

Der Übergang zu einer neuen Sicht auf Kognition: Der Konnektionismus und die Simulation neuronaler Prozesse

Wie oben ausgeführt, führte die Sichtweise der klassischen Kognitionswissenschaft zu einer starken Kritik, wobei in unserem Kontext ein zentraler Punkt ist, dass die oben skizzierten Systeme nicht lernen konnten. Mitte der 1980er Jahre kam es, ausgelöst durch eine in dem Doppelband „Parallel Distributed Processing“ von David E. Rumelhart und James F. McClelland (1986) veröffentlichte Sammlung von Einzelarbeiten, zu einem Siegeszug eines neurowissenschaftlich inspirierten Modells, das bislang vom Mainstream der Kognitionswissenschaft ignoriert worden war: die *Künstlichen neuronalen Netze* (KNN).

Ein KNN (in der Regel eine Computersimulation, es sind aber auch physische Umsetzungen möglich) besteht aus vielen sehr einfachen, identisch aufgebauten Einheiten, die als *Units* oder künstliche Neuronen bezeichnet werden und über sogenannte Gewichte (diese simulieren in sehr vereinfachter Weise die Funktion von Synapsen) untereinander verbunden sind. Typischerweise haben KNN, welche für die Modellierung kognitiver Leistungen herangezogen werden, eine Schicht von *Units*, der Stimuli präsentiert werden (engl. „input layer“), eine Schicht von *Units*, die etwas ausgeben (engl. „output layer“) sowie eine oder mehrere Schichten dazwischen (engl. „hidden layer“), die jeweils linear oder rekursiv miteinander verbunden sind.

Die Aufgabe/Funktion jeder einzelnen *Unit* besteht darin, die Aktivierungen der eingehenden Verbindungen zu integrieren und an die jeweils benachbarten weiterzugeben. Dies geschieht durch einfaches Aufsummieren der gewichteten Inputs und Weitergabe der eigenen Aktivierung, wenn diese einen bestimmten Schwellenwert überschreitet. Dies wird von allen *Units* parallel durchgeführt und konstituiert das Gesamtverhalten des KNN. Wesentlich ist, dass diese Netze in ihrer Architektur (meist) fest „verdrahtet“, die Gewichte aber veränderbar sind. In Kombination mit den Inputs aus der Umwelt sind die Gewichte für die Verhaltensdynamik des Netzwerks verantwortlich. Anstatt die Gewichte von Hand einstellen zu müssen, wurde in den frühen 1980er Jahren ein Algorithmus gefunden, der die schrittweise Veränderung der Gewichte in einem

Trainingsprozess in einer Weise durchführt, dass das Netz seine Aufgabe schließlich fast perfekt lösen kann: KNN können ohne Eingabe von Regeln und Symbolen, nur anhand von Beispielen, mit denen sie trainiert werden, lernen. Nach jeder Aufgabe bekommen sie ein Feedback, ob die Antwort richtig oder falsch war, indem ihre Gewichte ganz minimal in Richtung der korrekten Lösung verändert werden, bis sie fast zu 98 Prozent richtig liegen. Allerdings können sie nicht alle Aufgaben gleichermaßen gut lösen. Gut sind sie, kurz gesagt, bei Mustererkennung, Kategorisierungsaufgaben, Vorhersage von Wahrscheinlichkeiten, etc. Modelle von Aspekten menschlicher Kognition, die auf KNN basieren, weisen einige sehr charakteristische Eigenschaften auf (ausführlicher behandelt z.B. in Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996):

- Bei Kategorisierungsaufgaben kann ein KNN generalisieren. Trainiert man ein solches Netz zum Beispiel dazu, Bilder von Gesichtern zu erkennen und zeigt ihm ein Gesicht, das es nicht im Rahmen seines Trainings „gelernt“ hat, kann es dieses mit sehr, sehr hoher Wahrscheinlichkeit der Kategorie „Gesicht“ zuordnen.
- Sie können dieselben Fehler bei der Generalisierung machen wie Menschen. Zum Beispiel übergeneralisieren Kinder beim Spracherwerb häufig unregelmäßige Formen, wenn sie Grammatik lernen, sagen zum Beispiel „goed“ (gehte) statt „went“ (ging).
- Die Lernkurve gleicht häufig der, die bei Menschen gefunden wurde: KNN lernen zunächst sehr schnell, dann flacht die Lernkurve zusehends ab.
- Auch wenn das Netz richtige Antworten liefert, kann es sein, dass das, was es gelernt hat, nicht der Intention der Architekten/innen des Netzes entspricht. So unterschied ein KNN, das lernen sollte Gesichter voneinander zu unterscheiden, die gezeigten Bilder auf Basis des Haaransatzes voneinander.
- Das in einem KNN repräsentierte Wissen ist in zweifacher Weise robust: (1) Beim Lernen eines neuen Assoziationspaares „vergisst“ das Netz nicht das bereits Gelernte; (2) auch vergisst das Netz nicht schlagartig alles, wenn man einzelne Neuronen und Gewichte entfernt.

Mit diesen Eigenschaften stellten KNN noch keinen grundsätzlichen Widerspruch zur klassischen Sicht auf Kognition dar. Man konnte sie durchaus als eine Ergänzung begreifen, die eine Erklärung lieferte, wie durch Lernen (von Kategorien) Symbole „in den Kopf kommen“ können. Allerdings stellte sich die Frage, welcher Natur diese Symbole denn seien. In neuronalen Netzwerken sind Symbole und Regeln nicht sauber voneinander getrennt „abgespeichert“, vielmehr ist alles, was das Netz „weiß“ in der gesamten Architektur des Netzes, das heißt in allen Neuronen, allen Gewichten und deren Konfiguration, verteilt repräsentiert. Man spricht daher auch von einem *subsymbolischen Ansatz* (Rumelhart et al., 1986; Smolensky, 1998; Elman, 1990).

“

!

Eigenschaften Künstlicher Neuronaler Netze (KNN):

- KNN lernen anhand von Beispielen (Erfahrungslernen), ohne explizit eingegebene Regeln und Symbole.
- Sie können, kategorisieren, generalisieren und Muster erkennen.
- Die Repräsentation ist verteilt (subsymbolischer Ansatz) und robust.
- Sie machen ähnliche Fehler wie Menschen und sind biologisch plausibler, weil ihre Architektur von der Struktur natürlicher Neuronaler Netze inspiriert ist.

Konsequenzen für unsere Begriffe von Wissen und Lernen

Der erste Erfolg der Künstlichen Neuronalen Netze war zunächst, ein biologisch plausibles Modell dafür zu liefern, wie Symbole und Regeln gelernt werden können. In gewisser Weise setzen sie eine Ebene tiefer an als der symbolverarbeitende Ansatz: Sie bieten eine Alternative auf der subsymbolischen Ebene an (Smolensky, 1998). Konsequenz war aber ein neues Bild von Repräsentation und Eigenschaften kognitiver Systeme.

Damit erlauben die KNN eine fundamental andere Sichtweise auf Wissen (Peschl 1994; 1997). Zunächst ist klar: Das „Wissen im Kopf“ muss strukturell keineswegs identisch mit den in Symbolen und Regeln beschriebenen Strukturen der Welt sein. Nicht die korrekte Abbildung der Welt ist relevant, sondern das adäquate Ergebnis, also gewissermaßen die Handlung, die in die Struktur der Umwelt passen muss. Als eine Konsequenz der Aufgabe des Konzeptes der Abbildung sind die Inhalte der Repräsentation, im Gegensatz zu klassischen symbolverarbeitenden Systemen, nicht mehr unmittelbar verständlich; vielmehr bedarf es aufwändiger statistischer Verfahren, um herauszufinden, was so ein Netz eigentlich gelernt hat.

Eine weitere interessante Konsequenz der verteilten Repräsentation ist, dass, im Gegensatz zum klassischen Ansatz, keine Trennung zwischen Inhalt und Substrat besteht: das Netz *ist* sein Wissen und dieses Wissen ist in der Architektur verkörpert, zumindest potentiell, denn zumeist handelt es sich bei KNN ja um Computersimulationen (zum Beispiel Clark 1999, 2001). Damit gibt es auch keine leicht voneinander trennbaren Wissensobjekte mehr, vielmehr werden alle dem neuronalen Netzwerk präsentierten Stimuli (zum Beispiel Bilder) von allen Neuronen und allen Gewichten repräsentiert. Die Repräsentation des KNN kann man als einen Raum verstehen, in dem Inputs kategorisiert und dadurch in eine Beziehung (in diesen einfachen Modellen ist es Ähnlichkeit) gesetzt werden.

Die Analogien zu Bildungskontexten, insbesondere Frontalsituationen liegen auf der Hand: Die *Input-Output-Relation* ist dadurch bestimmt, dass Lernende durch Vortrag, Durcharbeiten eines Lernpfades, etc. einen Stoff präsentiert bekommen und in einer Prüfungssituation den gewünschten *Output* zu liefern haben. Doch Lernen ist kein Kopiervorgang von Wissensobjekten. Was gelehrt wird, muss noch lange nicht das sein, was gelernt wird. Nachdem

Lernen in unserem Bildungssystem häufig outputgetrieben ist („Was muss ich tun, um eine gute Note zu bekommen?“), liegt es daher nahe, Prüfungen so anzulegen, dass nicht isolierte Fakten abgefragt werden, sondern ein Verständnis der Kategorien und Bezüge des gesamten „Wissensraumes“ gefordert ist und eine *Deep-Learning*-Strategie auf Seiten der Lernenden notwendig wird.

“

?

Welche Prüfungssituationen, die Sie als Lernende erlebt oder als Lehrende gestaltet haben, haben Fakten abgefragt und welche Prüfungsmethoden sind ‚tiefer‘ gegangen?

Revision #1

Created 28 February 2025 21:15:17 by Bernd Grabner

Updated 28 February 2025 21:15:17 by Bernd Grabner