

Symbolische und konnektionistische Repräsentation

(Paul Natterer)

Die einflussreichsten und ausgearbeitetsten **seriellen** und **symbolischen** sowie **explizit regelgeleiteten** Neuroinformatik-Modelle zu Wahrnehmung, Gedächtnis und Kognition stammen von Newell: *Unified Theories of Cognition*, Harvard 1989, und Marr: *Vision: A computational investigation into the human representation and processing of visual information*, San Francisco 1982, der für die Diskussion der Wahrnehmung und ihrer Repräsentationsmedien bis heute maßgebliche Diskussionsrahmen. In Folge daher zunächst ein Überblick zu Marrs Modell:

- Die aufeinander folgenden Repräsentationen des Wahrnehmungsobjektes sind nach dieser einflussreichsten und experimentell abgestützten Theorie der Wahrnehmung und Objekterkennung die 2-dimensionale Primärskizze („primal sketch“) als (1) **grobe Primärskizze** („raw primal sketch“) und (2) **voll entwickelte Primärskizze** („full primal sketch“) auf der Ebene der Vorstellungskraft (imagery); (3) die **2,5-dimensionale Skizze** („2,5-D sketch“), eine vorläufige dreidimensionale, beobachterzentrierte Erzeugung statischer und bewegter Wahrnehmungsobjekte; (4) das objektzentrierte begriffliche und allgemeingültige **3-D Modell** der Wahrnehmungskonstrukte.
- Zu (1) Grobe Primärskizze: Hier geht es um **Konturentdeckung** und **Merkmalsextraktion**. Marr und Mitarbeiter gingen v.a. aus von der fundamentalen Hell-Dunkel- sowie Form-Wahrnehmung. Ihr Ansatz basiert auf der sog. **Raum-Frequenz-Analyse**. Mathematiker kennen die Technik unter dem Namen Fourieranalyse. Sie gehört zur mathematischen Analysis (Differential- und Integralrechnung) trigonometrischer Funktionen.
- Der Ausgangspunkt ist die Tatsache, dass man jedes elektromagnetische Reizmuster (etwa weiße oder farbige Lichtwellen) als eine Überlagerung unterschiedlicher Sinuswellen analysieren kann. Das Verfahren selbst wird **Marr-Hildreth-Algorithmus** genannt. Es besteht in der Anwendung von zwei und mehr sogenannter gaußscher Filter auf die von Wahrnehmungsobjekten reflektierten Lichtwellen. Diese Filter sind Funktionen in der Form gaußscher Normalverteilungen unterschiedlicher Frequenzen, welche also sensibel sind für unterschiedliche Graustufen oder Intensitätswerte.

- Die **Lokalisierung von Helligkeits- bzw. Intensitätswechseln** wird erzielt durch Feststellen der Nullstellen und Extrempunkte mittels der 1. und 2. Ableitungen, d.h. durch Differenzierung der gaußschen Funktionen. Die Summe der 2. Ableitungen durch den sogenannten Laplace-Operator erlaubt die orientierte Lokalisierung von Intensitätswechseln und eine Intensitätsskala in einem 2-dimensionalem Koordinatensystem.
- Die grobe Primärskizze ermittelt vier Faktoren: **Randsegmente, Balken, Enden, Punkte** als Bestandteile von Konturen und Oberflächenstrukturen einschließlich deren **Orientierung – Kontrast – Länge – Breite – Position**.
- Zu (2) Voll entwickelte Primärskizze Hier geht es um **Integration der Merkmale** bzw. Informationen aus Konturen, Oberflächenstruktur und Schattierung. Das Resultat sind **globale Strukturen** oder Figuren sowie deren innere und Oberflächenstruktur. Die bekannten **Gestaltgesetze** der Wahrnehmungspsychologie spielen dabei eine große Rolle. So die Prinzipien der **örtlichen Nähe** (*clustering*); der **ähnlichen Orientierung** und der **globalen Richtungstendenz** der Konturen usw.
- Zu (3) Die 2 1/2-dimensionale Skizze: Der wirkliche Dateneingang der visuellen Wahrnehmung ist nicht nur zweidimensional und statisch, sondern beidäugig (binokular) und damit **dreidimensional** sowie bewegt, **dynamisch**.
- Die visuelle Welt wird darüberhinaus „eingesammelt“ durch eine Serie von diskreten Fixationen oder Schnappschüsse, getrennt durch Augenbewegungen (Sakkaden), während denen keine Wahrnehmung stattfindet! Die **Integration** dieser aufeinander folgenden **Schnappschüsse von leicht veränderten Netzhautbildern** ist Aufgabe der Wahrnehmung bei beidäugig- gem Tiefensehen und den unterschiedlichen Formen echter und scheinbarer Bewegungswahrnehmung.
- Das Problem ist: Welche Aspekte des Netzhautbildes entsprechen welchen identischen Objekten? Wie geschieht der **Abgleich zwischen Netzhautbildern und Gegenständen** (*matching*)? Marr erfasst diese Wahrnehmungsleistungen durch Vektorfelder, wobei die Länge der Vektoren den Neigungsgrad der optischen Fläche, deren Orientierung die Richtung der Oberfläche, deren Skalarwert den Abstand vom Beobachter ausdrücken.
- Zu (4): 3-D-Repräsentation und begriffliche Objekterkennung. Marr schlug hier einen Satz elementarer Merkmale oder **3-D-Module in Zylinderform** vor, die sich – hierarchisch organisiert – auf verschiedenen Größenstufen wiederholen, jeweils gruppiert um eine Hauptachse (etwa die Rumpflinie) und Komponentenachsen (etwa die Arm- und Beinglied-

maßen). Das Ganze erinnert ein wenig an die mit Schnüren zusammengehaltenen Holzzylinder der Hampelmänner in Kinderzimmern.

- Dieser 3-D-Ansatz gilt als eher spekulativ und experimentell wenig gesichert. Andere, experimentell gestützte Forschungen belegen den Primat oder mindestens die Mitursächlichkeit der dynamischen, speziell biologischen **Bewegungswahrnehmung** bei der Objekterkennung. Jüngere Modelle arbeiten auch, besonders hinsichtlich rauer, unregelmäßiger, verzweigter Gebilde (Bäume, Wolken), mit den bekannten **fraktalen Strukturen** der Chaostheorie, die immer wieder dieselbe Form bei unterschiedlicher Größe erzeugen.
- Der Hauptkonkurrent serieller, symbolischer und regelgeleiteter Informatikmodelle zur Wahrnehmung ist das Modell der **impliziten parallel-verteilten Repräsentation und Verarbeitung** (*Parallel-Distributed-Processing*, PDP). Es kommt in vielen Aspekten der tatsächlichen Funktionsweise des Gehirns soweit bis heute erkennbar am nächsten. Es steht auch Pate bei der Entwicklung lernfähiger, leistungsstarker Schrift-, Gesichts- und Spracherkennungsprogrammen und elektronischen Zielsuchsystemen, die immer häufiger zum technischen Einsatz kommen.
- Die Leitidee und ihre mathematische Umsetzung erinnert an den Vektorraum der Quantenmechanik: Das Gehirn repräsentiert demnach die unterschiedlichen Merkmale realer Objekte durch deren Position in einem geeigneten **vieldimensionalen Zustandsraum** (Koordinatensystem, **Vektorraum**). Und das Gehirn führt Berechnungen an solchen Repräsentationen aus mit Hilfe von Koordinaten-Transformationen von einem Zustandsraum in den anderen.
- Das heißt, die primäre Aufgabe: die **neurale Repräsentation** oder Kodierung des sensorischen Dateneingangs (= Informationsaufnahme) erfolgt durch **sensorische Koordinatenräume in Form n-dimensionaler Vektorräume** oder Kodierungsvektoren. Die zweite Aufgabe: die rechnerischen Transformationen, die **neurale Komputation** (= Informationsverarbeitung) dieser elementaren mentalen Repräsentationen (Merkmale) durch deren assoziative Verknüpfung und abstrahierende Klassifikation vollzieht sich in Form linearer Abbildungen – **Vektortransformationen** – von n-dimensionalem Eingang zu m-dimensionalem Ausgang mittels der auch aus der Schulmathematik bekannten Matrizenmultiplikation.
- Die abstrahierende Klassifikation, d.h. die Unterordnung einer Menge von individuellen Objekten unter einen Allgemeinbegriff, vollzieht sich speziell so, dass jeder der vielen Vektoren des Dateneingangs in einen einzigen zusammenfassenden Ausgabevektor umgewandelt wird, der in

einem **prototypischen Vektorraum** gipfelt. Dieser verkörpert und definiert idealtypisch ein Objekt.

- Gleiches gilt für die sensomotorische Koordination und die motorische Ausgabekodierung und -berechnung. Jedem Muster, jeder Farbe, jeder Bewegung, jedem Begriff, jedem Laut usw. entspricht ein besonderes und plastisches, d.h. lernfähiges **Aktivationsmuster in einem neuronalen Netzwerk**. Die Konfiguration der Werte oder Gewichte (*weights*), d.h. der Stärken der synaptischen Verbindungen in diesem neuronalen Netzwerk ist ebenfalls in diesem vektoriellen Zustandsraum kodiert und repräsentiert das gespeicherte Wissen über eine Sache oder einen Vorgang.
- Es ist klar geworden: PDP bietet nicht zuletzt eine Technik des Aufeinanderabbildens verschiedener Koordinatensysteme, die der tatsächlichen visuellen und sonstigen Informationsverarbeitung gerecht wird. Die Neurowissenschaften zeigen dies besonders gut an den sog. kognitiven Landkarten oder **Referenzrahmen** der Wahrnehmung und ihrer Gegenstände, die aufeinander abgebildet und dabei immer maßstabsgetreuer und objektiver werden.
- Am Anfang steht ein (1) netzhautbezogener (retinotoper) Rahmen mit logarithmischer Auflösung vom Zentrum (Fovea) zu den Rändern, dann folgt (2) der egozentrische Rahmen mit dem eigenen Kopf und Körper als Zentrum, sowie (3) der umwelt- oder objektzentrierte Rahmen. Dazu kommt noch ein weiterer (4) erfahrungszentrierter Rahmen, der aus dem über die aktuelle Wahrnehmung hinausgehenden Weltwissen geographischer, astronomischer, kultureller, philosophischer Art besteht.
- Was das Verhältnis von explizit-symbolischer zur implizit-verteilter Repräsentation betrifft, so ist die vorherrschende Auffassung der PDP-Forschung, dass diese sich nicht widersprechen, sondern ergänzen. Man sagt gerne, dass Erstere die **Grobstruktur** der Kognition charakterisiert und Letztere deren Mikro- oder **Feinstruktur**. Beide Formate werden in der Forschung oft nebeneinander benutzt und die Grenzen sind fließend.
- Die Grenzen von PDP sind jedoch dann erreicht, wenn es zu den denknötigen logischen Gesetzen begrifflicher Bedeutungen, grammatischer Strukturen und formallogischer Ableitungen kommt, also zur Kognition im eigentlichen Sinn. Kritiker von PDP drücken dies so aus, dass es eine Theorie auf der Stufe der Hardware und des Maschinenkodes ist, aber keine Theorie der intelligenten Software und logisch strukturierter Programmiersprachen.

Derzeitiger Hauptkritiker des PDP-Ansatzes ist Steven Pinker (Harvard), einer der bekanntesten Sprach- und Kognitionswissenschaftler der Gegenwart und Erneuerer und Fortführer der aprioristischen Chomsky-Schule in der Linguistik. Pinkers Buch *The Blank Slate. The Modern Denial of Human Nature* (New York 2002) plädiert dafür, dass der Ansatz der parallel-verteilter Informationsverarbeitung (PDP, Konnektionismus)

v.a. assoziative, basale kognitive Leistungen erklären kann. Dieser Ansatz versagt aber, so Pinker, vor sprachlich-begrifflicher (symbolischer) Informationsverarbeitung mit ihren komplexen Datenbasen und Funktionen, kategorialen und logischen Relationen und deren Rekursivität, Kompositionalität und Quantifizierung (2002, 79–82).

Eine weitere ungelöste Hypothek des konnektionistischen Ansatzes, dessen Hochburg die UC in San Diego ist, ist die extrem lange überwachte oder auch selbstständige **Trainingsphase** in Simulationen der Kognition durch künstliche neuronale Netze, die in der Regel eine zusätzliche sorgfältige Aufbereitung des Datenmaterials erfordert. Sie steigt angesichts der tatsächlichen, extrem komplexen und anspruchsvollen menschlichen Wahrnehmungs- und Kognitionsleistungen rapide ins Unermessliche.

Dazu kommt die von vorne herein in Form von komplizierten *Algorithmen* in die neuronalen Netze einzubauende „künstliche“ Intelligenz: „Much of the interesting work has been done in setting up the innate wiring of the network“ (2002, 83). Bei der menschlichen Kognition erscheint diese auch hier **voraussetzende apriorische Intelligenz** nur durch genetische, angeborene Strukturen der Wissensrepräsentation und -verarbeitung erklärbar, nicht allein aus aktuellen selbstorganisierenden Prozessen: „Nothing comes out of nothing, and the complexity of the brain has to come from somewhere.“ (2002, 75)

Menschliche Erfahrung und Wissen, so Pinker, besteht nicht nur in bildlichen Wahrnehmungen (*perception*) und Vorstellungen (*imagery*), sondern erfordert kognitive Interpretation und Evaluation in Begriffen. Diese enthalten objektive Merkmale und **definitive Bedeutungen** als semantische Essenz (2002, 203–204). Sie sind ihrerseits in angeborene **intuitive Theorien** eingebettet. Begriffe wie Theorien folgen dabei **apriorischen Kategorien** (Objekt – Relation – Ursache – Wirkung – Zahl, Raum, etc.) und **Sprachkompetenzen** (2002, 210–221).