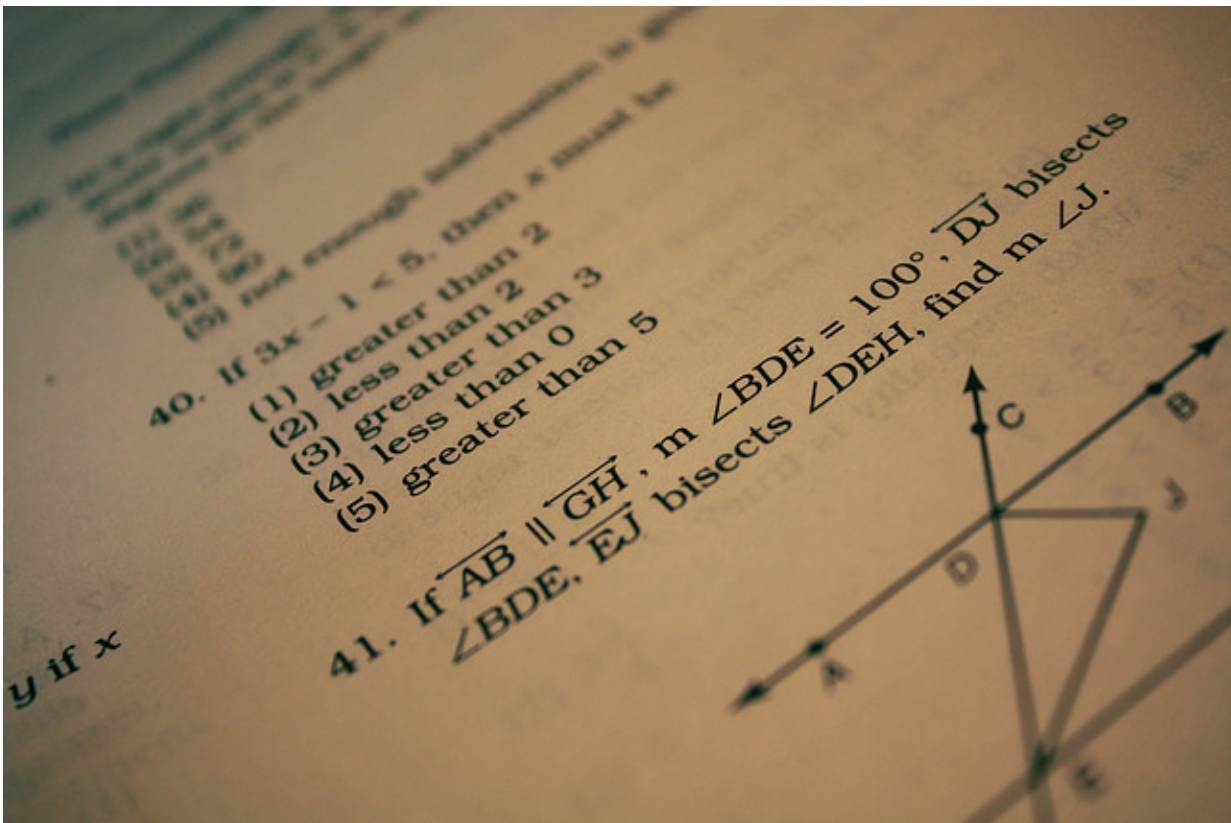


AI system solves SAT geometry questions as well as average human test taker

September 21 2015, by Jennifer Langston



Credit: Aaron Escobar, flickr

The Allen Institute for Artificial Intelligence (AI2) announced today it has created an artificial intelligence (AI) system that can solve SAT geometry questions as well as the average American 11th-grade student,

a breakthrough in AI research.

This system, called GeoS, uses a combination of computer vision to interpret diagrams, [natural language processing](#) to read and understand text, and a geometric solver to achieve 49 percent accuracy on geometry questions from the official SAT tests. If these results were extrapolated to the entire Math SAT test, the computer roughly achieved an SAT score of 500 (out of 800), the average test score for 2015.

A paper outlining the research, entitled "[Solving Geometry Problems: Combining Text and Diagram Interpretation](#)," was a joint effort between University of Washington Computer Science & Engineering department and AI2.

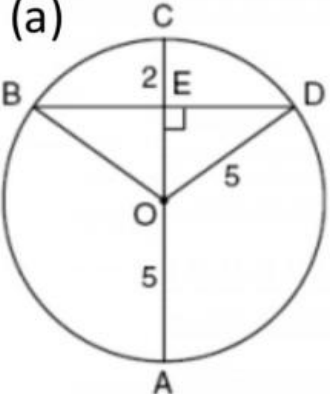
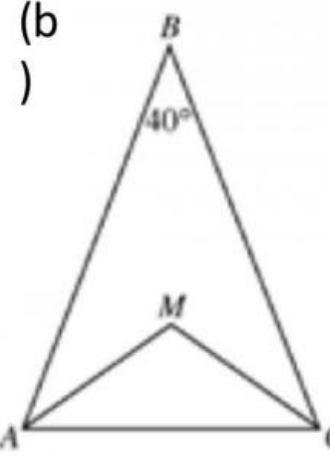
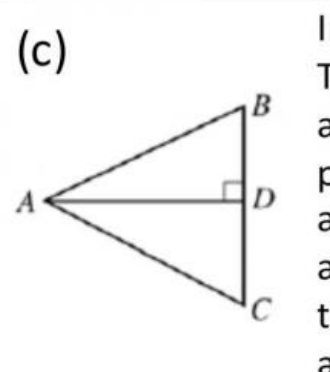
These results, presented at the Conference on Empirical Methods in Natural Language Processing (EMNLP) in Lisbon, Portugal, were achieved by GeoS solving unaltered SAT questions that it had never seen before and that required an understanding of:

- Implicit relationships
- Ambiguous references
- The relationships between diagrams and Natural-Language text

A demonstration of the system's problem solving is available here: geometry.allenai.org

"Unlike the Turing Test, standardized tests such as the SAT provide us today with a way to measure a machine's ability to reason and to compare its abilities with that of a human," said Oren Etzioni, CEO of AI2. "Much of what we understand from text and graphics is not explicitly stated, and requires far more knowledge than we appreciate. Creating a system to be able to successfully take these tests is challenging, and we are proud to achieve these unprecedented results."

Said Ali Farhadi, senior research manager for Vision at AI2 and assistant professor of computer science and engineering at UW, "We are excited about GeoS's performance on real-world tasks. Our biggest challenge was converting the question to a computer-understandable language. One needs to go beyond standard pattern-matching approaches for problems like solving geometry questions that require in-depth understanding of text, diagram, and reasoning."

Questions	Interpretations
<p>(a)</p>  <p>In the diagram at the left, circle O has a radius of 5, and $CE = 2$. Diameter AC is perpendicular to chord BD. What is the length of BD?</p>	<p><i>Equals(RadiusOf(O), 5)</i> <i>IsCircle(O)</i> <i>Equals(LengthOf(CE), 2)</i> <i>IsDiameter(AC)</i> <i>IsChord(BD)</i> <i>Perpendicular(AC), BD)</i> <i>Equals(what, Length(BD))</i></p> <p style="text-align: center;">correct</p> <p>a) 12 b) 10 c) 8 d) 6 e) 4</p>
<p>(b)</p>  <p>In isosceles triangle ABC at the left, lines AM and CM are the angle bisectors of angles BAC and BCA. What is the measure of angle AMC?</p>	<p><i>IsIsoscelesTriangle(ABC)</i> <i>BisectsAngle(AM, BAC)</i> <i>IsLine(AM)</i> <i>CC(AM, CM)</i> <i>CC(BAC, BCA)</i> <i>IsAngle(BAC)</i> <i>IsAngle(AMC)</i> <i>Equals(what, MeasureOf(AMC))</i></p> <p style="text-align: center;">correct</p> <p>a) 110 b) 115 c) 120 d) 125 e) 130</p>
<p>(c)</p>  <p>The bisector of angle BAC is perpendicular to BC at point D. If $AB = 6$ and $BD = 3$, what is the measure of angle BAC?</p>	<p><i>IsAngle(BAC)</i> <i>BisectsAngle(line, BAC)</i> <i>Perpendicular(line, BC)</i> <i>Equals(LengthOf(AB), 6)</i> <i>Equals(LengthOf(BD), 3)</i> <i>IsAngle(BAC)</i> <i>Equals(what, MeasureOf(BAC))</i></p> <p style="text-align: center;">correct</p> <p>a) 15 b) 30 c) 45 d) 60 e) 75</p>

An new AI system called GeoS was able to solve SAT questions such as these as well as the average American 11th grader. Credit: AI2/University of Washington

How GeoS Works

GeoS is the first end-to-end system that solves SAT plane geometry problems. It does this by first interpreting a geometry question by using the diagram and text in concert to generate the best possible logical expressions of the problem, which it sends to a geometric solver to solve. Then it compares that answer to the multiple choice answers for that question.

This process is complicated by the fact that SAT questions contain many unstated assumptions.

For example, for this sample SAT question:

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

There are several unstated assumptions, such as the fact that lines BD and AC intersect at E, that "circle O has a radius of 5" is the same as "circle O radius equals 5" and that the drawing may or may not be to scale.

GeoS' accuracy was much higher on questions it was confident enough to answer, which is an important dimension of learning. Today, GeoS can solve plane geometry questions; AI2 is moving to solve the full set of Math SAT questions in the next three years.

As part of AI2's commitment to sharing its research for the common good, all data sets and software are available for other researchers to use. See <http://www.allenai.org/data.html>.

AI2 is also building systems that can tackle science tests, which require a knowledge base that includes elements of the unstated, common sense knowledge that humans generate over their lives. This Aristo project is described here: <http://allenai.org/aristo.html>.

Provided by University of Washington

Citation: AI system solves SAT geometry questions as well as average human test taker (2015, September 21) retrieved 26 April 2024 from <https://phys.org/news/2015-09-ai-sat-geometry-average-human.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.