July 29, 2020

Intel, MIT and Georgia Tech Deliver Improved Machine-Programming Code Similarity System

SANTA CLARA, Calif.--(BUSINESS WIRE)-- **What's New:** Today, Intel unveiled a new machine programming (MP) system – in conjunction with Massachusetts Institute of Technology (MIT) and Georgia Institute of Technology (Georgia Tech). The system, machine inferred code similarity (MISIM), is an automated engine designed to learn what a piece of software intends to do by studying the structure of the code and analyzing syntactic differences of other code with similar behavior.

> *"Intel's ultimate goal for machine programming is to democratize the creation of software. When fully realized, MP will enable everyone to create software by expressing their intention in whatever fashion that's best for them, whether that's code, natural language or something else. That's an audacious goal, and while there's much more work to be done, MISIM is a solid step toward it."*
> – Justin Gottschlich, principal scientist and director/founder of Machine Programming Research at Intel

Why It Matters: With the rise of heterogeneous computing, hardware and software systems are becoming increasingly complex. This complexity, paired with a shortage of programmers who can code at an expert level across multiple architectures, spotlights a need for new development approaches. Machine programming, a term coined by Intel Labs and MIT in their "Three Pillars of Machine Programming" paper, aims to improve development productivity through the use of automated tools. A key technology to several of these emerging machine programming tools is code similarity, which has the potential to accurately and efficiently automate some of the software development process to meet this need.

Yet building accurate code similarity systems is a relatively unsolved problem. These systems attempt to determine whether two code snippets show similar characteristics or aim to achieve similar goals — a daunting task when having only source code to learn from. MISIM can accurately determine when two pieces of code perform a similar computation, even when those pieces use different data structures and algorithms. "This is an important step toward the grander vision of machine programming," Gottschlich said.

How It Works: A core differentiation between MISIM and existing code-similarity systems lies in its novel context-aware semantic structure (CASS), which aims to lift out what the code actually does. Unlike other existing approaches, CASS can be configured to a specific context, allowing it to capture information that describes the code at a higher level. CASS can provide more specific insight into what the code does rather than how it does it. Moreover, MISIM can do all of this without using a compiler, which translates humanreadable source code into computer-executable machine code. This has many benefits over existing systems, including the ability to execute on incomplete snippets of code that a developer may be currently writing – an important practical characteristic for recommendation systems or automated bug fixing.

Once the code's structure is integrated into CASS, neural network systems give similarity scores to pieces of code based on the jobs they are designed to carry out. In other words, if two pieces of code look very different in their structure but perform the same function, the neural networks would rate them as largely similar.

By bringing together these principles in a unified system, researchers found that MISIM was able to identify similar pieces of code up to 40x more accurately than prior state-of-the-art systems.

What's Next: While Intel is still expanding the feature set of MISIM, the company has moved it from a research effort to a demonstration effort, with the goal of creating a code recommendation engine to assist all software developers programming across Intel's various heterogeneous architectures. This type of system would be able to recognize the intent behind a simple algorithm input by a developer and offer candidate codes that are semantically similar but with improved performance.

Intel's Machine Programming Lab is also engaging with software groups at Intel to see how MISIM can be integrated into their day-to-day development. Gottschlich, who is also an adjunct assistant professor at the University of Pennsylvania, hopes to help them, and Intel at large, to improve productivity and eliminate some of the mundane parts of programming, like hunting down bugs. Gottschlich speculates, "I imagine most developers would happily let the machine find and fix bugs for them, if it could – I know I would."

More Context: MISM: An End-to-End Neural Code Similarity System | Why More Software Development Needs to Go to the Machines | Intel Labs (Press Kit) | Three Pillars of Machine Programming

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.

View source version on businesswire.com:
https://www.businesswire.com/news/home/20200729005192/en/

Alexa Korkos
415-706-5783

alexa.korkos@intel.com

Source: Intel Corporation