

GLOBE

FOKUS

ALLES DIGITAL

Wie Datenwissenschaft
die Welt verändert.

SEITE 14

Start-up: Ein Roboter
sticht ins Auge
SEITE 10

Wissenschaftsförderung
made in Europe
SEITE 34

Sabine Döbeli gestaltet
Finanzwirtschaft nachhaltig
SEITE 46

Die Umwelt in Big Data

Wälder, Wiesen und Böden speichern Kohlenstoff. Wie sie so das Klima beeinflussen, erforschen Umwelt- und Datenwissenschaften zusammen.

TEXT Florian Meyer

eine Art Korrekturprogramme für Code. Nutzer können damit ihre Programme überprüfen und sich Vorschläge anzeigen lassen, wie die Programme verbessert werden können, damit sie für Aussenstehende besser nachvollziehbar sind. Auch kann damit der Inhalt von Programmen entziffert werden, die absichtlich schwer verständlich programmiert wurden, etwa um Schadsoftware zu verschleiern. Mehr als 200 000 Entwickler und IT-Sicherheitsleute weltweit haben das Assistenzprogramm JS Nice bis jetzt verwendet.

Im vergangenen Jahr gründeten Vechev und sein früherer Doktorand Veselin Raychev den ETH-Spin-off Deepcode. Die Firma hat sich zur Aufgabe gesetzt, basierend auf Forschung aus Vechevs Labor neue Assistenzprogramme für Entwickler zu schaffen. Es sind denn auch weitere Anwendungen denkbar: Programme, die Programmierfehler finden und beheben.

«Langfristig möchten wir Software entwickeln, die intellektuell schwierige Herausforderungen beim Programmieren besser lösen kann als ein Mensch», sagt Vechev. «Vor wenigen Jahren waren wir unter den Ersten, die sich zum Ziel gesetzt haben, von «Big Code» zu lernen. Heute zeigen sich viele Kollegen und Softwarefirmen interessiert. Das Forschungsfeld wächst schnell.» //

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

SPIN-OFF DEEPCODE:
www.deepcode.ai



MARTIN VECHEV ist Assistenzprofessor mit Tenure Track am Department Informatik der ETH Zürich und leitet das Software Reliability Lab. Er interessiert sich besonders für die Anwendung von maschinellem Lernen auf Programmiersprachen.

Wälder, Wiesen und Ackerböden spielen eine wichtige Rolle im Klimawandel. Als natürliche Speicher können sie Kohlenstoff aus der Atmosphäre binden und damit der Klimaerwärmung entgegenwirken. Wie viel Kohlenstoff sie tatsächlich speichern, hängt aber ebenfalls mit der Klimaerwärmung zusammen. Sollten die Temperaturen in Zukunft weiter steigen, könnte es sein, dass Wälder und Böden vermehrt Treibhausgase an die Atmosphäre freigeben, während die Kohlenstoffspeicherung zurückgeht.

Wie genau der Klimawandel die Kohlenstoff-Speicherfähigkeit von Wäldern, Wiesen und Äckern erhöht oder vermindert, untersucht Nina Buchmann, ETH-Professorin für Graslandwissenschaften am Department Umweltsystemwissenschaften. Im Rahmen von zwei grossen, schweiz- und europaweit geführten Forschungsprojekten (Swiss FluxNet, ICOS) misst ihre Forschungsgruppe die Stoffflüsse von typischen Treibhausgasen wie Kohlendioxid (CO₂), Wasserdampf (H₂O), Methan (CH₄) und Lachgas (N₂O) von Wäldern, Wiesen und Weiden, aber auch von Ackerland.

Die Messungen werden an sechs Standorten in der Schweiz durchgeführt. Um herauszufinden, wie gross der Austausch an Treibhausgasen zwischen Ökosystem und Atmosphäre

ist, setzen die Umweltforschenden neuste Messtechniken ein. Diese erzeugen einen nahezu ununterbrochenen Strom an hochpräzisen und hochaufgelösten Daten. Im Projekt «Swiss FluxNet» etwa werden Windrichtung und -geschwindigkeit sowie Gaskonzentrationen 20 Mal pro Sekunde gemessen, auf ihre Qualität geprüft, mit klimatischen Daten verschnitten und die Treibhausgasflüsse berechnet. Ausserdem fotografieren Zeitrafferkameras, so genannte Phenocams, tagsüber automatisch alle ein bis zwei Stunden die Vegetation. Anhand der Bilder lassen sich allfällige Veränderungen des Bestandes feststellen.

NEUE DATEN, NEUE FRAGEN

Mit diesen Methoden gewinnen die Forschenden eine Fülle neuer Daten. Die können sie gut gebrauchen, denn die klimatischen, biologischen und geochemischen Wechselwirkungen sind vielschichtig: Bestandesstruktur und Bodenklima zum Beispiel beeinflussen die Stoffflüsse der Treibhausgase ebenso wie Windturbulenz, Sonneneinstrahlung, Temperatur und Feuchtigkeit. Auch die Pflanzen haben einen Einfluss, zum Beispiel über die Fotosynthese oder das Wurzelwachstum, ebenso die Mikroorganismen im Boden oder die Menschen, wenn sie die Böden düngen und das Ökosystem bewirtschaften.



Messstation auf der Chamau

Doch mit den neuen Daten kommen auch neue Fragen: «Wenn wir verstehen wollen, wie sich die Ökosysteme im Klimawandel verändern, dann müssen wir die vielen verschiedenen Daten miteinander verbinden», sagt Nina Buchmann. Nicht ganz einfach, denn die Daten sind zum Teil so unterschiedlich, dass sie sich nicht ohne Weiteres miteinander verknüpfen lassen. Im Rahmen von ICOS sind die Umweltforschenden nun europaweit daran, ihre Messinfrastruktur und ihr Datenmanagement zu vereinheitlichen.

Die riesigen Datenmengen mit Hunderten von Variablen sind aber manchmal eine Knacknuss für die Umweltforschung. Das menschliche Fassungsvermögen und die klassischen statistischen Methoden stossen hier an Grenzen. Um aus den komplexen Datensätzen neue wissenschaftliche Erkenntnisse herzuleiten, sind neue Ansätze der Datenanalyse gefragt. Zum Beispiel das maschinelle Lernen. Dabei suchen Algorithmen nach regelmässigen Mustern in den Daten.

GRENZEN ÜBERWINDEN

Darüber hat sich Nina Buchmann mit Andreas Krause ausgetauscht. Der Informatikprofessor ist ein Spezialist für lern- und anpassungsfähige Computersysteme. Zum Beispiel untersucht er im Nationalen Forschungsprogramm Big Data, wie man grosse

Datenmengen zusammenfassen kann, um effizientes und dennoch präzises maschinelles Lernen zu ermöglichen. Krause ist zudem wissenschaftlicher Co-Leiter des Swiss Data Science Center (SDSC), das die ETH Zürich und die ETH Lausanne im Februar 2017 eröffnet haben.

Dieses Zentrum bildet eine Brücke zwischen den Forschenden, die wie Nina Buchmann Daten produzieren, und jenen, die wie Andreas Krause neue Techniken der Datenanalyse und Datensysteme entwickeln. «Im Zentrum verbinden wir datenwissenschaftliche Methoden wie maschinelles Lernen, Statistik oder Informationstechnologie mit den Forschungskompetenzen von datenreichen Disziplinen wie Lebens- und Umweltwissenschaften», sagt Andreas Krause. Fragen, wie sie Nina Buchmann hat, werden fachübergreifend bearbeitet, damit Forschende aus allen Disziplinen auf die neusten Verfahren des Datenmanagements und der Datenanalyse zurückgreifen können. Forschungsfragen der Anwendung wiederum befeuern die Entwicklung neuer datenwissenschaftlicher Ansätze.

In seiner Forschung untersucht Krause Algorithmen für die effiziente interaktive Datenanalyse. Ein Beispiel ist das sogenannte aktive Lernen. Üblicherweise benötigen Lernverfahren grosse Mengen an Trainingsbeispielen. Dies führt oft zu hohen Kosten. Im



Seit April 2003 ist **NINA BUCHMANN** Professorin für Graslandwissenschaften am Institut für Agrarwissenschaften der ETH Zürich. Sie untersucht die Biogeochemie der Ökosysteme Wald, Grasland und Acker. >www.gl.ethz.ch



ANDREAS KRAUSE ist seit 2009 Professor für Informatik an der ETH Zürich. Er ist Spezialist für maschinelles Lernen und künstliche Intelligenz und Protagonist der Schweizer Data-Science-Initiative. >las.inf.ethz.ch

aktiven Lernen hingegen wählt der Lernalgorithmus selbstständig aus, welche Daten die Datenanalysten beachten müssen. Das beschleunigt das Lernen und senkt die Kosten. Ähnliche Probleme treten auch auf, wenn es darum geht, Lernverfahren auf extrem grossen Datensätzen anzuwenden. Hierbei geht es beispielsweise darum, repräsentative Fallbeispiele zu identifizieren.

Dieses Lernen umfasst auch Netzwerke, in denen Menschen und Computer Daten sammeln – zum Beispiel über Erdbeben oder Luftqualität. Solche Netze bieten die Chance, dass man auch die weniger verlässlichen Daten ausgleichen und für ein Gesamtbild nutzen kann. «Wie können Mensch und Maschine zusammen vollenden, was beide für sich nicht erreichen können?», fragt Andreas Krause, und Nina Buchmann ergänzt: «Vielleicht finden wir mit maschinellem Lernen ganz neue Zusammenhänge, die wir heute noch gar nicht kennen.» //

DATA SCIENCE CENTER:
datascience.ch