



KKC-KOLUMNE

Wie die Google-KI das alte Problem der Proteinfaltung löst

Bei der Entwicklung von Medikamenten gegen das Corona-Virus spielt die Simulation von Proteinfaltungen eine entscheidende Rolle. Am 10. Dezember 2020 veröffentlichte das Fachmagazin Nature einen Bericht mit der Schlagzeile „Es wird alles ändern: AI macht einen großen Sprung bei der Lösung von Proteinstrukturen.“ Was war passiert, dass die medizinische Fachwelt in große Aufregung versetzt wurde?

Das von Google entwickelte KI-Programm DeepMind wurde bislang als weltbeste Schach- und Go-Spieler bekannt. Die neue Software AlphaFold 2.0 schloss kürzlich bei einem Wettbewerb zur Vorhersage von Proteinfaltungen mit einem sensationellen Ergebnis ab. Mit 92,4 von 100 möglichen Punkten schlug die Google-KI den bisherigen Bestwert von 75 der konkurrierenden Forschergruppen. Sie übertraf sogar das momentan weltstärkste Computersystem Folding@home.

Die Proteinfaltung ist ein Kernproblem der Biologie, an dem sich seit mehr als 50 Jahren die Forscher versuchen. Dank der Genomforschung sind zwar 180 Millionen Sequenzen der Proteine bekannt, aber nur bei einem Tausendstel hat man die dreidimensionale Proteinstruktur entschlüsselt. Diese ist aber entscheidend für das Verständnis der Funktion sowie der Diagnose und Therapie von nahezu jeder Krankheit, darunter Krebs, Demenz, Diabetes und Covid-19. All diese Krankheiten sind auf eine fehlerhafte Proteinfaltung zurückzuführen. Mittels manipulierter Moleküle könnte man den Faltungsprozess eines Proteins unterstützen oder eine fehlerhafte Faltung zerstören.

Die Anzahl der theoretisch möglichen Varianten einer Faltung wird auf 10 hoch 300 geschätzt und ist somit selbst mit maximaler Computerpower nicht berechenbar. AlphaFold 2.0 nutzte für den Durchbruch eine neuartige KI-Architektur, den Aufmerksamkeits-Mechanismus. Den DeepMind-Entwicklern gelang es damit, die Fähigkeit des menschlichen Gehirns nachzubilden, welches in der Lage

ist, aus einer Fülle an Daten in kurzer Zeit die Information auszuwählen, die für eine Entscheidung relevant ist und dabei alles Unwichtige auszublenden. Die neue Transformer-Architektur der Bild- und Sprachanalyse erlaubt neuronalen Netzen das Lernen wesentlich komplexerer Zusammenhänge, indem sie über tausend Wörter im Kontext der gesamten Umgebungsdaten erfasst.

Allerdings benötigt DeepMind erhebliche Datenmengen als Lernmaterial. Bei seltenen Krankheiten und bei Einflüssen von pH-Wert, Temperatur oder Salzkonzentrationen auf Proteinstrukturen liegen zur Nutzung durch AlphaFold 2.0 noch keine systematisch gesammelten Daten vor.

Manfred Kindler
Präsident des Krankenhaus-
Kommunikations-Centrums e.V. (KKC)
Kontakt: m.kindler@kkc.info

Folding@home

Das im Jahre 2000 begonnene Projekt der Stanford-University für Simulationsanwendungen kopierte vor der Corona-Pandemie die verteilte Leistung von 30.000 Computern, meist von privaten PC-Besitzern. Am 13. April 2020 erreichte das System mit 1,4 Millionen Nutzern eine Rekordleistung von 2,43 Exaflops und war damit schneller als alle TOP500-Supercomputer der Welt zusammengenommen. Hierdurch konnte die Proteinfaltung auf atomarer Ebene über einen tausendmal längeren Zeitraum als bislang simuliert werden.

Die Forschung konzentriert sich neben der Behandlung von Virenerkrankungen wie SARS-CoV 2 auf Alzheimer- und Huntington-Krankheiten, Krebstherapien und die Glasknochenkrankheit.