

Helge Ritter

## Dynamik neuronaler Netze



Geboren 1958 in Naila, Oberfranken. Physik und Mathematikstudium an den Universitäten Bayreuth und Heidelberg. Promotion in theoretischer Physik 1988 an der Technischen Universität München. 1989 Gastaufenthalt am Department of Computer Science der Helsinki University of Technology, anschließend Research Assistant Professor am Department of Physics der University of Urbana-Champaign und dem Beckman Institute of Advanced Science and Technology, Illinois. Seit Oktober 1990 Professur an der Technischen Fakultät der Universität Bielefeld. Arbeitsgebiete: Informationsverarbeitung in neuronalen Netzen, Mustererkennung, Computersehen und Robotik. — Adresse: Arbeitsgruppe Neuroinformatik, Technische Fakultät, Universität Bielefeld, Postfach 100 131, D-33501 Bielefeld.

Ich erinnere mich noch gut des zunächst trüben Oktober-Freitags, mit dem mein Aufenthalt im Wissenschaftskolleg begann. Ich war mit dem Zug angereist und kam gerade rechtzeitig mittags an, um noch von Frau Sanders sogleich dem gemeinschaftlichen Essen zugeführt zu werden, bei dem ich erste Bekanntschaft mit den bereits seit einigen Tagen anwesenden anderen Fellows schließen konnte.

Nach dem Mittagessen führte mich Dr. Meyer-Kalkus durch die verschiedenen Teile des Hauses und machte mich mit einigen Mitarbeitern bekannt. Es war jetzt früher Nachmittag und draußen setzte sich strahlender Sonnenschein durch. Heute weiß ich, daß dies eine höchst passende Symbolik für die im Wissenschaftskolleg bevorstehende Zeit war.

Ich begann dies allerdings bereits am ersten Tag zu ahnen, als ich nachmittags in mein Appartement einzog, den netten Willkommenskorb vorfand (eine von den vielen scheinbaren „Kleinigkeiten“, die in ihrer Summe sehr zu der besonderen Atmosphäre des Hauses beitragen) und Dr. Lindenberg mir mein Büro zeigte und mich dort in die Benutzung einer eigens für meine Arbeit am Kolleg eingerichteten Sun-Workstation einwies.

Ursprünglich hatte ich etwas Sorge gehabt, ob am Wissenschaftskolleg eine für die Durchführung meiner Arbeiten geeignete Rechnerausstattung verfügbar sein würde; nun war ich sehr erfreut und zugleich

beruhigt, daß sich meine Bedenken als gänzlich gegenstandslos erwiesen und alles bestens vorbereitet war.

Diese gute Grundstimmung sollte auch im weiteren Verlauf des Jahres nicht getrübt werden. Die EDV-Ausstattung wurde von Herrn Lindenberg und seinem Stab in vorbildlicher Weise betrieben und die wenigen auftretenden Probleme wurden jedesmal schnell behoben, so daß das Arbeiten eine Freude blieb. Es konnte also losgehen. Mein Ziel war die Untersuchung von Eigenschaften der Dynamik einer gewissen Klasse neuronaler Netze. Die Besonderheit betraf dabei die Struktur der Verknüpfungen zwischen den einzelnen Neuronen. Biologische Neuronennetze weisen eine außerordentlich hohe Verschaltungsdichte auf. Modelle derartiger Netze sind aber mathematisch nur dann gut behandelbar, wenn über die Verschaltungsstruktur sehr einschränkende Annahmen getroffen werden. Die beiden wichtigsten Modellklassen ergeben sich aus zwei unterschiedlichen Vereinfachungsannahmen: im einen Fall fordert man eine Verschaltungsstruktur, bei der geschlossene Schleifen (also etwa Verbindungen von einem Neuron A zu einem Neuron B und von dort zurück nach A) verboten sind. Solche Netze besitzen dann keinerlei Rückkopplung. Auch wenn sich damit nur sehr begrenzte Aspekte realer Neuronennetze modellieren lassen, spielen derartige Netze heutzutage bereits in vielen technischen Anwendungen eine große Rolle, da wir für diesen Fall praktikable Trainingsverfahren kennen, mit denen sich anhand von Datenbeispielen die Reaktionen des Netzes auf Eingaben in gewünschter Weise trainieren lassen.

Die zweite Klasse von Netzwerken läßt Rückkopplungsschleifen zu, bedingt jedoch eine Einschränkung in Form einer Symmetrieforderung: wenn ein Neuron A auf ein Neuron B einwirkt, so muß B in genau gleich starker Weise auf A zurückwirken. Derart charakterisierte Netze gehen auf den Physiker Hopfield zurück und zeigen ein Verhalten, welches ihnen die Bezeichnung „Attraktornetzwerk“ eingebracht hat: die symmetrischen Kopplungen zwischen den Neuronen führen für jeden beliebigen Anfangszustand der Neuronenaktivitäten nach einiger Zeit zu einem Gleichgewichtszustand, in dem sich die Neuronenaktivitäten nicht mehr ändern. Dabei gibt es in der Regel mehrere mögliche Gleichgewichtszustände, die miteinander für die Herausbildung des Endzustandes der Neuronenaktivität miteinander in Wettbewerb stehen und dabei für das System gewissermaßen wie Anziehungspunkte in einem (abstrakten) Zustandsraum wirken. Das Anfangsmuster der Neuronenaktivitäten bestimmt dabei, welcher Gleichgewichtszustand den Wettbewerb gewinnt und sich am Ende durchsetzt. Die allmähliche Umwandlung des Anfangsmusters in eines der Gleichgewichtsmuster läßt sich als ein vom Anfangsmuster ausgelöster, „assoziativer Abruf“

des Gleichgewichtsmusters deuten, und diese Interpretation begründet die Bedeutung dieses Ansatzes als Gehirnmodell.

Trotz der (in zweifachem Sinne!) attraktiven Eigenschaften solcher symmetrisch verschalteter Netze bleiben aus der Sicht realer Neuronenetze mehrere Gesichtspunkte unbefriedigend: Zum einen ist die Annahme symmetrischer Verbindungen biologisch immer noch sehr unrealistisch. Zweitens ist die beschriebene, stets auf die Ausbildung eines Gleichgewichtszustandes ausgerichtete Systemdynamik für viele Funktionsaspekte des Gehirns zu begrenzt: beispielsweise ist bereits die Erzeugung eines einfachen rhythmischen Verhaltens, wie es für viele Motorikfähigkeiten unverzichtbar ist, nicht möglich, von komplexerem zeitlichen Verhalten ganz zu schweigen.

Zur Überwindung der genannten Schwierigkeiten wurden in der Vergangenheit zahlreiche Lösungsansätze erprobt. Viele davon lassen die beschriebene Grundstruktur der Attraktornetze im wesentlichen unverändert und versuchen, durch eine geeignete Hinzufügung externer Steuermechanismen die gewünschten zusätzlichen Eigenschaften zu erzeugen. Zwar lassen sich so durchaus Netze mit einem gewünschten, komplexeren Zeitverhalten erzeugen, jedoch trägt ein solcher Ansatz wenig zu unserem Verständnis von Netzen mit weniger regulären, nicht-symmetrischen Verknüpfungen bei.

Diese Situation bildete die Motivation für die Wahl meines Themas am Wissenschaftskolleg: die Untersuchung der Dynamik neuronaler Netze mit Verschaltungsmustern, die Rückkopplung und Nichtsymmetrie zulassen. Hierzu existieren erst vergleichsweise wenige Arbeiten, da ein analytisches Verständnis solcher Netze noch auf erhebliche Schwierigkeiten stößt und viele Eigenschaften bisher nur in Simulationen demonstriert werden können. Im Zentrum meines Interesses standen dabei Netze, deren Verschaltungsstruktur sich statistisch möglichst einfach beschreiben läßt: Eine Verknüpfung von einem Neuron A zu einem Neuron B wird nur mit einer gewissen Wahrscheinlichkeit eingerichtet, die allein vom Abstand der beiden Neuronen abhängt, und die mit zunehmendem Abstand immer kleiner wird. Die Stärke einer auf diese Weise zustandekommenden Verbindung wird dann auf einen Wert gesetzt, der durch zufällige Wahl aus einem vorgegebenen Variationsbereich (beispielsweise das Intervall  $[-s/2, s/2]$  um den Nullpunkt) festgelegt wird.

Diese Regeln sind sehr einfach. Dennoch können die damit erzeugten, statistischen Verbindungsstrukturen zu einem höchst komplexen Verhalten des Netzes führen. Eine wichtige Rolle spielt dabei die Breite  $s$  des Bereichs, in dem die Kopplungsstärken zwischen den Neuronen statistisch variieren.

Für kleine Werte von  $s$  zeigt sich ein Verhalten, das dem der weiter oben beschriebenen Attraktornetze stark ähnelt: in diesem Falle streben auch hier die Neuronenaktivitäten stets einem Gleichgewichtszustand zu, und es lassen sich ähnlich wie bei den weiter oben geschilderten symmetrischen Attraktornetzen vorgegebene Zielmuster „speichern“, indem die Verbindungsstärken zwischen den Neuronen durch eine „Lernregel“ vom Hebbischen Typus geeignet verändert werden. Die Abbildung zeigt das Beispiel eines Netzes aus  $120 \times 120$  Neuronen, in dem auf diese Weise 10 Muster eingelesen wurden. Neun davon sind zufällig gewählt, das zehnte Muster zeigt die Bibliothek des Kollegs. Die Bildfolge illustriert sechs Zeitschritte der Entwicklung der Neuronenaktivierungen des nicht-symmetrischen Netzes, wenn als Anfangszustand (linkes oberes Bild) ein das obere Drittel umfassende Teilfragment der Ansicht der Bibliothek gewählt wird. Wie man gut erkennt, wird das Fragment in wenigen Zeitschritten zu dem richtigen Gesamtmuster „assoziativ“ ergänzt.

Erhöht man die Variationsbreite der Kopplungen zwischen den Neuronen, so ergibt sich folgender Befund: Zunächst wird das überschaubare Attraktorverhalten von einer Abfolge zunehmend komplexerer periodischer Muster abgelöst. Oberhalb einer Schwelle zeigt sich zuerst ein einfaches oszillierendes Verhalten, das in symmetrischen Netzen nicht vorkommen kann. Wenig oberhalb der Schwelle nimmt daran nur eine kleine Teilpopulation aller Neuronen teil. Eine Erhöhung der Variationsbreite führt zum „Erwachen“ weiterer oszillierender Teilpopulationen mit zum Teil anderen Oszillationsfrequenzen. Damit ergibt sich ein zunehmend komplexeres Zeitverhalten, das für noch größere Variationsbreiten der Kopplungsstärken in deterministisches Chaos übergeht.

Kann dieses Auftreten von deterministischem Chaos für die Modellierung von Gehirnfunktionen von Interesse sein? Deterministisches Chaos besitzt die besondere Eigenschaft der „Sensitivität auf kleine Störungen“, die auch für den „Schmetterlings-Effekt“ beim Wetter verantwortlich ist: Der anfänglich sehr kleine Einfluß des Flügelschlags eines Schmetterlings wird immer weiter verstärkt, bis sich schließlich eine makroskopische Auswirkung auf das Wettergeschehen ergibt. Damit stehen Systeme, deren Dynamik deterministisches Chaos zeigt, gewissermaßen am Rande deterministischer Gesetzmäßigkeit: zwar gehorcht ihre Entwicklung genau angebbaren, deterministischen Regeln. Dennoch entziehen sie sich einer effektiven deterministischen Vorhersagbarkeit, indem sie die Welten unbeobachtbarer, mikroskopischer Unwesentlichkeiten und beobachtbarer, makroskopischer Geschehnisse unauflösbar miteinander verbinden. Mit dieser Eigenschaft bieten

chaotische Systeme die Chance, den scheinbaren Widerspruch des Versuchs einer deterministischen Beschreibung für das in vieler Hinsicht nicht besonders determiniert erscheinende System „Gehirn“ aufzulösen.

In der Tat zeigten meine Simulationen auch für das hier betrachtete Modell diese Eigenschaft der Sensitivität auf kleine Störungen. Dies führte zu einer Anschlußfrage: Kann man diese Sensitivität ausnutzen, das Verhalten eines chaotischen neuronalen Systems durch sehr kleine Signale zu steuern? Die Antwort auf diese Frage kann wichtig für unsere Vorstellung darüber sein, wie Informationen in neuronalen Signalen kodiert werden: wenn bereits kleine Signale sensitiv genutzt werden können, so können selbst in einzelnen Spikes viele Informationen „versteckt“ sein. Ferner könnten sich daraus neue Mechanismen für eine Auswertung schwacher sensorischer Signale ergeben.

Meine Arbeiten zu diesem Aspekt chaotischer neuronaler Systeme sind noch nicht abgeschlossen. Ausgehend von bekannten Verfahren der „Chaoskontrolle“ konnte ich mittels eines prädiktiven Steuerungsansatzes ein chaotisches System gekoppelter Neuronen durch sehr kleine Triggersignale durch eine vorgegebene Mustersequenz steuern. Der dabei benutzte Ansatz erfordert jedoch einige Rechenschritte, die mit unbiologisch hoher Präzision ausgeführt werden müssen, um die Zielfolge zu realisieren. Es ist noch ein wichtiger offener Punkt, hierfür robustere Verfahren zu finden, die Aussicht auf eine biologische Implementierbarkeit besitzen.

Die seitens des Wissenschaftskollegs gebotenen „Störungen“ bei diesen Arbeiten waren vielfältiger Natur und machten sich in der Regel in irgendeiner Form stets angenehm bemerkbar. Dies begann mit der hilfreichen Unterstützung der Raumpflege, die selbst nach vorangegangener langer Arbeitsnacht bereits am Morgen eine Fortsetzung meiner in der Nacht doch nicht mehr ganz fertiggewordenen Arbeit gewährleistete, und die mich dabei durch ein blitzblank aufgeräumtes Appartement besonders anspornte. Die Störungen setzten sich fort mit den in höchst erfreulicher Form ausgeübten, lebenserhaltenden Maßnahmen der ausgezeichneten Küche, fanden einen allwöchentlichen Zwischengipfel in den dienstags abgehaltenen Fellow-Kolloquien und endeten (in dieser nicht allumfassenden Aufzählung) bei den eingestreuten Abendveranstaltungen, von denen mir viele als Höhepunkte des Jahres in guter Erinnerung bleiben werden.

Zu einem besonderen Erlebnis wurde das Jahr durch die Gelegenheit, mit Fellows aus den unterschiedlichsten Fachgebieten zusammenzutreffen, einen längeren Zeitraum in derselben Umgebung zu verbringen und über alle möglichen Dinge Gedanken auszutauschen. Gerade

zwischen entfernten Fachgebieten erwies sich dies oft als eine größere Herausforderung als es auf den ersten Blick erscheinen mag. Besonders die Rolle der Biologie in Hinblick auf die Sozial- und Geisteswissenschaften bildete einen häufig auftretenden Diskussionsgegenstand und führte — neben weiteren Einzelthemen — zu einer Reihe regelmäßiger Zusammentreffen eines kleinen Kreises von Fellows, über deren Diskussionsergebnisse an anderer Stelle näher berichtet wird. Ich empfand diese kleinen informellen Arbeitstreffen als eine angenehme Bereicherung, da hier im Vergleich zu den plenarischen Dienstagskolloquien eine wesentlich intensivere Diskussionsmöglichkeit bestand.

Zu Beginn des Fellow-Jahres wurde uns mitgeteilt, daß am Ende des Jahres Verbesserungsvorschläge für die Organisation des Kollegs dankbar entgegengenommen würden. Leider bot das Kolleg für die Fellows hinsichtlich der Bearbeitung dieser Frage nur geringe Denkanstöße. *But nobody is perfect*. Ich wünsche dem Wissenschaftskolleg (auch für den nachrückenden Jahrgang!), daß dieser Kritikpunkt auch künftig bestehen bleiben wird.

