

Künstliche Intelligenz beflügelt Proteomforschung

Maschinelles Lernen macht Proteomikforschung leistungsfähiger

Mit dem Einsatz von künstlicher Intelligenz ist es Forschenden der Technischen Universität München (TUM) gelungen, die massenhafte Analyse von Eiweißen aus beliebigen Organismen deutlich schneller als bisher und praktisch fehlerfrei zu machen. Dieser neue Ansatz wird das Forschungsfeld der Proteomik stark verändern, da er sowohl in der Grundlagen- als auch in der klinischen Forschung angewandt werden kann.

Das Genom jedes Organismus enthält die Baupläne für Tausende von Eiweißen, die praktisch alle Funktionen des Lebens steuern. Fehlerhafte Proteine führen zu schweren Krankheiten wie Krebs, Diabetes oder Demenz. Eiweiße sind somit auch die wichtigsten Angriffspunkte für Medikamente. Damit man Lebensvorgänge und Erkrankungen besser verstehen und passendere Therapien entwickeln kann, muss man möglichst viele Proteine gleichzeitig analysieren. Aktuell wird hierzu die Massenspektrometrie genutzt, die in der Lage ist, Art und Menge der Eiweiße in einem biologischen System zu bestimmen. Jedoch machen die derzeitigen Verfahren der Datenanalyse noch viele Fehler.

Einem Team der TU München um den Bioinformatiker Mathias Wilhelm und den Biochemiker Bernhard Küster ist es nun gelungen, massenhaft erhobene proteomische Daten zu nutzen, um ein neuronales Netzwerk so zu trainieren, dass es Proteine deutlich schneller und praktisch fehlerfrei erkennen kann.

Lösung für ein gravierendes Problem

Massenspektrometer messen Proteine nicht direkt, sondern analysieren kleinere Teile, bestehend aus Aminosäuresequenzen mit bis zu 30 Bausteinen. Die gemessenen Spektren dieser Ketten werden mit Datenbanken abgeglichen, um sie einem bestimmten Protein zuzuordnen. Die Auswertesoftware kann jedoch nur einen Teil der enthaltenen Informationen nutzen. Daher werden manche Proteine nicht oder falsch erkannt.

„Das ist ein gravierendes Problem“, sagt Küster. Das neuronale Netzwerk, das das TUM-Team entwickelt hat, nutzt alle Informationen der Spektren für die Identifizierung. „Dadurch verpassen wir weniger Proteine und es passieren 100 mal weniger Fehler“, so Bernhard Küster, Professor für Proteomik und Bioanalytik an der TU München.

Auf alle Organismen anwendbar

„Prosit“, wie die Forscher die KI-Software nennen, ist „auf alle Organismen dieser Welt anwendbar, auch wenn man deren Proteome vorher nie untersucht hat“, sagt Mathias Wilhelm. „Das ermöglicht Untersuchungen, die vorher nicht denkbar waren.“

Der Algorithmus ist mit Hilfe von 100 Millionen Massenspektren so umfangreich angelernt worden, dass er ohne erneutes Trainieren für alle gängigen Massenspektrometer eingesetzt werden kann. „Unser System ist hier weltweit führend“, sagt Küster.

Ein Milliardenmarkt

Kliniken, Biotechunternehmen, Pharmafirmen und die Forschung nutzen solche Hochleistungsgeräte; es ist schon jetzt ein Milliardenmarkt. Mit „Prosit“ können zukünftig noch leistungsfähigere Geräte entwickelt werden. Ebenso werden Forscher und Mediziner besser und schneller nach Biomarkern im Blut oder Urin von Patienten suchen oder Therapien hinsichtlich der Wirksamkeit überwachen können.

Auch für die Grundlagenforschung versprechen sich die Forscher viel. „Mit dem Verfahren kann man neuen Regulationsmechanismen in Zellen auf die Spur kommen“, so Küster. „Wir erhoffen uns hier einen erheblichen Erkenntnisgewinn, der sich mittel- und langfristig in der Behandlung von Erkrankungen von Mensch, Tier und Pflanze niederschlagen wird.“

Auch Wilhelm erwartet, dass „KI Methoden wie Prosit schon bald das Forschungsfeld der Proteomik nachhaltig verändern werden, da sie in nahezu allen Bereichen der Proteinforschung eingesetzt werden können.“

Publikationen:

Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning
Siegfried Gessulat, Tobias Schmidt, Daniel Paul Zolg, Patroklos Samaras, Karsten Schnatbaum, Johannes Zerweck, Tobias Knaute, Julia Rechenberger, Bernard Delanghe, Andreas Huhmer, Ulf Reimer, Hans-Christian Ehrlich, Stephan Aiche, Bernhard Küster und Mathias Wilhelm
Nature Methods, 27.05.2019 – [DOI: 10.1038/s41592-019-0426-7](https://doi.org/10.1038/s41592-019-0426-7)

Mehr Informationen:

Die Studie wurde durch das vom BMBF geförderte Projekt ProteomeTools realisiert, hier arbeiten die Autoren mit den Firmen JPT (Berlin), SAP (Potsdam) und ThermoFisher Scientific (Bremen) zusammen. Prosit ist über ProteomicsDB verfügbar, welches im Rahmen des DIAS Projekts vom BMBF gefördert wird.

<http://www.proteomicsdb.org>

<http://www.proteomicsdb.org/prosit>

<http://www.proteometools.org/>

[Weitere aktuelle Meldung zum Einsatz von künstlicher Intelligenz in der Genomforschung](#)