

Software Scientist



With a little data, Eureka generates fundamental laws of nature

By Rachel Ehrenberg

Games such as chess have long been mastered by thinking machines. But weightier intellectual feats, such as deducing the laws of nature, have remained the domain of living, breathing brainiacs — until now.

A new computer program called Eureka comes up with fundamental mathematical laws, the great equations of textbooks and history, from scratch. Feed Eureka a mess of raw data, and it will find the underlying rules describing the observations.

Consider the laws of motion and conservation of energy. Eureka's creators, Cornell engineer and computer scientist Hod Lipson and then-graduate student Michael Schmidt, used a motion-tracking camera to capture the chaotic swings of two pendulums linked together. After measuring the pendulums' angles and velocities, the researchers fed the numbers to Eureka. Without any knowledge of geometry or physics, the program came up with Newton's second law of motion and other equations governing the double pendulum's behavior.

Now Eureka is analyzing even messier

systems, those involving living things. Eureka has figured out a set of seven known equations describing how the amounts of chemical compounds and enzymes fluctuate rhythmically in yeast cells deprived of oxygen, Lipson, Schmidt and colleagues reported in the October *Physical Biology*.

The yeast project came about after Lipson appealed to other scientists to give him projects for Eureka to tackle. He wanted to show that a range of scientific disciplines can benefit from the program, ultimately demonstrating Eureka's potential to uncover nature's basic rules rather than just re-create them. "It's not an artificial intelligence thing that someday will do something," Lipson says. "It can be applied to real problems."

Evolution embracer

John Wikswow of Vanderbilt University in Nashville and Jerry Jenkins of the HudsonAlpha Institute for Biotechnology in Huntsville, Ala., had a good test for Eureka. They gave Lipson the seven equations that describe yeast's glucose metabolism, but for all the Cornell scientists knew, the math could have

represented the movements of celestial bodies. Lipson and Schmidt generated a data set from the equations and added a healthy dose of random error, a realistic bit of messiness that would be found in ordinary experimental data.

Eureka took the information and started bumbling through equation space, creating and assessing different ways to describe the yeast results mathematically. Once it was done with the task, Lipson met with Wikswow and Jenkins at Vanderbilt, projecting the equations Eureka had come up with onto a conference-room screen.

Jenkins was on the lookout for one mathematical term in particular that relates to how quantities of a certain molecule limit the transport of sugar in the cell. "That was the first thing I was looking for, since it is very difficult to estimate in practice," he says. "They had nailed it."

Part of what makes Eureka so impressive is that the program assesses dynamic systems, which change over time and have parts that often change together. There are very clever "thinking machines" in existence today, such as Watson, the IBM computer that conquered *Jeopardy!* last year. But next to Eureka, Watson is merely a glorified search engine.

"It's a tour de force," says David Waltz, director of the Center for Computational Learning Systems at Columbia University. What Eureka does "is vastly more complicated than anything out there."

For all of its cold, hard computing power, Eureka is surprisingly biological, embracing the random, stupid beauty of evolution. A key component of the program's power is replication with variation, an idea borrowed from life that has become popular among scientists using machine learning and artificial intelligence to solve all sorts of problems.

Say you feed Eureka a spreadsheet. It may have 10 data points, or hundreds of thousands. You choose some mathematical operations, such as add, subtract, multiply or divide. (There's a default set of operations if you don't want to pick your own.) Then Eureka looks at a chunk of the data and starts throwing together mathematical building blocks into equations.

“You get a lot of random equations,” Schmidt says. “At first, they are really poor hypotheses. They are all junk.”

But by comparing those equations with a data sample, Eureqa can discern that some of the equations are ever so slightly less junky than others. Those slightly better equations are recombined: Terms are switched around and others are added, generating a new population of equations the way genetic material is recombined to make fresh versions of life.

As Eureqa comes up with new equations, it identifies and uses the data that will help it rule out bad equations most quickly. This process is repeated over and over again as Eureqa gets ever closer to equations that agree with observations. It can also request experiments that will generate helpful data, so some refer to Eureqa as a “robot scientist.”

Cross-disciplinary queries

Lipson and Schmidt published the yeast and pendulum work, but Eureqa has many more papers with many different authors to its name. The program is openly available online and has been downloaded more than 25,000 times. In November Schmidt brought the program to the cloud, that computer in the sky that takes advantage of servers across the country to crunch data in moments rather than the hours it might take on a home machine.

Physicists have used Eureqa’s brainpower to improve the resolution of their particle accelerators, while other scientists are evaluating speech recognition programs. Users have also queried Eureqa regarding stock market data, the growth of plant hairs and aircraft

$$\frac{dS_1}{dt} = 2.5 - \frac{100 \cdot A_3 S_1}{1 + 13.68 \cdot A_3^4}$$

$$\frac{dS_1}{dt} = 2.53 - \frac{98.79 \cdot A_3 S_1}{1 + 12.66 \cdot A_3^4}$$

$$\frac{dS_2}{dt} = \frac{200 \cdot A_3 S_1}{1 + 13.68 \cdot A_3^4} - 6 \cdot S_2 - 6 \cdot S_2 N_2$$

$$\frac{dS_2}{dt} = \frac{200.23 \cdot A_3 S_1}{1 + 13.80 \cdot A_3^4} - 6.87 \cdot S_2 - 6.87 \cdot N_2 + 0.95$$

$$\frac{dS_3}{dt} = 6 \cdot S_2 - 6 \cdot N_2 S_2 - 64 \cdot S_3 + 16 \cdot A_3 S_3$$

$$\frac{dS_3}{dt} = 6.00 \cdot S_2 - 6.00 \cdot N_2 S_2 - 64.16 \cdot S_3 + 16.08 \cdot A_3 S_3$$

Original equations
Eureqa-generated equations

Yeast laws redone After being fed data on yeast glucose metabolism, Eureqa reproduced the seven known equations that govern the process (three shown above). Notice that the formulas generally take the same form, for the most part differing only in their numerical constants.

SOURCE: M. D. SCHMIDT ET AL./PHYSICAL BIOLOGY 2011

tire dynamics. One user is apparently feeding Eureqa information about his daily life, such as how many e-mails he receives, with the hope that the program will be able to mathematically discern happy days from sad ones.

By one yardstick, Eureqa has even discovered the answer to the ultimate question of life, the universe and everything, a problem tackled by the fictional computer Deep Thought in Douglas Adams’ *The Hitchhiker’s Guide to the Galaxy*. As part of the pendulum problem, Lipson and Schmidt were trying to ask Eureqa what in the system wasn’t changing.

“Science is about finding the laws despite the complexity around us,” says Lipson. “We want to know what, in the apparent chaos, is always constant.”

Plenty of things don’t change in a given experimental setup, such as the temperature of the room or the color of the walls, but the researchers were looking

for meaningful invariance. In an early round, Eureqa failed to get the point. The way Lipson tells the story, the program replied to the invariance question with the same answer Deep Thought gave for the meaning of everything: 42.

Among all that Eureqa has tackled, its truest tests are those that involve coming up with meaningful unknowns. Gurol Suel, a biologist at the University of Texas Southwestern Medical Center at Dallas, fed Eureqa data on the genetics of cell division and growth of a particular bacterium. While Suel came up with his own possible equation to describe the system, Eureqa found an even simpler one. Scientists are now trying to figure out what Eureqa’s equation really means. ■

Explore more

■ For more on Eureqa, visit: creativemachines.cornell.edu/eureqa

Rules of self

Before designing Eureqa, a computer program that can come up with rules about the outside world, Cornell’s Hod Lipson was focused on introspective robots. He and then-graduate student Victor Zykov, along with Josh Bongard of the University of Vermont, created a four-legged robot and tasked it with learning to walk. The robot (shown) didn’t know what its own body looked like—whether it had four, eighteen or no legs, nor how those legs were arranged. But by interacting with

the environment, the robot generated hypothetical models of itself and then carried out actions that would test those models. The more the robot learned, the more directed its actions became. By the 16th trial, the machine figured out it had four legs and how to lurch forward. “Even a wrong model is good because it can allow you to make decisions in the right direction,” says Lipson, who has five Roomba vacuuming robots at home. —Rachel Ehrenberg

