# Grundlagen der Kognitiven Informatik

## Wissensrepräsentation und Logik

**Ute Schmid**

Kognitive Systeme, Angewandte Informatik, Universität Bamberg

letzte Änderung: 14. Dezember 2010

# Computermodelle des Denkens

## Inferenzprozesse

- Schlußfolgerndes Denken oder Inferenz bezeichnet einen Prozess, bei dem aus gegebenen Fakten und Regeln Schlüsse gezogen werden.
- Ein System, was zum schlussfolgernden Denken in der Lage sein soll, benötigt eine **Wissensbasis** und einen **Inferenzmechanismus**.
- In der Wissensbasis wird das Wissen über einen Gegenstandsbereich gespeichert.
- Der Inferenzmechanismus greift darauf zurück, um Schlußfolgerungen zu ziehen.
- Während das Wissen je nach Anwendung immer wieder neu in die Wissensbasis eingebracht – erfasst und repräsentiert – werden muss, sind die Inferenzregeln allgemeingültig.
- In der KI wird üblicherweise mit **logischen Formalismen** gearbeitet, um Inferenzprozesse auf dem Rechner zu modellieren.

# Logische Schlüsse mit Syllogismen

- Ein bekanntes Beispiel für einen logischen Schluss ist der *modus ponens*.
- Dies ist eine spezielle Inferenzregel beim syllogistischen Schließen.

- Links in der Tabelle ist der Syllogismus umgangssprachlich dargestellt, daneben in Form einer logischen Repräsentation.
- Geht man von der Annahme aus, dass die Behauptung, dass alle Menschen sterblich sind, wahr ist, und erhält man ausserdem die Information, dass Sokrates ein Mensch ist, so kann man folgern, dass Sokrates sterblich ist.

| | | |
|---|---|---|
| Alle Menschen sind sterblich. | $\forall$ x Mensch(x) $\rightarrow$ Sterblich(x) | **Axiom/Prämisse** |
| Sokrates ist ein Mensch. | Mensch(Sokrates) | **Fakt/Beobachtung** |
| | ———————————– | |
| Also ist Sokrates sterblich. | sterblich(Sokrates) | **Schluss** |

# Anmerkungen

- Der *modus ponens* ist eine allgemeingültige Schlussregel, das heißt, auf zwei beliebige Aussagen der obigen Form angewendet ist die erzeugte Schlussfolgerung immer korrekt.
- Das heißt, logische Schlüsse sind wahrheitserhaltend.
- Logische Schlussregeln können in Algorithmen umgesetzt und mechanisch ausgeführt werden.
- Die Wahrheit der Aussagen, auf die die Schlussregel angewendet wird, kann dagegen nicht innerhalb eines logischen Systems geprüft werden.
- Die Schlussregel würde beispielsweise aus den Aussagen *Alle Dozenten sind gute Tänzer.* und *Peter ist ein Dozent.* ableiten, dass Peter ein guter Tänzer ist. Diese Ableitung ist nur dann wahr, wenn beide Aussagen wahr sind.

# Drei Arten von Inferenz

- Das bisher betrachtete Beispiel illustriert sogenannte deduktive Inferenz.
- **Deduktion** bezeichnet *logische* Inferenz, also den Fall, bei dem korrekte Schlüsse aus vorhandenem Wissen gezogen werden.
- Dabei sind die Schlüsse stets *spezieller* als die vorgegebenen Axiome.
- Deduktion produziert also kein „neues"Wissen, sondern hilft, bereits Bekanntes zu explizieren.
- Zwei andere Arten von Inferenzproessen sind Induktion und Abduktion.

# Induktion

- Eine andere Art der Inferenz ist die **Induktion**.
- Hier wird aus Hintergrundwissen und Beobachtungen (Beispielen) eine Hypothese über einen generalisierten Sachverhalt aufgebaut.
- Mit Induktion wird also neues, aber ungesichertes Wissen erzeugt.
- Induktion ist deshalb der zentrale Inferenzmechanismus beim **maschinellen Lernen**.
- Natürlich wäre es höchst gewagt, aus nur einer einzigen Beobachtung bereits eine generalisierte Hypothese aufzubauen.

| | |
|---|---|
| Sokrates ist ein Mensch. | **Hintergrundwissen** |
| Sokrates ist sterblich. | **Beobachtung** |
| | |
| Alle Menschen sind sterblich. | **Hypothese** |

# Abduktion

- Eine dritte Form der Inferenz ist die **Abduktion**, wo aus einer gegebenen Theorie und beobachteten Sachverhalten auf Prämissen geschlossen wird.
- Diese Art des Schließens ist typisch für Diagnosesysteme, etwa im Bereich der Medizin.
- Beispielsweise kann ein Arzt aus seinem theoretischen Wissen, dass bei Grippe Fieber und Kopfschmerzen vorliegen, bei einem Patienten, der über diese Symptome klagt, schließen, dass der Patient an Grippe erkrankt ist.

| | |
|---|---|
| Alle Menschen sind sterblich. | **Theorie** |
| Sokrates ist sterblich. | **Beobachtung** |
| | |
| Sokrates ist ein Mensch. | **Diagnose** |

# Abduktion

- Auch mit abduktiver Inferenz können keine „wahren“Diagnosen abgeleitet werden.
- Es existiert keine allgemeingültige logische Schlussregel, mit der aus $A \rightarrow B$ (lies $A$ *impliziert* $B$) und dem Vorliegen von $B$ die Gültigkeit der Prämisse $A$ gefolgert werden kann!

In der Psychologie wie in der KI wird unter schlussfolgerndem Denken insbesondere deduktives Schliessen betrachtet.

# Logischen Deduktion in der Psychologie

- In der Psychologie gibt es zahlreiche Belege dafür, dass Menschen ihre Schlüsse häufig nicht nach den Gesetzen der logischen Deduktion ziehen.
- Beispiele sind das Umdrehen einer irrelevanten Karte bei der Wason-Aufgabe (Wason, 1972) oder die Bevorzugung einer Konjunktion von Fakten gegenüber einem einzelnen dieser Fakten als Schlussfolgerung (Tversky, 1983).
- Die Kenntnis von Grundlagen der deduktiven Logik ist aber dennoch auch für Psychologen sinnvoll, da logische Kalküle als normatives Modell für korrektes logisches Schliessen herangezogen werden können.
- Zudem basieren Repräsentationsformalismen, die auch in der Psychologie Verwendung finden – wie semantische Netze oder Schema-Hierarchien – , auf logischen Grundlagen.

# Grundlagen logischer Deduktion

- Um Inferenzen automatisch, auf einer Maschine, ablaufen lassen zu können, benötigt man einen sogenannten **Kalkül**.
- Ein Kalkül ist eine Menge syntaktischer Transformationsregeln, die für eine bestimmte Sprache definiert sind.
- Das dargestellte Kaffee-Dose-Problem ist ein Beispiel für ein sehr einfaches Kalkül:

  Die Sprache besteht aus allen möglichen Abfolgen der Buchstaben $S$ und $W$. Die Transformationsregeln ersetzen Buchstabenmuster durch andere Buchstabenmuster. Dies geschieht rein syntaktisch, ohne Berücksichtigung der Bedeutung der Buchstabenfolgen.
- Für die Automatisierung logischer Schlüsse ist die Sprache beispielsweise Aussagenlogik oder Prädikatenlogik erster Stufe. Ein bekanntes Kalkül ist das **Resolutionskalkül** (Robinson, 1965), das nur auf einer einzigen Regel, der Resolutionsregel, basiert.

# Logische Repräsentation

- Eine Logik ist eine formale Sprache.
- Wie bei einer natürlichen Sprache wird festgelegt, was die syntaktisch korrekten Ausdrücke sind, die zu der Sprache gehören.
- Die Menge aller syntaktisch korrekten Ausdrücke definiert die wohlgeformten Formeln (analog zu syntaktisch korrekten Sätzen einer natürlichen Sprache).
- Wesentlich ist, dass den Ausdrücken *Bedeutung* zugewiesen werden kann.
- Zu einer Logik gehört also auch eine Menge von bedeutungszuweisenden Regeln.
- Schließlich muß für eine Logik ein Beweiskalkül existieren, mit dem aus gegebenen Aussagen neue, wahre Aussagen abgeleitet werden können.
- Zwei der bekanntesten und am meisten eingesetzten Logiken sind Aussagenlogik (*propositional logic*) und Prädikatenlogik erster Stufe.

# Aussagenlogik

- Die atomaren syntaktischen Bausteine der Aussagenlogik sind Aussagen.
- Mithilfe einer Menge von Junktoren (*und*, *oder*, *impliziert*, *genau dann wenn*) können Aussagen zu komplexeren Aussagen verknüpft werden.
- Die Definition der Syntax der Aussagenlogik ist induktiv – ausgehend von den atomaren Elementen der Sprache wird festgelegt, wie aus bereits syntaktisch korrekten Ausdrücken komplexere neue, ebenfalls syntaktisch korrekte Aussagen gebaut werden können.
- Der Abschlusssatz in der Definition verhindert, dass auch andere syntaktische Ausdrücke als diejenigen, die durch die Konstruktionsanleitung definiert werden, zur Menge der wohlgeformten aussagenlogischen Formeln gezählt werden können.

# Syntax der Aussagenlogik

- *True* und *False* sowie Symbole {P, Q, R , ... } sind Formeln der Aussagenlogik. (atomare Formeln)
- Wenn $P$ und $Q$ aussagenlogische Formeln sind, dann auch
  $\neg P$ (nicht P),
  $P \wedge Q$ (P und Q), $P \vee Q$ (P oder Q)
  $P \rightarrow Q$ (P impliziert Q) $P \leftrightarrow Q$ (P genau dann wenn Q).
- Das sind alle aussagenlogischen Formeln.

# Bedeutung logischer Formeln

- Beispielsweise ist die Aussage $P \rightarrow Q$ eine aussagenlogische Formel.
- Diese Formel könnte für die Aussage "Wenn Mensch, dann sterblich" ("Mensch sein impliziert sterblich sein") stehen.
- Die **Semantik** aussagenlogischer Formeln ist durch den Wahrheitswert gegeben, für den sie stehen.
- Dieses Konzept von Bedeutung ist vielleicht zunächst etwas befremdlich.
- Vielleicht hilft folgendes Beispiel zur Veranschaulichung: Die Aussage "Das ist ein Baum" ist wahr, genau dann, wenn ich dabei auf einen Baum zeige und sonst falsch.
- Der Wahrheitswert "wahr" wird also "wahren Aussagen" zugewiesen.
- Die syntaktische Konstante "True" erhält immer den Wahrheitswert "wahr" (notiert mit 1), die Konstante "False" den Wahrheitswert "falsch" (notiert mit 0).

# Bedeutung logischer Formeln

- Die Bedeutung der Junktoren wird im folgenden definiert.
- Komplexe Formeln können mithilfe der Bedeutung der Junktoren zu Wahrheitswerten verkürzt werden.
- Formeln, die unabhängig von den Wahrheitswerten der Atome immer zu wahr auswerten, heissen **Tautologien**.

| P | Q | $\neg$ P | P $\wedge$ Q | P $\vee$ Q | P $\rightarrow$ Q | P $\leftrightarrow$ Q |
|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| 0 | 1 | | 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | | 1 | 1 | 1 | 1 |

# Tautologien

- Betrachten wir die Tautologie $P \rightarrow (Q \vee \neg Q)$.
- Die Disjunktion $Q \vee \neg Q$ ist immer wahr: Wenn $Q$ wahr ist, ist $\neg Q$ falsch und umgekehrt.
- Eine Disjunktion, bei der ein Element wahr ist, ist wahr.
- Die Implikation "Wenn $P$, dann wahr" ist ebenfalls immer wahr, unabhängig davon, ob $P$ wahr oder falsch ist.
- Damit ist gezeigt, dass beispielsweise die Bauernregel "Wenn der Hahn kräht auf dem Mist ($P$), dann ändert sich das Wetter ($Q$) oder bleibt wie es ist ($\neg Q$)" eine allgemeingültige Aussage ist.

# Äquivalenzumformungen

Wie bei mathematischen Ausdrücken, können logische Formeln durch Nutzung von Äquivalenzen umgeformt werden.

Beispielsweise gelten die Regeln von de Morgan:

$$\neg(P \wedge Q) \equiv (\neg P \vee \neg Q)$$
$$\neg(P \vee Q) \equiv (\neg P \wedge \neg Q)$$

und eine Implikation kann durch eine Disjunktion ersetzt werden:

$$(P \rightarrow Q) \equiv (\neg P \vee Q).$$

Die obige Aussage $P \rightarrow (Q \vee \neg Q)$ ist also gleichbedeutend mit der Aussage $\neg P \vee (Q \vee \neg Q)$.

- Solche Äquivalenzumformungen sind wesentlich für die Anwendung syntaktischer Beweiskalküle, wie das unten dargestellte Resolutionskalkül. Durch rein syntaktische Umformungen wird versucht, eine gegebene Menge von Formeln so zu transformieren, dass am Ende ein Wahrheitswert als Ergebnis steht.

# Prädikatenlogik erster Stufe

- In der Aussagenlogik steht eine atomare Formel für eine komplette Aussage.
- In der Prädikatenlogik (erster Stufe) erhalten die Formeln dagegen interne Struktur.
- Die Aussage "Sokrates ist ein Mensch" ist eine Formel ($P$) in der Aussagenlogik.
- In der Prädikatenlogik können wir dagegen schreiben *mensch(Sokrates)*.
- Dabei heißt *mensch* **Prädikat** und *Sokrates* ist eine **Konstante**, die als Argument des Prädikates dient.

# Prädikatenlogik erster Stufe

- Die Aussage “Der Vater von Sokrates ist ein Mensch” kann ausgedrückt werden als *mensch(vater-von(Sokrates))*.
- Dabei ist *vater-von* eine **Funktion**, die für ihr Argument einen konstanten Wert zurückliefert.
- Für *vater-von(Sokrates)* = *Sophroniskus* kann der funktionale Ausdruck durch sein Ergebnis, ersetzt werden, die Formel kann also reduziert werden zu *mensch(Sophroniskus)*.
- Schließlich können Prädikate auch über **Variablen** definiert werden. Dies ist insbesondere interessant für All- und Existenzaussagen.
- Die Aussage “Alle Menschen sind sterblich” kann mit dem **Allquantor** ($\forall$) dargestellt werden als $\forall x$ *mensch(x)* $\rightarrow$ *sterblich(x)*.
- Die Aussage, “Es existiert ein glücklicher Mensch” kann mit dem **Existenzquantor** ($\exists$) dargestellt werden als $\exists x$ *mensch(x)* $\wedge$ *glücklich(x)*.

# Prädikatenlogik erster Stufe

- Allgemein sind die Argumente von Prädikaten Terme.
- Atomare Formeln können dann, wie in der Aussagenlogik, durch Junktoren zu komplexeren Formeln zusammengefügt werden.
- Zusätzlich gibt es bei der Prädikatenlogik erster Stufe noch die beiden genannten Quantoren – den Allquantor und den Existenzquantor.
- Auch die Prädikatenlogik erster Stufe kann induktiv definiert werden.
- Dabei ist es wesentlich, den Unterschied zwischen **Termen** und **Formeln** zu beachten.
- Terme werden, wie in der Mathematik üblich, zu Werten ausgewertet, etwa $2 + 5$ zu 7.
- Prädikate und Formeln dagegen, wie auch in der Aussagenlogik, zu ganz speziellen Werten, nämlich Wahrheitswerten.

# Syntax der Prädikatenlogik erster Stufe

- Terme:
  - Variablen $x \in X$ sind Terme.
  - Wenn $f$ ein Funktionssymbol mit $n$ Argumenten ist und $t_1 \dots t_n$ Terme sind, dann ist auch $f(t_1, \dots, t_n)$ ein Term. (Konstanten sind dabei Funktionen mit 0 Argumenten).
  - Das sind alle Terme.
- Formeln:
  - Wenn $P$ ein Prädikatsymbol mit $n$ Argumenten ist und $t_1 \dots t_n$ Terme sind, dann ist $P(t_1, \dots, t_n)$ eine Formel. (atomare Formel)
  - Wenn $P$ und $Q$ Formeln sind, dann auch
    $\neg P$ (nicht P),
    $P \wedge Q$ (P und Q), $P \vee Q$ (P oder Q)
    $P \rightarrow Q$ (P impliziert Q) $P \leftrightarrow Q$ (P genau dann wenn Q).
  - Wenn $x$ eine Variable ist und $P$ eine Formel, dann sind auch $\exists x\ P$ und $\forall x\ P$ Formeln.
  - Das sind alle Formeln.

# Bedeutung prädikatenlogischer Formeln

- Wie in der Aussagenlogik wird die Bedeutung von Formeln durch Wahrheitswerte festgelegt.
- Allerdings kann dies hier nicht direkt geschehen, da zunächst die Bedeutung der Argumente der Prädikate bestimmt werden muß.
- Dies geschieht bezüglich einer Menge von Objekten mithilfe einer **Interpretationsfunktion**.
- Objektmenge und Interpretationsfunktion gemeinsam definieren eine Struktur.
- Die formale Definition der Semantik der Prädikatenlogik findet sich in Logiklehrbüchern (z.B. Schöning, 1992) und, etwas anschaulicher, auch in KI-Lehrbüchern (Russel & Norvig, 2002).

# Illustration

Formel:
$(\exists x \; on(x, A) \rightarrow \neg clear(A)) \wedge \exists y \; clear(topof(y))$

Interpretation der Symbole:

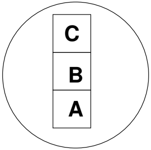
Konstantensymbol *A*: Block in einer Blockwelt

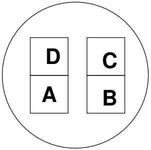
Einstelliges Funktionssymbol *topof(x)*: Funktion, die den Block liefert, der direkt auf Block *x* liegt

Einstelliges Prädikatsymbol *clear(x)*: wahr, wenn kein Block auf Block *x* liegt, falsch sonst

Zweistelliges Prädikatsymbol *on(x, y)*: wahr, wenn Block *x* auf Block *y* liegt, falsch sonst

Mögliche Strukturen:



# Illustration

- Wenn wir die Bedeutung der Symbole der gegebenen Formel bezüglich einer Blockwelt interpretieren, so kann die Formel der folgenden Aussage entsprechen:
- Wenn ein Block auf Block A liegt, so ist A nicht frei und es existiert ein Block, der auf einem anderen Block steht, aber auf dem selbst kein weiterer Block steht.
- An den zwei möglichen graphischen Veranschaulichungen sehen wir, dass Block $A$ nicht frei ist, weil in beiden Strukturen ein Block auf $A$ liegt.
- Ferner gilt in der ersten Veranschaulichung für $B$ und in der zweiten für $A$ oder $B$, dass der Block, der auf diesen Blöcken liegt, frei ist.
- Der erste Teil der Konjunktion ist in jeder legalen Blockwelt wahr; der zweite Teil ist in jeder Blockwelt wahr, in der ein Turm aus mindestens zwei Blöcken existiert.

# Erfüllbarkeit und Allgemeingültigkeit

- Eine Struktur, in der eine Formel zu “wahr” ausgewertet werden kann, heisst auch **Modell der Formel**.
- Wenn jede mögliche Struktur ein Modell der Formel ist, heißt die Formel **allgemeingültig**.
- Wenn mindestens ein Modell existiert, heißt die Formel **erfüllbar**.
- Beispielsweise ist die Formel $\forall P(x) \vee \neg P(x)$ gültig. Egal, wie wir die Formel interpretieren, ob als “Das Wetter ist regnerisch oder das Wetter ist nicht regnerisch” oder “Ein Block ist frei oder nicht frei”, oder mit einer beliebigen anderen Semantik, in der klassischen Logik ist die Disjunktion eines Prädikates mit seiner Negation immer wahr.
- Die oben angegebene Formel ist erfüllbar – wir haben zwei mögliche Modelle angegeben.

# Logische Folgerung

- Eine Formel $G$ heißt logische Folgerung aus einer Menge von Formeln $F = \{F_1 \ldots F_n\}$, wenn jedes Modell von $F$ auch ein Modell von $G$ ist.
- Da alle Formeln in $F$ als wahr angenommen werden, entspricht die Menge der Konjunktion $F_1 \wedge F_2 \wedge \ldots F_n$, abgekürzt notiert als $\wedge_{i=1}^{n} F_i$.
- Wenn gilt, dass $(\wedge_{i=1}^{n} F_i) \rightarrow G$ gültig ist (eine Tautologie ist), ist $G$ eine logische Folgerung aus $F$.
- Äquivalent dazu ist dann die Formel $(\wedge_{i=1}^{n} F_i) \wedge \neg G$ nicht erfüllbar (ein Widerspruch).

# Historisches

- Während die Aussagenlogik bereits in der Antike eigeführt wurde, wurde die Prädikatenlogik erst im neunzehnten und zwanzigsten Jahrhundert entwickelt.
- Zunächst formalisierte George Boole 1847 die Aussagenlogik.
- Darauf aufbauend entwarf Gottlob Frege 1879 die Prädikatenlogik.
- Das Konzept der Interpretation von Formeln bezüglich Objekten einer realen (oder Modell-) Welt hat Alfred Tarski dann im 20. Jahrhundert eingeführt.