

KIT REPORT 115

Zur Entstehung hierarchischer Kategorienstrukturen
selbstorganisierenden neuronalen Netz

Daniel Durstewitz

Interdisziplinäres Forschungsprojekt "Kognition und Kontext"
Sekt. FR 5-12, Franklinstr. 28/29, D - 10587 Berlin
Technische Universität Berlin

März 1994

Ich danke Prof. Klaus Eyferth und Dr. Stephanie Kelter für die Durchsicht und viele wertvolle Ratschläge.

Diese Arbeit ist im Rahmen des von der TU Berlin geförderten IFP "Kognition und Kontext" entstanden.

Zusammenfassung

Verschiedene Forscher vertreten die Annahme, daß natürliche Kategorie implizit und weitgehend unabhängig von Rückmeldungen aus der Umwelt a Natürliche Kategorien werden nach ökonomischen Prinzipien strukturier Merkmalsähnlichkeiten zwischen den Objekten zunutze machen. Der Vorte Kategoriensystems ist, daß unbeobachtete Objektmerkmale und Eigenscha reliabel vorhergesagt werden können. Ausgehend von psychologischen Be vorliegenden Arbeit ein selbstorganisierendes neuronales Netz vorgest unüberwachten Lernprozeß ein hierarchisches Kategoriensystem nach Pri die auch die Struktur menschlicher Kategoriensysteme bestimmen. Anhan Simulationsergebnissen wird demonstriert, daß die vom Netzwerk aufgekl Hierarchieebenen viele wesentliche Eigenschaften mit dem Basic Level, und Subordinate Level natürlicher Kategoriensysteme gemeinsam haben.

Evolution of hierarchical category structures in a self-organizing ne

Abstract

Various researchers have proposed that natural categories are in the implicitly and without environmental feedback. The development of nat is governed by economical principles which work on the basis of featu the objects. The advantage of such a category system is that unobserv and features of objects never seen before can be predicted reliably. with a self-organizing neural net which is capable of forming hierarc in an unsupervised manner thereby using principles known from human c shown by simulation results the hierarchical levels of the network ca important properties with the basic level, superordinate level and su category systems. The neural net presented here is therefore suggeste model for unsupervised human category learning.

1. Zur Struktur natürlicher Kategoriensysteme

Eine Verdichtung der Umweltinformation in Form kategorialer Strukturen ist Voraussetzung für alle höheren kognitiven Prozesse, insbesondere für (Bischof, 1987; Rosch, 1978; Smith, 1988). Kategoriensysteme haben das Wissen über konkrete Objekte und Situationen hinaus generalisieren und Anwendung auf neue Situationen (Smith, 1988) oder die Vorhersage unbekannter Objekteigenschaften (Anderson, 1991) erlauben. Um diesen Zielen gerecht zu werden, ist es notwendig, daß möglichst viel Information über die ursprüngliche Objekteigenschaften im Kategoriensystem erhalten bleibt und Vorhersagewahrscheinlichkeiten maximiert werden (Anderson, 1991).

Im ersten Teil dieser Arbeit werden zunächst einige empirische Ergebnisse zur Struktur natürlicher Kategoriensysteme besprochen. Es sind vor allem die umfangreichen Forschungsergebnisse von Rosch und Mitarbeitern (Rosch & Mervis, 1975; Gray, Johnson & Boyes-Braem, 1976a; Rosch, Simpson & Miller, 1976b), die zeigen, daß natürliche Kategorien nach bestimmten Prinzipien 'kognitiver Ökonomie' gebildet werden. Im zweiten Teil wird dann ein selbstorganisierendes sog. Competitive Classification Machine (CCM), vorgestellt, die Kategorien externaler Lehrer aufbaut mit genau denselben Eigenschaften, wie sie für natürliche Kategoriensysteme gefunden wurden. Im dritten Teil soll anhand von Simulationen die Mächtigkeit der CCM als psychologisches Erklärungsmodell für die Struktur natürlicher Kategorien demonstriert werden.

In den meisten experimentellen Untersuchungen zum Kategorielernen werden den Probanden vom Forscher definierte künstliche Kategorien in einem sequentiellen Verfahren über Fehlerrückmeldung erlernt. Die Bedeutung von Fehlerrückmeldungen oder vorgegebene Kategorienamen auch für natürliche Kategorien, sicherlich eine Rolle spielen, wird von verschiedenen Forschern bezweifelt. Die Faktoren für den Aufbau natürlicher Kategoriensysteme sind entscheidend sind (Rosch, 1984; Smith, 1981; Anderson, 1991). So wird angenommen, daß Wissen über

Kategorien zunächst implizit und unbewußt erworben wird (Anderson, 1978), daß es der Sprachentwicklung z.T. vorausgeht (Rosch et al., 1978). Strategien bei der Klassifikation von Objekten verwendet werden (Smit Smith, Shoben & Rips, 1974). Fried und Holyoak (1984) konnten Beispiele für Kategorien allein durch Beobachtung lernen können, ohne vorab Beispiele aus verschiedenen Kategorien vorliegen, bzw. sogar ohne dies überhaupt. Darüber hinaus sprechen theoretische Überlegungen dafür, daß Kategorien allein durch Rückmeldung aus der Umwelt erlernt werden können: Für natürliche Kategorien gibt es in der Regel keine definierenden Merkmale im Sinne hinreichend notwendiger Bedingungen für die Kategorienzugehörigkeit (Rosch & Mervis, 1987) und die Anzahl aller möglichen Partitionen der Objektumwelt geht in die Höhe (vgl. Abschnitt 2.1). Unter diesen Bedingungen ist schwer vorstellbar, daß ein Apparat beim Erlernen von Objekt-Kategorie-Zuordnungen ohne ein vorbestimmtes internes Prinzip auskommt. Rückmeldungen können aber z.B. dazu dienen, die Attribute aufmerksam zu machen, und so die Struktur von Kategorien befestigen (Mervis, 1987).

Natürliche Kategorien werden im wesentlichen durch ein Prinzip zusammengehalten, das von Wittgenstein (1953) als Familienähnlichkeit bezeichnet wurde. Die wechselseitigen Überlappungen in den Merkmalsmustern von Objekten zusammen, daß viele natürliche Kategorien nur unscharfe Grenzen besitzen (Smith, 1988; Flanagan, Fried & Holyoak, 1986) und daß es innerhalb natürlicher Kategorien graduelle Unterschiede gibt: Personen können Objekte reliabel hinsichtlich ihrer Zugehörigkeit zu einer Kategorie beurteilen (Rosch, 1975; Rosch & Mervis, 1975; Smith, 1988; vgl. aber Barsalou, 1987). Aufgrund dieser Ratings lassen sich die Objekte einer Kategorie bilden. Verschiedene Autoren nehmen an, daß für erfahrener Objekte einer Kategorie ein Prototyp als eine Art zentrale Repräsentation abstrahiert bzw. so konstruiert wird, daß er maximale Gemeinsamkeit mit den Objekten der Kategorie besitzt (Rosch, 1978; Rosch et al., 1976b; Posner & Smith, 1982). Dafür spricht, daß Prototypen auch dann nicht schlechter als bekannte

werden, wenn sie selbst nicht Teil der Lernmenge waren (Posner & Keel Bransford, 1971). Pbn lernen um so schneller, Objekte richtig zu klas diese für eine Kategorie sind bzw. je größer ihre Ähnlichkeit mit dem al., 1976b). Über die Klassenzugehörigkeit typischer Objekte kann sch werden, und sie werden mit höherer Wahrscheinlichkeit reproduziert al Objekte (Rosch et al., 1976b). Typikalität beeinflusst auch die Stärke Paarvergleichs-Aufgaben (Rosch, 1975).

TypikaÄtt und Familienähnlichkeit lassen sich über das Konzept de formaler fassen (Rosch & Mervis, 1975; Rosch et al., 1976a; Rosch, 19 eines Merkmals ist die bedingte Wahrscheinlichkeit einer Klasse, gege Ein Objekt ist um so typischer für eine Klasse, je mehr Merkmale es m der gleichen Klasse teilt (die Familienähnlichkeit) und je weniger Me hat mit Objekten anderer Klassen (die interkategoriale Ähnlichkeit), Validities der Objektmerkmale sind. Rosch und Mervis (1975) konnten z entsprechend berechnete 'objektive' Typikalitätswerte für natürliche Typikalitätseinschätzungen der Pbn korrelierten. Darüber hinaus fand Experiment mit artifiziellen Kategorien, daß sowohl die Familienähnli entgegengesetzter Weise die interkategoriale Ähnlichkeit zur Typikali beitragen. Smith (1981) konnte nachweisen, daß Kinder verschiedener A Erwachsene in einer freien Sortieraufgabe Kategorien nach einem Famil konstruieren, sofern die Stimuli hinreichend komplex sind.

Von verschiedenen Autoren wird angezweifelt, daß für das Kategor Familienähnlichkeitsprinzip ausschlaggebend ist (Medin & Schwanenflug Wattenmaker & Michalski, 1987). Inwieweit sich die Ergebnisse dieser Erwerb natürlicher Kategorien übertragen lassen, ist jedoch zweifelha Objekt-Kategorie-Zuordnungen vorgegeben wurden, zum anderen, weil die der Kategorien sehr unnatürlich war (vgl. Rosch, 1977, für eine ähnli Untersuchungen). Darüber hinaus wurden in den genannten Studien Stimu verglichen mit der Komplexität natürlicher Objekte eine sehr einfache

perzeptuell besonders saliente Merkmale (Ahn & Medin, 1992) aufwiesen mag eine analytische Strategie durchaus praktikabel sein.

Innerhalb menschlicher Begriffssysteme gibt es eine ausgezeichnete Kategorisierung, den sog. Basic Level. Objekte in Basic Level-Kategorien besitzen wesentlich mehr gemeinsame Merkmale als Objekte in Superordinate Level-Kategorien, andererseits mehr diskriminative, d.h. die Kategorien unterscheidende Merkmale auf dem Subordinate Level (Rosch et al., 1976a) (Beispiele für Superordinate Level- und Subordinate Level-Objekte sind 'Möbel', 'Stuhl', 'Bürostuhl', 'Vogel', 'Amsel'). Während Objekte in Superordinate Level-Kategorien gemeinsame Merkmale besitzen, die jedoch zumeist diskriminativ sind, besitzen Subordinate Level-Kategorien nur wenig mehr gemeinsame Merkmale als die Objekte, von denen die meisten auch auf Objekte kontrastierender Kategorien (Tab. 3). Diese Tendenz zeigt sich in von Pöhlmann generierten Merkmalslisten, in Formähnlichkeiten bildhafter Darstellungen der Objekte und in den Motorischen Aktionen, die in bezug auf die bzw. mit den Objekten durchgeführt werden (Rosch et al., 1976a). Das Verhältnis zwischen Homogenität innerhalb der Kategorien und dem Überschneidungsgrad zwischen den Kategorien ist also auf dem Basic Level am günstigsten.

Rosch (1978; Rosch et al., 1976a) überprüfte die Gültigkeit der Gültigkeit für Objekte auf dem Basic Level am höchsten und diese deshalb besonders einfach kodierbar sind. Diese besonderen Eigenschaften scheinen die Organisation der Wahrnehmungswelt, in der Sprache und in der kognitiven Entwicklung zuzukommen. Kinder bilden erste valide Kategorien auf dem Basic Level (Rosch et al., 1976a). Insbesondere ergab sich in diesen Untersuchungen, daß die Sprachkategorienamen für die unteren Altersstufen signifikant schlechter waren als die Klassifikationsfähigkeiten selbst. Die Befunde legen somit nahe, daß die Kategorienstrukturen entwickeln, bevor das Kind über entsprechende sprachliche Unterscheidungen verfügt. Namen für Basic Level-Kategorien sind in der Sprache obligatorisch als Namen für Kategorien der anderen Ebenen (Rosch et al., 1976a).

am häufigsten benutzt und in der Kindheit als erste erlernt (Rosch et al. 1976a). Die hier vorgestellten Level-Kategorien scheinen es anders als für Superordinate Level-Kategorien noch eine direkte Verbindung zu visuellen Codes zu geben: In Priming- und beschleunigten vorgegebenen Basic Level-Kategorienamen den Paarvergleichen die Erkennung von Objekten ebenso wie Subordinate Level-Kategorienamen (Rosch et al. 1976a). Interessanterweise zeigt sich der Basic Level sowohl dem Superordinate Level als auch dem Subordinate Level signifikant überlegen, wenn die kategoriale Zuordnung möglichst schnell zu beurteilen ist (repliziert von Murphy & Smith, 1982). Da hier zudem Superordinate Level-Kategorien zu schnelleren Reaktionszeiten als Subordinate Level-Kategorienamen führen (Rosch et al. 1976a), daß für die Einordnung von Objekten auf dem Superordinate Level und dem Subordinate Level unterschiedliche Prozesse eine Rolle spielen. In der Regel werden wahrgenommene Objekte zunächst auf dem Basic Level eingeordnet, wenn die Zugehörigkeit zu einer Superordinate Level-Kategorie geschlossen wird. Erst nach Beobachtung zusätzlicher diskriminativer Merkmale wird die passende Subordinate Level-Kategorie bestimmt.

Wir können zusammenfassen, daß natürliche Kategorien intern organisiert sind und durch Familienähnlichkeitsbeziehungen zusammengefaßt werden können ohne externen Lehrer aufgebaut werden und sind vermutlich zunächst unabhängig von der Sprachentwicklung. Auf den verschiedenen hierarchischen Abstraktionsebenen besitzen Kategorien des Basic Level besondere Formen, die sich in verschiedenen kognitiven Phänomenbereichen nachweisen lassen.

2. Die Competitive Classification Machine: Ein selbstorganisierendes Klassifikationsnetz

2.1. Architektur und Lernalgorithmus der Competitive Classification Machine

Im folgenden wird ein Netzwerk-Modell, die Competitive Classification vorgeschrieben, das gemäß den Annahmen über die Entstehung natürlicher Kategorien unüberwachten Aufbau einer hierarchischen Kategorienstruktur beschreibende Lernalgorithmus entspricht im Kern dem Competitive Learning (Grossberg, 1976, 1987a, 1987b; Rumelhart & Zipser, 1985; Kohonen, 1988) mit einigen wichtigen Modifikationen und zusätzlichen Lernfunktionen.

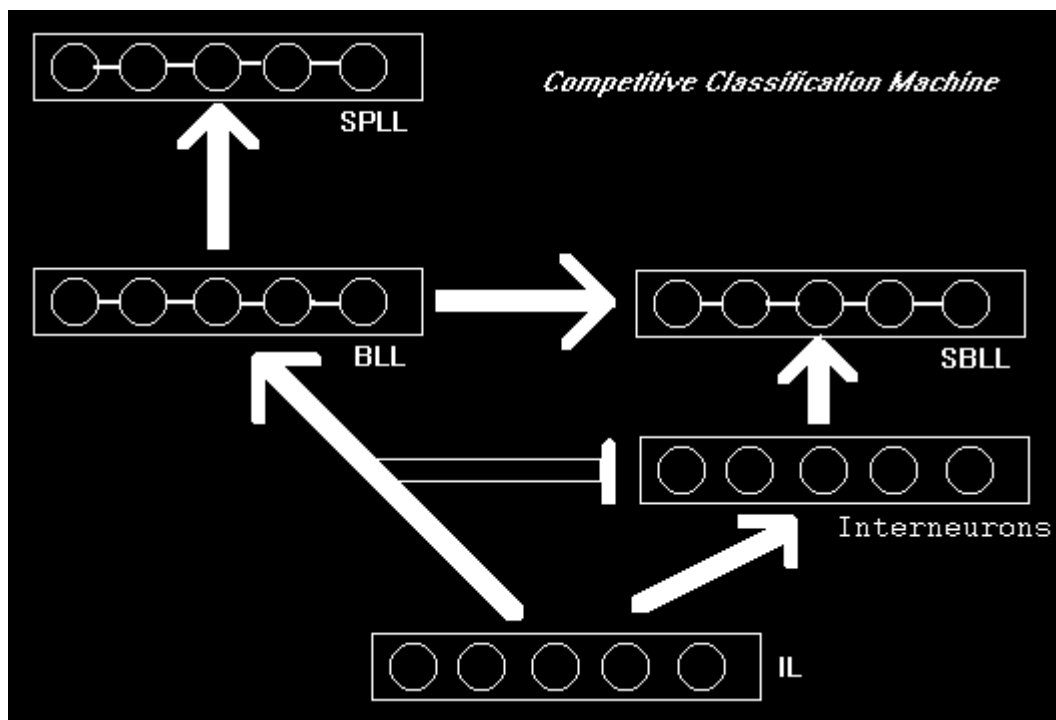


Abb. 1. Architektur der Competitive Classification Machine (SPLL = Subordinate Level Layer, BLL = Basic Level Layer, SPLL = Subordinate Level Layer, IL = Input Layer)

In Abb. 1 ist die Architektur der Competitive Classification Machine dargestellt. CCM besitzt neben dem Input Layer (IL) vier neuronale Schichten, von denen die oberste (SPLL) die endgültige Klassifikation vornimmt. Die Schichten sind hierarchisch angeordnet, wobei die unterste Schicht (IL) die Eingangsdaten verarbeitet und die oberste Schicht (SPLL) die endgültige Klassifikation vornimmt.

Level und dem Superordinate Level der Kategorienhierarchie identifizierten entsprechenden Schichten werden im folgenden mit Basic Level Layer (BLL), Subordinate Level Layer (SLL) und Superordinate Level Layer (SPLL) bezeichnet. Die interneuronale Verbindung ist eher aus Gründen der Veranschaulichung in die Grafik eingezeichnet. Ihre Funktion erfordert nicht unbedingt die Annahme einer zusätzlichen Schicht. Die Striche kennzeichnen die Richtung der neuronalen Verbindungen und damit die Richtung der Aktivationsausbreitung im Netzwerk. Alle Neurone einer vorhergehenden Schicht sind jeweils mit allen Neuronen der nachfolgenden Schicht verbunden, wobei die Verbindungen innerhalb des BLL, des SLL und des SPLL rein inhibitorisch sein können. Die Verbindungen zwischen den Interneuronen ab, die in Abb. 1 durch die Striche innerhalb der Kästen angezeichnet sind, sind ebenfalls rein inhibitorisch. Die Gewichte dieser Verbindungen sind dem Betrag nach äquivalent zu denen der Verbindungen zwischen den Interneuronen ab, die in Abb. 1 durch den offenen Balken symbolisiert sind.

Im folgenden werden die gerichteten Verbindungen zwischen zwei Schichten als w_{ij} bezeichnet und als $w_{ij} = \pm 1$ (für $i, j = 1, 2, \dots, n$) angegeben.

Die Lernumwelt des Netzwerkes besteht aus einer Menge binärer Mustervektoren $\mathbf{x} = (x_1, x_2, \dots, x_n)$, wobei ein Vektorbit codiert, ob ein bestimmtes Merkmal vorhanden ist. Die Mustervektoren werden als Exemplare bezeichnet und stehen für Objekte in der Lernphase. In der Lernphase werden einzelne Mustervektoren nach einer gegebenen Wahrscheinlichkeitsverteilung aus der Menge aller Exemplare ausgewählt und dem IL gebracht. Der IL hat gerade so viele Neurone, wie die Mustervektoren besitzen, und jedes Neuron nimmt den Zustand des korrespondierenden Vektors an. Das Netzwerk kann mit einer beliebigen Anzahl von Neuronen auf dem BLL, SLL und SPLL und einer Ausgangskonfiguration zufällig vorbesetzter Gewichte gestartet werden. In jedem Zeitschritt werden die Neurone der jeweils nachfolgenden Schicht aktiviert. Der Aktivationswert eines Neurons ergibt sich einfach aus dem Skalarprodukt des Eingangsvektors mit dem zugehörigen Gewichtsvektor:

$$(1) a(u_j) := \overline{xw}_j ,$$

wobei $a_j(u)$ der Aktivationswert der Unit j (im folgenden jeweils a_j) ist, \overline{x} der Eingabevektor (d.h. das Aktivitätsmuster der vorangehenden Schicht) und \overline{w}_j der j gehörige Gewichtsvektor ist. Werden Eingabe- und Gewichtsvektor auf normiert, hängt der Aktivationswert eines Neurons wegen

$$(2) \overline{xw}_j = \|\overline{x}\| \|\overline{w}_j\| \cos \varphi$$

zunächst nur vom Winkel zwischen Eingabe- und Gewichtsvektor ab.

Man kann sich nun vorstellen, daß der Aktivationswert weiter in Inhibition von den Verbindungen zwischen den Neuronen innerhalb einer Schicht überträgt wird, und zwar so, daß im wesentlichen nur das Neuron mit dem höchsten Aktivationswert 'überlebt' und somit für einen Lernfortschritt selektiert wird. Diese 'überlebende' Unit' genannt. Es steht für die Kategorie auf der entsprechenden Klasse gegebene Mustervektor auf IL als Objekt zugeordnet wird. Formal ist dies über (3) definiert.

$$(3) u_k \overline{xw}_k \geq \overline{xw}_j \text{ für alle } j.$$

In der Literatur zur statistischen Mustererkennung wird ein Klassifikator aus einer Menge linearer Diskriminanzfunktionen (1) und der entsprechenden Entscheidungsregel (3) resultiert, als lineare Maschine bezeichnet (vgl. Sie teilt die Menge von Mustervektoren in linear trennbare Untermengen).

Ziel des Lernvorganges ist es, iterative Anpassung der Gewichte zu einer Menge von Diskriminanzfunktionen zu finden, die die Exemplarmenge partitionieren. Zum anderen sollten die Gewichtsvektoren die Mitglieder der Klasse optimal repräsentieren, so daß in der internen Repräsentation

viel Information über die einzelnen Exemplare erhalten bleibt und Schunbeobachtete Eigenschaften neuer Exemplare ableitbar sind. Worin ein der Exemplarmenge besteht, ist nicht einfach anzugeben und dürfte sich spezifischen Aufgabenstellung bzw. von den Ansprüchen an das System an & Hart, 1973). Lernziel kann z.B. sein, die Wahrscheinlichkeiten, mit Objektmerkmale korrekt vorhergesagt werden können, zu maximieren (And Definition einer angemessenen Kostenfunktion garantiert noch nicht, d Optimum dieser Funktion gefunden werden kann (vgl. Duda & Hart, 1973) möglichen Partitionen auf ihre Güte prüfen, so steigt der Rechenaufwa Kardinalität der Exemplarmenge (vgl. auch Anderson, 1991, für die Dis Problems). Bei einer Zahl von 81 Exemplaren, wie sie in den später be Simulationen verwendet wird, liegt die Anzahl aller möglichen Aufteil leere disjunkte Klassen bereits etwa in $5 \cdot 10^7$ Größenordnungen Arbeit geht es darum, die Tauglichkeit der CCM als psychologisches Mo demonstrieren. Die hier angesprochenen Probleme werden darum nicht we Ritter und Schulten (1988, zit. in Hertz, Krogh & Palmer, 1991, Kostenfunktion für den Competitive Learning-Algorithmus vorgeschlagen clusteranalytischen Verfahren verwendet wird (vgl. Eckes & Roßbach, 1 1973):

$$(4) E(P_1) = \frac{1}{2} \sum_k \sum_{i \in S_k} \|\bar{x}_i - \bar{w}_k\|^2$$

Für jede Partition der Exemplarmenge werden die quadrierten euklidischen zwischen den Gewichtsvektoren der Klassen und den Mustervektoren der Objekte $i \in S_k$ aufsummiert. Eine Lernregel, die nach einem inkrementellen Gradientenabstiegsverfahren entlang der Kostenfunktion (4) vorgeht, 1

$$(5) \Delta \bar{w}_k = \gamma a_k (\bar{x}_i - \bar{w}_k) \quad \gamma \in]0, 1[\quad ,$$

$$\bar{w}_k = \frac{\bar{w}_k + \Delta \bar{w}_k}{\|\bar{w}_k + \Delta \bar{w}_k\|} .$$

mit $a_k = 1$, wenn die Winner-Unit ist, 0 sonst. In einem Lernschritt wird euklidische Abstand zwischen Eingabevektor und Gewichtsvektor verringert. Gewichtsvektor hinterher wieder normalisiert. In allgemeiner (positiver) Lernrate η für die meisten Simulationen etwa bei 0.05 liegt und der notwendig für Gewichtsvektoren 'konvergiert'.

Der Lernalgorithmus (5) findet lokale Minima der Kostenfunktion (4). Stabilität einer Lösung (Partition) kann allerdings nur für den Fall erreicht werden, wenn der maximale Abstand zwischen zwei Mustervektoren eines Klusters K_l kleiner ist als der minimale Abstand zwischen Mustervektoren eines beliebigen anderen Klusters K_m (vgl. Grossberg, 1987a). Da ein Mittelzentrum μ_l so stabiler ist, je kleiner der zugehörige Wert $E(P_l)$ ist, werden bei Darbietung zufälliger Muster die lokalen Minima der Kostenfunktion (4) mit höherer Wahrscheinlichkeit verlassen als ein globales Minimum.

Die Kostenfunktion (4) hat den Nachteil, dass die Wahrscheinlichkeit, ein globales Minimum über alle Partitionen mit beliebiger Musteranzahl zu erreichen, wenn jedem Kategorienneuron genau ein Exemplar zugeordnet ist, mit zunehmender Kategorienzahl abfällt. Das Competitive Learning-Modell von Rumelhart und Zipser (1985) ist die einfachere Variante, bei der die Kategorienzahl einfach über die Anzahl der Kategorienneurone fest vorgegeben ist. In Grossbergs Adaptiver Resonanz Theorie (Grossberg, 1987a, 1987b) wird die Kategorienzahl über einen Vigilanzparameter reguliert, der festlegt, wie groß die Abstände zwischen Mustervektoren und dem nächstgelegenen Gewichtsvektor mindestens sein müssen, bevor eine neue Kategorie gebildet wird. Diese beiden Vorgehensweisen erfordern a priori ein bestimmtes Wissen über die Struktur der Exemplarmenge.

In der CCM wird eine Beschränkung der Kategorienzahl dadurch erreicht, dass die Aktivationswerte der Non-Winner-Units kleine negative Werte annimmt:

$$(6) a_k = \begin{cases} a_k & \text{wenn } a_k = \sup_j a_j \\ \tau, & \text{sonst} \end{cases}$$

mitte $[0,1]$. Während die Winner-Unit ihren Gewichtsvektor auf den Einzugsbereich zubewegt, lernen die Non-Winner-Units, ihre Gewichtsvektoren ein wenig in die Gegenrichtung zu verschieben. Für Neurone, die nur selten Winner-Unit sind, ist die Wahrscheinlichkeit groß, daß sich ihre Gewichte immer weiter in den n-dimensionalen Eingaberaum verschieben, bis die Neurone schließlich nicht mehr positiv auf Vektoren im Eingaberaum reagieren und somit quasi absterben. Liegen mehrere Neuro-Objekt-Cluster, so wird dasjenige, in dessen Einzugsbereich am meisten andere Neurone schließlich aus dem Cluster 'hinausdrängen'. Die Rate ist abhängig von der Verteilung der Exemplare im Vektorraum, ohne daß spezielle Vorkehrungen nötig sind. Außerdem wird vermieden, daß eigentlich homogene Cluster aufgeteilt werden, wie es z.B. in dem von Rumelhart und Zipser (1985) entwickelten Modell vorkommen kann.

Die Normierung der Gewichtsvektoren wird durchgeführt, damit die Gewichte aller Neurone unabhängig von der Länge ihrer Gewichtsvektoren sind. In diesem Kontext gewinnt jedoch die Normierung noch eine andere Bedeutung. Verschiedene Untersuchungen zeigen, daß die Interaktion von Merkmalen innerhalb eines Cluster eine wesentliche Rolle beim Klassifikationslernen und der Generalisierung spielen kann (Gluck & Bower, 1988; Medin & Edelson, 1988; Estes, Camp & Hurwitz, 1989). Anders betrachtet, werden durch die Normierung die Gewichte, die zur Aufteilung auf die einzelnen Merkmalsdimensionen zur Verfügung stehen, pro Neuron limitiert. Infolgedessen stehen die Merkmale in ihrer Vorhersage einer Kategorie miteinander in Konkurrenz, so daß sich innere kompetitive Merkmalsinteraktionen ergeben. In Verbindung mit anderen Befunden lassen sich so einige Befunde der oben erwähnten Autoren plausibel erklären (Durstewitz, 1993).

2.2. Das Zusammenspiel der CCM-Hierarchieebenen

Im folgenden wird beschrieben, wie die neuronalen Schichten der CCM z insbesondere, wie die feinere Partitionierung der Exemplarmenge auf d entsteht und welche Funktionen die Entwicklung von Klasseninklusionsb Wie in Abschnitt 1 berichtet, konnten Rosch et al. (1976a) zeigen, da Klassifikationsschemata, die sich entwickeln, auf dem Basic Level lie wahrgenommene Objekte zunächst dort eingeordnet werden. Während Basic Subordinate Level einen direkten Zugriff auf Informationen aus dem Wa bzw. auf visuelle Codes zu haben scheinen, ist dies für den Superordi der Fall. Rosch et al. (1976a) nahmen aufgrund ihrer experimentellen der Einordnung eines Objektes auf dem Basic Level auf die Zugehörigke Superordinate Level-Kategorie geschlossen wird und dadurch, daß weite Merkmale entdeckt werden, die Einordnung auf dem Subordinate Level ge Abschnitt 1). Bauplan und Mechanismus der CCM entsprechen diesen Anna BLL und der SPLL besitzen eigenständige Verbindungen zum IL, während BLL aufsetzt und somit Objekte erst nach der Klassifikation auf dem E eingeordnet werden können. Da das Aktivitätsmuster auf dem BLL nach A den Input für die SPLL-Neurone darstellt, ist die Anzahl der Klassen, bilden können, abhängig von der Anzahl aktiver Neurone auf dem BLL.

Die Entwicklung einer Klassenstruktur auf dem SPLL ist ebenfalls Prozessen auf dem BLL, allerdings auf andere Art und Weise. Die Erken diskriminativer Merkmale für die Einteilung auf dem Subordinate Level erreicht, daß Merkmale, die für die Klassenbildung auf dem Basic Leve relevant gewesen sind, für die SPLL-Neurone über eine Exponentialfunk Muster herausgefiltert werden. Die Filterung geschieht über die inhib von den Verbindungen zwischen dem IL und dem BLL zu den Interneuronen Abb. 1). Das binäre Input-Muster wird zunächst identisch auf der Inte abgebildet. Die Winner-Unit j des BLL bestimmt die aktiven Gewichte,

Interneurone gehemmt werden. Je größer ein Gewicht von einem Neuron d Unit des BLL ist, desto stärker wird in exponentiell ansteigender Wei Interneuron gehemmt und die entsprechende Merkmalskomponente somit ge so gefilterte Muster dient als Input für den SBLL.

Ein anderer Aspekt des SBLL betrifft die Zuordnung von SBLL-Neur Neuronen des BLL. Man könnte jedem BLL-Neuron mehrere Neurone des SBL zuordnen. Damit würde jedoch ein Teil der Flexibilität und der adapti Ressourcen auf dem SBLL verlorengehen. Statt dessen werden in der CCM Zuordnungen über die Verbindungen vom BLL zum SBLL gelernt. Weil die Klasseninklusionsbeziehungen mit dem jeweiligen Entwicklungsstand auf abgestimmt werden muß, kann der Lernparameter nicht fest gewählt werden. Veränderung der Gewichte in den beiden Verbindungsstrukturen (IL,SBLL (BLL,SBLL) wird über Parameterfunktionen reguliert, die von Aktivationswerten des SBLL und dem höchsten Aktivationswert a_k auf der

$$(7) \quad \text{für } w \in (IL, SBLL) \quad \tau(a_j) = \gamma \exp\left[-\alpha(a_j - a_k)^2\right] - \eta \quad \alpha, \eta \in \mathbb{R}, \eta < \gamma$$

$$(8) \quad \text{für } w \in (BLL, SBLL) \quad \tau(a_j) = \eta - \exp\left[-\alpha(a_j - a_k)^2\right] \quad \alpha, \eta \in \mathbb{R}, 0 \leq \eta < 1$$

Wie die Parameterfunktionen (7) und (8) zusammenwirken, läßt sich etw verstehen:

Der Gewichtsvektor $(\bar{w}_1^t, \bar{w}_2^t)$ eines SBLL-Neurons ist zusammengesetzt aus z Teilvektoren $\bar{w}_1 \in (IL, SBLL)$ und $\bar{w}_2 \in (BLL, SBLL)$. Wesentlich ist zunächst, daß der gesamte Vektor \bar{w} nach jedem Lerndurchgang auf die Länge 1 normiert wird, beiden Teilvektoren voneinander abhängig sind. Neurone, die auf gegeben mit ähnlich hohen Aktivationswerten reagieren, bewegen \bar{w}_2 auf (8) den T der Winner-Unit zu, den \bar{w}_1 (7) jedoch gleichzeitig von der Winne

¹Da die Gewichte vom IL zu den Interneuronen selbst nicht verändert werden und man auch ein negatives Feedback auf den IL annehmen kann, wird im folgenden die Schicht weggelassen und einfach (IL,SBLL) geschrieben.

Dadurch haben Neurone, die auf ähnliche Eingabemuster reagieren, zum sich an dieselbe BLL-Unit zu koppeln, während die Tendenz für im Vekt liegende Neurone genau gegenläufig ist. Zum anderen haben sie jedoch unterschiedliche Teilmengen von Mustervektoren innerhalb derselben Ok spezialisieren. Da nach der Phase lateraler Inhibition nur ein Neuron wachsen die Gewichte vom BLL zum SBLL schneller als die vom IL zum SB verhindert, daß Mustervektoren verschiedener Basic Level-Kategorien, Merkmalsfilterung ein ähnliches Aussehen haben, in dieselbe Subordina eingeordnet werden. Bei günstiger Wahl der Parameter (8) werden die Basic Level-Kategorien auf dem Subordinate Level reliabel in Teilmenge am Ende 'Inklusionsfehler' auftreten.

3. Entstehung hierarchischer Kategorienstrukturen in der CCM

3.1. Interne Struktur der CCM-Kategorien

Wie Durstewitz (1993) zeigen konnte, entspricht die interne Struktur einem Familienähnlichkeitsprinzip. Von der CCM werden kategoriale Strukturen auch wenn keine definierenden Merkmale für die einzelnen Kategorien gegeben sind. Bezugnehmend auf die Untersuchungen von Rosch und Mervis (1975; vgl. 1976a), kann man die objektive Typikalität eines Exemplars als die Summe der Validities der Objektmerkmale definieren und so einen Zusammenhang zwischen Aktivationswerten herstellen. Die Exemplare einer CCM-Kategorie lassen sich nach den Aktivationswerten, die sie auf dem BLL generieren, in der gleichen Weise nach den Typikalitäts-Ratings menschlicher Proben beschreiben. Wie von Rosch und Mervis (1976a) beschrieben, ist der Aktivationswert eines Kategorie-Neurons um so höher, je größer die Merkmalsübereinstimmung des Mustervektors mit Mustervektoren derselben Kategorie und je geringer die Übereinstimmung mit Mustervektoren anderer Kategorien ist (Durstewitz, 1993). Es ist nicht allz

Ergebnis aus der Aktivationsfunktion (1) und der Lernregel (5), (6) f Gewichtsvektor nach abgeschlossenem Lernvorgang (d.h. Eintrittezustand für alle k) im Mittel die höchste Ähnlichkeit zu allen Mustervektoren aufweist, läßt sich die Familienähnlichkeit eines Mustervektors über Gewichtsvektor, also über den entsprechenden Aktivationswert, messen. besonders nah am Gewichtsvektor einer Kategorie liegen, lassen sich a betreffenden Kategorie auffassen. Typikalität und Nähe zum Prototypen zusammen (Rosch, 1975).

Im CCM-Modell lassen sich einige sehr spezielle empirische Bef plausibel erklären. Medin und Edelson (1988) fanden z.B., daß Pbn unt Bedingungen neue Exemplare so klassifizieren, daß Vorhersagen nach de in systematischer Weise umgekehrt werden. In ihren Experimenten lernt Symptommuster fiktiven Krankheiten zuzuordnen. Im einfachst denkbaren Kategorie nur ein Symptommuster zugeordnet, z.B. einer Kategorie A da 1 0) und einer Kategorie B das Merkmalsmuster (0 1 1). Während das er Merkmal jeweils eindeutige Prädiktoren sind, ist das zweite Merkmal a Kategorie A tritt mit einer a priori-Wahrscheinlichkeit von 0.75 und von 0.25 auf. Medin und Edelson (1988) fanden nun, daß vorher nicht a Merkmalsmuster wie

(0 1 0) entsprechend der höheren a priori-Wahrscheinlichkeit Kategori Merkmalsmuster wie (1 0 1), die eindeutige Prädiktoren beider Kategor jedoch erstaunlicherweise signifikant häufiger der Kategorie mit der Wahrscheinlichkeit (in diesem Fall B) zugeordnet. Läßt man die Pbn hi eindeutigen Merkmalen unter ansonst gleichen Bedingungen lernen, so v Effekt. Weiter konnten Medin und Edelson (1988) zeigen, daß der 'inve um so stärker wird, je mehr ambige Prädiktoren sich in den Merkmalsmu Simuliert man den oben beschriebenen einfachen Fall in einem CCM (entsteht nach kurzer Lernphase die in Tab. 1 dargestellte Gewichtsver

Tab. 1 Gewichte der Kategorienneurone A und B nach einer Simulation des Edelson-Experiments (1988).

Merkmal	Neuron A	Neuron B
1	0,7112	-0,3112
2	0,7029	0,4981
3	-0,0083	0,8093

Wie man sieht, liegt auf dem eindeutigen Prädiktor 3 der geringere Wahrscheinlichkeitswert für Neuron B ein höheres Gewicht als auf dem entsprechenden Prädiktor 1 der Kategorie A. Der Mustervektor (0 0 1) einen höheren Aktivationswert für das Neuron B als der Vektor (1 0 0) für das Neuron A. Genau umgekehrt verhält es sich für den ambigen Prädiktor 2, daß das Muster (0 1 0) gemäß der höheren a priori-Wahrscheinlichkeit zugeordnet wird. Zu diesem Ergebnis kommt es, weil Kategorie A dreimal häufiger als Kategorie B und das Gewicht auf dem zweiten Merkmal für Kategorie B normiert (6) entsprechend häufiger gehemmt wird. Durch die Normierung der Gewichte verschiebt sich nun die Gewichtsverteilung für Neuron B stärker zugunsten des Prädiktors als für Neuron A. Wie man sich leicht überlegen kann, wird dieser Effekt in der Tat um so stärker, je mehr ambige Merkmale sich im Lernmaterial befinden. Der Befund von Medin und Edelson (1988) läßt sich so zumindest teilweise

erklären. Bevor die Eigenschaften der durch die CCM gebildeten Hierarchieeigenschaften werden können, ist es wichtig, auf die Konstruktion der Lernumgebung

3.2. Konstruktion einer Lernumgebung

Rosch u.a. nehmen an (Rosch, 1977, 1978; Smith, 1988; Anderson, 1991) daß die Strukturen eine in der Umwelt (aus der Perspektive des Menschen) vorgegebene adäquat abbilden. Das bedeutet, daß die multivariate Verteilung von Merkmalen in der Umwelt, so wie sie vom Menschen wahrgenommen wird, inhärent bestimmte Eigenschaften besitzen muß, die menschlichen Kategoriensystemen zugrunde

1978). Gegeben eine Menge extrahierter Merkmale, müssen die Exemplare hinsichtlich ihrer Merkmalsähnlichkeiten hierarchisch in Mengen unter Aufgabe des kognitiven Apparates ist, die optimale Partition aufzufin Lernumgebung für das Netzwerk zu konstruieren, sind daher einige sinn über die Merkmalsverteilungen notwendig. Eine plausible Annahme in be Struktur natürlicher Kategorien ist, daß Exemplare um den Centroid ei normalverteilt sind (Flannagan, Fried & Holyoak, 1986).

Die Lernumgebung die CCM wurde über drei ineinandergeschachtelte Binomialverteilungen erzeugt. Die Lernmenge umfaßte insgesamt 81 Must binären Merkmalskomponenten, bei denen der Erwartungswert für die Anz Komponenten ungefähr 8 betrug (also 1/3 aller Komponenten). Die Menge Exemplaren läßt sich genau dreimal auf nicht triviale Weise in Unterm verschiedenen Verteilungen zerlegen, so daß der Erwartungswert für di übereinstimmender Merkmale für Vektoren derselben Klasse jeweils in a ist und mit jeder feineren Partition größer wird. Die gröbste Partiti zerlegt die Exemplarmenge in drei Kategorien mit jeweils 27 Exemplare teilt jede der drei Kategorien wieder dreimal in Klassen zu je neun E feinste Partition teilt wiederum \times je drei Klassen auf. Daraus ergibt s in Tab. 2 angegebene Ähnlichkeitsstruktur der erzeugten Lernumwelt. F Partitionsstufe, die sich aus dem dreistufigen Erzeugungsprozeß ergibt erwartende Merkmalsähnlichkeit für Mustervektoren derselben Kategorie

Tab. 2 Ähnlichkeitsstruktur der künstlichen Lernumwelt (vgl. Erläuteru

Mustervektoren	Erwartungswert für Merkmalsähnlichkeit
----------------	--

verschiedener Kategorien 1.	2.46
derselben Kategorie 1. Stufe	3.49
derselben Kategorie 2. Stufe	6.23
derselben Kategorie 3. Stufe	7.47

Man beachte, daß die Lernumwelt stochastisch generiert wurde und daß Partitionen, die auf jeder Stufe die Exemplare entsprechend der zugru Verteilungen aufteilen, nicht notwendigerweise auch optimal sein müssen.

Tab. 3 zeigt zum Vergleich eine von Rosch et al. (1976a) empirisch Ähnlichkeitstruktur, die die durchschnittliche Anzahl gemeinsamer Merkmale derselben Superordinate Level-, Basic Level- und Subordinate Level-Kategorie zeigt, daß die Relation der Abstände zwischen den zwei jeweils aufeinanderfolgenden Hierarchieebenen mit der Relation der Abstände zwischen erster und zweiter und dritter Stufe in der künstlichen Lernumwelt etwa übereinstimmt. In der zweiten Stufe in der künstlichen Lernumwelt sollte darum eine in natürlichen Kategoriensystemen etwa entsprechen.

Tab. 3 Durchschnittliche Anzahl gemeinsamer Merkmale aller Objekte in Superordinate, Basic und Subordinate Level (nach Rosch et al. 1976a,

Kategorie	Superordinate	Basic Level	Subordinate Level
Musikinstrumente	1	6.0	8.5
Früchte	7	12.3	14.7
Werkzeuge	3	8.3	9.7
Kleidung	3	10.0	12.0
Möbel	3	9.0	10.3
Fahrzeuge	4	8.7	11.2

3.3. Simulationsergebnisse

In jeder Epoche wurde einer der 81 Merkmalsvektoren zufällig und ohne Auswahl ausgewählt, so daß jeder Vektor nach jeweils 81 Lerndurchgängen der C präsentiert wurde. In Tab. 4 ist die Parameterkonfiguration angegeben (zur Bedeutung der Parameter vgl. Abschnitte 2.1 und 2.2).

Tab. 4 Parameterfestlegungen für die im Text beschriebene CCM-Simulation (Abschnitte 2.1 und 2.2).

Funktion	Parameter
(5)	$\gamma = 0.05$
(6)	$\tau = 0.038$
(7)	$\eta = 0.0, = 7.$
(8)	$\eta = 0.0, = 7.$

Die Simulation lief mit 20 Neuronen auf dem BLL, 35 auf dem SBL und 50 auf dem SPL. Die Anzahl der Lerndurchgänge (Epochen) betrug 3000.

Auf dem Superordinate Level sind insgesamt 3, auf dem Basic Level 28 und auf dem Subordinate Level 28 Kategorien entstanden, wobei alle Basic Level-Kategorien 3 Objekte enthalten und die meisten Subordinate Level-Kategorien 3 Objekte. Dies entspricht damit die vom Netzwerk gefundene Struktur in etwa der aufgegebenen.

Im Lernverlauf zeigt sich, daß das erste auf dem BLL, nämlich nach ca. 1000 Epochen, eine stabile Klassenstruktur ausgebildet hat. Da, wie in Abschnitt 2.1, die Lernvorgänge auf dem SPL und SBL immer abhängig von Veränderungen auf dem BLL sind, ist nicht verwunderlich, daß auf diesen Ebenen die Entwicklung des Kategoriensystems erst später abgeschlossen ist (nach 1600 bzw. 2000 Epochen). Sollte man nach den entwicklungspsychologischen Befunden von Rosch et al. diesen Verlauf erwarten. Nach 2200 Epochen war auch die Bedingung der Klasseninklusion für alle SBL-Kategorien erfüllt.

Bevor einige Simulationsergebnisse genauer betrachtet werden können, ist es notwendig, einige Begriffe zu präzisieren:

Definition Kontrastierende Kategorien

Kontrastierende Kategorien sind alle Kategorien, deren Objekte derselben Menge S_k angehören, d.h. alle Mengen S_i , die Teilmenge derselben Obermenge sind.

Definition Objektvalidität

Sei f_i die relative Häufigkeit eines Merkmals i innerhalb einer Kategorie S_k und f_i' die relative Häufigkeit des Merkmals i über alle kontrastierenden Kategorien. Die Objektvalidität eines Objektes x_j für eine Kategorie S_k ist wie folgt definiert:

$$O(S_k | x_j) := h(S_k) \sum_i m_i (f_i / f_i') \quad ,$$

wobei $m_i = 1$, wenn Objekt x_j Merkmal i besitzt, und $h(S_k)$ die relative Häufigkeit der Kategorie S_k ist. Für die CCM-Simulationen ist diese definiert als die Anzahl von Objekten in S_k dividiert durch die Anzahl von Objekten in allen kontrastierenden Kategorien (vgl. Anderson, 1991). Die Objektvalidität bildet etwa die Sicherheit, mit der eine Klasse zugeordnet werden kann.

Definition Familienähnlichkeit

Die Familienähnlichkeit eines Objektes x_j in der Kategorie S_k ist die gemittelte Merkmalsähnlichkeit mit allen anderen Objekten der Klasse S_k :

$$F(x_j) := \sum_i m_i f_i \quad \text{mit } m_i \in \{0, 1\}.$$

Definition Interkategoriale Ähnlichkeit

Die interkategoriale Ähnlichkeit eines Objektes x_j in der Kategorie S_k ist die gemittelte Merkmalsähnlichkeit mit allen Objekten der kontrastierenden Kategorie $S_{k'}$. Die interkategoriale Ähnlichkeit ist dann definiert als

$$I(x_j) := \sum_i m_i f_i'' \quad \text{mit } m_i \in \{0, 1\}.$$

Definition 3 entspricht dem von Rosch und Mervis (1975) zur Vorhersage von Typikalitätseinschätzungen verwendeten Maß für Familienähnlichkeit, n

daß Rosch und Mervis (1975) statt der relativen absolute Häufigkeiten Barsalou (1985) definiert die Familienähnlichkeit eines Exemplars als Ähnlichkeit zu allen Exemplaren derselben Kategorie und seine mittleren Exemplaren kontrastierender Kategorien. Hier geht die interkategoriale Definition von Familienähnlichkeit mit ein. Seine Definition entspricht aber ungefähr der Definition 2 der Objektvalidität. Um das zu sehen, sieht man sich die Summe in Definition 2 auch als Skalarprodukt des Exemplarvektors mit dem Mittelwertsvektor der Kategorie an. Die Komponenten m_i und eines Vektors (f_i/f_i) , den Cue Validities nach Definition von Rosch und Mervis (1975), auffassen läßt. Nehmen wir die Komponenten des Vektors die relativen Häufigkeiten f_i nach Definition 2 ist der Mittelwertsvektor der Kategorie. Man weiter, daß das Skalarprodukt zweier binärer Mustervektoren gerade die Anzahl übereinstimmender Merkmale ist. Das Produkt die mittlere Ähnlichkeit des Exemplars mit allen Exemplaren der Kategorie. Da die Komponenten des Vektors den Quotienten f_i/f_i bilden werden, geht in den Ausdruck die Unähnlichkeit zu Exemplaren kontrastierender Kategorien mit ein.

Die definierten Maße scheinen adäquat, Eigenschaften des CCM-Kategorien zu bestimmen. Zunächst wurden getrennt nach den drei Ebenen Mittelwert Familienähnlichkeiten und interkategorialen Ähnlichkeiten aller Objekte. Die Ergebnisse sind in Abb. 2 und Tab. 5 dargestellt. Wie von Rosch et al. beschrieben, sind sowohl die durchschnittliche Familienähnlichkeit innerhalb der CCM-Kategorien um so geringer, je höher die Ebene in der Hierarchie liegt. Objekte einer Superordinat haben also weniger gemeinsame Merkmale als Objekte einer Basic Level- auch weniger Merkmale gemeinsam mit Objekten kontrastierender Kategorien.

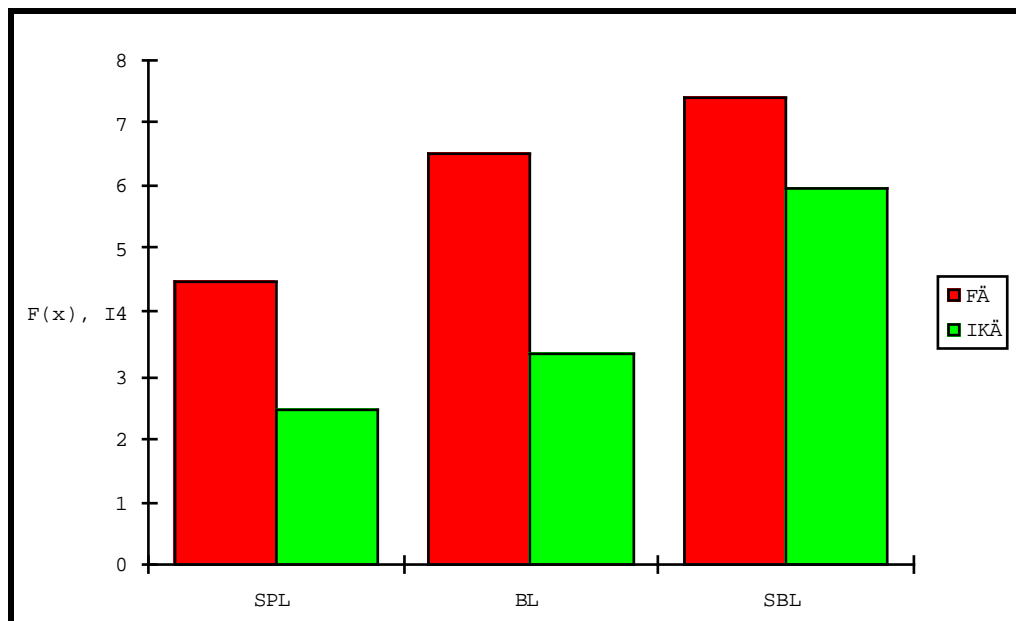


Abb. 2. Durchschnittliche Familienähnlichkeiten (FÄ) und interkategor (IKÄ) von Objekten der drei Klassifikationsebenen (SPL = Superordinat Level, SBL = Subordinate Level).

Im Vergleich der drei Ebenen weisen Objekte aus BLL-Kategorien im Mit Verhältnis von Familienähnlichkeit zu interkategorialer Ähnlichkeit a entnehmen kann. Das Kategoriensystem des BLL besitzt also information eine besonders günstige Struktur, da die Kategorien intern eine hohe und sich gleichzeitig gut differenzieren lassen von den kontrastieren genau die Eigenschaften, die Rosch und ihre Mitarbeiter für natürlich festgestellt haben (Rosch et al., 1976a; Rosch, 1978).

Tab. 5 Mittelwerte, gekennzeichnet durch ein Dach (^), über verschiedene Verhaltensmaße in Abhängigkeit der drei Klassifikationsebenen. In Klammern Standardabweichungen angegeben.

Ebene/ Parameter	Superordinate	Basic Level	Subordinate Le
$\hat{F}(x)$	4.45 (0.96)	6.50 (1.33)	7.38 (1.67)
$\hat{I}(x)$	2.46 (0.70)	3.38 (0.71)	5.94 (1.36)
$\hat{O}(S x)$	3.95 (0.80)	4.60 (1.63)	4.06 (1.61)
\hat{a}	0.61 (0.01)	0.83 (0.08)	0.73 (0.13)

In Tab. 5 sind weitere Kennwerte eingetragen, die den guten Zusammenhänge zwischen den Befunden bestätigen. Dies sind zum einen die gemittelten Objektvaliditäten (Def. 2), zum anderen ein Verhaltensmaß des Netzwerkes, nämlich gemittelter Aktivationswert

Um das Aktivationsverhalten der drei Schichten, die sehr unterschiedliche Eingabemuster erhalten, miteinander vergleichen zu können, werden die Aktivationsvektoren für ein Objekt über alle Neurone kontrastierender Kategorien erzeugt, zu dem Aktivationswert der Winner-Unit durch die Euklidische Norm des Aktivationsvektors geteilt, der sich aus den Aktivationswerten der kontrastierenden Kategorien zusammensetzt. Daraus ergibt sich, gemittelter Aktivationswert $\hat{a}(x)$, das indiziert, wie gut die Kategorienzugehörigkeit der Exemplare erkannt wird bzw., der Interpretation aus Abschnitt 3.1 folgend, wie gut die Exemplare für die Kategorien einer Ebene sind. Wie sich Tab. 5 entnehmen läßt, ist dies in dieser Hinsicht am besten ab. Dies zeigt sich auch in anderen Maßen, die sich aus den Aktivationswerten gewinnen lassen (Durstewitz, 1993). Aus dem Zusammenhang zwischen Aktivationswerten und Familienähnlichkeiten und interkategorialen Ähnlichkeiten (vgl. 2.1) hätte man dieses Ergebnis

In allen Maßen zeigt sich also eindeutig die Überlegenheit des B Kategoriensysteme der anderen bei den Ergebnissen sind robust gegenüber Variationen der Lernumwelt, d.h., verschiedene nach dem in 3.2. angekonstruierte Lernumwelten liefern dieselben Resultate, und zwar über Simulationen hinweg.

Zusammenfassend kann man sagen, daß dem Netzwerk gebildete hierarchische Kategorienstruktur und das Aktivationsverhalten des Netzwerkes in gut mit empirischen Belegen stehen. Die von der CCM auf den drei Hierarchiepartitionen der Exemplarmenge weisen informationstechnische Eigenschaften Rosch et al. (1976a) für natürliche Kategorien beschrieben wurden. Diese bestehen in dem unterschiedlichen Verhältnis von Familienähnlichkeiten interkategorialen Ähnlichkeiten der Exemplare in bezug auf das entsprechende Kategoriensystem. Der Basic Level spielt in dieser Hinsicht sowohl in den Kategorienstrukturen als auch in der von der CCM erzeugten Struktur eine Rolle. Wie beim Menschen (Rosch et al., 1976a; Rosch, 1978, 1977; vgl. sich die Besonderheiten des Basic Level im Verhalten der CCM ab.

5. Allgemeine Diskussion

Empirische Untersuchungen von Rosch u.a. (Rosch, 1975; Rosch & Mervis et al., 1976a; Rosch et al., 1976b) haben gezeigt, daß natürliche Kategorien auf informationalen Kriterien aufgebaut und strukturiert werden, die verschiedene Zielsetzungen genügen. Um bekannte und neue Objekte und Ereignisse schnell klassifizieren zu können, werden im Entwicklungsverlauf Kategorien zu verschiedenen Ebenen organisiert, auf der die Objekte in ihrer Klassenzugehörigkeit unterscheidbar sind. Das dabei angewendete Strukturierungsprinzip macht die Familienähnlichkeiten und interkategorialen Ähnlichkeiten der Objekte

²Wertet man die Daten in einer Meßwiederholungsanalyse mit Layer als dreigestuftem werden die Unterschiede für alle Variablen hoch signifikant ($p < .001$). Da jedoch die Familienähnlichkeiten voneinander abhängig sind und auch die Aktivationswerte Familien reflektieren, ist fraglich, ob ein solches Verfahren überhaupt zulässig ist.

reflektiert in der prototypen-zentrierten internen Struktur der Kategorie gleichzeitig möglichst kohärente Kategorien geschaffen werden, können Objekteigenschaften mit hoher Trefferquote vorhergesagt werden. Erst über- und untergeordnete Ebenen, so daß je nach Kontextbedingungen grobe Einteilungen möglich werden. Der Basic Level behält seine dominante Funktion in der Organisation der Umwelt in der Wahrnehmung und Sprache aber auch bei, wenn bereits vollständige Kategorienhierarchien ausgebildet haben (Rosch et al., 1989).

Die im ersten Teil dieser Arbeit geschilderten psychologischen Befunde haben die Konstruktion eines selbstorganisierenden neuronalen Modells, der Competitive Classification Machine, in dem der Basic Level eine zentrale Rolle hat in der CCM als einziger einen direkten bzw. unbeeinflussten Kontext Merkmalsebene und lernt anhand der Objektähnlichkeiten ein Kategoriensystem, das von seiner Struktur her in wesentlichen Aspekten mit dem menschlichen übereinstimmt. Auf dem SPLL und dem SPLL entstehen Kategoriensysteme, die vom BLL-System abgeleitet bzw. in ihrer Entwicklung von diesem abhängig sind, scheinen gut mit psychologischen Befunden und psychologisch plausiblen Ergebnissen übereinzustimmen. Insbesondere zeigt sich das Kategoriensystem des Basic Level als überlegen, was sich auch im Aktivierungsverhalten der CCM widerspiegelt.

Die Idee, den unüberwachten Aufbau von Kategorienstrukturen mit selbstorganisierenden neuronalen Netzen zu modellieren, ist nicht neu. Kohonen (1982, 1989) und Schyns (1991) haben Modelle entwickelt, die als Grundlage des Kategorisierungsprozesses ebenfalls Ähnlichkeiten (bzw. Distanzen) zwischen Objekten nutzen. Der Lernalgorithmus ist eine Erweiterung des Competitive Learning (Kohonen, 1982, 1989) und organisiert die Kategorienneurone zusätzlich. Die Leistungen dieser Netze wurden allerdings bislang nicht systematisch mit psychologischen Befunden in Beziehung gesetzt und es ist nicht klar, inwieweit diese hierarchischen Strukturen psychologisch adäquat beschreiben können (vgl. Kohonen, 1989).

Wie die in Abschnitt 3.1 berichtete Untersuchung zeigt, lassen sich CCM-Modells über die beiden Prinzipien der intra- und interkategorial einige sehr spezielle Befunde zu Merkmalsinteraktionen erklären. Ich Möglichkeiten aufzeigen, wie sich auch andere wichtige empirische Phänomene menschlichen Klassifikationsverhaltens mit Hilfe der CCM modellieren. Vorhersage von Objekteigenschaften beispielsweise läßt sich über den Übergang von den Kategorienneuronen zu den Merkmalsneuronen modellieren. Die Kontextabhängigkeit von Klassifikationen und Typikalitätseinschätzung (Smith, Shoben & 1987) läßt sich eventuell darüber modellieren, daß alle Neurone einen Bias erhalten. Kontexteinflüsse könnten direkt über eine Veränderung oder indirekt über zusätzliche Aktivationsquellen wirken, indem sie das Auftreten bestimmter Merkmale oder Kategorien verändern. Das Phänomen, daß Ähnlichkeitsurteile oft nicht symmetrisch sind (vgl. Smith, Shoben & 1977), könnte möglicherweise darüber erklärt werden, daß in Abhängigkeit von der Darbietungsreihenfolge eines zu beurteilenden Paares unterschiedliche Kategorienneurone entstehen. Das wäre zunächst deshalb so, weil je nach Reihenfolge ein einzigartiges Aktivationsmuster auf der Kategorienebene erzeugt. In der anschließenden Phase inhibitorischer Aktivität könnte dies zu unterschiedlichen Urteilen führen. Schließlich ist bekannt, daß die Häufigkeit, mit der ein Exemplar einer spezifischen Kategorie auftritt, seine Typikalität und die interne Struktur beeinflussen (Barsalou, 1985; Nosofsky, 1988). Derartige Häufigkeitseffekte lassen sich auch in der CCM einstellen, da häufigere Exemplare den Gewichtsvektor für diese Kategorie zentraler machen und die Tendenz einer Kategorie stärker in die Richtung des entsprechenden Exemplars ziehen.

Damit sind einige Möglichkeiten angedeutet, wie sich das CCM-Modell anwenden ließe. Andere Fragen bleiben jedoch zunächst ungeklärt. So wurde die Theorie als gegeben vorausgesetzt und nicht erklärt, um was für Merkmale es sich handelt, wie komplex diese sind, welche Zusammenhänge unter ihnen bestehen und auf

kognitiven Prozesse und Kriterien sie aus der Umwelt extrahiert wurde als wären einige dieser Fragen auch psychologisch noch weitgehend un-

Als nächster Schritt ist ein Experiment geplant, in dem genauer über ob Pbn eine Objektwelt, die nach dem gleichen Algorithmus wie die künstliche der CCM konstruiert wurde (vgl. Abschnitt 3.2), auf ähnliche Weise parallel hierarchisch abbilden.

Literatur

Ahn, W.-K. & Medin, D.L. (1992). A Two-Stage Model of Category Construction. *Science*, 16, 81-121.

Anderson, J.R. (1991). The Adaptive Nature of Human Categorization. *Psychological Review*, Vol. 98, 3, 409-429.

Barsalou, L.W. (1985). Ideals, Central Tendency, and Frequency of Instance Determinants of Graded Structure in Categories. *Journal of Experimental Learning, Memory and Cognition*, Vol. 11, 4, 629-654.

Barsalou, L.W. (1987). The instability of graded structure: implications for concepts. In: U. Neisser (Ed.), *Concepts and conceptual development* (Cambridge, London: Cambridge University Press).

Bischof, N. (1987). Zur Stammesgeschichte menschlicher Kognition. *Schweizerische Zeitschrift für Psychologie*, 46, 77-90.

Brooks, L.R. (1978). Nonanalytic concept formation and memory for instances. In: B.B. Lloyd (Eds.), *Cognition and Categorization* (pp.169-211). New York: Wiley & Sons.

Duda, R.O. & Hart, P.E. (1973). *Pattern Classification and Scene Analysis*. London: John Wiley & Sons.

Durstewitz, D. (1993). Selbstorganisation von Kategorienstrukturen in neuronalen Klassifikationsnetz. Forschungsbericht aus dem Institut für Technische Universität Berlin.

Eckes, T. & Roßbach, H. (1980). Clusteranalysen. Stuttgart, Berlin: V

Estes, W.K., Campbell, J.A., Hatsopoulos, N. & Hurwitz, J.B. (1989). Category Learning: A Comparison of Parallel Network and Memory Storage Models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 556-571.

Flanagan, M.J., Fried, L.S. & Holyoak, K.J. (1986). Distributional Induction of Category Structure. *Journal of Experimental Psychology: Cognition*, Vol.12, 2, 241-256.

Franks, J.J. & Bransford, J.D. (1971). Abstraction of Visual Patterns. *Psychology*, Vol. 90, 1, 65-74.

Fried, L.S. & Holyoak, K.J. (1984). Induction of Category Distributional Classification Learning. *Journal of Experimental Psychology: Learning and Cognition*, Vol.10, 2, 234-257.

Gluck, M.A. & Bower, G.H. (1988). From Conditioning to Category Learning: A Network Model. *Journal of Experimental Psychology: General*, 117, 3, 234-257.

Grossberg, S. (1976). Adaptive Pattern Classification and Universal Representation Development and Coding of Neural Feature Detectors. *Biological Cybernetics*, 15, 167-202.

Grossberg, S. (1987a). Competitive Learning: From Interactive Activation to Resonance. *Cognitive Science*, 11, 23-63.

Grossberg, S. (1987b). *The Adaptive Brain I: Cognition, Learning, and Rhythm*. Amsterdam: North Holland.

Hertz, J., Krogh, A. & Palmer, R.G. (1991). *Introduction to the Theory of Neural Computation*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43, 59-69.

Kohonen, T. (1989). *Self-Organization and Associative Memory*, 3rd Ed. New York: Springer-Verlag.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago, London: The University of Chicago Press.

- Medin, D.L. & Edelson S.M. (1988). Problem Structure and the Use of B from Experience. *Journal of Experimental Psychology: General*, Vol.117
- Medin, D.L. & Schwanenflugel, P.J. (1981). Linear Separability in Cla *Journal of Experimental Psychology: Human Learning and Memory*, Vol. 7
- Medin, D.L., Wattenmaker, W.D. & Michalski, R.S. (1987). Constraints Inductive Learning: An Experimental Study of Human and Machine Perform Science, 1, 299-339.
- Mervis, C.B. (1987). Child-basic object categories and early lexical Neisser (Ed.), *Concepts and conceptual development* (pp. 101-140). Cambridge University Press.
- Murphy, G.L. & Smith, E.E. (1982). Basic-Level Superiority in Picture *Journal of Verbal Learning and Verbal Behavior*, 21, 1-20.
- Nosofsky, R.M. (1988). Similarity, Frequency, and Category Representa *Experimental Psychology: Learning, Memory and Cognition*, Vol. 14, 1,
- Posner, M.I. & Keele, S.W. (1968). On the Genesis of Abstract Ideas. *Psychology*, Vol.77, 3, 353-363.
- Ritter, H. & Kohonen, T. (1989). Self-Organizing Semantic Maps. *Biolo* 241-254.
- Rosch, E. (1975). Cognitive Representations of Semantic Categories. *J Psychology: General*, Vol.104, 3, 192-233.
- Rosch, E. (1977). Human Categorization. In: N. Warren (Ed.), *Advances Psychology*, Vol.1 (pp. 1-49). London: Academic Press.
- Rosch, E. (1978). Principles of Categorization. In: E. Rosch und B.B. and Categorization (pp.27-48). New York, Toronto: John Wiley & Sons.
- Rosch, E. & Mervis, C.B. (1975). Family Resemblances: Studies in the Categories. *Cognitive Psychology* 7, 573-605.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. & Boyes-Braem, P. Objects in Natural Categories. *Cognitive Psychology*, 8, 382-439.

Rosch, E., Simpson, C. & Miller, R.S. (1976b). Structural Bases of Type of Experimental Psychology: Human Perception and Performance, Vol.2,

Rummelhart, D.E. & Zipser, D. (1985). Feature Discovery by Competitive Cognitive Science, 9, 75-112.

Schyns, P.G. (1991). A Modular Neural Network Model of Concept Acquisition Science, 15, 461-508.

Smith, E.E. (1988). Concepts and Thought. In: R.J. Sternberg & E.E. Smith Psychology of Human Thought (pp.19-49). Cambridge, New York: Cambridge Press.

Smith, E.E., Shoben, E.J. & Rips, L.J. (1974). Structure and Process Psychology Review, Vol. 81, 3, 214-241.

Smith, L.B. (1981). Importance of the Overall Similarity of Objects for Classifications. Journal of Experimental Psychology: Human Perception Vol. 7, 4, 811-824.

Tversky, A. (1977). Features of Similarity. Psychological Review, 84,

Wittgenstein, L. (1984). Philosophische Untersuchungen, Werkausgabe, von Anscombe, G.E.M., von Wright, G.H. & Rhees, R., 1953, Blackwell, a.M.: Suhrkamp.