

Press release**Max-Planck-Institut für Biochemie****Dr. Christiane Menzfeld**

05/28/2019

<http://idw-online.de/en/news716555>Research results, Scientific Publications
Biology, Chemistry, Information technology, Mathematics
transregional, nationalmax-planck-institut
für biochemie**Aus spektralen Informationen lernen**

Die Proteinzusammensetzung von Zellen hängt von deren Funktion und aktuellen Zustand ab. Mit der Massenspektrometrie (MS) lassen sich Identität und Menge der in einer Probe gefundenen Proteine ermitteln. Allerdings ist die Analyse der entstehenden Daten zeit- und ressourcenintensiv. Forscher am Max-Planck-Institut für Biochemie (MPIB) haben gemeinsam mit Datenspezialisten von Verily in den USA einen Ansatz des maschinellen Lernens entwickelt. Die sich kontinuierlich selbst verbessernden Algorithmen vereinfachen die Analyse solcher MS-Daten. Ihr Programm, das zudem zur Entdeckung neuer chemischer Muster in Proteinen geführt hat, wird in der Fachzeitschrift *Nature Methods* veröffentlicht.

Die letzte Station für alte Autos ist oft der Schrottplatz, auf dem diese Fahrzeuge in ihre möglicherweise noch verwertbaren Einzelteile zerlegt werden. Ein erfahrener Mitarbeiter des Schrottplatzes könnte durch einen Blick auf die Gesamtheit aller Teile die Identität des verschrotteten Fahrzeugs ermitteln. Massenspektrometer (MS), die die in einer Probe enthaltenen Proteine identifizieren und quantifizieren, funktionieren quasi wie molekulare Schrottplätze. Zunächst werden die Proteine in kleinere Fragmente, die Peptide, zerlegt. Die Informationen über Identität und Menge der Peptide lassen sich als Spektren erfassen. Um die Informationen über die Proteine aus der analysierten Probe zu rekonstruieren, werden die Merkmale dieser Spektren dann mit früher erfassten Proteinbibliotheken verglichen. Dieser Prozess erfordert eine enorme Rechenleistung.

Maschinelles Lernen unterstützt Datenanalyse

In Kooperation mit Verily, dem Life-Sciences-Unternehmen von Alphabet, haben Forscher am MPIB jetzt das Modell DeepMass:Prism entwickelt, das die Interpretation von MS-Daten vereinfacht. Sie verwendeten Methoden des maschinellen Lernens, um Algorithmen darin zu trainieren, Proteine in MS-Spektren zu „übersetzen“. Die Übersetzung solcher abstrakten Daten ist eine echte Herausforderung für die künstliche Intelligenz und gelingt am besten mit „tief lernenden“ Deep Learning Algorithmen. Vergleichbare Ansätze werden auch bei der automatischen Übersetzung von Sprachen eingesetzt. Aber statt beispielsweise eine Übersetzung vom Englischen ins Deutsche oder umgekehrt durchzuführen, ist DeepMass:Prism darin geschult, Übersetzungen zwischen Proteinen und Spektren, wie sie üblicherweise in einer MS-Analyse generiert werden, zu erstellen.

„Der Schlüssel zum Erfolg dieses Projektes war die Synergie unserer eingehenden Erfahrung in der Massenspektrometrie mit den Kenntnissen von Verily in Deep Learning, insbesondere in den Bereichen Biologie und Life Sciences“, sagt Jürgen Cox, unabhängiger Gruppenleiter am MPIB. Das von den Forschern entwickelte Programm DeepMass:Prism wurde mit über 60 Millionen Peptidspektren aus öffentlich zugänglichen Datenbanken trainiert. Das Programm erkennt jetzt Muster aus den Trainingsspektren und wendet diese auf die Analyse neuer Proben an.

Der Systembiologe Cox weist darauf hin, dass DeepMass:Prism verschiedene Anwendungen der Massenspektrometrie verbessert. Eine potenzielle Einsatzmöglichkeit für die MS ist die Charakterisierung von Proben, deren Zusammensetzung völlig unbekannt ist. Dabei können die neuen Algorithmen dazu beitragen, die Anzahl der Peptide zu erhöhen, die mit diesem Verfahren identifiziert werden. Oder man kann große Mengen von Proben einer grundsätzlich

ähnlichen Zusammensetzung auf individuelle Unterschiede in der Proteinmenge hin vergleichen. So sind die Blutproben von Patienten beispielsweise grundsätzlich durch eine ähnliche Protein-zusammensetzung charakterisiert. Aber für gezielte Krankheitsdiagnosen ist es wichtig, veränderte Proteingehalte zu erkennen. „Auf diesem Gebiet hat unser Programm DeepMass:Prism die größten Fortschritte vorzuweisen“, erläutert Cox. „Statt die Referenzbibliotheken, mit denen die Proben verglichen werden, experimentell zu bestimmen, kann das Programm diese jetzt vorhersagen – eine Abkürzung also, die sich zeit- und ressourcensparend auswirkt.“

Die Nadeln im Peptid-Heuhaufen finden

Die über 200 verschiedenen Zelltypen des menschlichen Körpers unterscheiden sich nicht nur durch das Vorkommen unterschiedlicher Proteine, sondern auch durch den unterschiedlichen Gehalt der identifizierten Proteine. Für die MS-Analysen stellen insbesondere die unterschiedlichen Proteinmengen eine Herausforderung dar. Jürgen Cox verdeutlicht die Bedeutung der Messung von Proteinmengen noch einmal anhand einer Analogie aus der Autowelt: „Zerlegt man Autos komplett in ihre Einzelteile, gleichen sich die Haufen mit den Komponenten auf den ersten Blick ziemlich stark. Deshalb liefert die Ermittlung der Menge bestimmter Teile bei der Identifizierung eine wertvolle Hilfe. Wenn man in einem solchen Teilehaufen sechs Zylinder findet, kann man daraus schließen, dass die Teile nicht von einem Auto mit einem Vierzylindermotor stammen.“ Gibt es in einem Haufen nur drei Reifen, ist von einem möglichen Defekt des Fahrzeugs auszugehen. Das gleiche Prinzip lässt sich auf die Analyse von Zellen oder Geweben anwenden. Denn Krankheiten können bewirken, dass bestimmte Proteine in größeren oder geringeren Mengen vorhanden sind als in gesunden Kontrollproben.

Viele diagnostische Verfahren stützen sich auf die massenspektrometrische Auswertung von Proteinen in Patientenproben. „Für die Entdeckung neuer Biomarker als Indikatoren für Krankheiten brauchen wir eine hochgenaue Massenspektrometrie. Manchmal kann schon eine geringe Abweichung bei einem bestimmten Biomarker ein Anzeichen für das Fortschreiten der Erkrankung sein. Deshalb muss die Vorhersage präzise und reproduzierbar sein“, sagt Peter Cimermanic, Senior Scientist bei Verily. Mit DeepMass:Prism konnten die Wissenschaftler die Korrelation zwischen den prognostizierten und den tatsächlich gemessenen Spektren erheblich verbessern. Er ist deshalb zuversichtlich, dass das Modell zur Entwicklung neuer Diagnoseinstrumente führen wird.

Obwohl DeepMass:Prism nicht mit chemischen Informationen geschult wurde, entdeckte das Programm neue chemische Regeln, die bestimmen, wie die Peptide in kleinere Fragmente zerlegt werden. „Mit dem früheren bibliotheksbasierten Ansatz konnte nur reproduziert werden, was bereits bekannt war. DeepMass:Prism kann dagegen Informationen kombinieren und eigene Rückschlüsse ziehen, um neues Wissen erzeugen. Das ist ein sehr spannendes Ergebnis“, sagt Cox. „Das ist so, als ob der Mitarbeiter eines Schrottplatzes verstehen würde, wo ein bestimmtes Teil des Autos montiert war, obwohl er diesen Autotyp nie zuvor gesehen hat. Die Vorhersagen durch DeepMass:Prism haben die Identifizierung einer neuen Art der Interaktion innerhalb von Proteinen ermöglicht. Diese Entdeckung ist nach unserer Einschätzung erst der Beginn dessen, was Deep Learning für die Forschung im Bereich Life Sciences in Zukunft bedeuten kann.“ DeepMass:Prism wird zum Herunterladen über eine Google Cloud zur Verfügung stehen. [CW]

contact for scientific information:

Prof. Jürgen Cox, PhD
Computational Systems Biochemistry
Max-Planck-Institut für Biochemie
Am Klopferspitz 18
82152 Martinsried/München
E-mail: cox@biochem.mpg.de
www.biochem.mpg.de/cox

Original publication:

S. Tiwary*, R. Levy*, P. Gutenbrunner*, F.S. Soto, K. Palaniappan, L. Deming, M. Berndl, A. Brant, P. Cimermanic und J. Cox: High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis.

