consider another modality (Zettlemoyer and Collins, 2005; Kate and Mooney, 2007; Poon and Domingos, 2009; Kwiatkowski et al., 2013; Flanigan et al., 2014; Reddy et al., 2014; Berant et al., 2014; Cowie and Lehnert, 1996; Culotta and Sorensen, 2004)

Challenge #2: Lexical Ambiguity

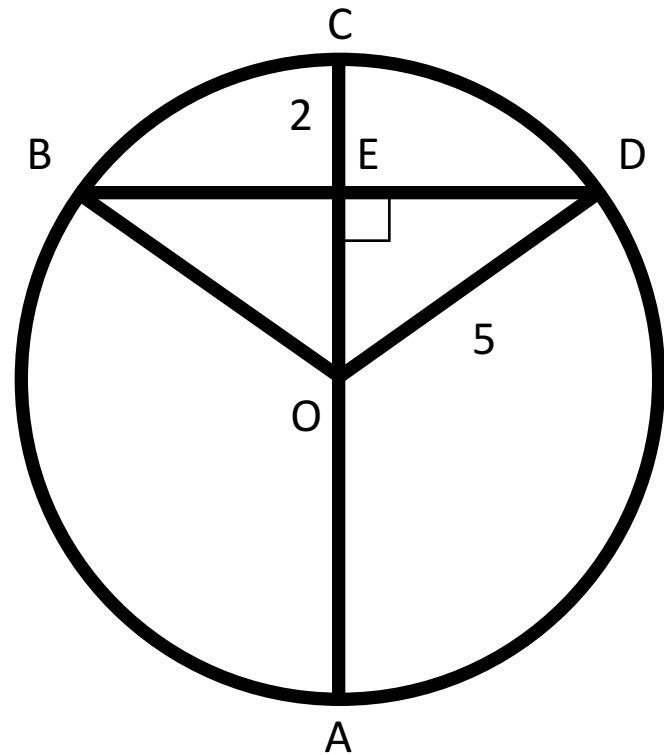
Line OC bisects line AB,
and line OC bisects angle
AOB.



Challenge #3: Implication

Circle O has a radius of 5.

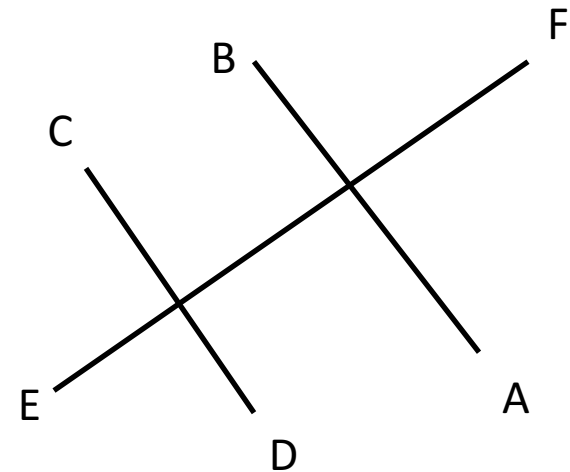
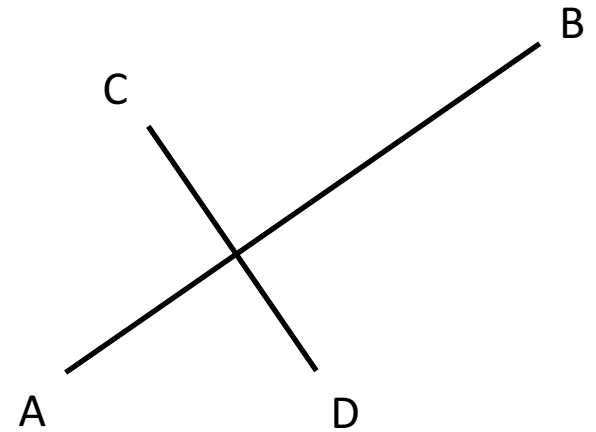
Equals (RadiusOf (O) , 5)



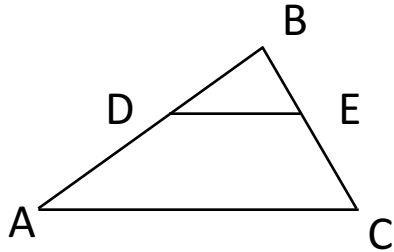
Challenge #4: Syntactic Complication

AB and CD are perpendicular.

AB and CD are perpendicular **to**
EF.



GeoS: Overview



$PointLiesOnLine(D, AB) \wedge$
 $PointLiesOnLine(E, BC)$

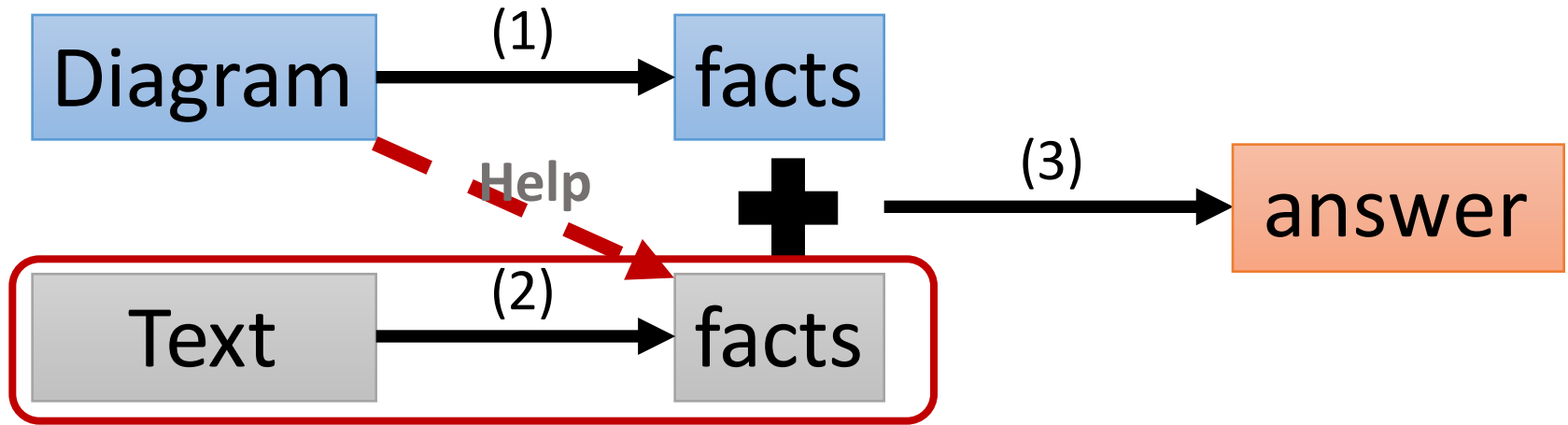
+ \rightarrow **(d) 15**

In triangle ABC, line DE is parallel with line AC, DB equals 4, AD is 8, and DE is 5. Find AC.

- (a) 9 (b) 10 (c) 12.5
(d) 15 (e) 17

$IsTriangle(ABC) \wedge$
 $Parallel(AC, DE) \wedge$
 $Equals(LengthOf(DB), 4) \wedge$
 $Equals(LengthOf(AD), 8) \wedge$
 $Equals(LengthOf(DE), 5) \wedge$
 $Find(LengthOf(AC))$

GeoS: Overview



(1) Diagram understanding (Seo et al., 2014)

(2) Text parsing

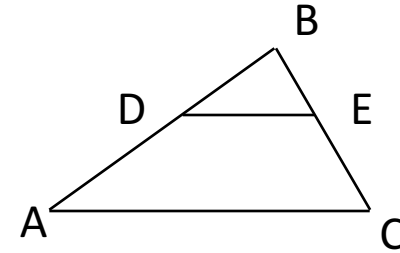
(3) Solving

Diagram-aided text parsing

*Text
Input*

In triangle ABC, line DE is parallel with line AC, DB equals 4, AD is 8, and DE is 5. Find AC.

(a) 9 (b) 10 (c) 12.5 (d) 15 (e) 17



*Logical
form*

$IsTriangle(ABC) \wedge$ $Parallel(AC, DE) \wedge$
 $Equals(LengthOf(DB), 4) \wedge Equals(LengthOf(AD), 8) \wedge$
 $Equals(LengthOf(DE), 5) \wedge Find(LengthOf(AC))$

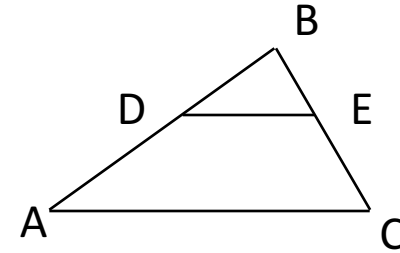
Difficult to directly map text to a long logical form!

Diagram-aided text parsing

Text
Input

In triangle ABC, line DE is parallel with line AC, DB equals 4, AD is 8, and DE is 5. Find AC.

(a) 9 (b) 10 (c) 12.5 (d) 15 (e) 17



Our
method

Over-generated literals

IsTriangle(ABC)

Parallel(AC, DE)

Parallel(AC, DB)

Equals(LengthOf(DB), 4)

Equals(LengthOf(AD), 8)

Equals(LengthOf(DE), 5)

Equals(4, LengthOf(AD))

...

Text scores

0.96

0.91

0.74

0.97

0.94

0.94

0.31

...

Diagram scores

1.00

0.99

0.02

n/a

n/a

n/a

n/a

...

Selected subset

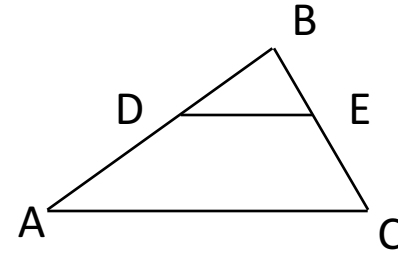
Logical
form

IsTriangle(ABC) \wedge Parallel(AC, DE) \wedge
 Equals(LengthOf(DB), 4) \wedge Equals(LengthOf(AD), 8) \wedge
 Equals(LengthOf(DE), 5) \wedge Find(LengthOf(AC))

Step 1. Literal over-generation

In triangle ABC, line DE is parallel with line AC, DB equals 4, AD is 8, and DE is 5. Find AC.

(a) 9 (b) 10 (c) 12.5 (d) 15 (e) 17



Over-generated literals

IsTriangle(ABC)
Parallel(AC, DE)
Parallel(AC, DB)
Equals(LengthOf(DB), 4)
Equals(LengthOf(AD), 8)
Equals(LengthOf(DE), 5)
Equals(4, LengthOf(AD))
...

Step 1. Generating literals

“Lines AB and CD are perpendicular to EF”



```
IsLine(AB)
IsLine(CD)
IsLine(EF)
Perpendicular(AB, CD)
Perpendicular(CD, EF)
Perpendicular(AB, EF)
```

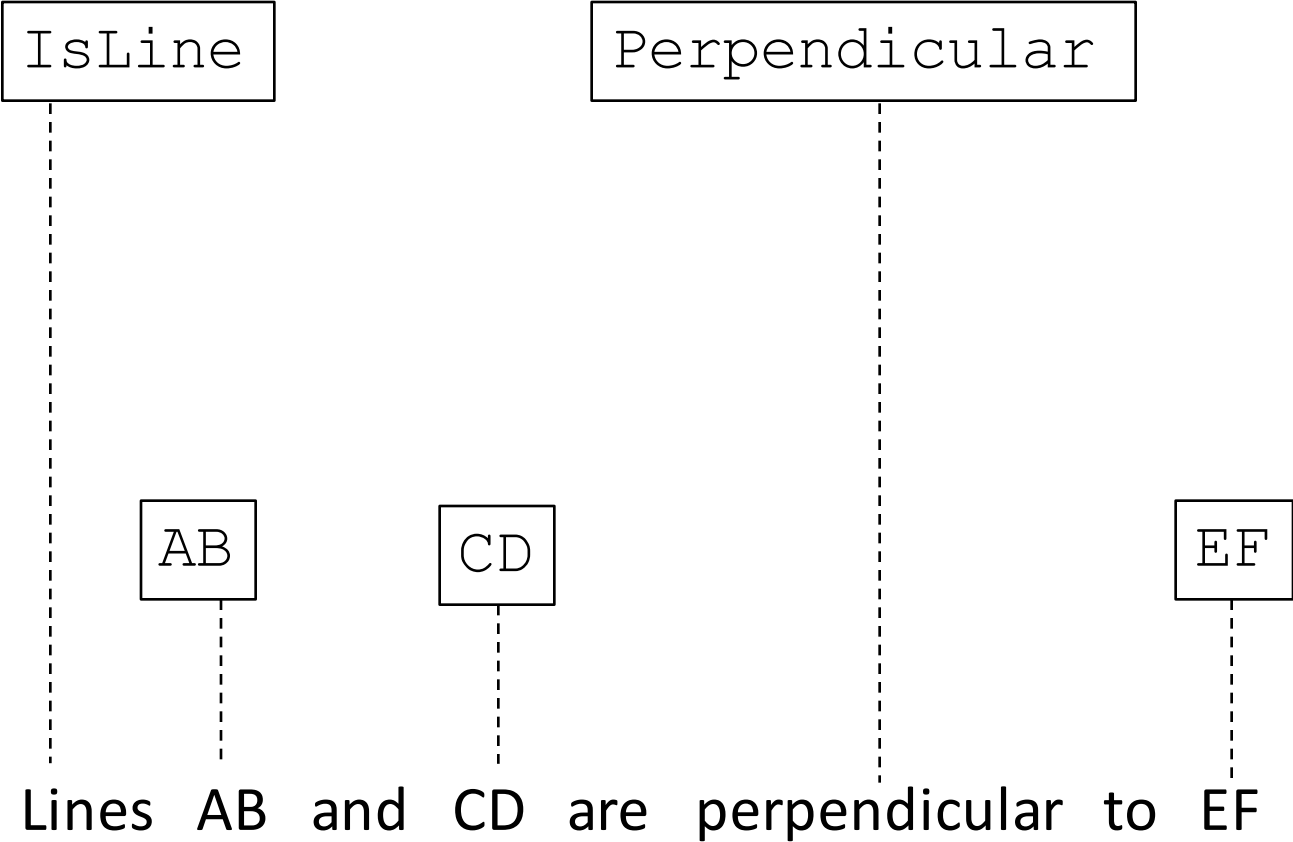

Step 1. Generating literals

“Lines AB and CD are perpendicular to EF”

`IsLine (AB)`
`IsLine (CD)`
`IsLine (EF)`
`Perpendicular (AB, CD)`
`Perpendicular (CD, EF)`
`Perpendicular (AB, EF)`

*Red literals
are false.*

Concepts

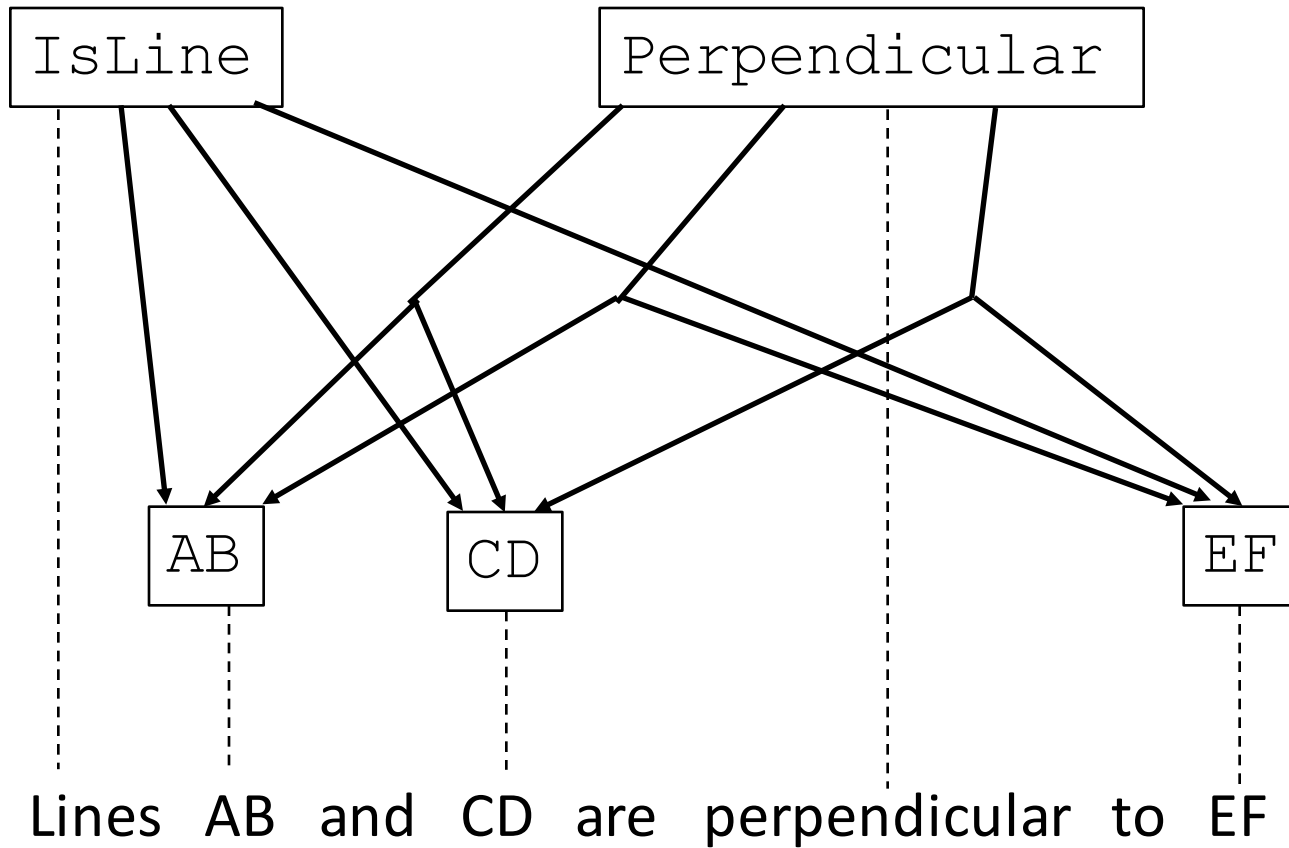


Lexicon

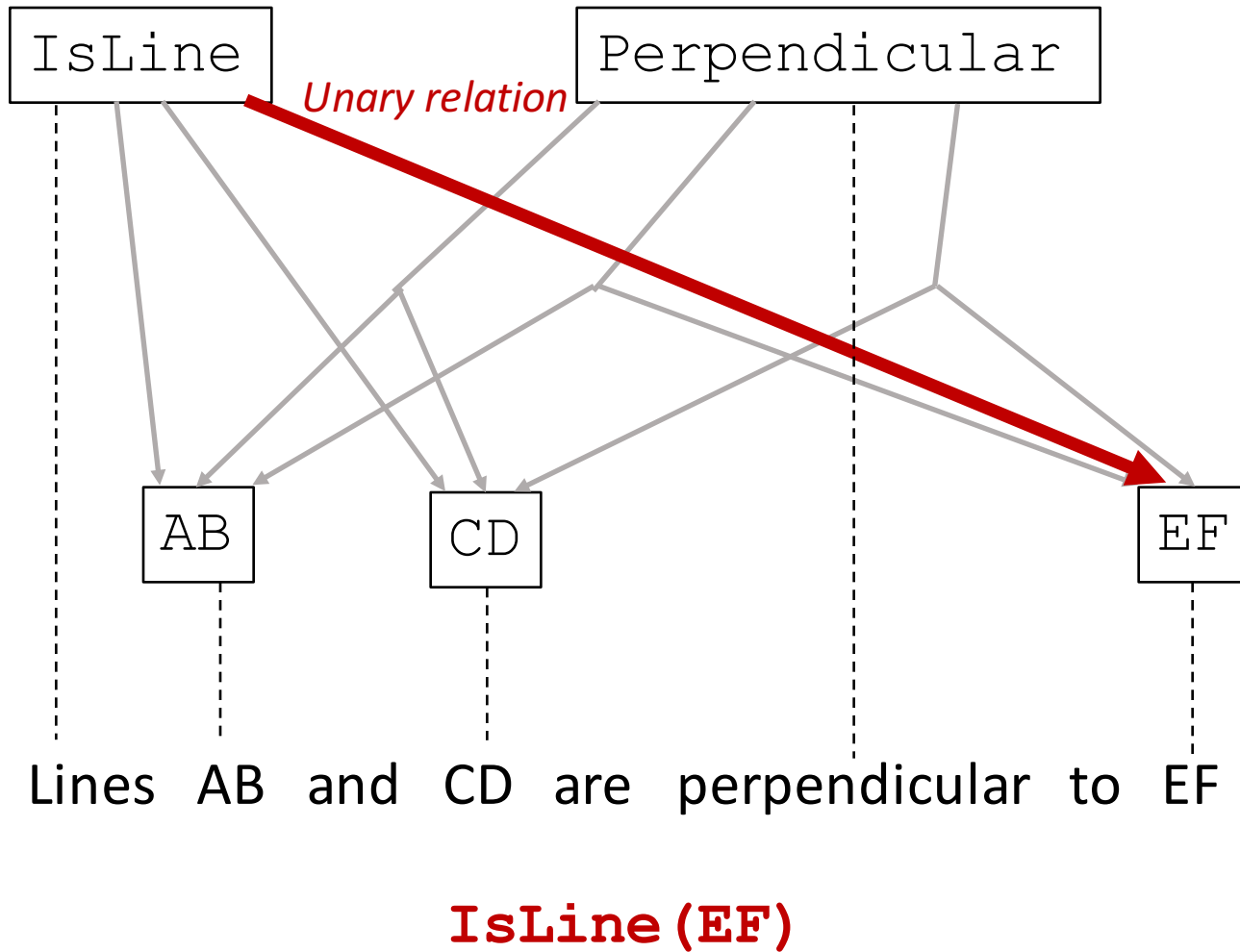
- We built lexicon from training data and textbooks
- Lexicon maps geometry-related words (or phrases) to *concepts*
- Some concepts are obtained via simple regular expressions
- Single word can map to two or more concepts

Word or phrase	Concept
“Perpendicular”	Perpendicular
“Lies on”	PointLiesOnLine, PointLiesOnCircle
“CD”	line, arc
“ABC”	triangle, angle

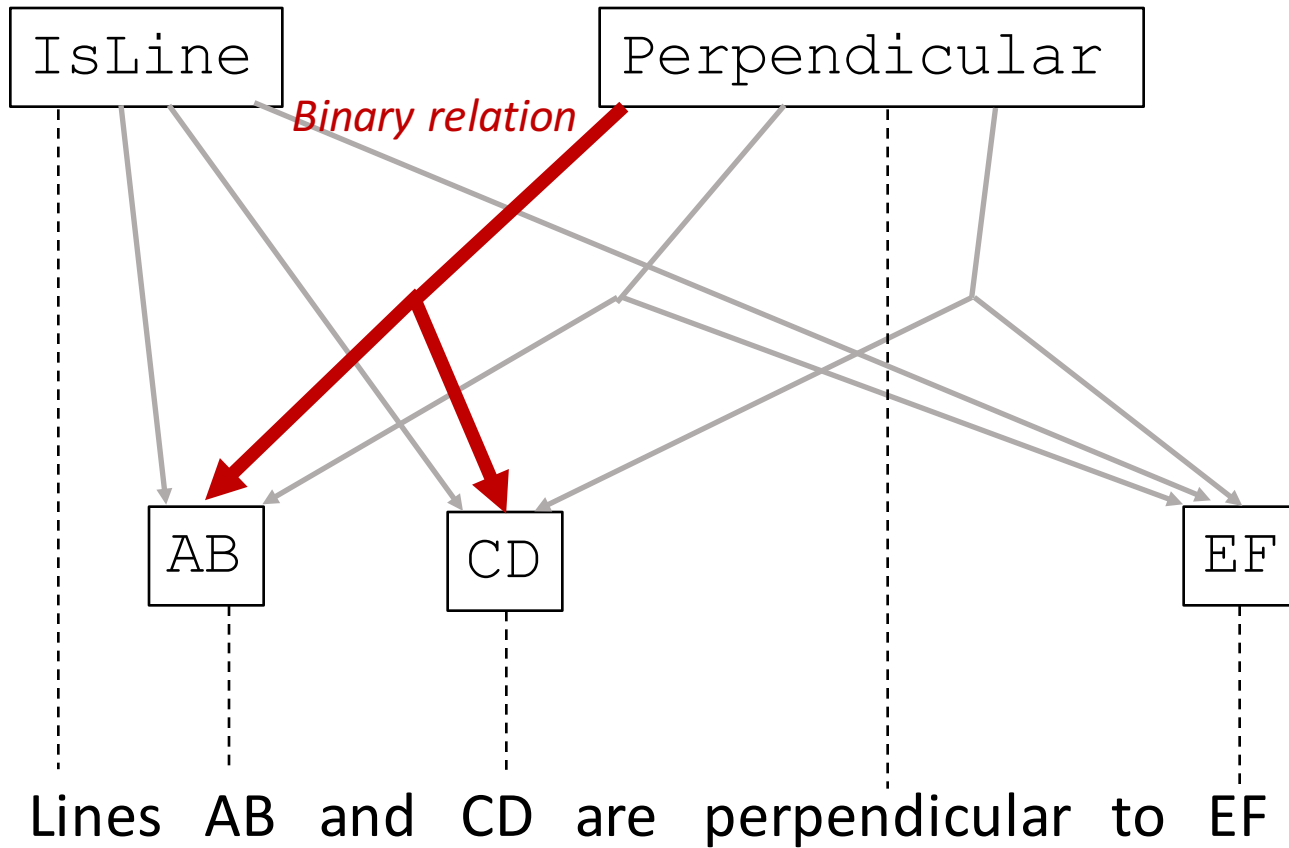
Relations



Relations



Relations

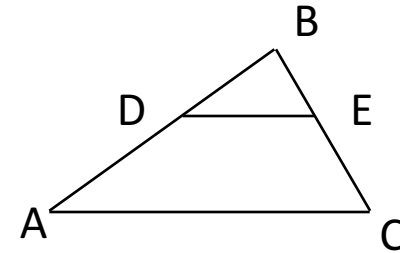


Perpendicular (AB, CD)

Step 2. Text scores of literals

In triangle ABC, line DE is parallel with line AC, DB equals 4, AD is 8, and DE is 5. Find AC.

(a) 9 (b) 10 (c) 12.5 (d) 15 (e) 17



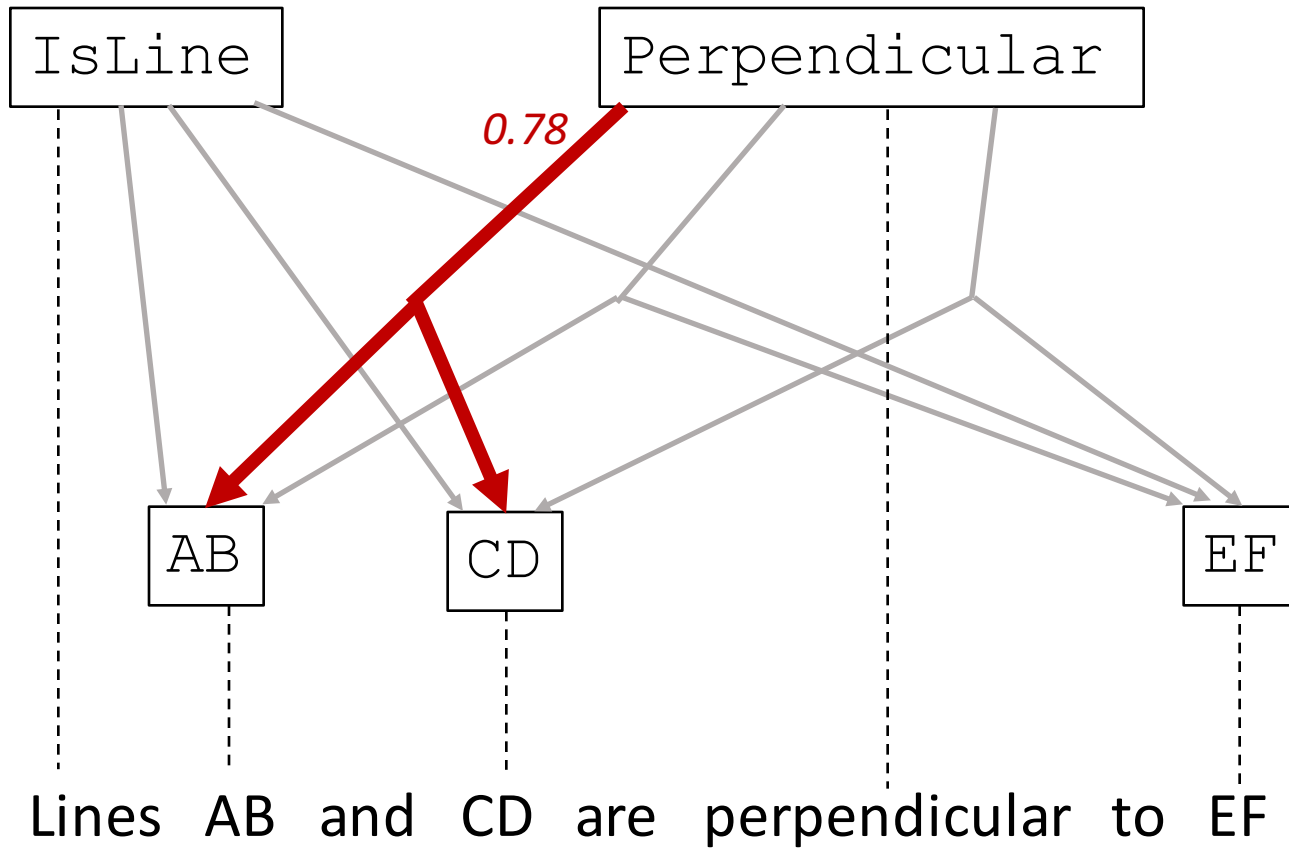
Over-generated literals

IsTriangle(ABC)
Parallel(AC, DE)
Parallel(AC, DB)
Equals(LengthOf(DB), 4)
Equals(LengthOf(AD), 8)
Equals(LengthOf(DE), 5)
Equals(4, LengthOf(AD))
...

Text scores

0.96
0.91
0.74
0.97
0.94
0.94
0.31
...

Relation score

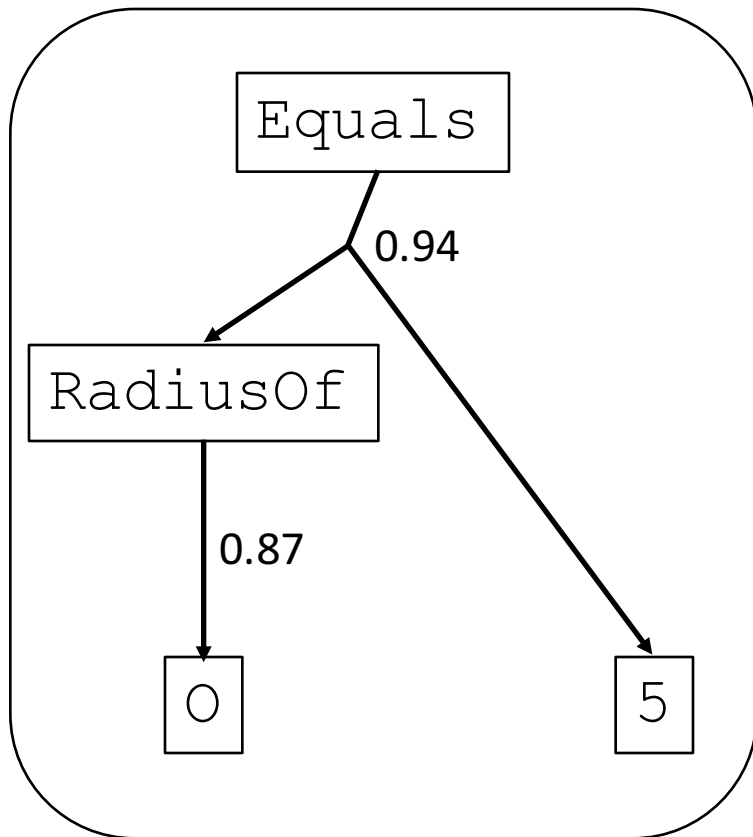


Relation classification

- **Supervision:** annotated logical forms
- **Training data:** all possible relations from training questions
 - Relations found in annotations: **positive**
 - All others: **negative**
- Logistic regression with L2 regularization
- Features:
 - Stanford dependency parse
 - Part of speech tags
 - Type of concept (line, circle, triangle, predicate, etc.)

```
IsLine->AB  
IsLine->CD  
IsLine->EF  
Perpendicular->AB, CD  
Perpendicular->CD, EF  
Perpendicular->AB, EF
```

Text scores of literals



“Circle O has radius of 5”

$$\mathcal{A}_{text}(l) = \sum \log P_{\theta}(y_i = 1 | r_i, t)$$

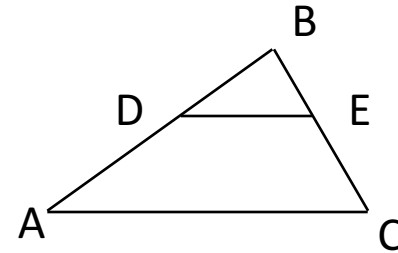
- l Literal
- y_i Label for edge
- r_i Edge (relation)
- t Question text

- θ Logistic regression parameters to be learned

Step 3. Diagram scores of literals

In triangle ABC, line DE is parallel with line AC, DB equals 4, AD is 8, and DE is 5. Find AC.

(a) 9 (b) 10 (c) 12.5 (d) 15 (e) 17



Over-generated literals

<i>IsTriangle(ABC)</i>	0.96
<i>Parallel(AC, DE)</i>	0.91
<i>Parallel(AC, DB)</i>	0.74
<i>Equals(LengthOf(DB), 4)</i>	0.97
<i>Equals(LengthOf(AD), 8)</i>	0.94
<i>Equals(LengthOf(DE), 5)</i>	0.94
<i>Equals(4, LengthOf(AD))</i>	0.31
...	...

Text scores

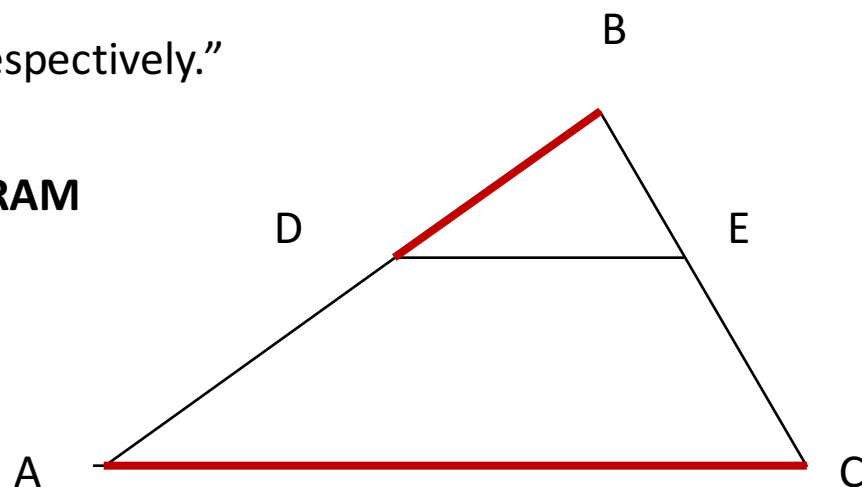
Diagram scores

1.00
0.99
0.02
n/a
n/a
n/a
n/a
...

Step 3. Diagram scores of literals

“AC and DB are parallel with DE and AD, respectively.”

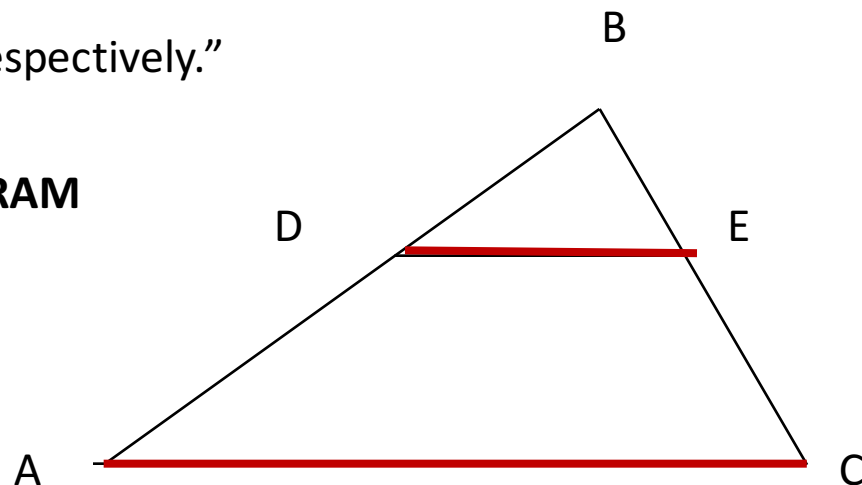
	TEXT	DIAGRAM
<code>Parallel (AC, DB)</code>	0.74	0.02



Step 3. Diagram scores of literals

“AC and DB are parallel with DE and AD, respectively.”

	TEXT	DIAGRAM
<code>Parallel(AC, DB)</code>	0.74	0.02
<code>Parallel(AC, DE)</code>	0.78	0.99



Step 3. Diagram scores of literals

“AC and DB are parallel with DE and AD, respectively.”

	TEXT	DIAGRAM
<code>Parallel (AC, DB)</code>	0.74	0.02
<code>Parallel (AC, DE)</code>	0.78	0.99

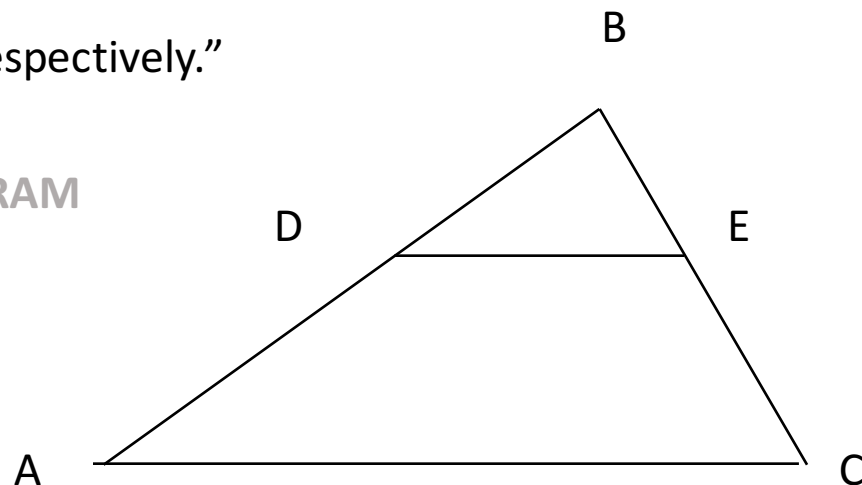
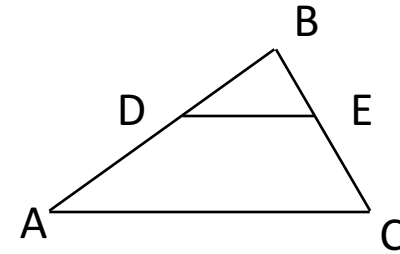


Diagram understanding in geometry questions (Seo et al., 2014)

Step 4. Subset selection

In triangle ABC, line DE is parallel with line AC, DB equals 4, AD is 8, and DE is 5. Find AC.

(a) 9 (b) 10 (c) 12.5 (d) 15 (e) 17



Over-generated literals

IsTriangle(ABC)
Parallel(AC, DE)
Parallel(AC, DB)
Equals(LengthOf(DB), 4)
Equals(LengthOf(AD), 8)
Equals(LengthOf(DE), 5)
Equals(4, LengthOf(AD))
 ...

Text scores

0.96
 0.91
 0.74
 0.97
 0.94
 0.94
 0.31
 ...

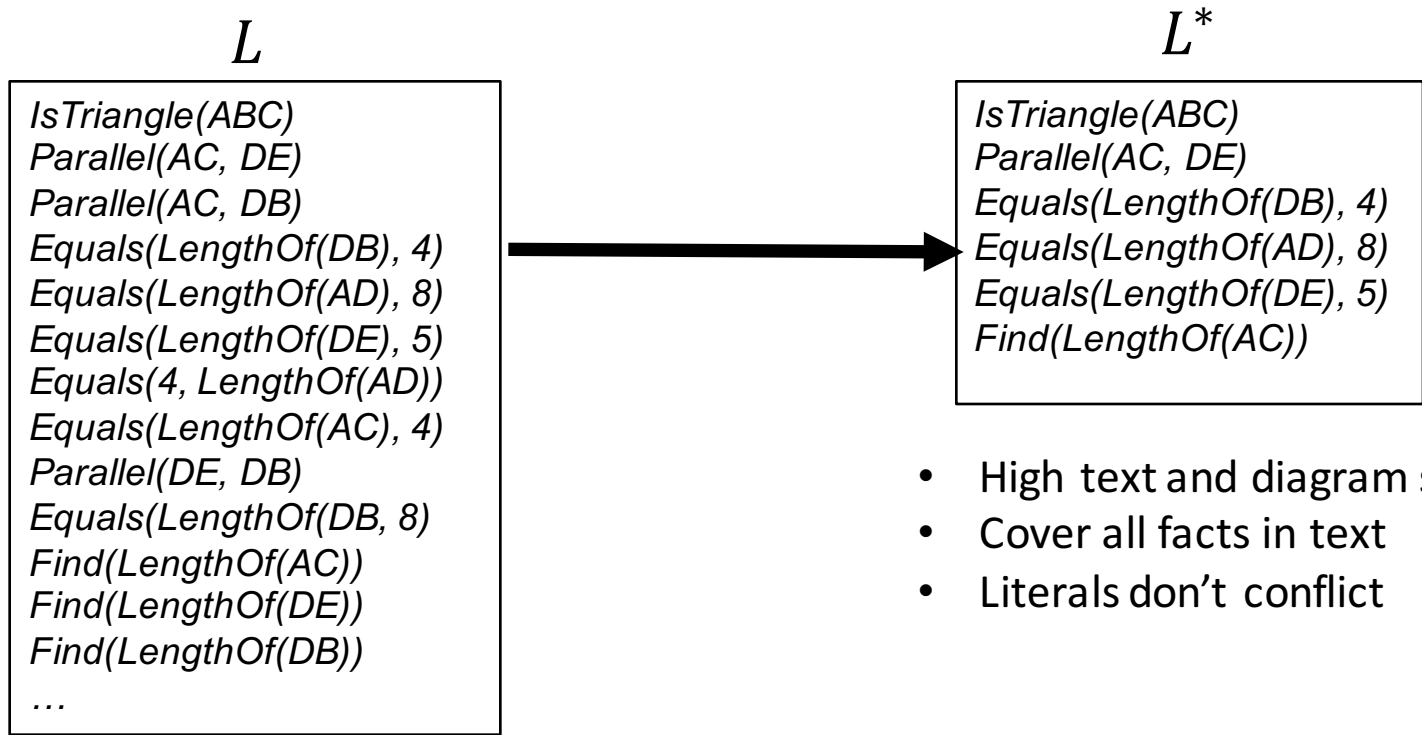
Diagram scores

1.00
 0.99
 0.02
 n/a
 n/a
 n/a
 n/a
 ...

Selected subset

IsTriangle(ABC) \wedge *Parallel(AC, DE)* \wedge
Equals(LengthOf(DB), 4) \wedge *Equals(LengthOf(AD), 8)* \wedge
Equals(LengthOf(DE), 5) \wedge *Find(LengthOf(AC))*

Step 4. Subset selection



- High text and diagram scores
- Cover all facts in text
- Literals don't conflict

$$L^* = \operatorname{argmax}_{L' \subset L} \mathcal{F}(L')$$

$$\mathcal{F}(L') = \underbrace{\lambda \mathcal{A}(L', t, d)}_{\substack{\text{Affinity} \\ \text{(diagram+text)}}} + \underbrace{\mathcal{H}(L', t)}_{\substack{\text{Coherence} \\ \text{(covers all facts,} \\ \text{no conflict)}}$$

Optimization algorithm

$$L^* = \operatorname{argmax}_{L' \subset L} \underbrace{\lambda \mathcal{A}(L', t, d)}_{\text{affinity}} + \underbrace{\mathcal{H}(L', t)}_{\text{coherence}}$$

Bad news: combinatorial optimization is NP-hard

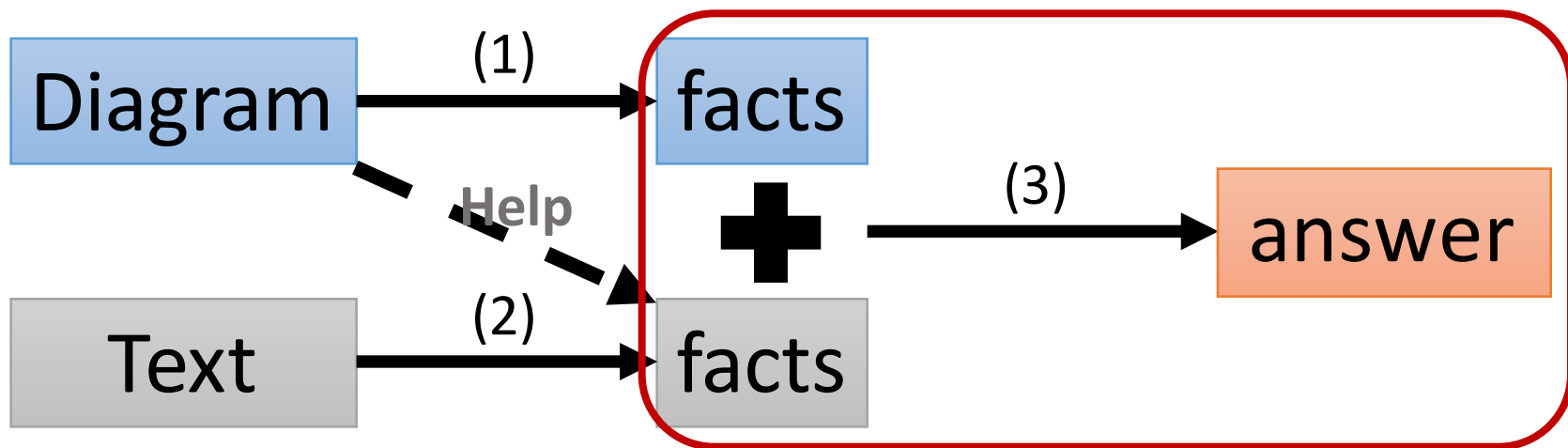
Good news: objective function is *submodular*

Greedy algorithm efficiently finds a solution with bounded distance to the optimum.

Starting from empty set, greedily add the next best literal to the set.

$$l_j = \operatorname{argmax}_{l_j \in L \setminus L'} \mathcal{F}(L' \cup \{l_j\}) - \mathcal{F}(L')$$

Solving



Numerical solver

- Translate literals to numeric equations

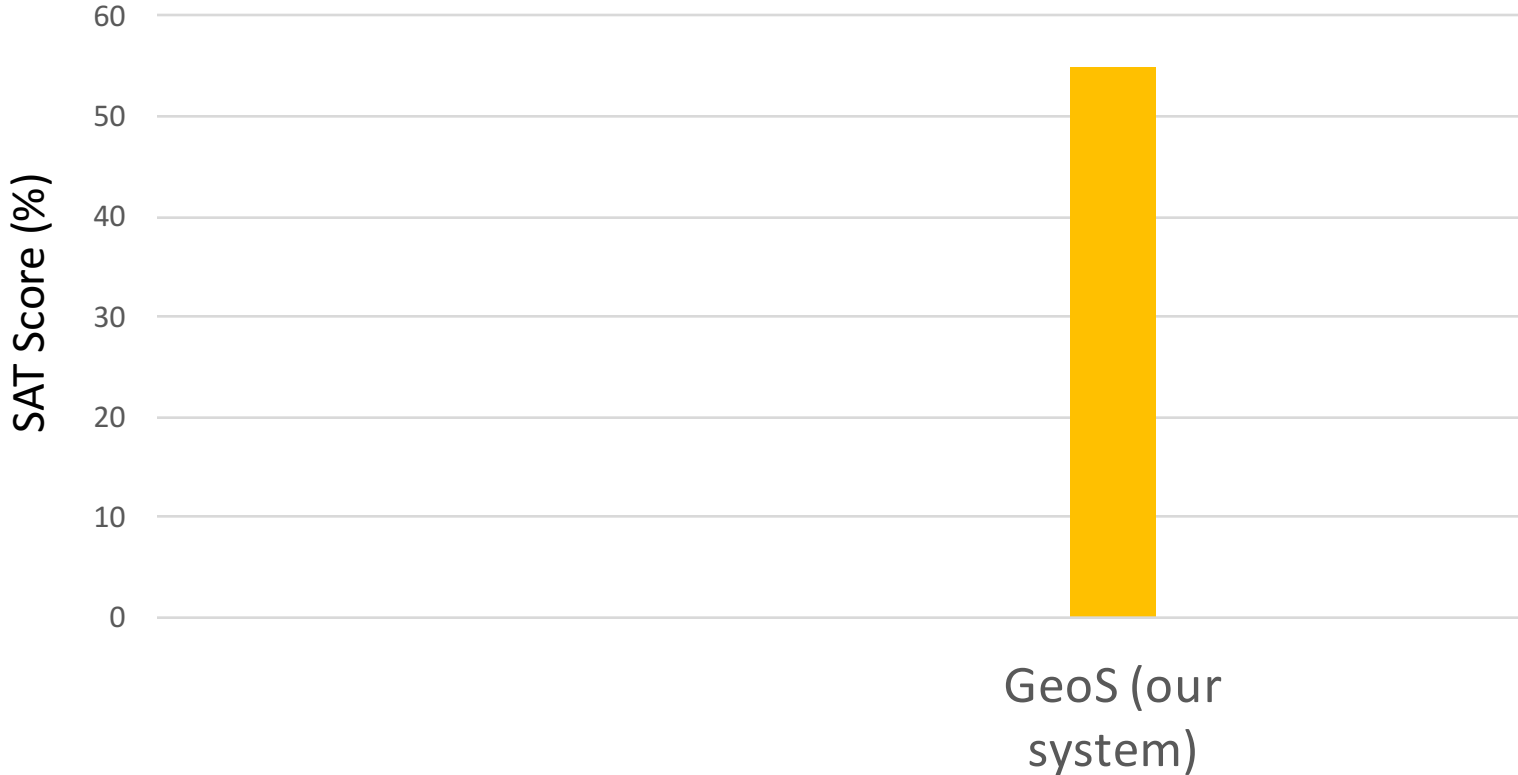
Literal	Equation
Equals(LengthOf(AB),d)	$(A_x - B_x)^2 + (A_y - B_y)^2 - d^2 = 0$
Parallel(AB, CD)	$(A_x - B_x)(C_y - D_y) - (A_y - B_y)(C_x - D_x) = 0$
PointLiesOnLine(B, AC)	$(A_x - B_x)(B_y - C_y) - (A_y - B_y)(B_x - C_x) = 0$
Perpendicular(AB,CD)	$(A_x - B_x)(C_x - D_x) + (A_y - B_y)(C_y - D_y) = 0$

- Find the solution to the equation system
- Use off-the-shelf numerical minimizers (Wales and Doye, 1997; Kraft, 1988)
- Numerical solver can choose not to answer question

Dataset

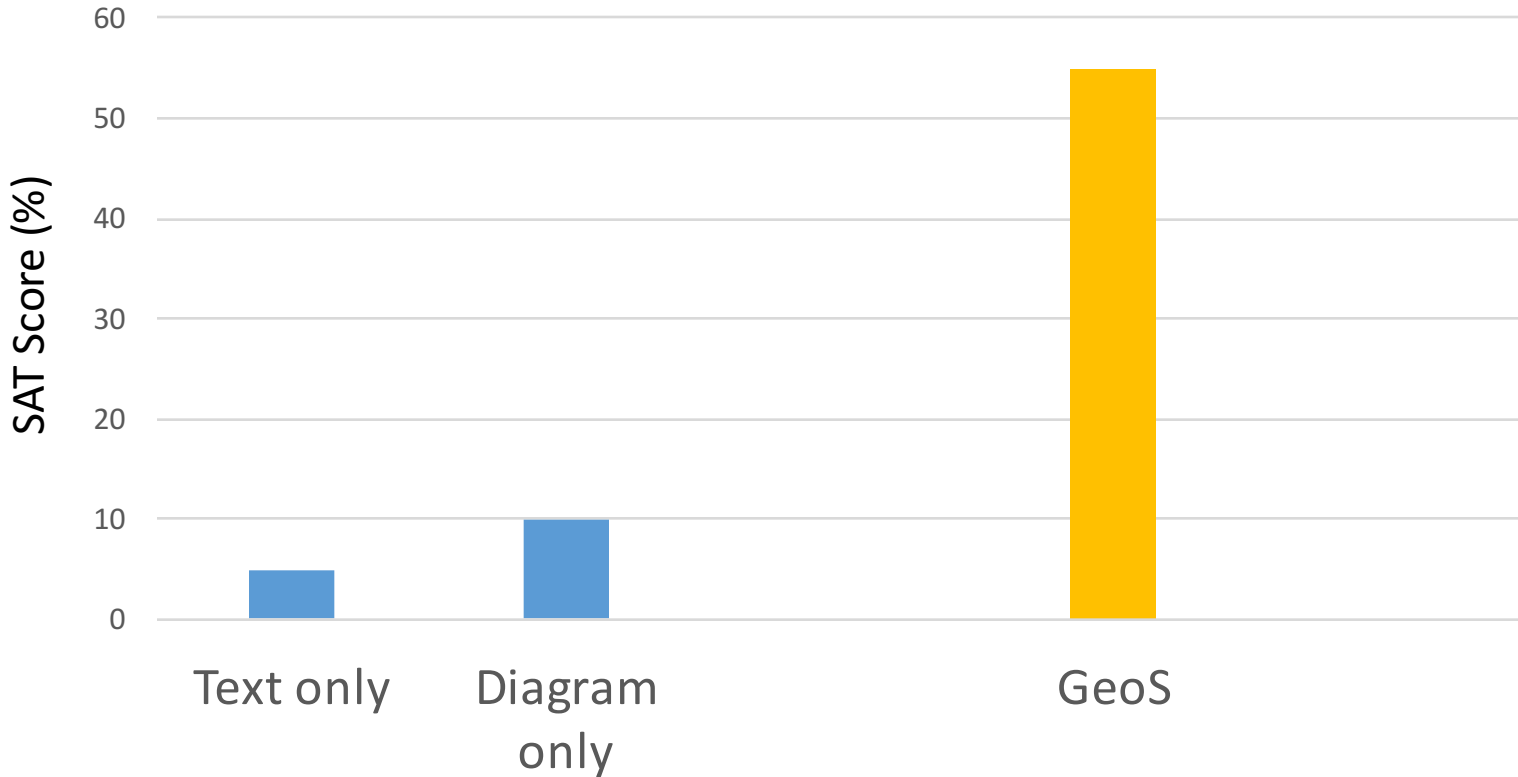
- **Training questions** (67 questions, 121 sentences)
 - Seo et al., 2014
 - High school geometry questions
- **Test questions** (119 questions, 215 sentences)
 - We collected them
 - SAT (US college entrance exam) geometry questions
- We manually annotated the text parse of all questions
- Dataset is publicly available at:
`geometry.allenai.org`

Experiment 1: answering questions



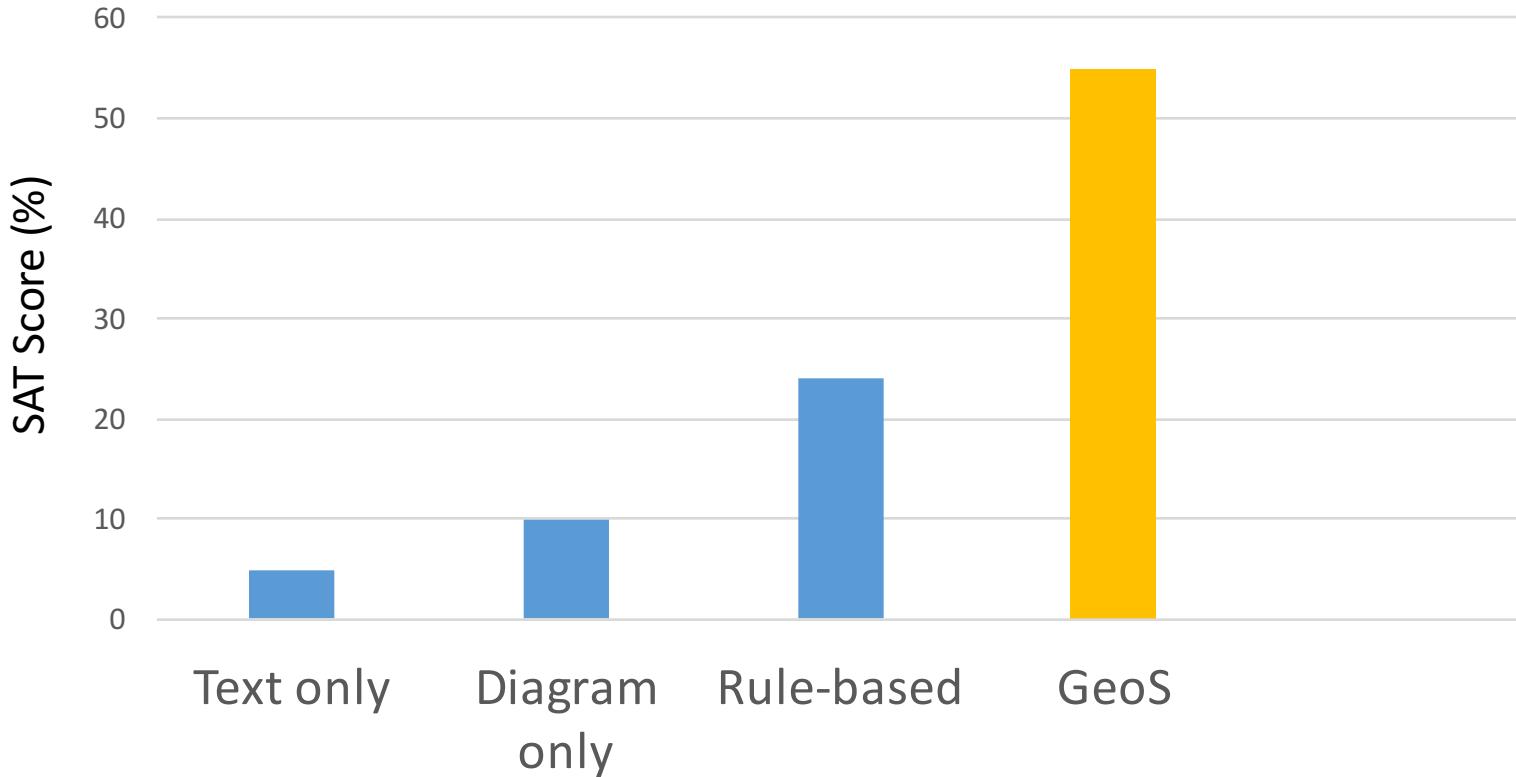
*** 0.25 penalty for incorrect answer

Experiment 1: answering questions



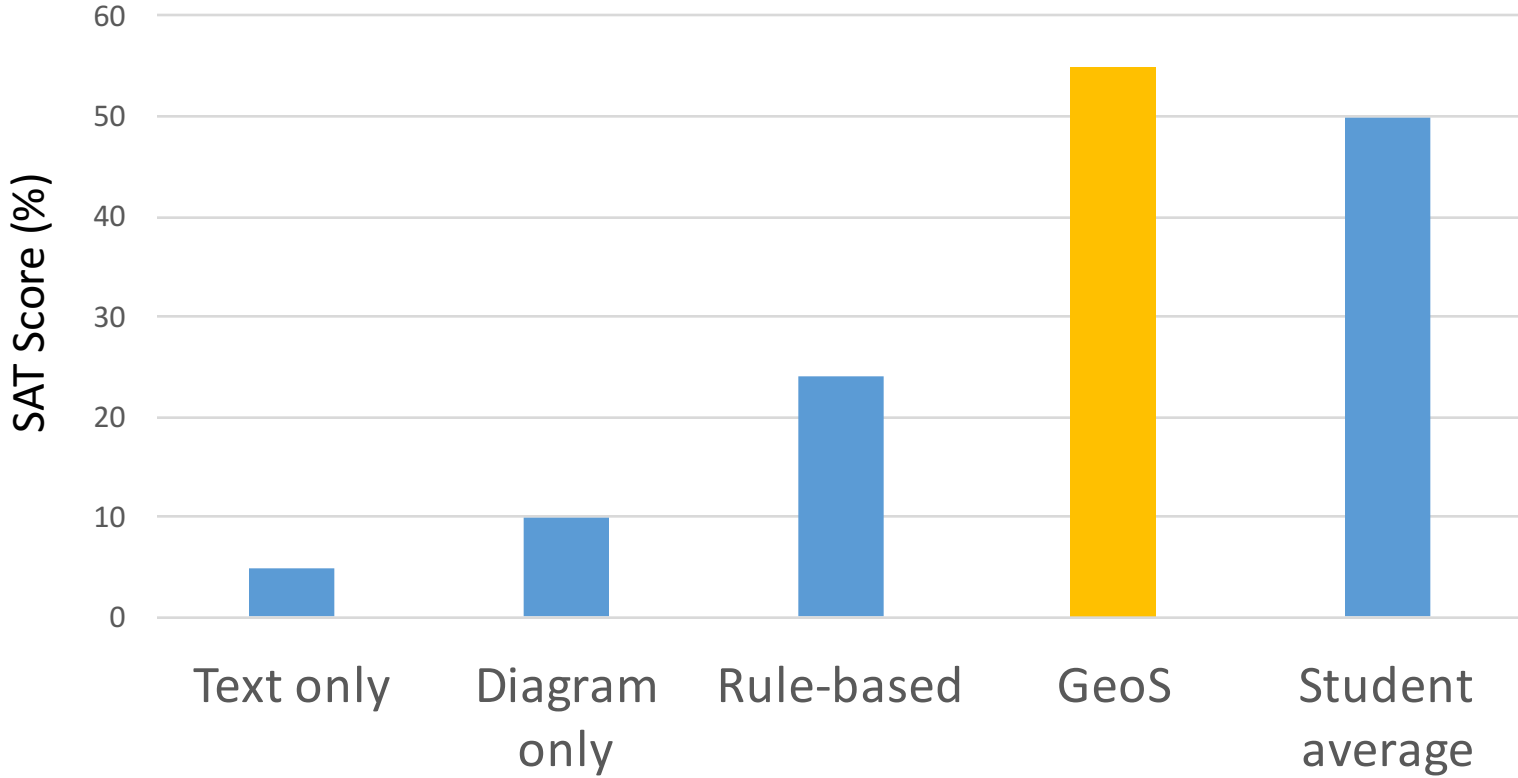
*** 0.25 penalty for incorrect answer

Experiment 1: answering questions



*** 0.25 penalty for incorrect answer

Experiment 1: answering questions

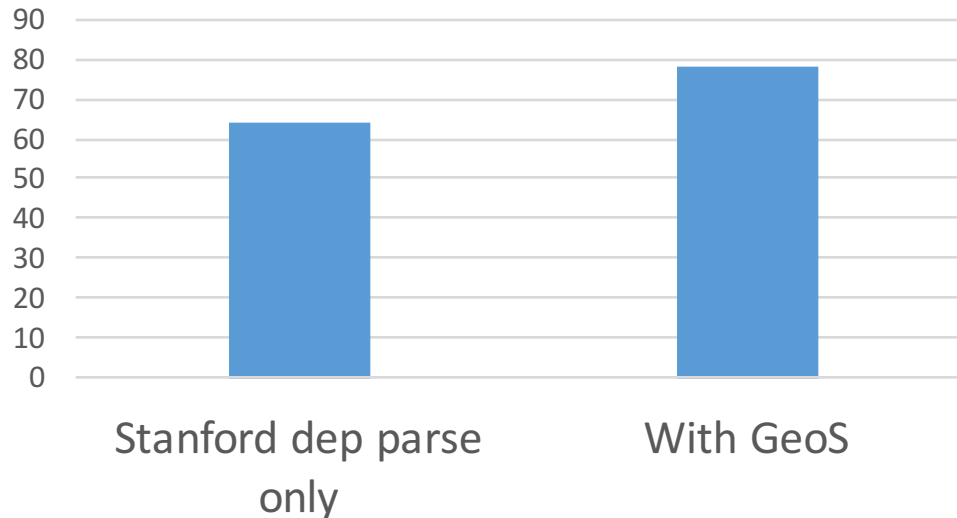
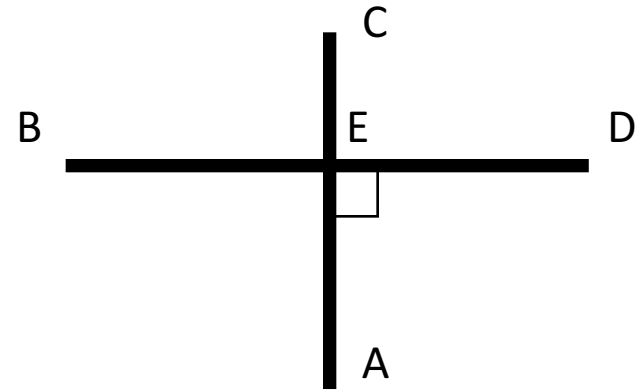


*** 0.25 penalty for incorrect answer

Experiment 2: Improving dependency parsing

“BD is perpendicular to AC at point E.”

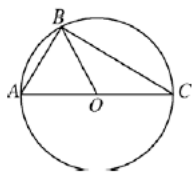
Obtain top-k dependency parses, and re-rank them based on GeoS result



Demo (geometry.allenai.org/demo)



In the figure to the left, triangle ABC is inscribed in the circle with center O and diameter AC. If $AB=AO$, what is the degree measure of angle ABO?



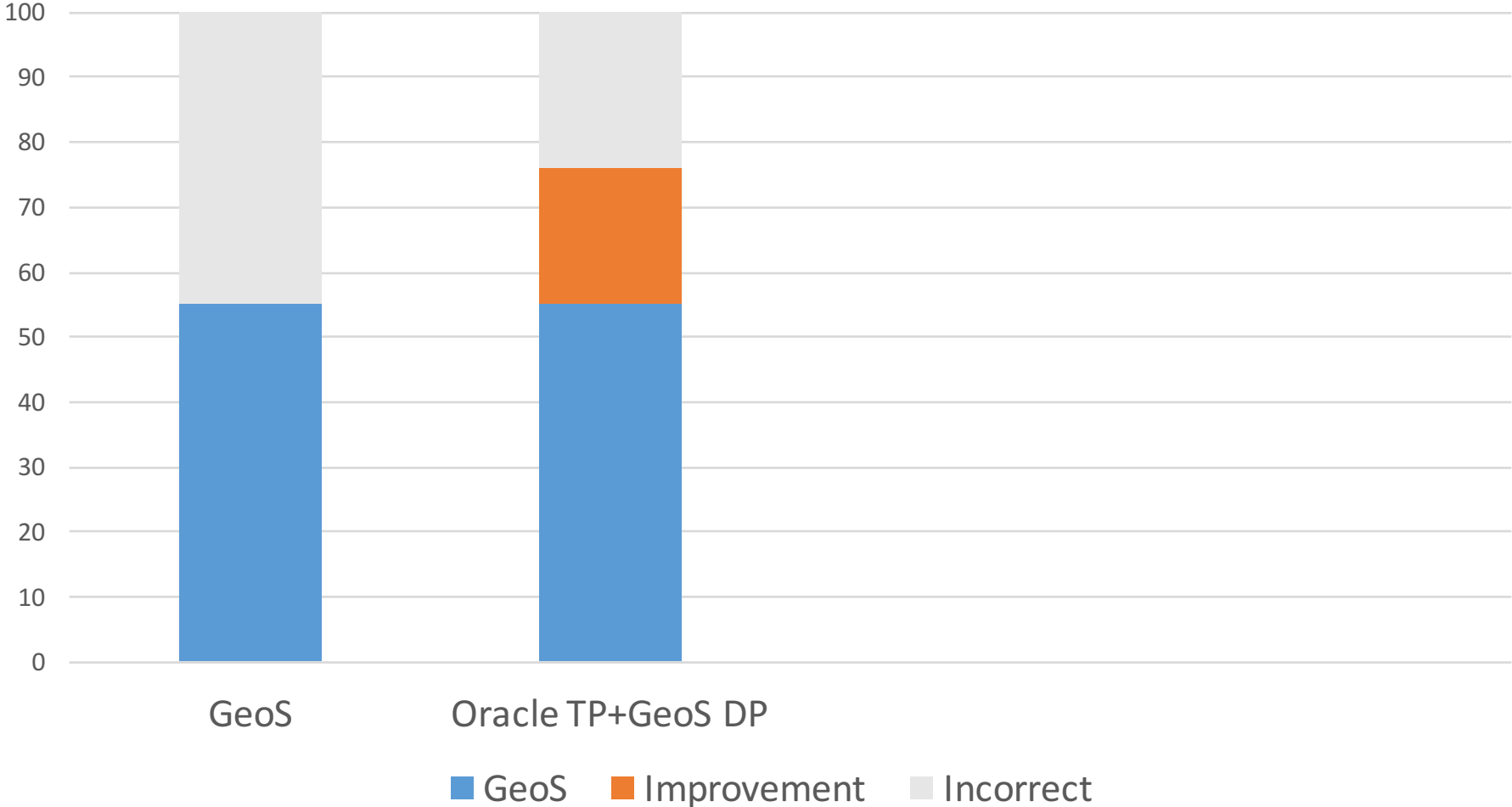
- (A) 15°
- (B) 30°
- (C) 45°
- (D) 60°
- (E) 90°

Solve Problem

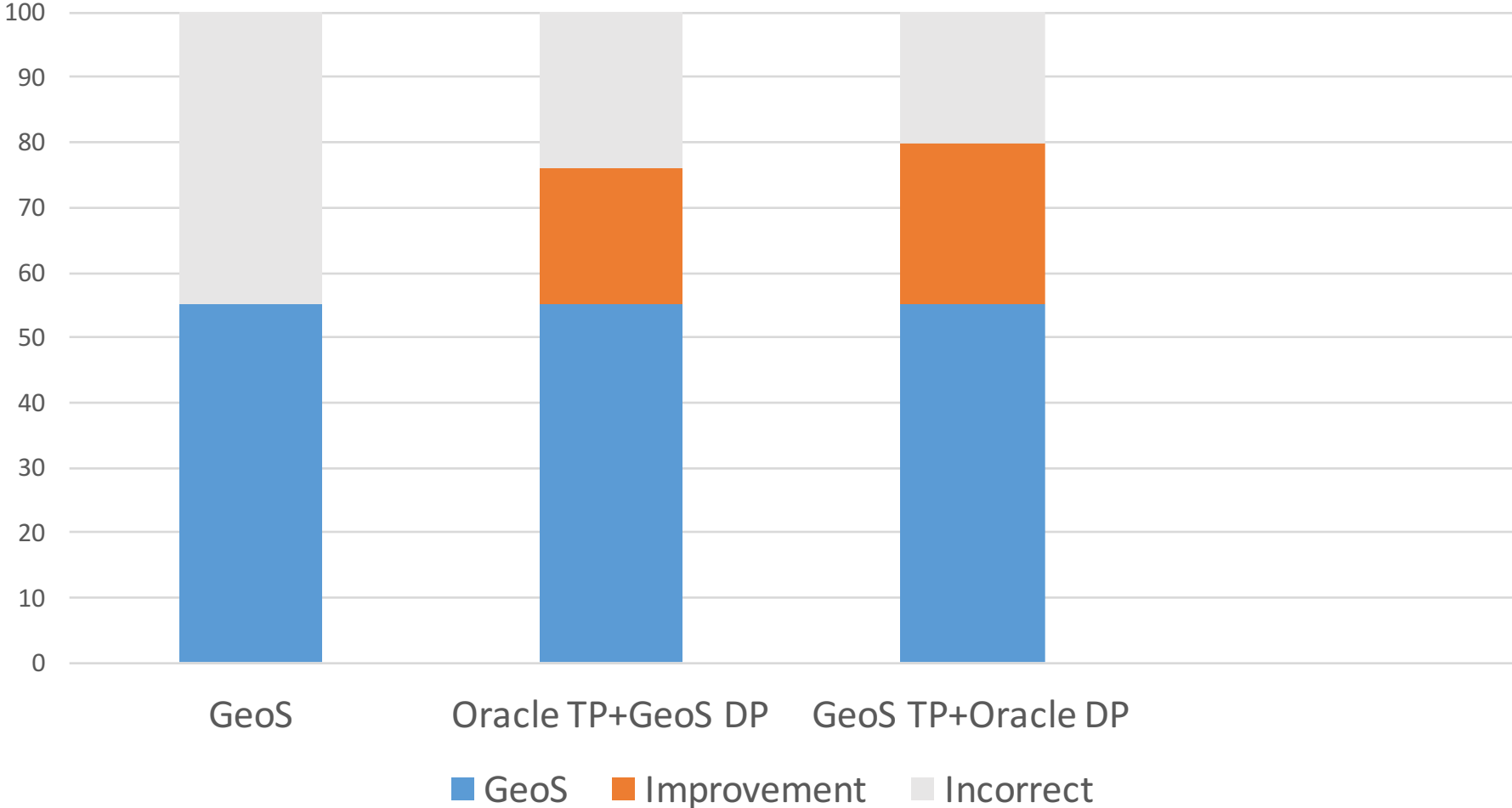
Oracle Studies



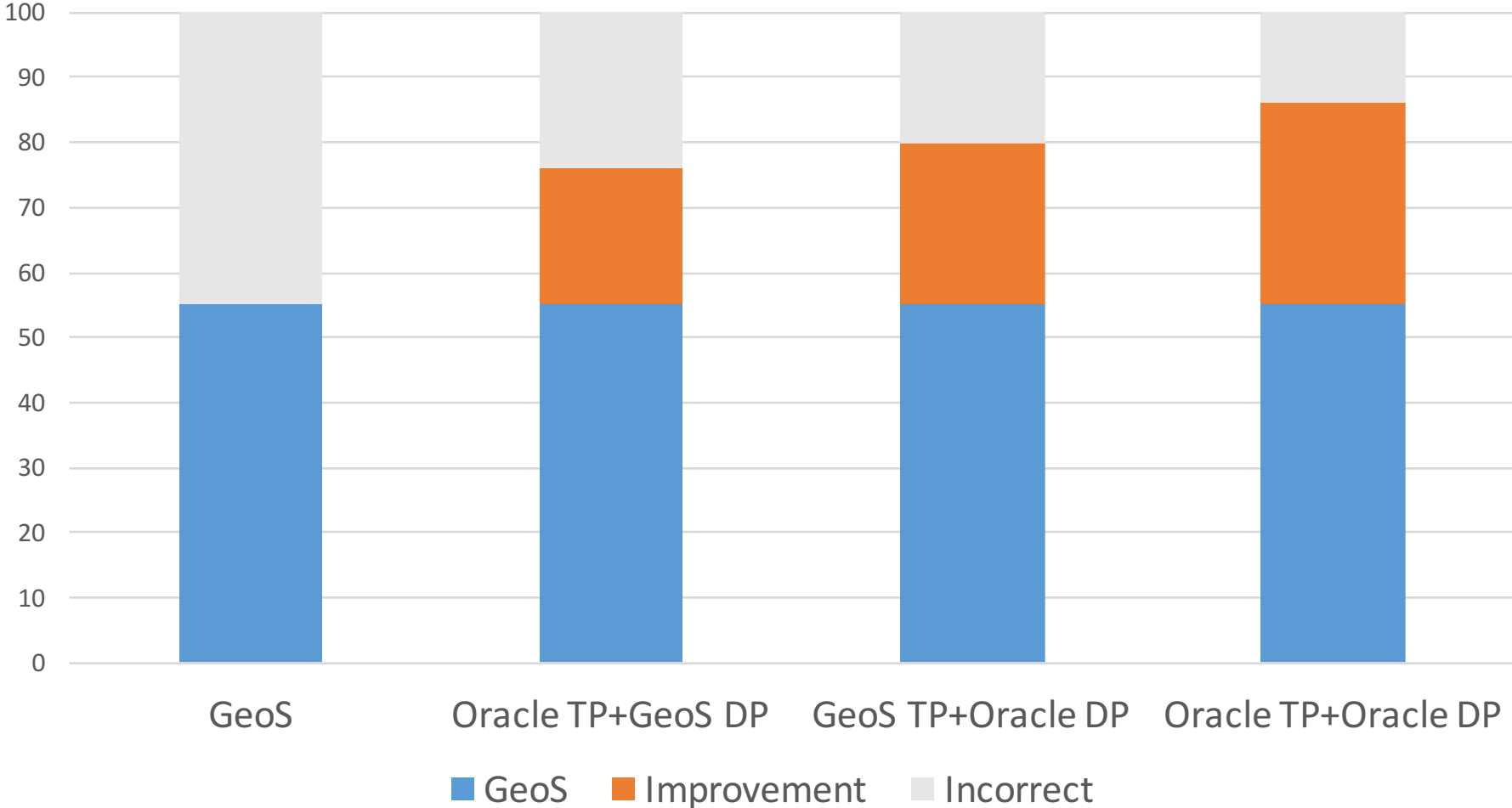
Oracle Studies



Oracle Studies

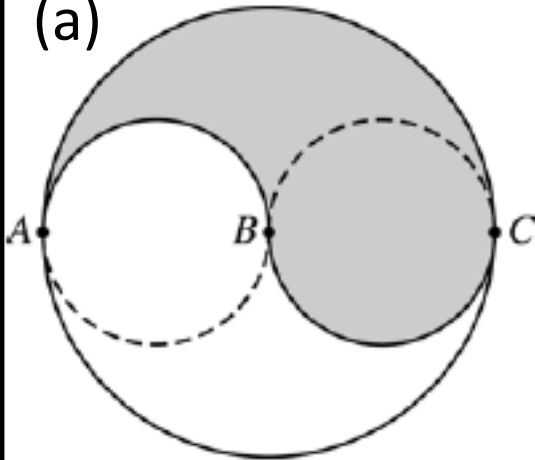


Oracle Studies



Failure Modes

(a)

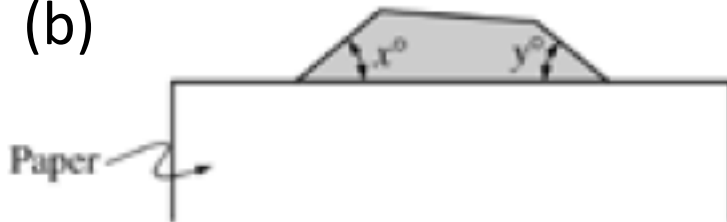


In the figure at the left, the smaller circles each have radius 3. **They** are tangent to the larger circle at points A and C, and are tangent to **each other** at point B, which is the center of the larger circle. What is the perimeter of the shaded region?

Fails to resolve "they" to "each other"

(a) 6π (b) 8π (c) 9π (d) 8π (e) 15π

(b)



***Requires complex reasoning:
Cannot understand that the polygon
is "hidden"***

In the figure at the left, a shaded polygon which has equal angles is partially covered with a sheet of blank paper. If $x+y=80$, how many sides does the polygon have?

(a) 10 (b) 9 (c) 8 (d) 7 (e) 6

Summary

- First end-to-end system for solving high school geometry problems
- Achieved 55% on official and practice SAT geometry questions
- Text parsing in the presence of diagram

Future work

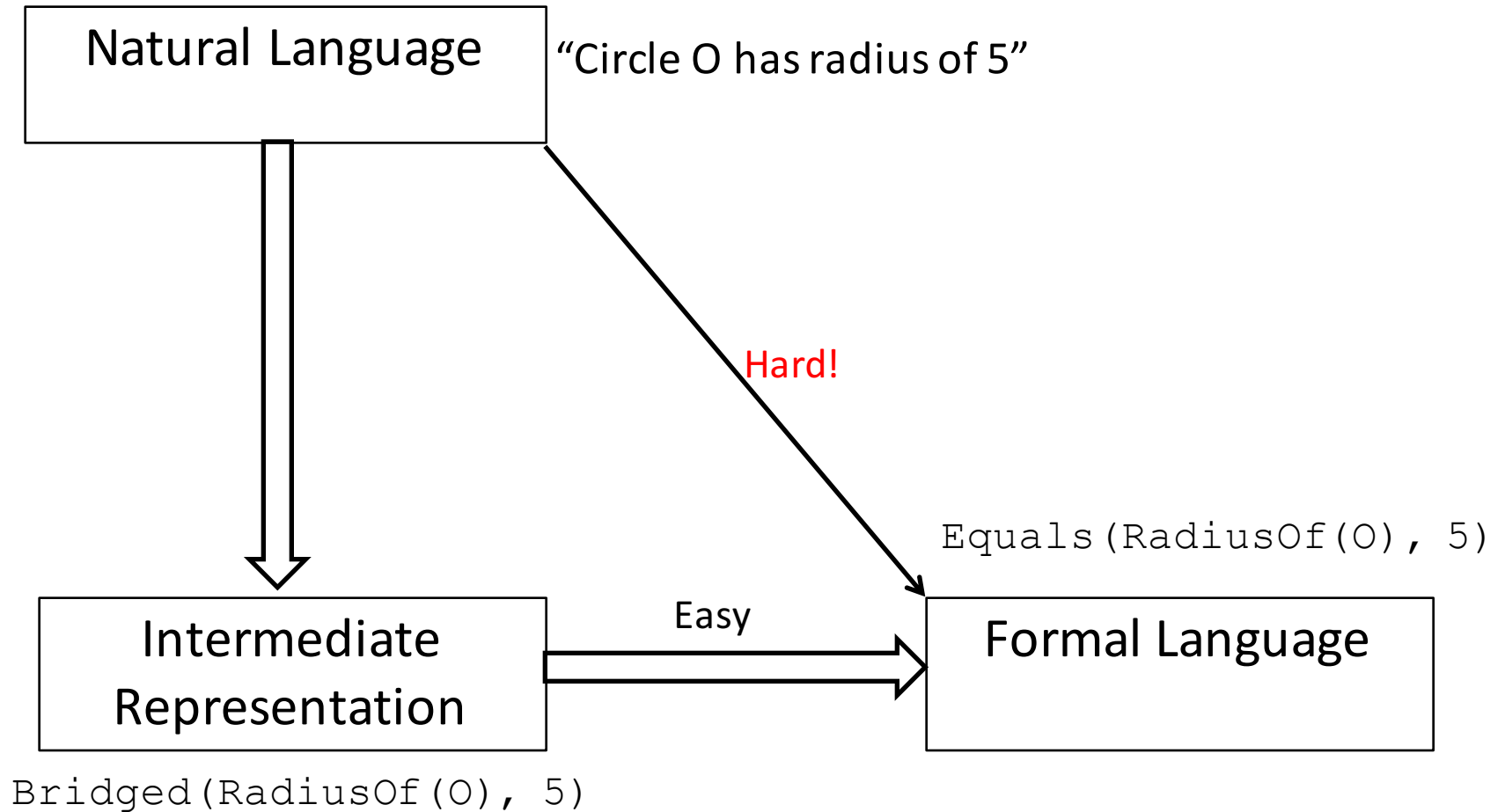
- Expand text parsing algorithm to other grounded language acquisition domains
- Improvements in solving geometry problems:
 - Increase data size
 - Weakly-supervised learning
 - More interaction between text and diagram
 - More transparent solver
 - Numerical solver is black box
 - Logical solver: gives more feedback from the solution

Thank you!

For more information, please visit:

`geometry.allenai.org`

Two-stage parsing



(Kwiatkowski et al., 2013)

Two-stage parsing: examples

Natural language “Circle O has a radius of 5.”

Intermediate `Bridged(RadiusOf(O), 5)`

Formal language `Equals(RadiusOf(O), 5)`

Natural language “AM and CM bisect BAC and BCA.”

Intermediate `Bisects(AM, BAC) ^
CC(AM, CM) ^ CC(BAC, BCA)`

Formal language `Bisects(AM, BAC) ^
Bisects(CM, BCA)`

Affinity score function

- Each literal has text score and diagram score
- Affinity score is the sum of text and diagram scores of literals

$$\mathcal{A}(L', t, d) = \sum_{l'_j \in L'} [\mathcal{A}_{text}(l'_j, t) + \mathcal{A}_{diagram}(l'_j, d)]$$

Coherence score function

“DE is parallel with AB, and EF equals 5.”

Parallel (DE, EF)
Equals (AB, 5)
Equals (EF, 5)

High coverage, high redundancy

Parallel (DE, AB)

Low coverage, low redundancy

Parallel (DE, EF)
Equals (AB, 5)

High coverage, low redundancy

Parallel (DE, AB)
Equals (EF, 5)

High coverage, low redundancy

$$\mathcal{H}(L', t) = \underbrace{N(L')}_{\text{coverage}} - \underbrace{R(L')}_{\text{redundancy}}$$

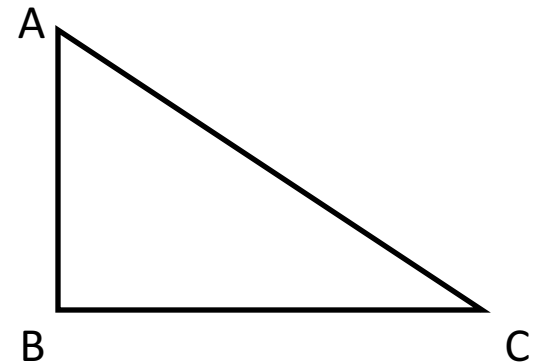
Numerical solver

AB is perpendicular to BC, $AB = 3$ and $BC = 4$.
What is the length of AC?

a) 3 b) 4 c) 5 d) 6 e) 7

```
Perpendicular(AB, BC)
Equals(LengthOf(AB), 3)
Equals(LengthOf(BC), 4)
Equals(LengthOf(AC), What)
```

- 2 variables for each point (x, y)
- 1 variable for unknown (What)
- One equation for each literal
- Simultaneously satisfy 4 equations with 7 variables (3 variables are free due to translation and rotation)



Experiment 2: semantic parsing

