'Learning database' speeds queries from hours to seconds

August 31 2017, by Gabe Cherry

Credit: CC0 Public Domain

A tool that makes large databases work smarter, not harder, could unlock the potential of big data to drive medical research, inform business decisions and speed up a slew of other applications that today are mired in a worldwide data glut.

University of Michigan researchers developed software called Verdict that enables existing databases to learn from each query a user submits, finding accurate answers without trawling through the same data again and again. Verdict allows databases to deliver answers more than 200

times faster while maintaining 99 percent accuracy. In a research environment, that could mean getting answers in seconds instead of hours or days.

When speed isn't required, it can be set to save electricity, using 200 times less than a traditional [database](#). This could lead to substantial power savings, the researchers say, as data centers gobble up a growing share of the world's electricity.

Verdict is believed to be the first working example in a new field of research called "database learning."

"Databases have been following the same paradigm for the past 40 years," said Barzan Mozafari, the Morris Wellman Faculty Development Assistant Professor of Computer Science and Engineering. "You submit a query, it does some work and provides an [answer](#). When a new query comes in, it starts over. All the work from previous queries is wasted."

Verdict changes that. It relies on advanced statistical principles, using past question-and-answer pairs to infer where the answers to future queries are likely to lie.

Big data bottleneck

The researchers say the innovation can't come soon enough, as the digital world is up to well over 1 billion gigabytes of stored data—everything from genomic data to hospital records and online shopping histories. And new data is streaming in far more quickly than systems can process it. Increased processing power won't solve the problem, as the rate of new data generation is increasing faster than processing power.

Meanwhile, data has become a driver for life-saving [medical research](#) and sophisticated business decision-making. It's increasingly being

tasked with not just finding answers, but also uncovering new ideas that can drive the direction of research. Medical researchers are turning databases loose on massive stockpiles of patient data to find buried connections between health status and disease. Retailers like Amazon are taking a similar approach to find precisely what motivates customers to buy and how to optimize supply chains, while online ad firms use data-driven algorithms to serve up the right ad at the right moment.

Such research can involve hundreds or thousands of simultaneous queries, and waiting hours for an answer is more than just an inconvenience. Studies have shown that even a short delay can hamper productivity and stifle innovation.

How Verdict works

Verdict is what's known as a "thin layer"—a small, nimble piece of software that can be placed in front of any existing database. At first, it simply stores queries that go in and out of the database, compiling them into what's called a query synopsis.

After storing a given number of queries, it goes into action, breaking each query up into component parts called snippets and using them to build a mathematical model of questions and answers. When a new query comes in, it uses that model to point the database to a certain subset of data where the answer is likely to be found. In some cases, it can even find an answer using only the model—without looking at the database at all.

Verdict itself uses minimal computing resources, and Mozafari, along with research fellow Youngjoo Park, has demonstrated that it doesn't slow performance. It also enables users to tailor the balance between speed and accuracy to fit individual applications. Mozafari believes a commercial product is likely a few years off.

"We've really just scratched the surface of what database learning can do," he said. "The important thing is that we've turned the mechanics of the database upside down. Instead of just additional work, each query is now an opportunity to learn and make the database work better."

The project is detailed in a study titled "[Database learning: Toward a database that becomes smarter every time](#)."

  More information: Read the paper: Database Learning:Toward a Database that Becomes Smarter Every Time: [web.eecs.umich.edu/~mozafari/p … s/dbl_techreport.pdf](#)

Provided by University of Michigan

Citation: 'Learning database' speeds queries from hours to seconds (2017, August 31) retrieved 4 October 2023 from https://techxplore.com/news/2017-08-database-queries-hours-seconds.html