

Linzer Forscher errechnen Giftigkeit von Stoffen



Die JKU-Bioinformatiker Thomas Unterthiner und Sepp Hochreiter

Foto: JKU Linz

Infobox

► krone.at/Digital ist auf Facebook - werden Sie jetzt Fan!

10.000 chemische Stoffe auf ihre Giftigkeit zu überprüfen - für Menschen eine unlösbare Aufgabe, für die Computer von Professor Sepp Hochreiter, Thomas Unterthiner und ihrem Team vom Institut für Bioinformatik der Uni Linz eine Herausforderung. Mit ihrer "Deep Learning"-Methode lösten sie diese Aufgabe besser als jeder andere - und gewannen die "Tox21 Challenge" der amerikanischen Gesundheitsbehörde.

"Deep Learning"-Methoden erlauben Computern, gewaltige Datenmengen auszuwerten und - ähnlich dem Gehirn - mit riesigen neuronalen Netzwerken Analysen durchzuführen. Die Grundlagen wurden von Hochreiter bereits vor 20 Jahren entwickelt, aber erst moderne Technologie konnte diese Grundlagen zur Anwendung bringen. "Uns liegen erst jetzt ausreichend große Datenmengen vor - die bekannte 'Big Data'-Entwicklung - und vor allem gibt es nun erst eine Hardware, die damit auch umgehen kann", erklärt der Experte für Maschinelles Lernen.

Moderne Sprachsteuerungen oder Objektdarstellungen, mit denen Unternehmen wie Google oder Facebook arbeiten, basieren alle auf diesem von Hochreiter entwickelten Prinzip. So konnte beispielsweise die Fehlerquote bei der Spracherkennung durch Computer Dank "Deep Learning" von 24 auf 16 Prozent verringert werden, nachdem die Quote mehr als ein Jahrzehnt stagniert hatte.

"Deep Learning": In diesem Bereich "ist die Hölle los"
"Im Bereich 'Deep Learning' ist weltweit die Hölle los", beschreibt Hochreiter den Run auf die Big-Data-Auswertung, die der Wissenschaft völlig neue Wege eröffnet. Einer des wichtigsten Kompetenzzentren für diese Technologie ist die Linzer Johannes-Kepler-Uni, wie der Bewerb der US-Gesundheitsbehörde bewies. "Wir mussten die Giftigkeit von Stoffen in der Umwelt, also in Nahrung, Haushalt, Medikamenten et cetera mit Computermethoden aus der chemischen Struktur vorhersagen", erläutert Hochreiter die Aufgabenstellung.

Die Aufgabe lag den Linzer Wissenschaftlern: Am Institut für Bioinformatik ist das Know-how sowohl für Informatik als auch für Naturwissenschaften wie Chemie vereint. Millionen von chemischen Merkmalen wurden extrahiert, ein eigens entwickeltes quantenmechanisches Modell erlaubte moderne Sprach- und Mustererkennung. Der Computer - er setzt auf Tesla-K40-GPUs von Nvidia - kombinierte diese Merkmale auf 12.000 parallel geschalteten Prozessoren miteinander und erlaubte präzise Rückschlüsse, sodass die US-Gesundheitsbehörde die "hohe Qualität des Modells" hervorhob.