

METHODEN DER KÜNSTLICHEN INTELLIGENZ

Ipke Wachsmuth, Universität Bielefeld

Ergänzender Kurztext zu Teil 1: Kognitive Kategorien (für besonders Interessierte)

Dieser Kurztext befasst sich etwas eingehender mit der Wissensmodellierung, als Teilaufgabe der Wissensrepräsentation. Unter einer Wissensrepräsentation versteht man ein symbolisch dargestelltes Modell eines Wissensbereichs aus Objekten, Fakten und Regeln in operationaler Form für einen Handlungsträger (Agenten) mit symbolverarbeitender Kompetenz. Die Aufgabe der Wissensrepräsentation umfasst zweierlei: (1) den Entwurf von Repräsentationsformalismen, die Beschreibungsmittel für die symbolische Darstellung von Wissen bereitstellen und die automatische Berechnung von Inferenzen über der Menge kodierten Wissens ermöglichen; (2) die Bereichsmodellierung (Modellierung bestimmter Fachgebiete und Weltbereiche: Domänen), bei der insbesondere folgende Fragen zu beantworten sind:

- *Welche* Kategorien sind zur Einordnung von Objekten und Ereignissen eines Wissensbereichs zu wählen?
- *Welche* Eigenschaften und Annahmen sind ihnen zuzuordnen?
- *Wie* werden sie untereinander in Beziehung gesetzt?
- *Welche* Folgerungen sollen im Kontext bestimmter Annahmen möglich sein?

Die Weise, wie ein Mensch die äußere Welt wahrnimmt und "versteht" (d.h. auf interne Modelle abbildet, *intern repräsentiert*), wird wesentlich von Kategorien seines Wissenssystems – den kognitiven Kategorien – bestimmt. Kategorien fassen einfach ausgedrückt Mengen von Objekten unter bestimmten Aspekten zusammen. Sie ermöglichen einerseits die Einordnung wahrgenommener Dinge, Sachverhalte, Situationen (*Klassifikationsaspekt*) und stellen andererseits weitere Information über Welt Dinge und ihre Zusammenhänge untereinander zur Verfügung, welche bei Bedarf ein umfassenderes Verständnis der beobachteten Situation ermöglicht (*Zuschreibungs- bzw. Elaborationsaspekt*).

Bei der Modellierung von Wissen für ein KI-System möchte man ähnliches erreichen; das ist der Grund, warum wir uns hier damit beschäftigen. Will man zum Beispiel ein natürlich-sprachliches System konzipieren, also ein System, das den Bedeutungsgehalt von sprachlich übermittelter Information erschließen kann, so erfordert dies konkrete Annahmen betreffs der Kategorien, über die Sprecher/Hörer einer natürlichen Sprache verfügen. Eine Aufgabe für die Wissensrepräsentation ist es somit, solche Kategorien und ihre typischen Bedeutungsmerkmale zu modellieren. Aufschlüsse über kognitive Kategorien erbringen empirische Untersuchungen, z.B. Kategorisierungsexperimente, auf die unten eingegangen wird.

Da Kategorien in der Regel sprachlich benannt sind (das beginnt schon bei "gut" und "böse"), lassen sich die Ebenen von Sprache und Wissen nicht immer scharf trennen. Dies wird in der unterschiedlichen Verwendung des Terminus *Kategorie* deutlich. Als philosophischer Grundbegriff wird er bereits bei Aristoteles in doppeltem Sinn gebraucht: Einerseits drücken Kategorien die Struktur der Sprache aus, mit der wir über die Welt sprechen, andererseits werden sie als allgemeinste Formen des Seins selbst interpretiert (ontologische Auffassung). Der Terminus "Kategorie" wird oft auch im Sinne von "Art" (Spezies; engl.: *kind*) gebraucht.

Normen für Begriffe

Mit Hilfe von Kategorien werden also grundsätzlich Mengen von Objekten organisiert. Häufig spricht man auch von Konzepten oder Begriffen (wobei es zuweilen Streit darüber gibt, ob das verschiedene Dingen sind; im Englischen gibt es für beides das Wort "concept"; lateinisch heißt *conceptus* "Gedanke" im Sinne von "Begriff"). Zunächst sei einmal festgestellt, dass es sogar DIN-Normen für Begriffe gibt – das heißt aber überhaupt nicht, dass damit für die Wissensrepräsentation bereits alles gelöst ist, wie wir sehen werden. Dennoch ist es sinnvoll, diese Normen genauer anzusehen, mit denen sich in scharf abgrenzbaren Bereichen durchaus erfolgreich operieren lässt.

Als Hilfsmittel zum Gebrauch begrifflicher Werkzeuge stellen die DIN-Normen 2330 und 2331 allgemeine Grundsätze zur Benennung und Darstellung von Begriffen und Begriffssystemen bereit, die aus der philosophischen Begriffslehre übernommen sind. Wegen der Schwierigkeit, dass Umfang und Inhalt eines Begriffes oft nicht vollständig angegeben werden können (Bsp. "Mensch"), wird der Vorschlag gemacht, sich auf die an Gegenständen interessierenden Merkmale zu beschränken. Die Definition von Begriff und Begriffssystem erfolgt daher in bezug auf einen festgelegten Kontext in einem mengensprachlichen Modell. Dieses läuft darauf hinaus, dass ein Begriff eines Kontextes durch Nennung aller umfassten Gegenstände (Umfang/Extension) bzw. aller Merkmale (Inhalt/Intension) eindeutig festgelegt ist. Ein *Unterbegriff* eines Begriffes ist dadurch gekennzeichnet, dass er *weniger* Gegenstände bzw. *mehr* Merkmale hat (umgekehrt für *Oberbegriff*). Die Menge aller Begriffe eines Kontextes bildet dann mit der hierarchischen Beziehung "Unterbegriff – Oberbegriff" einen vollständigen Verband ("Begriffsverband des Kontextes"). Das hat den Vorteil, dass Methoden und Ergebnisse der Ordnungs- und Verbandstheorie für die Begriffsanalyse nutzbar gemacht werden können.

Für dieses Vorgehen ist festzustellen, dass alles vom Kontext (der betrachteten "Welt") abhängt, mit anderen Worten *kontextspezifisch* ist: Die Merkmale sind definitiv; weitere evtl. interessante Merkmale können bei einem festgelegten Kontext nicht berücksichtigt werden. Alle formal bildbaren Merkmalsbündel können bei diesem Ansatz einen Begriff bilden, aber nicht alle "machen Sinn" (haben nichtleere Extensionen). Alle Begriffsinstanzen ("Exemplare") haben gleichen Status; jeder Unterbegriff "erbt" Merkmale der Oberbegriffe und hat weitere spezifizierende Merkmale; "abstraktere" (hierarchiehöhere) Begriffe werden durch *Weglassen* von Merkmalen gewonnen. Gerade hier gibt es aber Konflikte mit der "kognitiven Realität"! Sie sollen im Folgenden etwas genauer beleuchtet werden.

Begriffsbildung: Zwei Aspekte des Abstrahierens

Ein wichtiges Prinzip der Ausprägung von Begriffen ist die *Abstraktion*, das heißt das Weglassen bestimmter Inhalte, von denen zugunsten anderer Inhalte abgesehen, 'abstrahiert' wird. Abstraktion ist aber in der Regel mit einer Hervorhebung von interessierenden Merkmalen verbunden, die unter einem bestimmten Gesichtspunkt als 'wesentlich' für eine Menge von Gegenständen betrachtet werden. Bei der begrifflichen Abstraktion geht es also sowohl um (1) die Reduktion von spezifischen Merkmalen wie um (2) das Hervorheben des Wesentlichen. Deshalb kann es sinnvoll sein, nicht einfach Attribute (Merkmale/ features) zur Begriffskennzeichnung zu verwenden, wie in der DIN-Norm und bestimmten Logiken (sog. feature-Logiken), sondern zwischen Attributnamen (Slots) und Attributwerten (Fillers) zu unterscheiden, das heißt etwa folgende Beschreibungen für "Metall" oder "Messer" vorzusehen:

METALL

Farbe: <Farbe>
Beschaffenheit: hart, dicht
Textur: glänzend

MESSER

haspart: Klinge
haspart: Griff
material: Metall
function: schneiden

Genau hingesehen: Dies sind Beschreibungen der Dinge, wie sie "begriffen" sind: Begriffe (engl. *concepts*). Sie sind jeweils in Beziehung zu setzen mit Weltobjekten bestimmter Arten (engl. *kinds*), und sie haben eine sprachliche Benennung durch Artnamen (engl. *kind terms*). Und es fehlt auch etwas, was die Dinge sonst noch haben mögen – davon wurde abstrahiert. Schließlich sind die beiden Beispiele noch dadurch verschieden, dass Metalle etwas Natürliches, vom Menschen in der Welt Vorgefundenes sind, und Messer Artefakte, etwas vom Menschen Geschaffenes. Das ist wichtig, denn Artefakte haben immer eine Funktion, für die sie geschaffen sind (oder Sozialarbeiter eine Funktion, die sie ausüben); deshalb findet man hier einen Attributtyp *function*. (Man spricht im ersten Fall von *natural categories* und im zweiten von *artifactual* oder allgemeiner von *social categories*.)

Kategorisierungsexperimente – Basisbegriffe – Prototypen

Nähere Aufschlüsse über kognitive Kategorien wurden durch empirische Untersuchungen gefunden. Sie zeigen, dass menschliche Kategorienstrukturen differenzierter als Verbandsstrukturen sind und von Begriffsauffassungen der philosophischen Logik abweichen. In sog. Kategorisierungsexperimenten mussten Probanden zum Beispiel Objekte Begriffen zuordnen, Begriffe Objekten zuordnen, gegebene Zuordnungen beurteilen, Beispiele von Objekten eines Begriffs auflisten, Merkmale von Objekten eines Begriffs auflisten und beurteilen, wie typisch vorgelegte Begriffsbeispiele sind. Ein bekanntes Ergebnis ist der Nachweis des 'Typikalitätsphänomens' (z.B. ist ein *Spatz* ein typischerer *Vogel* als ein *Pinguin*): Für solche typischen Beispiele wird schneller über Begriffszugehörigkeit entschieden als für untypische; bestimmte Merkmale sind wichtiger für den Begriff von einem Objekt als andere. Weitere Studien zeigen, dass Vererbungsbeziehungen teils komplexen Bedingungen genügen. So ist für Menschen der Begriff *Brot* in bezug auf die mitgedachten Attribute dem Oberbegriff *Nahrungsmittel* näher als dem Oberbegriff *Backwaren*, obwohl *Brot* in einer Vererbungshierarchie nach logischen Kriterien unter *Backwaren* und nicht unmittelbar unter *Nahrungsmittel* anzuordnen wäre.

Aus diesen Untersuchungen folgt zunächst, dass Beziehungen zwischen Begriffen unterschiedlicher Abstraktheit innerhalb einer Hierarchie *nicht* nur durch Hinzufügen oder Weglassen von Merkmalen bestimmt werden. Ferner scheinen innerhalb von Begriffshierarchien bestimmte Begriffe wichtiger als andere zu sein. Eine besondere Rolle spielen dabei sog. *Basisbegriffe*: Darunter versteht man grob gesprochen solche Begriffe, bei denen die Zahl gemeinsamer Merkmale am stärksten ansteigt, wenn man von einem Oberbegriff (wie *Möbel*) zu einem "mittleren" Begriff (wie *Stuhl*) übergeht, verglichen mit dem Übergang vom mittleren zu einem Unterbegriff (wie *Küchenstuhl*): Das Bild eines Küchenstuhls wird spontan als *Stuhl*, nicht als *Möbel* oder *Küchenstuhl* benannt. Für Basisbegriffe besteht ein günstiges Verhältnis von Detailreichtum zu Sparsamkeit: Sie sind einerseits detailliert genug, um deutlich von anderen Begriffen der gleichen Ebene unterschieden zu werden, andererseits detailarm genug, um ökonomisch zu sein. Basisbegriffe sind also eine Art "kognitiver Referenzpunkte". In einer Wissensrepräsentation sollten zumindest die Basisbegriffe einer Domäne aufgenommen sein.

Weitere Anhaltspunkte zur Begriffsrepräsentation erbringen die sog. *Prototyptheorien*. Sie beschreiben die Hervorhebung der 'wesentlichen' Begriffseigenschaften durch die Auswahl eines 'typischen' Exemplars und durch Bezug auf Basisbegriffe. Nach Ansicht des Kognitionswissenschaftlers George Lakoff sind die empirisch festgestellten Prototyp-Effekte Ausdruck besonderer Eigenschaften kognitiver Modelle, von denen der Mensch eine Vielzahl besitzt. Es gibt also Effekte, die sich *eben nicht* mit Merkmalsbündeln (als Begriffsrepräsentationen) erklären lassen. Eine wichtige Einsicht ist, dass viele Eigenschaften, die für die Charakterisierung kognitiver Kategorien relevant sind, wesentlich auf der Interaktion zwischen Menschen und Dingen beruhen ("interactional properties", auch "affordances"), zum Beispiel dass ein Stuhl etwas ist, auf dem man sitzen kann.

Zwischenfazit

Im Hinblick auf die Aufgaben der Wissensmodellierung soll es hier nicht darum gehen, ob die eine oder andere Repräsentationsform plausibler ist, sondern ob sie menschliche Intelligenzleistungen zutreffend charakterisiert. Neuere Formalismen der Wissensrepräsentation, zum Beispiel KL-ONE und Feature-Logik, benutzen Konzepthierarchien zur Repräsentation kategorialen Wissens. Sie versuchen, einige der beim Menschen festgestellten Prinzipien in einem Kalkül aufzugreifen, dessen Berechnungsverhalten ähnliche Eigenschaften und Effizienz wie die begriffliche Informationsverarbeitung beim Menschen aufweist (z.B. bei der Objektklassifikation, Bedeutungszuschreibung, Begriff/Unterbegriff-Beziehung).

Ideal wäre es, wenn sich *Typen* von Merkmalen für Teilbereiche einer Begriffshierarchie bestimmen ließen, d.h. dass zur Modellierung einzelner Konzepte nicht potentiell alle, sondern nur wohlabgegrenzte Teilmengen von Features in Betracht gezogen werden müssen. Und es gibt auch hier noch Erkenntnisse, die in Modellierungsansätzen ausgenutzt werden könnten.

Merkmalstypen in Kategoriensystemen

In dem 1988 erschienenen Buch "Naive Semantics for Natural Language Understanding" der Wissenschaftlerin Kathleen Dahlgren wird eine Modellierung des Common-Sense-Wissens für natürlichsprachliche Systeme vorgelegt (*Common sense* – "gesunder Menschenverstand": Grundbegriff der Erkenntnistheorie). Dieser Ansatz macht plausibel, dass *nicht* alle formal bildbaren Merkmalsbündel einen Begriff bilden können, denn nicht alle "machen Sinn" (haben nichtleere Extensionen). Es ist zum Beispiel im Alltagsleben unsinnig, dass man eine Kategorie für 'Vögel mit Rädern' benötigte.

Auf Basis von empirischen Untersuchungen unterbreitet Dahlgren eine These, nach der sich die Anzahl möglicher Merkmalskombinationen – durch Typisierung und Korrelation – beschränken lässt. Die feature types für ein Objekt sind durch die ontologische Klassifikation des Objekts festgelegt, zum Beispiel werden NATURAL kinds (wie Zitrone, Wasser, Gold, Tiger) primär mit perzeptuellen Merkmalen beschrieben, ARTIFACTUAL kinds (wie Messer, Lastwagen, Haus) haben darüber hinaus funktionale Merkmale, SENTIENT kinds (wie Sekretärin, Bergmann, Pfleger) haben u.a. Charakterzüge (traits). Hiermit wird das Problem der Komplexität kognitiver Kategorien somit systematisch angegangen, wenn der Ansatz auch erst nur eine Richtung andeutet, als dass damit das Modellierungsproblem für Alltagswissen grundsätzlich gelöst wäre.