



OPEN DATA IMPACT AWARD 2020

Es gibt immer noch zu wenige Forscherinnen und Forscher, die ihre wissenschaftlichen Daten für andere bereitstellen. Um dies zu ändern, hat der Stifterverband in diesem Jahr erstmals den „Open Data Impact Award“ ausgelobt. „Wir fördern damit die strategische Öffnung von Datensätzen für alle Formen der Nachnutzung. Davon profitiert nicht nur die Wissenschaft, sondern die Gesellschaft insgesamt. Denn aus Daten entstehen Innovationen, die uns helfen, die großen Herausforderungen der Zukunft gemeinsam zu lösen“, so Volker Meyer-Guckel, stellvertretender Generalsekretär des Stifterverbandes. Der mit insgesamt 30 000 Euro dotierte Preis wurde am 7. Oktober in Kooperation mit der DUZ an drei Preisträger vergeben



1. PREISTRÄGER

EIN PLATZ AM FENSTER

Der Sieger des diesjährigen „Open Data Impact Award“ ist Michael Schmitt von der Hochschule München. Der Professor für Angewandte Geodäsie und Fernerkundung wurde für seinen Datensatz zur Entwicklung von KI-Verfahren in der Erdbeobachtung ausgezeichnet

TEXT: BENJAMIN HAERDLE

Wenn Michael Schmitt in den Flieger steigt, ist eines gewiss: Der 36-jährige Professor für Angewandte Geodäsie und Fernerkundung an der Hochschule München braucht einen Platz am Fenster: „Ich liebe es, von oben auf die Erde zu schauen.“ Schon als Kind hatte er ein Faible für Landkarten, unterstützt durch seinen Großvater, der ihm ADAC-Karten schenkte. Die Leidenschaft für den Blick aus der Vogelperspektive wird in Michael Schmitt aber nicht nur geweckt, wenn er fliegt, sondern auch, wenn er in seinem Münchner Büro am Rechner sitzt. Schmitt forscht auf dem weiten Feld der Ferner-

kundung an der Schnittstelle zwischen der Sensorik, der technischen Ableitung von Geodaten aus Satelliten- oder Radarbildern, und der angewandten Fernerkundung, wie also mit dem vorhandenen Kartenmaterial relevante gesellschaftliche Fragen etwa zur Entwicklung der Abholzung der tropischen Regenwälder beantwortet werden. Seine Expertise liegt genau dazwischen: Schmitt entwickelt Algorithmen, mit denen sich die Geodaten auswerten und schließlich in Informationen transformieren lassen, um so aus Satellitenbildern Geoinformationssysteme wie etwa Karten zu machen.

Der Spaß am Umgang mit großen Datenmengen hat ihm nun den ersten Platz in dem vom Stifterverband ausgelobten Open Data Impact Award eingebracht. Schmitt, der bei der Bewerbung für den Preis noch an der Technischen Universität (TU) München forschte und seit Anfang September an der Hochschule München beschäftigt ist, gelang es, einen Datensatz zu erstellen, mit dem sich in der Erdbeobachtung Modelle der Künstlichen Intelligenz für die automatisierte Auswertung von Satellitendaten entwickeln lassen. Er nutzte dafür drei Arten von Bilddaten: Radarsatellitenbilder der Sentinel-1-Mission, die über die Satellitenmission der Europäischen Weltraumorganisation alle sechs Tage von jedem Ort der Erde aufgenommen werden und die selbst bei Wolkenbedeckung im Unterschied zu optischen Satellitenbildern Aufnahmen produzieren; optische Satellitenbilder, die im Rahmen der Sentinel-2-Mission überall auf dem Globus alle fünf Tage mit einem Multispektral-Sensor Farbfotos ähnelnde Aufnahmen machen sowie relativ grobe Landbedeckungskarten aus früheren Satellitenmissionen, die jedem Pixel eine Landbedeckungskategorie wie etwa Nadelwald, Wasser oder Ödland zuweisen.

Diese verschiedenen Datensätze zu verknüpfen, war nun der eigentliche Clou von Schmitts Leistung: Denn die Daten haben eine unterschiedliche räumliche Auflösung, die Größe eines Pixels pro Meter variiert. Die verschiedenen Auflösungen müssen mittels Resampling vereinheitlicht werden. Bei Radarbildern liegt die Pixelgröße bei 5 bis 20 Metern, bei den optischen Bildern je nach Spektralband von 10, 20 oder 60 Metern pro Pixel, bei den älteren Landbedeckungskarten bei 500 Metern je Pixel. Hinzu kommt: „Jedem Satellitenbild müssen über mathematische Methoden für jedes Pixel Geokoordinaten zugeordnet werden, damit man genau weiß, wo das Pixel eigentlich liegt“, erklärt Schmitt.

Mit dieser Georeferenzierung und dem Resampling lassen sich die drei unterschiedlichen Arten von Bilddaten gut verschneiden. Dafür nutzte der Fernerkundler Google Earth Engine – eine frei verfügbare, cloudbasierte Analyseplattform für raumbezogene Daten, auf der beispielsweise Wissenschaftler Satellitenbilder sehen und auswerten können, mit deren Hilfe er den von ihm neu entwickelten Datensatz SENI2MS nun nutzbar machte. „Wie weltweit Ozeane oder Wälder verteilt sind, lässt sich aus den bisherigen Satellitendaten relativ leicht ablesen. Unser Ziel war es aber, ein wissenschaftliches Konzept für einen Datensatz zu entwickeln, mit dem sich zum Beispiel Wiesen von Heidelandschaften oder verschiedene Waldtypen wie Laub-Nadel- oder Mischwald untereinander abgrenzen lassen“, sagt er. Möglich sei damit aber auch, globale Prozesse wie den Klimawandel, die Urbanisierung oder die Entwicklung von Slums zu analysieren.

Um möglichst genaue Geoinformationssysteme zu bekommen, feilen Wissenschaftler schon seit Längerem daran, die Automatisierung der Datenauswertung stetig zu verfeinern. Dabei hilft der Einsatz von Künstlicher Intelligenz (KI). „Die heutigen KI-Verfahren in der Fernerkundung sind zwar häufig schon deutlich besser als frühere Verfahren, aber noch lange nicht perfekt“, sagt Schmitt. Mit seinem Datensatz sollen künftig bestehende KI-Verfahren trainiert und diese so verbessert werden. Der von ihm entwickelte Datenfundus stärkt aber auch die KI-Grundlagenforschung. Denn KI-Algorithmen lassen sich oft nicht gut trainieren, wenn deren Beispieldaten schlecht sind oder sie in der Auflösung weit auseinander liegen. „Label noise“ nennen das die Experten, das sogenannte Rauschen in den Trainingsdaten. „Wie man aus schlechten Ausgangsdaten gute KI-Algorithmen entwickeln kann, ist in der Forschung zu maschinellem Lernen eine wichtige Fragestellung“, sagt er. Auch dazu soll sein Datensatz beitragen.

Wie in vielen anderen Forschungsbereichen hat die KI-Forschung des maschinellen Lernens in der Fernerkundung und damit auch bei Michael Schmitt in den vergangenen Jahren einen immer breiteren Raum eingenommen. So war Schmitt einer der Wissenschaftler der TU München, der



Foto: privat

erfolgreich am Antrag für das Zukunftslabor für Künstliche Intelligenz in der Erdbeobachtung mitschrieb, das das Bundesforschungsministerium im Mai bewilligte – und dies mit einer Fördersumme von fünf Millionen Euro verband. „Mit Forschung nur zur Erdbeobachtung ist es derzeit schwierig, so hohe Fördersummen zu bekommen, doch im Verbund mit KI steigen die Chancen“, sagt Schmitt. Der Ingenieurwissenschaftler stand dem Zentrum in den ersten sechs Monaten als kommissarischer Geschäftsführer vor, ehe er dann eine Gelegenheit erhielt, die er nicht ausschlagen konnte, obwohl er dafür seine akademische Heimat verlassen musste. Das zeigt ein Blick in seinen Lebenslauf: Denn Michael Schmitt ist eigentlich ein Kind der TU München. Dort studierte er, machte seinen Diplomingenieur auf dem Gebiet der Geodäsie und Geoinformationswissen-

schaften, promovierte in der Fernerkundung und habilitierte schließlich im Jahr 2018 auf dem Gebiet der Datenfusion. Als sich jetzt die Möglichkeit für eine eigene Professur an der Hochschule München ergab, konnte er sich dieser kaum entziehen. „Irgendwann will man Professor werden und zu dem Angebot, jetzt eine unbefristete Fernerkundungs-Professur zu übernehmen, konnte ich nicht nein sagen“, sagt er. Hinzu kommt: Die Hochschule München hat die Forschung auf seinem Fachgebiet stark ausgebaut. Erst im Juli gründeten mehrere Fakultäten neue „Institut für Anwendungen des maschinellen Lernens und intelligenter Systeme“. Es passte also alles für Schmitt für den Neustart im Herbst. Und das Beste: Für seine Professur muss er nicht mal die Stadt wechseln – der gebürtige Münchener kann seiner Heimatstadt München treu bleiben. //

PROF. DR. MICHAEL SCHMITT: MEINE FORSCHUNG

DIE HERAUSFORDERUNG

Die Erde unterliegt einem ständigen Wandel und muss immer wieder neu vermessen und kartiert werden. Die Satellitenfernerkundung bildet dafür ein wichtiges Werkzeug. Da es aber immer mehr Satellitendaten mit unterschiedlichen Eigenschaften gibt, müssen immer ausgefeiltere Algorithmen entwickelt werden, die eine optimale Informationsgewinnung aus diesen Daten erlauben.

DROHENDE GEFAHREN

Wenn es keine Wissenschaftler gäbe, die an der Auswertung von Satellitendaten und deren Automatisierung arbeiteten, würden unvorstellbare Datenmengen ungenutzt in den Archiven landen und noch immer relativ große Investitionen keine Früchte abwerfen. Wertvolle Beobachtungen unserer Umwelt blieben unbemerkt.

MEIN BEITRAG

In meiner Forschung untersuche ich neuartige Datenquellen und entwickle Algorithmen der Signalverarbeitung und des maschinellen Lernens, die optimal an die neuen Datenquellen angepasst sind und eine Umwandlung von Satellitendaten in Karten und andere Geodaten erlauben. Wegen der heute verfügbaren großen Datenmengen

kommen viele der Methoden aus dem Bereich der Künstlichen Intelligenz.

OFFENE FRAGEN

Die Ära der Künstlichen Intelligenz hat auch in der Fernerkundung Einzug gehalten. Gleichwohl ist es nicht möglich, Standardverfahren, wie sie etwa in der Robotik oder dem autonomen Fahren verwendet werden, eins zu eins auf Satellitendaten anzuwenden. Ein wichtiger Beitrag ist die zielgenaue Entwicklung und Anpassung von Verfahren für komplexe Datentypen. Und es gilt, die KI-Verfahren nicht nur als sogenannte „Black Boxes“ anzuwenden, sondern zu verstehen, wie diese zu ihren Entscheidungen kommen.

MEIN NÄCHSTES PROJEKT

Mit Kollegen arbeite ich derzeit an einem Datensatz und KI-basierten Verfahren zur Entfernung von Wolken in Satellitenbildern, um eine zeitlich engmaschigere Beobachtung der Erdoberfläche zu ermöglichen. Darüber hinaus plane ich einige Projekte, die sich mit der Entwicklung von automatischen Kartierungsverfahren auf Basis von frei verfügbaren Satellitendaten beschäftigen, zum Beispiel für die Überwachung von Wäldern zur Früherkennung von Schäden oder die Kartierung von schutzwürdigen Wildnisgebieten.