Team turns deep-learning AI loose on software development

April 25 2018

Rice University computer scientists (from left) Vijay Murali, Chris Jermaine, Swarat Chaudhuri and Letao Qi created a software-coding application called Bayou that can help human programmers by writing chunks of code in response to keywords. Credit: Jeff Fitlow/Rice University

Computer scientists at Rice University have created a deep-learning, software-coding application that can help human programmers navigate the growing multitude of often-undocumented application programming

interfaces, or APIs.

Known as Bayou, the Rice application was created through an initiative funded by the Defense Advanced Research Projects Agency aimed at extracting knowledge from online source code repositories like GitHub. A paper on Bayou will be presented May 1 in Vancouver, British Columbia, at the Sixth International Conference on Learning Representations, a premier outlet for deep learning research. Users can try it out at askbayou.com.

Designing applications that can program computers is a long-sought grail of the branch of computer science called artificial intelligence (AI).

"People have tried for 60 years to build systems that can write code, but the problem is that these methods aren't that good with ambiguity," said Bayou co-creator Swarat Chaudhuri, associate professor of computer science at Rice. "You usually need to give a lot of details about what the target program does, and writing down these details can be as much work as just writing the code.

"Bayou is a considerable improvement," he said. "A developer can give Bayou a very small amount of information—just a few keywords or prompts, really—and Bayou will try to read the programmer's mind and predict the program they want."

Chaudhuri said Bayou trained itself by studying millions of lines of human-written Java code. "It's basically studied everything on GitHub, and it draws on that to write its own code."

Bayou co-creator Chris Jermaine, a professor of computer science who co-directs Rice's Intelligent Software Systems Laboratory with Chaudhuri, said Bayou is particularly useful for synthesizing examples of code for specific software APIs.

"Programming today is very different than it was 30 or 40 years ago," Jermaine said. "Computers today are in our pockets, on our wrists and in billions of home appliances, vehicles and other devices. The days when a programmer could write code from scratch are long gone."

Bayou architect Vijay Murali, a research scientist at the lab, said, "Modern software development is all about APls. These are system-specific rules, tools, definitions and protocols that allow a piece of code to interact with a specific operating system, database, hardware platform or another software system. There are hundreds of APIs, and navigating them is very difficult for developers. They spend lots of time at question-answer sites like Stack Overflow asking other developers for help."

Murali said developers can now begin asking some of those questions at Bayou, which will give an immediate answer.

"That immediate feedback could solve the problem right away, and if it doesn't, Bayou's example code should lead to a more informed question for their human peers," Murali said.

Jermaine said the team's primary goal is to get developers to try to extend Bayou, which has been released under a permissive open-source license.

"The more information we have about what people want from a system like Bayou, the better we can make it," he said. "We want as many people to use it as we can get." Bayou is based on a method called neural sketch learning, which trains an artificial neural network to recognize high-level patterns in hundreds of thousands of Java programs. It does this by creating a "sketch" for each program it reads and then associating this sketch with the "intent" that lies behind the program.

When a user asks Bayou questions, the system makes a judgment call

about what program it's being asked to write. It then creates sketches for several of the most likely candidate programs the user might want.

"Based on that guess, a separate part of Bayou, a module that understands the low-level details of Java and can do automatic logical reasoning, is going to generate four or five different chunks of code," Jermaine said. "It's going to present those to the user like hits on a web search. 'This one is most likely the correct answer, but here are three more that could be what you're looking for.'"

More information: Neural Sketch Learning for Conditional Program Generation, arxiv.org/abs/1703.05698

A live web demo is available at: www.askbayou.com

Provided by Rice University