

Neue Mobilitätskonzepte wie autonomes Fahren erzeugen riesige Datenmengen.

Futter für die Zahlenfresser

Vom Klimawandel bis zum Brexit: Der Umgang mit **Big Data** prägt zunehmend auch Fragestellungen in der Wissenschaft.

Von **Achim Killer**



Die Simulation beanspruchte den Supercomputer weit über eine Woche lang: Bigger Data geht nicht.

Number Cruncher“ werden Supercomputer von Enthusiasten oft liebevoll genannt, Zahlenfresser. Und „Big Data“ heißt ein neuartiges Futter für Zahlenfresser, große Datenmengen, wie sie etwa im Internet oder durch den zunehmenden Einsatz von Sensoren entstehen.

Da Unternehmen und Wissenschaftler diese großen Datenmengen oft schnell durchforsten müssen – weil sie sonst etwa zeitlich überholt sind –, benötigt man immer stärkere Rechner. Und unter den stärksten der Welt sind die SuperMUCs des Leibniz-Rechenzentrums. SuperMUC-NG heißt der Jüngste, er ging vor kurzem in Betrieb.

Big Data in der Wissenschaft

Bei ihren aufwändigsten Berechnungen hatten es die Wissenschaftler am LRZ allerdings nicht mit kurzlebigen Daten zu tun. Im Gegenteil: 2015 simulierten sie im Rahmen des „Magneticum Pathfinder“-Projekts auf dem SuperMUC die Entwicklung des Kosmos über 12,5 Milliarden Jahre. Die Simulation umfasste zehn Prozent des sichtbaren Universums. Sie beanspruchte den Supercomputer weit über

eine Woche lang, alle 86.016 Prozessorkerne und über 150 Terabyte Arbeitsspeicher: Bigger Data geht nicht. Heraus kamen 320 Terabyte wissenschaftliche Daten, die über das Web zugänglich sind.

2016 und 2017 wurde am LRZ die Klimaentwicklung in Bayern berechnet, und zwar von 1950 bis 2100, also über 150 Jahre hinweg, mit 50 verschiedenen Szenarien. Das ergibt 7.500 Jahre. ClimEx nennt sich das Projekt: Klimawandel und hydrologische Extremereignisse. Klima- und wasserwirtschaftliche Daten wurden kombiniert, um extreme Wetterereignisse wie das Jahrhunderthochwasser von 1999 prognostizieren zu können. Dazu mussten spezielle Wetterlagen in den 7.500 Jahresszenarien automatisch erkannt werden: stabile Tiefdruckgebiete südlich der Alpen, die sich gegen den Uhrzeigersinn ums Gebirge drehen. Die bringen oft besonders große Regenmassen nach Bayern.

Neuronale Netze trainieren

Für solche Aufgaben werden üblicherweise künstliche neuronale Netze trainiert. Sie heißen so, weil ihre Struktur dem menschlichen Gehirn nachempfunden



Das Donauhochwasser im Juni 2013 am Autobahnkreuz bei Deggendorf. Der Klimawandel und extreme Wetterereignisse stehen im Mittelpunkt des ClimEx-Projekts.

ist. Der Name täuscht allerdings. Die Metapher beschreibt kein physisches Netz, sondern Software: viele relativ einfache Berechnungen, deren Ergebnisse ineinander eingehen. Sie bilden ein logisches Netz, das sich verändern kann, indem die Rechenknoten ihre Eingangswerte verschieden gewichten: Das Netz lernt. Bei einem biologischen Gehirn verändert sich dabei auch die Hardware, also die Vernetzung der Neuronen, bei künstlichen neuronalen Netzen nur die Software.

Allerdings gibt es Computer-Hardware, die sich besonders gut zum Anlernen von neuronalen Netzen eignet: Graphikprozessoren beispielsweise, wie sie in einfacher Form in jedem PC oder Smartphone stecken. Sie verfügen über etliche kleine Rechenkerne zur Aufbereitung dessen, was auf dem Computer-Bildschirm oder dem Handy-Display zu sehen ist. Im LRZ stehen mehrere Systeme aus Graphikprozessoren mit jeweils 28.672 Kernen. Darauf lief unter anderem das neuronale Netz, das in den ClimEx-Daten nach Tiefdruckzonen gesucht hat. Eine Erkenntnis aus den Suchläufen: Jahrhunderthochwasser drohen in Bayern künftig häufiger aufzutreten als nur alle 100 Jahre.

Umgang mit Big Data im Studium

Im Rahmen des Data Innovation Lab der TU München arbeiten Studentinnen und Studenten mit Daten aus der Wirtschaft an den Computersystemen des LRZ. So hat etwa gerade eine Gruppe ein Projekt abgeschlossen, bei dem es darum ging, maschinell zu erkennen, ob der Fahrer eines Kraftfahrzeugs abgelenkt ist oder aufmerksam. Die Studenten trainierten dazu ein neuronales Netz, das die Blickrichtung des Fahrers analysiert. Mit 11.000 zuvor gesichteten und bewerteten Fotos lernte das Netz zu erkennen, wohin ein Mensch schaut.

Eine andere Gruppe analysierte die Brexit-Stimmung der Briten. Über 32 Millionen Tweets der Plattform Twitter wurden dazu verarbeitet. Weil so viel kein Mensch lesen kann und mag, wurden die Tweets zunächst in mathematische Objekte, Vektoren, überführt, die durch den Kontext einzelner Wörter gebildet werden – ein gängiges Verfahren der Computerlinguistik. Ein neuronales Netz wurde trainiert und der Einsatz einer Support Vector Machine getestet, um Tweets von Brexiteers und EU-Befürwortern zu unterscheiden. Auch

bei dieser Maschine handelt es sich um reine Software. Im Ergebnis zeigt der zeitliche Verlauf des Gezwitschers, dass die anfängliche Euphorie der Brexiteers schnell nachgelassen hat.

Die Garching Supercomputer sind Universalrechner, die lernende Software verarbeiten können, aber nicht speziell dafür ausgelegt sind. Allerdings verfügen einige Rechenknoten über fürs Maschinelle Lernen besonders geeignete Graphikprozessoren, die ergänzt werden durch die speziellen Graphiksysteme mit Tausenden Verarbeitungseinheiten am LRZ. Bei der nächsten Generation soll die Lernfähigkeit des SuperMUC-Computers noch weiter ausgebaut werden.

Achim Killer

arbeitet als freier Journalist für den Bayerischen Rundfunk, diverse ARD-Magazine und den Deutschlandfunk. Sein Spezialgebiet ist das Internet.
