ETH zürich

GLOBE

NO. 2 / 2017

FOCUS

EVERYTHING GOES DIGITAL

How data science is changing the world

PAGE 14

available online. These are a sort of correction program. Developers can use them to check their programs and to obtain suggestions on how to transform that program so that it is more easily understood by outsiders. The tools can also be used to decipher algorithms that were deliberately programmed in an obfuscating way, for example in order to obscure malware. More than 200,000 developers and IT security professionals worldwide have used JS Nice since its release. Last year, Vechev and his former PhD student Veselin Raychev founded DeepCode, an ETH spin-off. The company has set itself the task of creating new assistance programs for developers based on a direction explored in Vechev's SRL laboratory. This opens up opportunities for further applications down the line, such as programs that detect programming errors and propose fixes.

"We're seeking to develop software capable of solving challenges in software better than people can," says Vechev. "A few years ago, we were one of the first groups to set ourselves the goal of learning from 'big code'. Today, this area is attracting interest from many colleagues as well from various software companies. It's an interesting and fast-growing field of research." //

SOFTWARE RELIABILITY LAB:
www.srl.inf.ethz.ch

SPIN-OFF DEEPCODE:
www.deepcode.ai

MARTIN VECHEV is a tenure-track Assistant Professor in the Department of Computer Science at ETH Zurich and heads up the Software Reliability Lab. His primary area of interest is the application of machine learning to programming languages.

Big data for the environment

Forests, meadows and arable soils are natural stores of carbon. Researchers investigate the impacts.

TEXT Florian Meyer

Forests, meadows and farmland have an important role to play in climate change. As natural carbon stores, they can sequester carbon from the atmosphere, thereby counteracting the effects of global warming. However, the amount of carbon they are able to store is dependent on the environment around them. If temperatures continue to rise, it is possible that woods and soils would begin to release more greenhouse gases into the atmosphere while storing less and less carbon.

Investigating precisely how climate change impacts the carbon storage capacity of forests, meadows and soil is Nina Buchmann, ETH Professor of Grassland Sciences in the Department of Environmental Systems Science. As part of two major research projects – one for Switzerland (Swiss FluxNet), one for Europe (ICOS) – her research group is measuring the exchange fluxes of typical greenhouse gases such as carbon dioxide (CO_2), water vapour (H_2O), methane (CH_4) and nitrous oxide (N_2O) of forests, meadows and pastures, as well as farmland, with the atmosphere.

Measurements are being taken at six locations within Switzerland to determine the extent of the exchange of greenhouse gases between the ecosystem and the atmosphere. In doing so, the environmental researchers have employed the latest measurement techniques, which generate an almost uninterrupted stream of high-quality, high-resolution data. In the Swiss FluxNet project, for instance, wind speed, wind direction and gas concentrations are measured 20 times per second. The resulting data are then assessed for quality, combined with climatic data and ultimately used to calculate the fluxes of greenhouse gases between the ecosystems and the atmosphere. On top of this, time lapse cameras known as phenocams automatically photograph the vegetation at intervals of one to two hours throughout the day. These images can be used to identify any changes in the ecosystem.

NEW DATA, NEW QUESTIONS

Thanks to these new techniques, researchers have a mass of new data to work with. It is sorely needed, since the climatic, biological and geochemical factors at play interact in complex ways; for example, the ecosystem structure and soil climate influence the flux of greenhouse gases, as do wind turbulence, sun exposure, temperature and humidity. Plants also have an impact, such as through photosynthesis or root growth, and the same goes for microorganisms in the soil and for the people who fertilise the soil and farm the land.

Yet with this stream of new data comes a host of new questions. "If we want to understand the impact of climate change on ecosystems, we need to find a way to connect all the different data we have," says Buchmann. No easy task, since the data can be so divergent that it cannot possibly be reconciled without further work. In the ICOS project, environmental researchers are now collaborating across Europe to standardise their measurement infrastructure and data management systems.

Even so, the huge quantities of data involved and the hundreds of variables are still a sticking point in environmental research. There is a limit to what the human brain can process, and to what can be achieved with traditional statistical methods. Obtaining fresh insight from the complex data sets involves calls for new techniques of data analysis. One example of this is machine learning, in which algorithms search for recurring patterns in the data.

PUSHING THE BOUNDARIES

The issue of data analysis is what brought Buchmann to Andreas Krause. Krause is a computer science professor who specialises in machine learning and adaptive systems. In the "Big Data" national research programme (NRP 75), for instance, he is exploring how to draw together large quantities of data to facilitate efficient yet accurate machine learning. Krause is also academic co-director of the Swiss Data Science Center (SDSC), launched by ETH Zurich and EPFL in February 2017.

The SDSC forms a bridge between researchers who are producing large quantities of data, like Buchmann, and those who are developing new data systems and techniques of data analysis, like Krause. "At the SDSC, we combine the techniques of data science such as machine learning, statistics and information technology with the research skills of data-heavy disciplines such as life and environmental sciences," says Krause.

Questions such as those presented by Buchmann are addressed with an interdisciplinary approach, so that researchers from all disciplines are able to access the latest techniques in data management and data analysis. In turn, research questions inspired by real-life applications drive forward the development of new techniques in data science.

In his research, Krause tests algorithms that can be used for efficient, interactive data analysis. One example is a concept known as active learning. The learning process usually requires large quantities of training examples, which is often costly. Active learning takes a different approach, since it is the learning algorithm itself that selects the data that data analysts must consider. This accelerates learning and cuts costs. Similar issues arise when scientists look to apply learning techniques in extremely large data sets – for example, identifying representative use cases.

Active learning is also applicable to networks in which people and computers collect data together – for instance, data on earthquakes or air quality. These scenarios afford the opportunity to compensate for less reliable data and still use it to build up an overall picture. "How can humans and machines work together to achieve what neither of them could on their own?" wonders Krause, while Buchmann adds: "Maybe machine learning will open up a whole new perspective that we have never considered before."

DATA SCIENCE CENTER:
datascience.ch

Measurement station at Chamau

NINA BUCHMANN has held the post of Professor of Grassland Sciences in the Institute of Agricultural Sciences at ETH Zurich since April 2003. Her research focuses on the biogeochemistry of terrestrial ecosystems, namely forest, grassland and cropland.
›www.gl.ethz.ch

ANDREAS KRAUSE has held the post of Associate Professor in Computer Science at ETH Zurich since 2009. He is a specialist in machine learning and artificial intelligence, and a prominent figure in the Swiss Data Science Initiative.
›las.inf.ethz.ch

Image: Giulia Marthaler

Image: Grassland Sciences Group; Giulia Marthaler (2)