

Digitalisierung

Künstliche Intelligenz als Chemiker

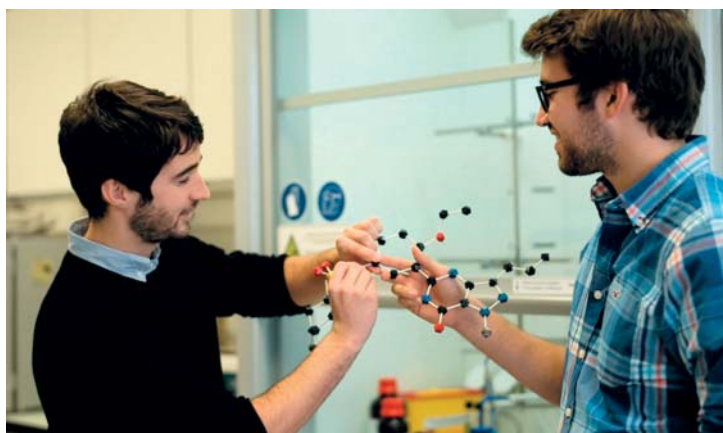
Computer schlagen Menschen bei Brettspielen wie Schach und Go oder bei Strategie-Videospielen und müssen zuvor nicht mal die Regeln kennen. Aber damit nicht genug: Inzwischen entwerfen sie sogar Retrosynthesen.

Seit den 1960er Jahren versuchen Forscher, Programme zu entwickeln, die organische Synthesen planen. Der Erfolg war lange mäßig, denn Computer konnten nur das wiedergeben, was sie von Menschen erhalten hatten. Ein Mensch musste sie programmieren, Regeln aufstellen, mitteilen, was zu tun ist. Beim Schach war das machbar, aber Chemie ist komplexer.

Der Inhalt chemischer Literatur etwa ist für eine Software schwer zu verarbeiten. Viele Reaktionen funktionieren nur unter bestimmten Bedingungen. Wissenschaftsjournalist Derek Lowe beschrieb es in *Nature* so: „Die funktionelle Gruppe X wird zu Y, es sei denn, das Edukt enthält die funktionelle Gruppe Z. Dann läuft die Reaktion nur, wenn Q anwesend ist – aber auch das nur, wenn der pH unter einem bestimmten Wert liegt, die Temperatur hoch genug oder kein Wasser anwesend ist.“¹⁾ Dieses Wechselspiel war einem Computer kaum verständlich zu machen. Zudem werden ständig neue Reaktionen veröffentlicht, chemisches Wissen ist also nicht statisch.

Ein Computerprogramm für Retrosynthesen zu entwickeln, das auf Algorithmen basiert, ist vor allem eins: mühsam. Jede Namensreaktion, ihre Substrate, tolerierte und nicht tolerierte funktionelle Gruppen sind per Hand einzugeben und auf dem Laufenden zu halten.

Über zehn Jahre lang arbeitete ein Team um Bartosz Grzybowski, Chemiker am Ulsan National Institute of Science and Technology in Süd-



Philippe Schwaller (rechts) und Theophile Gaudin von IBM Research wollen chemische Reaktionen mit dem Computer vorhersagen. Abbildungen: IBM Research

korea, genau daran. Chematica heißt das Ergebnis, eine Software, die auf einer Datenbank chemischen Wissens aufbaut und Syntheserouten für Moleküle vorschlägt, die für die Pharmaforschung interessant sein könnten. Merck kaufte das Programm im Jahr 2017. Ein Jahr später hatte Grzybowski acht der Syntheserouten, die sein Algorithmus vorgeschlagen hat, im Labor nachgekocht – mit dem vom Computer vorhergesagten Ergebnis.²⁾ Mittlerweile haben Grzybowski und sein Team dem Programm beigebracht, patentrechtlich geschützte Reaktionsstufen zu vermeiden.³⁾

Begreifen statt auswendig lernen

Inzwischen muss man dem Computer nicht mehr alles sagen, was er wissen muss – dank neuronaler Net-

ze und maschinellen Lernens, speziell Deep Learning (Glossar S. 36). Ein Computersystem generiert selbst Wissen aus Erfahrungen, erkennt in Beispielen Gesetzmäßigkeiten und verallgemeinert die Beispiele. Die künstliche Intelligenz (KI) war geboren. In der Chemie bedeutet das: Nachdem sich eine KI tausende chemische Reaktionen aus der Literatur angesehen hat, kann sie Reaktionen vorhersagen, die sie zuvor noch nie gesehen hat.

Im Jahr 2016 fütterten Chemiker und Informatiker von der Harvard-Universität in Cambridge ein System mit 3400 Reaktionen, die halogenierte Kohlenwasserstoffe, Alkene oder beides enthalten.⁴⁾ Die KI sagte nach dem Training neue Reaktionen desselben Typs mit einer Genauigkeit von mindestens 80 Prozent vorher; je mehr Trainingsbeispiele das System bekam, desto besser wurde es.

Die Sprache der Chemie

Forscher von IBM Research in Zürich entwickelten die webbasierte KI RXN for Chemistry, die jedem kostenlos organische Reaktionen vorher-sagt.^{5,6)} IBM veröffentlichte das Programm im August 2018 während des Herbstmeetings der American Chemical Society. Der Nutzer zeichnet in der Webseite die Edukte, und kurz darauf erscheint das wahrscheinlichste Produkt (Abbildung rechts).

„Das System hat kein chemisches Wissen an sich, es macht keine Quantensimulationen, um die Eigenschaften der Moleküle zu berechnen“, erklärt Materialwissenschaftler Philippe Schwaller, Doktorand an der britischen Universität Cambridge und gleichzeitig bei IBM Research. „Das Modell hat auch keinen Zugang zu einer Datenbank mit Regeln. Es hat einfach nur aus Beispielreaktionen gelernt.“

Diese Beispielreaktionen kommen aus einer Datenbank, die auf US-Pa-

zenten aus den Jahren 1976 bis 2016 basiert.^{7,8)} Diese nutzt das simplified molecular-input line-entry system, kurz Smiles: Damit lassen sich Atome als Buchstaben, Moleküle als Wörter und chemische Reaktionen als Sätze darstellen. „Die Sprache der Chemie ist nicht viel anders als eine natürliche Sprache“, sagt Schwaller. „Ändert man ein Wort in einem Satz, kann sich der Sinn des Satzes völlig verändern – zum Beispiel, wenn man ein ‚nicht‘ in einen Satz einfügt. Bei einer chemischen Reaktion kann eine neue funktionelle Gruppe den Ausgang der Reaktion völlig verändern.“ Ähnlich funktionieren Deep-Learning-Systeme für Sprache, etwa Übersetzungsprogramme wie Deepl: Sie lernen, wie wichtig einzelne Wörter für die Bedeutung eines Satzes sind.⁹⁾

RXN for Chemistry liege inzwischen bei 90,4 Prozent aller Vorhersagen richtig, sagt Schwaller. Viel besser werde man nicht werden, schon weil die Datenbank, die allem

zugrunde liegt, viele falsche Reaktionen enthalte. Bereits beim Verfassen der Patente passierten Fehler, ab und zu fehle ein Kohlenstoffatom, und beim Extrahieren und Umschreiben in computergeeignete Datenbanken gingen Infos verloren.

Die Forscher wollen der KI weitere Reaktionen beibringen, etwa aus wissenschaftlichen Artikeln statt aus Patenten. Bisher sind zudem keine Reaktionsbedingungen berücksichtigt. „Als wir anfangen, mit experimentellen Chemikern zusammenzuarbeiten, wurde uns schnell bewusst, dass man noch viel mit den Reaktionsbedingungen herumspielen kann“, sagt Schwaller und lacht. In einem nächsten Schritt wollen sie die KI daher damit vertraut machen, wie unterschiedliche Reaktionsbedingungen wirken.

Rückwärts gewandt

Langfristiges Ziel bei RXN for Chemistry sei ein System, das auch Retrosynthesen beherrscht, sagt Schwaller. Eine solche KI entwickelte vor kurzem das Team um den Chemiker Marwin Segler, ehemaliger Doktorand am organisch-chemischen Institut der Universität Münster und jetzt leitender Forscher beim Londoner Unternehmen Benevolent AI.

Zusammen mit Chemiker Mark Waller und Wirtschaftsinformatiker Mike Preuß entwickelte Segler ein Programm, das 12,4 Millionen einstufige Reaktionen kennt und für ein gewünschtes Molekül eine Syntheseroute vorschlägt.¹⁰⁾ Es plant die Synthese so weit zurück, bis es bei gut zugänglichen Ausgangsprodukten angekommen ist. Für Routen mit zehn oder mehr Schritten braucht das System nur ein paar Minuten, und „je länger man wartet, desto mehr Optionen spielt der Rechner durch“, sagt Segler. Auch hier lernt die KI nur aus den Daten selbst, sie greift nicht auf eine Datenbank mit Regeln zurück, wie das beispielsweise Chematica tut.

GLOSSAR: Maschinen zum Lernen bringen

Künstliche Intelligenz (KI): Teilgebiet der Informatik, das versucht, menschliche Intelligenz mit Maschinen zu simulieren. Eine KI ist ein Computersystem, das lernt, Schlussfolgerungen zieht und sich korrigiert. Der Begriff ist nicht eindeutig abgrenzbar, da bereits Intelligenz nicht genau definiert ist.

Maschinelles Lernen: Wissen künstlich aus Erfahrungen generieren. Ein künstliches System lernt aus Beispielen und kann diese nach Beendigung der Lernphase verallgemeinern. Mögliche Anwendungen sind automatisierte Diagnoseverfahren, Erkennen von Kreditkartenbetrug, Sprach-, Bild- und Texterkennung.

Deep Learning: Bereich des maschinellen Lernens, das neuronale Netze und große Datenmengen nutzt. Die Funktionsweise ist vom Lernen im menschlichen Gehirn inspiriert. Das

System kann das Erlernte immer wieder mit neuen Inhalten verknüpfen und dadurch lernen.

Künstliche neuronale Netze: Computermodelle, die sich an den Organisationsprinzipien und den Lernprozessen des menschlichen Gehirns orientieren. Wie neuronale Netze in einem menschlichen Gehirn werden Netzwerke künstlicher Neuronen angeordnet, meist in hintereinander liegenden Schichten. Die Zahl der Neuronen und der Neuronenschichten sowie der Verbindungsmöglichkeiten der Neuronen verschiedener Schichten bestimmt, wie komplex das neuronale Netz ist und wie fähig, Probleme zu lösen.

Algorithmus: Handlungsvorschrift, um ein Problem zu lösen, bestehend aus definierten Einzelschritten. Hierzu braucht es weder KI noch maschinelles Lernen.

„Das Tool eignet sich gut für Verbindungen, wie sie in der medizinischen und agrochemischen Wirkstoffentwicklung vorkommen“, erläutert Segler. „Komplexe Naturstoffe und exotischere Systeme mit mehreren anellierten Ringen funktionieren noch nicht so gut.“ Allerdings sei die Retrosynthese dieser Verbindungen derzeit auch für Menschen schwierig. Meistens schlägt der Rechner laut Segler Routen mit Standardreaktionen vor, die auch ein Chemiker bevorzugen würde. Manchmal allerdings greife der Computer zu selteneren Reaktionen – „der Computer erscheint oft mutiger als der Mensch.“

Medizin und Toxikologie

Künstliche Intelligenz eignet sich auch für die Kandidatensuche in der Medikamentenforschung. Das Unternehmen Atomwise in San Francisco etwa hat das Programm Atomnet entwickelt, das vorhersagt, welche kleinen Moleküle wie fest an welche Proteine binden.¹¹⁾ Es generiert ein dreidimensionales Modell für jedes Protein-Liganden-Paar und lernt, welche Eigenschaften die Bindung beeinflussen, ohne dass es ihm jemand gesagt hat.

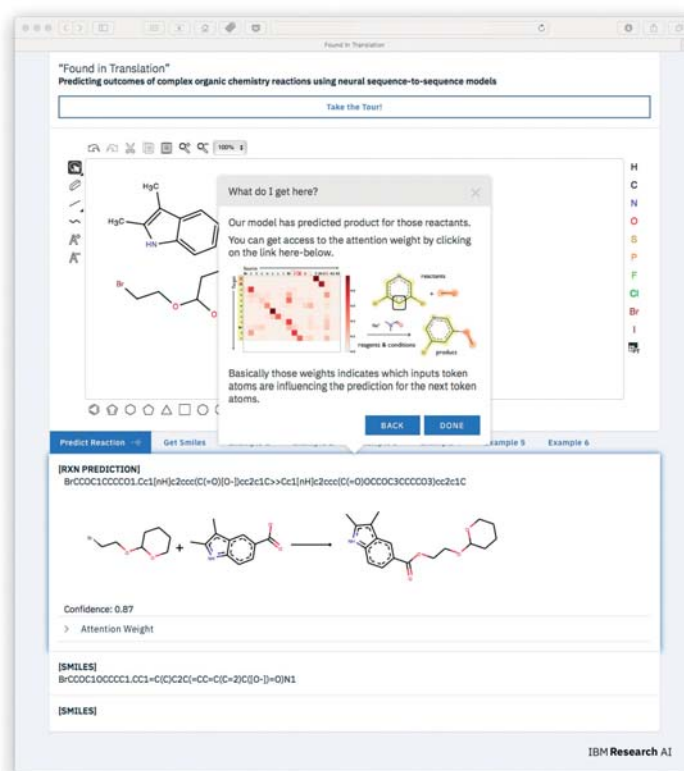
Ein Team um Thomas Hartung von der Johns Hopkins University im US-amerikanischen Baltimore wiederum hat ein System darauf trainiert, die Toxizität unbekannter Verbindungen vorherzusagen.¹²⁾ „Künstliche Intelligenz basierend auf einem Big-Data-Satz an Giftstoffen identifiziert giftige Substanzen sogar besser als Tierversuche“, sagte Thomas Hartung auf dem Euroscience Open Forum in Frankreich im Juli 2018.

Chemiker bald arbeitslos?

Wenn die HPLC sich selbst steuert, Roboter ohne menschliche Hilfe pipettieren und KI Retrosynthesen plant – sind dann noch Chemiker nötig? Selbstverständlich, meint Marwin Segler. Sein Retrosynthese-

Werkzeug etwa sei lediglich ein Assistent, der einem Chemiker helfe, so schnell wie möglich von A nach B zu kommen. „Navigationsgeräte haben vielleicht Landkarten aus Papier überflüssig gemacht – aber nicht den Autofahrer, der entscheidet, wo es hingehet.“

Auch Philippe Schwaller von IBM Research sieht Vorhersagewerkzeuge wie RXN for Chemistry lediglich als Unterstützung, um schneller eine Synthese zu entwerfen – „damit Chemiker sich auf wichtigere Fragen und kreativere Aufgaben konzentrieren können.“ Statt darüber nachzudenken, wie man ein Molekül macht, bleibt mehr Zeit, um zu überlegen, welches Molekül man braucht und wofür. Aber auch wenn künstliche Intelligenz Chemiker vermutlich nicht so bald arbeitslos machen wird – in einem sind sich die Experten einig: Sie wird mit Sicherheit die Arbeitswelt verändern. <<



So funktioniert RXN for Chemistry: Das Programm ermittelt, welche Atome der Ausgangsmoleküle am wahrscheinlichsten an der Reaktion beteiligt sind (oben) und wirft dann das wahrscheinlichste Reaktionsprodukt aus (unten).

Die promovierte Chemikerin **Brigitte Osterath** arbeitet als freie Wissenschaftsjournalistin nicht nur in Deutschland.

- 1) D. Lowe, *Nature* 2018, 555, 592
- 2) T. Klucznik, B. Mikulak-Klucznik, M. P. McCormack et al., *Chem* 2018, 4, 522
- 3) K. Molga, P. Dittwald, B. A. Grzybowski, *Chem* 2019, 5, 460
- 4) J. N. Wei, D. Duvenaud, A. Aspuru-Guzik, *ACS Cent. Sci.* 2016, 2, 725
- 5) <https://rxn.res.ibm.com/>
- 6) P. Schwaller, T. Gaudin, D. Lányi, C. Bekasa, T. Laino, *Chem. Sci.* 2018, 9, 6091
- 7) W. Jin, C. W. Coley, R. Barzilay, T. Jaakkola, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA. <https://arxiv.org/abs/1709.04555>
- 8) D. M. Lowe, Ph.D. thesis, University of Cambridge, 2012
- 9) www.deepl.com
- 10) M. H. S. Segler, M. Preuss, M. P. Waller, *Nature* 2018, 555, 604
- 11) www.atomwise.com/our-technology
- 12) T. Luechtefeld, D. Marsh, C. Rowlands, T. Hartung, *Tox. Sci.* 2018, 165, 198