AI System Solves SAT Geometry Questions At Same Level As Average Human Test Taker

Breakthrough Achieved by
Allen Institute for Artificial Intelligence and University of Washington

Seattle, WA (Sept. 21, 2015) --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. The technical paper is available here: geometry.allenai.org

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."

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?*
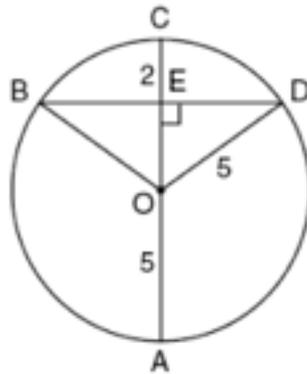
There are several unstated assumptions, such as the fact that lines BD and AC intersect at E, that "circle O has a radius of 5" is the same as "circle O radius equals 5" and that the drawing may or may not be to scale. GeoS' accuracy was much higher on questions it was confident enough to answer, which 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 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.

About AI2

AI2 was founded in 2014 with the singular focus of conducting high-impact research and engineering in the field of artificial intelligence, all for the common good. AI2 is the creation of Paul Allen, Microsoft cofounder, and is led by Dr. Oren Etzioni, a renowned researcher in the field of AI. AI2 employs more than 35 top-notch researchers and engineers, attracting individuals of varied interests and backgrounds from across the globe. AI2 prides itself on the diversity and collaboration of this team, and takes a results-oriented approach to complex challenges in AI.

About University of Washington Computer Science & Engineering (UW CSE):

UW CSE educates tomorrow's innovators, conducts high-impact research, transfers new discoveries to society and creates opportunities for faculty and students to push the boundaries of a rapidly expanding field while developing solutions to humanity's greatest challenges. Visit us online at www.cs.washington.edu.

##

Media Contacts:

Hamilton McCulloh
GreenRubino
206.957.4260
hamiltonm@greenrubino.com

Jennifer Langston
University of Washington Office of News and Information
206-543-2580
jlangst@uw.edu