

METHODEN DER KÜNSTLICHEN INTELLIGENZ

Ipke Wachsmuth, Universität Bielefeld

Kurztext zu Teil 1: Symbolische Repräsentation

Der Gegenstand der Künstlichen Intelligenz ist die Untersuchung von Berechnungsverfahren, die es ermöglichen, *wahrzunehmen*, *schlusszufolgern* und zu *handeln*. Nach diesem von P.H. Winston vorgeschlagenen Standpunkt unterscheidet sich die Künstliche Intelligenz von der Psychologie wegen der grösseren Betonung des Aspekts der Berechnung, und von der Informatik wegen der Betonung von Wahrnehmung, Schlussfolgern und Handeln. Ein Kernpunkt hierbei ist es, dass das schlussfolgernde Denken ("reasoning") von essentieller Bedeutung für höhere Intelligenzfunktionen ist; es beinhaltet kognitive Verarbeitungsprozesse, die es einem Individuum ermöglichen, darüber nachzudenken, was die beste Weise zu handeln ist, bevor tatsächlich gehandelt wird. Dabei ist wesentlich, dass Wissen über die Welt und über *alternative* Möglichkeiten des Handelns in der Welt herangezogen werden kann.

Der Begriff "Künstliche Intelligenz" lässt sich auf das Jahr 1956 zurückverfolgen. Unter den Teilnehmern der in diesem Jahr durchgeführten Dartmouth Conference, die als Gründungsereignis der KI gewertet wird, waren Allen Newell und Herbert Simon, die zusammen mit John Shaw gerade die Arbeiten an ihrem *Logic Theorist* abgeschlossen hatten, einem Programm, das mathematische Sätze in Whitehead und Russell's "Principia Mathematica" beweisen konnte. Dieses Programm verkörperte schon den "Informationsverarbeitungsansatz" – dass Theorien des *willentlichen* menschlichen Verhaltens auf der Basis von informationsverarbeitenden Systemen formuliert werden können, also Systemen, die sich aus Speichern, Prozessoren und Steuerstrukturen zusammensetzen und die auf Datenstrukturen arbeiten. Zentrale Grundposition dieses Ansatzes ist es, dass im Hinblick auf intelligentes Verhalten der Mensch als ein solches System betrachtet werden kann: aktiv, autonom, regelgeleitet und mit beschränkten Ressourcen.

Das wissenschaftliche Ziel der Künstlichen Intelligenz ist, zu bestimmen, welche Annahmen über die Repräsentation und Verarbeitung von Wissen und den Aufbau von Systemen verschiedenste Aspekte der Intelligenz erklären können. Der vorherrschende Gedanke in den Fassungen des Intelligenzbegriffs in der KI ist, dass Intelligenz aus der Interaktion vieler einfacher Prozesse "im Konzert" emergiert und dass Prozessmodelle intelligenten Verhaltens mit Hilfe des Computers detailliert untersucht werden können. Als akademische Disziplin will KI menschliche Intelligenz dadurch verstehen und erklären, dass es gelingt, Effekte der Intelligenz – intelligentes Verhalten – zu produzieren. Zur KI-Methodik gehört also, dass Erkenntnisfortschritte aufgrund lauffähiger Systeme angestrebt werden: Synthese vor Analyse. Drastisch gesprochen ist es *nicht* das Ziel der Künstlichen Intelligenz, intelligente Maschinen zu bauen, *nachdem* ein Verständnis natürlicher Intelligenz erlangt wurde, sondern natürliche Intelligenz *durch* das Bauen intelligenter Maschinen verstehen zu lernen.

Intelligenter Agent, symbolische Repräsentation

Zentrale Aufmerksamkeit gilt in der KI der internen Repräsentationen und Verarbeitung von *Symbolen* als die (begriffliche) Basis interner Prozesse, die man als "Denken" bezeichnet. Symbole werden dabei als bezeichnende Objekte verstanden, die den Zugriff auf Bedeutungen – Benennungen, Beschreibungen, Assoziationen – und damit verbunden eine Betrachtung

höherer mentaler Funktionen wie Problemlösen und Schlussfolgern ermöglichen. Der wichtigste Aspekt einer solchen Betrachtung von geistigen Fähigkeiten des Menschen ist es, dass sie sich auf der symbolischen Ebene – unabhängig von der Betrachtung neuronaler Architekturen und Prozesse – anhand symbolischer Beschreibungen und beschreibungsmanipulierender Prozesse untersuchen lassen. Das bedeutet aber, dass Gegenstand der KI-Forschung nicht das Gehirn und Prozesse des Abrufs von Gedächtnisbesitz ist, sondern vielmehr die Bedeutung, die sich einem Prozess vermöge symbolischer Beschreibungen zuordnen lässt.

Ein zentrales Paradigma der KI ist die 1972 von Newell und Simon formulierte Beschreibung des "intelligenten Agenten". Hiermit nehmen sie einen *funktionalen* Standpunkt über den Gedächtnisbesitz des Individuums ein und seine Fähigkeit, sein Handeln in der Welt darauf aufzubauen; sie bezeichnen diese funktionale Qualität mit *Wissen*. Der intelligente Agent verfügt über Sensoren, zur Wahrnehmung seiner Umgebung, und über Effektoren, mit denen er die äußere Welt beeinflussen kann (eine vereinfachte Darstellung zeigt Abb. 1). Spezifisch für diese Auffassung ist, dass der Agent zu "Probehandlungen" fähig ist: Bevor er in der Welt handelt und sie dadurch möglicherweise irreversibel verändert, manipuliert er eine interne Repräsentation der Außenwelt, um den Effekt alternativer Methoden zu explorieren. Diese Methoden sind ihm in einem internen Methodenspeicher verfügbar, und ihre Exploration wird durch ebenfalls intern verfügbares Wissen geleitet. Mit der Hypothese einer *Wissensebene* hat Newell Wissen als eine abstrakte Kompetenz – als ein Potenzial zu handeln – bezeichnet; es muss aber auf einer *Symbolebene* an eine konkrete symbolische Repräsentation gebunden sein, um einsatzfähig zu sein. Newell und Simon postulieren, dass kognitive Verarbeitungsprozesse Transformationen von Symbolstrukturen beinhalten und dass ein geeignetes "physikalisches" Symbolsystem zur Ausstattung eines jeden intelligenten Agenten gehört.

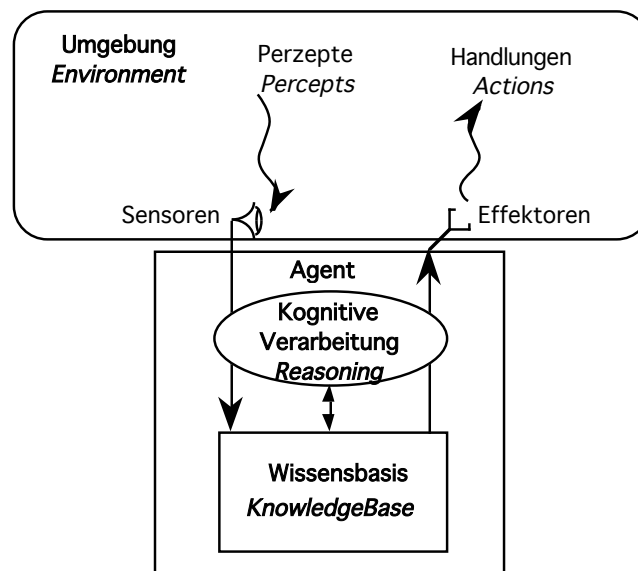


Abb. 1 Modell des intelligenten Agenten (angelehnt an Russell & Norvig 1995, Kap. 2)

Wie lässt sich Wissen symbolisch repräsentieren, was bedeutet ein Symbol?

So wie es bei der Konzeption des intelligenten Agenten offengelassen bleibt, ob man dabei an einen Menschen oder ein maschinelles System denkt, ist es auch bei der Vorstellung eines physikalischen Symbolsystems: Als Informationsträger müssen Symbole in irgendeiner Form physisch, also materiell, realisiert sein. Beim Menschen ist ein Symbol als physikalischer

Erregungszustand im Hirn vorstellbar, wie er auftritt, wenn man zum Beispiel "Hund" denkt. Im Computer wird das Symbol "Hund" durch ein elektronisch erzeugtes Bitmuster materiell realisiert. *Die Bedeutung eines Symbols wird durch die in seinem Kontext möglichen Schlussfolgerungen konstituiert*, beim Symbol "Hund" also etwa "ist ein Tier", "kann bellen" etc. Solches Wissen um die Bedeutung von Symbolen greift also auf andere Symbole (wie "Tier", "bellen") zurück, es kann ebenfalls symbolisch repräsentiert werden. Selbst persönliches Wissen (wie "an meinem ersten Schultag hat mich ein Hund gebissen") kann in Symbolen ausgedrückt werden, ganze Schlussketten wie "da bellt was" => "ein Hund" => "könnte mich beißen" => "ich halte Abstand" verlaufen in Symbolen, und man kann sie mit den Mitteln logischer Formeln und Schlussfolgerungen beschreiben und handhaben.

Einer Kernpunkt dieses Ansatzes ist, dass Logikformalisten ein fundamentales Werkzeug für Analysen auf Wissensebene sind und als Repräsentationsmittel für Wissen dienen können. Der Wissensebenen-Ansatz ist somit gewissermaßen ein Versuch der Mathematisierung bestimmter Intelligenzaspekte (unabhängig von der Realisierung durch physikalische Symbole). Dies betrifft vor allem die Aspekte des rationalen Verhaltens, Problemlösens und Schlussfolgerns. Logikformalisten werden so vielfach in der KI benutzt, um eine explizite Menge von Überzeugungen (*beliefs*) eines Agenten auszudrücken. Solch eine Menge von *beliefs* – gegeben in einer symbolischen Repräsentation – ist das, was man typischerweise als *Wissensbasis* bezeichnet. Diese logikorientierte (Wissensebenen-)Auffassung hat zur Klärung der Debatten, die bis gegen Anfang 1980 um den Begriff der Repräsentation geführt wurden, beigetragen. Unterschiedliche Ansätze der Wissensrepräsentation – wie semantische Netze und Frames – wurden als notationelle Varianten herausgestellt, soweit es die Ausdrucks- und Schlussfähigkeit anbelangt. Die Prominenz, die diese alternativen Notationen in vielen Anwendungsfeldern haben, leitet sich aus ihrer "Objekt-Zentriertheit" ab, die den Vorteil des einfachen Indexierens des mit einem Symbol assoziierten Wissens bietet, und das gesamte Feld der objektorientierten Programmierung ist in Verbindung damit groß geworden.

Interne Symbole müssen eindeutig sein

Eine symbolische Repräsentation ist eine idealisierte Darstellung eines Ausschnitts der Welt in einer einheitlichen (Symbol-)Sprache, in der ein Agent Aussagen über die Welt intern ausdrücken und manipulieren kann. Damit das funktioniert, müssen die intern verwendeten Symbole eindeutig sein. Wenn wir miteinander sprachlich kommunizieren, sind verschiedene Arten der Mehrdeutigkeit möglich. Wenn mir z.B. gesagt wird, morgen kommt Dave, und ich kenne zwei Daves, muss ich für mich klären, welcher von beiden gemeint ist. Im Denken sind also intern zwei individuelle Symbole (etwa dave-1 und dave-2) zu unterscheiden, an die sich verschiedenes Wissen knüpft. So etwas nennt man referentielle Mehrdeutigkeit: Auf welchen der beiden Daves bezieht sich die Aussage? Beim Übergang zu einer internen Symbolrepräsentation muss die referentielle Mehrdeutigkeit aufgelöst werden. Eine referentielle Mehrdeutigkeit liegt auch vor bei einer Mitteilung wie "Der Hund saß auf dem Tisch. Er bellte." Worauf bezieht (referenziert) nämlich das "er"? Die Lösung dieser Frage basiert auf unserem Wissen über die Welt: Zwar kann ein Hund (r1) bellen, nicht aber ein Tisch (r2), also ist r1 der eindeutige Bezug (Referent) für das sprachliche "er".

Es gibt weitere Arten von Mehrdeutigkeiten, nämlich semantische und funktionale. Eine semantische Mehrdeutigkeit finden wir im Satz: "Hans bringt das Geld auf die Bank." Das Wort "Bank" gibt es in zwei Bedeutungen, wie wir wissen, als Finanzinstitution und als Sitzgegenstand, und wieder hilft uns das Weltwissen bei der Auflösung dieser Mehrdeutigkeit (wer würde schon sein Geld auf einem Sitzgegenstand deponieren?). Im Denken unterschei-

den wir deshalb zum Beispiel interne Symbole "Geldbank" und "Sitzbank" und knüpfen unterschiedliche Semantik daran. Eine funktionale Mehrdeutigkeit liegt zum Beispiel für den Satz "Die Keule fängt Petra" vor. Hier gibt es zwei syntaktisch korrekte Lesarten des Satzes: Subjekt = Die Keule, Prädikat = fängt, Objekt = Petra, aber auch Subjekt = Petra, Prädikat = fängt, Objekt = Die Keule, denn in der deutschen Grammatik ist auch die Voranstellung des Objekts zulässig. Weil wir aber wissen, daß ein toter Gegenstand wie eine Keule nichts fangen kann, kann nur die zweite Lesart sinnvoll sein, in der Petra die Funktion des Subjekts und die Keule die Funktion des Objekts einnimmt.

Semantische Netze, Frames, Wissensmodellierung

Gegenüber prädikatenlogischen Repräsentationen bieten andere Repräsentationsformate wie semantische Netze und Frames (= Slot-and-Filler Notation) die Möglichkeit, unmittelbar die funktionale Struktur auszudrücken und Vererbungshierarchien zu beschreiben (beides lässt sich jedoch in prädikatenlogischen Repräsentationen rekonstruieren), und einiges mehr, wie zugeordnete Prozeduren und Erwartungswerte (Defaults). In Vererbungshierarchien erbt ein Frame allgemeine Eigenschaften von seinem Vorgänger in der Hierarchie, so dass nur die individuellen Eigenschaften beim Frame selbst abgespeichert werden müssen. Zur Flexibilitätssteigerung verwendet man auch Vererbungsheterarchien, bei denen ein Frame Eigenschaften von mehreren Vorgängern übernimmt. Es ist möglich, die Vererbung bestimmter Eigenschaften gezielt zu unterdrücken.

Zugeordnete Prozeduren sind kleine Programme, die einer Eigenschaft eines Frames zugeordnet sind und bei einem Lese- oder Schreibzugriff auf dessen Wert ausgeführt werden, z.B. um den Wert zu berechnen oder Konsequenzen aus einer Wertänderung zu propagieren. Auch zugeordnete Prozeduren lassen sich wie Werte vererben. Erwartungswerte (Defaults) sind Vorbelegungen von Werten, die als Standardannahme meistens, aber nicht immer gelten, z.B. "alle Vögel können fliegen". Erwartungswerte gelten nur dann, wenn kein genaueres Wissen verfügbar ist. Da sie gegebenenfalls korrigiert werden müssen, wird bei Verwendung von Erwartungswerten ein System zum nicht-monotonen Schließen (siehe Teil 4) benötigt.

Wenn man ein gut strukturiertes Framesystem mit Vererbungshierarchien hat, ist es ein naheliegender Gedanke, dass das System ein neues Objekt gemäß seiner Eigenschaften selbstständig in die richtige Position in der Hierarchie einfügt. Das erfordert eine strenge Definition, inwiefern ein Frame ein Unterframe eines anderen Frames ist, und natürlich das Verbot von Erwartungswerten, da sie willkürlich überschrieben werden können. Dieser Gedanke ist in der KL-ONE-Sprachfamilie realisiert, deren Kernstück ein "Classifier" zur Einordnung neuer Frames ist.

Ein Formalismus zur Wissensrepräsentation muss 1. Beschreibungsmittel für die symbolische Darstellung und Verarbeitung von Wissen bieten und 2. deren Anwendung zur Modellierung eines bestimmten Wissensbereichs (Domäne) gestatten. Bevor man Fakten über die Welt – oder einen Ausschnitt davon – repräsentiert und überlegt, wie daraus Inferenzen zu ziehen sind, muss erstmal eine "Ontologie" der existierenden Dinge begrifflich erfasst werden (Konzeptualisierung); dazu dient z.B. die T-Box in KL-ONE-Sprachen. Hierbei sollen Objekte mit gemeinsamen Merkmalen in einem Konzept abzulegen sein, um eine ökonomische Datenehaltung zu erreichen. Dabei abstrahiert man von anderen Merkmalen, die im betrachteten Kontext irrelevant sind. Wesentlich ist der Ökonomiegesichtspunkt: Es sind Modellierungen anzustreben, die Vererbung möglichst oft und das Einführen gleicher Merkmale bei Konzepten, die nicht in einer Erbbeziehung stehen, möglichst selten mit sich bringen.