Studientag zum Kurs 1698 –
Grundlagen der Künstlichen Intelligenz

Rainer Osswald

Intelligente Informations- und Kommunikationssysteme
Fakultät für Mathematik und Informatik
FernUniversität in Hagen

25. August 2007, 10:00 – 18:00

Charakterisierung von Intelligenz [1.1]

- Behavioristischer Ansatz: Turing-Test
- Konstruktiver Ansatz: Inhärente Merkmale von Intelligenz

C1 Fähigkeit zum Operieren mit Symbolen
C2 Besitz eines inneren Weltmodells (Wissensrepräsentation)
C3 Fähigkeit, gespeichertes Wissen zielgerichtet einzusetzen

C4 Fähigkeit zum logischen Schlussfolgern
C5 Fähigkeit zur Abstraktion und Spezialisierung
C6 Fähigkeit zum Transfer von Wissen auf neue Situationen
C7 Problemlösefähigkeit
C8 Anpassungsfähigkeit an verschiedene Problemsituationen
C9 Lernfähigkeit, mit Fähigkeit zur Selbsteinschätzung
C10 Beherrschung unvollständig beschriebener Situationen
C11 Fähigkeit zur Mustererkennung und Interaktion
C12 Kommunikationsmittel vergleichbar der menschlichen Sprache

Methodologisches Vorgehen der KI [1.6]

- Symbolische vs. neuronale KI

  Symbolische KI:

  "top down"
  Vorbild: Sprache, Logik

  Neuronale (bzw. subsymbolische) KI:

  "bottom up"
  Vorbild: Gehirn

- Simulationsmethode vs. phänomenologische Methode

  Simulationsmethode:

  Simulation der Mechanismen (menschlicher) Intelligenz

  Phänomenologische Methode:

  Rekonstruktion intelligenten Verhaltens ($\leadsto$ Turing-Test)

Philosophische Probleme der KI [1.7]

- Ontologische Gesichtspunkte

  Embodiment-Aspekt

  Trennung von Struktur und Funktion

  John Searles “Chinesisches Zimmer”

- Erkenntnistheoretische Gesichtspunkte

  Gödelscher Unvollständigkeitssatz

  Argument der fehlenden Fähigkeiten

  Argument von Ada Lovelace

Künstliche neuronale Netze [2.2]

- Neuronenfunktion

$$o_i \;=\; s\Big(\sum_{j=1}^{n} w_{ij}\cdot e_j - \theta\Big)$$

$s \;=\;$ Schwellwertfunktion, Sigmoidfunktion, ...

- Klassifikationsmerkmale künstlicher neuronaler Netze

  Topologie, Trainingsart, Neuronenfunktion, Dynamik

- Netzwerktypen (Auswahl)

  Perzeptron (lineare Separierbarkeit der Musterklassen)

  Backpropagation-Netz (Feed-Forward, Hidden Layers, Fehlerfunktion)

  Hopfield-Netz (Spinglasmodell, Energiefunktion)

  Kohonen-Netz (Unüberwachtes Lernen, Topologieerhaltung)

Typische KI-Systeme [1.5]

- Expertensysteme

  Zentrale Komponenten:

  Wissensbasis
  Inferenzsystem
  Interface zum Nutzer
  Erklärungskomponente
  Wissenserwerbskomponente

- Natürlichsprachliche Systeme

  Beispiele:

  Frage-Antwort-Systeme
  Systeme zu maschinellen Übersetzung
  Natürlichsprachliche Interfaces
  Textgenerierungssysteme

- Roboter

- Tutorielle Systeme

Logische Grundlagen / Aussagenlogik [5.3]

- Logische Verknüpfungen zwischen Aussagen:

  Negation: $\neg p$ (nicht p)
  Konjunktion: $p \wedge q$ (p und q)
  Disjunktion: $p \vee q$ (p oder q)
  Konditional: $p \rightarrow q$ (wenn p dann q)
  Bikonditional: $p \leftrightarrow q$ (p genau dann, wenn q)

- Bedeutung/Semantik via Wahrheitswerte (Aussagen sind wahr oder falsch)

p	q	$\neg p$	$p \wedge q$	$p \vee q$	$p \rightarrow q$	$p \leftrightarrow q$
F	F	T	F	F	T	T
F	T	T	F	T	T	F
T	F	F	F	T	F	F
T	T	F	T	T	T	T

Beispiel für wahre Aussage:

Wenn der Mond größer als die Erde ist, dann ist die Sonne kleiner als der Mond.

Aussagenlogik (Forts.) [5.3]

- Interpretation (formal)

  Vorgegeben: Menge P von Aussagevariablen
  Belegung/Interpretation: Funktion $I : P \longrightarrow \{\mathtt{F}, \mathtt{T}\}$
  Interpretation zusammengesetzter Aussagen gemäß Wahrheitswertetabelle

- Erfüllbarkeit, Allgemeingültigkeit, Inkonsistenz

  Aussage (genauer: aussagenlogischer Ausdruck) ϕ ist

  erfüllbar gdw. $I(\phi) = \mathtt{T}$ für eine Interpretation I (z.B. $p \wedge q$)
  allgemeingültig gdw. $I(\phi) = \mathtt{T}$ für alle Interpretationen I (z.B. $p \rightarrow p \vee q$)
  inkonsistent gdw. $I(\phi) = \mathtt{T}$ für keine Interpretation I (z.B. $p \wedge \neg p$)

- Äquivalenz, Implikation (Folgerung)

  ϕ und ψ sind äquivalent gdw. $I(\phi) = I(\psi)$ für alle Interpretationen I
  ϕ impliziert ψ gdw. mit $I(\phi) = \mathtt{T}$ stets $I(\psi) = \mathtt{T}$ für alle Interpretationen I

  Beispiele: p impliziert $p \vee q$, $p \rightarrow q$ ist äquivalent zu $\neg p \vee q$

  Beobachtungen: ϕ und ψ äquivalent gdw. $\phi \leftrightarrow \psi$ allgemeingültig
  ϕ impliziert ψ gdw. $\phi \rightarrow \psi$ allgemeingültig

Aussagenkalkül [5.3]

- Kalkül (allgemein)

  Ausdrucksmenge Σ (über Alphabet)

  Ableitungsrelation $\subseteq \wp(\Sigma) \times \Sigma$ bzw. Ableitungsfunktion $a : \wp(\Sigma) \rightarrow \wp(\Sigma)$

  "Startmenge" von Ausdrücken (Axiome)

  Forderungen: Monotonie, Einbettung, Abgeschlossenheit, Endlichkeitsaxiom

- Ableitungskalkül der Aussagenlogik (Beispiel!)

  Ausdrucksmenge: Aussagen mit $\vee$ und $\neg$

  Ableitungsrelation $\vdash$: Verkürzung, Ausdehnung, Assoziativität,
  Schnittregel: $A \vee B, \neg A \vee C \;\vdash\; B \vee C$

  Axiom: $\neg A \vee A$ (tertium non datur)

- Begriffe: Beweisbarkeit, Ableitbarkeit, Theorem, modus ponens/tollens

- Wichtig: Korrektkeit und Vollständigkeit eines Ableitungskalküls

- Deduktionstheorem: $A \vdash B$ gdw. $\vdash A \rightarrow B$

Logisch-deduktives Schließen (Aussagenlogik) [7.2]

- Folgerungs-/Widerlegungstheorem:

  B folgt aus $A_1, \ldots, A_n$

  gdw. $A_1 \wedge \ldots \wedge A_n \rightarrow B$ allgemeingültig

  gdw. $A_1 \wedge \ldots \wedge A_n \wedge \neg B$ inkonsistent

- Konjunktive/disjunktive Normalform, Klauselform

  Beispiel: $p \rightarrow q \wedge r$

  KNF: $(\neg p \vee q) \wedge (\neg p \vee r)$

  Klauselform: $\{\{\neg p, q\}, \{\neg p, r\}\}$ (Menge von Literalmengen)

- Resolventenregel:

  $\{A\} \cup K_1, \{\neg A\} \cup K_2 \vdash K_1 \cup K_2$ (K_1, K_2 Klauseln)

  (speziell: $\{A, B\}, \{\neg A, C\} \vdash \{B, C\}$)

- Resolutionsprinzip:

  Widerlegung durch Ableitung der leeren Klausel mittels sukzessiver Resolventenbildung

Prädikatenlogische Formalisierung [5.4]

- Individuen(konstanten), Prädikate, Funktionen

  $\mathrm{INFORMATIKERIN}(\mathrm{Maria})$ (*Maria ist eine Informatikerin*)

  $\mathrm{SCHWESTER}(\mathrm{Maria}, \mathrm{Peter})$ (*Maria ist eine Schwester von Peter*)

  $\mathrm{Hans} = \mathrm{Vater\_von}(\mathrm{Maria})$ (*Hans ist der Vater von Maria*)

- Quantoren und Variablen

  Existenzquantifizierung

  $\exists x\ P(x)$ *es gibt ein x derart, dass $P(x)$*

  Allquantifizierung

  $\forall x\ P(x)$ *jedes x ist derart, dass $P(x)$*

- Beispiel (für Skopusambiguität):

  Jeder Student kennt ein Logikbuch

  $\forall x(\mathrm{STUDENT}(x) \rightarrow \exists y(\mathrm{LOGIKBUCH}(y) \wedge \mathrm{KENNEN}(x,y)))$

  $\exists y(\mathrm{LOGIKBUCH}(y) \wedge \forall x(\mathrm{STUDENT}(x) \rightarrow \mathrm{KENNEN}(x,y)))$

Prädikatenlogik erster Stufe [5.4]

- Symbole, Terme, Formeln

  Symbole (außerlogische):

  Variablen, Konstanten, Funktionssymbole, Prädikate (Stelligkeit!)

  Terme:

  Variablen- und Konstantensymbole

  $f(t_1, \ldots, t_n)$, wobei t_i Terme, f n-stelliges Funktionssymbol

  Formeln:

  atomar: $P(t_1, \ldots, t_n)$, wobei t_i Terme, P n-stelliges Prädikat
  $t_1 = t_2$, wobei t_1, t_2 Terme

  $\exists x\, \phi$, $\forall x\, \phi$, wobei ϕ Formel

  Aussagenlogische Verknüpfungen (inkl. Negation) von Formeln

Modelltheoretische Semantik der Prädikatenlogik [5.4]

- Universum U: nichtleere Menge

  Interpretationsfunktion I:

  Konstanten $\rightsquigarrow$ Elemente von U
  n-stellige Prädikate $\rightsquigarrow$ n-stellige Relationen über U
  n-stellige Funktionssymbole $\rightsquigarrow$ n-stellige Funktionen über U

  Variablenbelegung b:

  Funktion $V \to U$ (V Menge der Variablen)

- Wert eines Ausdrucks (bzgl. I und b)

  Terme: $[c]_b^I = c^I$ (c Konstante)
  $[v]_b^I = b(v)$ (v Variable)
  $[f(t_1,\ldots,t_n)]_b^I = f^I([t_1]_b^I,\ldots,[t_n]_b^I)$

  Formeln: $[t_1 = t_2]_b^I = \mathtt{T}$ gdw. $[t_1]_b^I = [t_2]_b^I$
  $[P(t_1,\ldots,t_n)]_b^I = \mathtt{T}$ gdw. $\langle [t_1]_b^I,\ldots,[t_n]_b^I \rangle \in P^I$
  $[\exists x P(x)]_b^I = \mathtt{T}$ gdw. $[P(x)]_{b_{x/i}}^I = \mathtt{T}$ für ein $i \in U$
  ...

Semantik der Prädikatenlogik (Forts.) [5.4]

- Erfüllbarkeit, Allgemeingültigkeit

  ϕ erfüllbar gdw. $[\phi]^I_b = \mathtt{T}$ für ein I, b

  ϕ allgemeingültig bzgl. I gdw. $[\phi]^I_b = \mathtt{T}$ für alle b

  ϕ allgemeingültig gdw. $[\phi]^I_b = \mathtt{T}$ für alle I, b

- Modell für Formelmenge T:

  Interpretation I derart, dass jedes $\phi \in T$ allgemeingültig bzgl. I

- Implikation (Folgerung)

  T impliziert ϕ gdw. jedes Modell von T auch Modell von ϕ

  Beispiel: $\forall x\, P(x)$ impliziert $\exists x\, P(x)$

Prädikatenkalkül [5.4]

- Ausdrucksmenge:

  Prädikatenlogische Formeln ohne $\wedge$, $\rightarrow$, $\forall$
  ($\forall\phi$ äquivalent zu $\neg\exists\neg\phi$)

- Ableitungsrelation $\vdash$:

  Ableitungsregeln des Aussagenkalküls
  $\exists$-Einführung: $A \rightarrow B \vdash \exists x A \rightarrow B$ (x nicht frei in B)

- Axiome:

  Tertium non datur
  Einsetzungsaxiom
  Identitätsaxiom
  Gleichheitsaxiom

- Korrektheit, Vollständigkeit!

Prädikatenlogische Resolution [7.2]

- Konjunktive Pränexnormalform

  $Q_1 x_1 \ldots Q_n x_n \phi$, $\quad Q_i$ Quantor, ϕ in KNF

  Skolemnormalform

  Beseitigung von Existenzquantoren durch Einführung von Skolemtermen (implizite Allquantifikation)

  Klauselform ($=$ Menge von Literalmengen)

- Erfüllbarkeitstheorem:

  Formel inkonsistent gdw. ihre Skolemnormalform inkonsistent

- Resolventenregel:

  $\{A_1\} \cup K_1, \{\neg A_2\} \cup K_2 \ \vdash \ K_1\sigma \cup K_2\sigma$
  (σ allgemeinster Unifikator von A_1, A_2)

- Resolutionsprinzip:

  Widerlegung durch Ableitung der leeren Klausel mittels sukzessiver Resolventenbildung

Beschränkungen der Prädikatenlogik [5.3/6.2]

- Doppelte Negation

  Die Studientagsteilnehmer waren nicht unvorbereitet.

- Konditionalsätze

  Wenn der Himmel wolkenlos ist, dann sind die Sterne gut sichtbar.

- Definite Kennzeichnungen

  Der gegenwärtige König von Frankreich hat eine Glatze.

- "Esel-Sätze"

  Jeder Bauer, der einen Esel besitzt, schlägt ihn.

- Modalausdrücke

  Vielleicht ist die Prädikatenlogik für die Wissensrepräsentation geeignet.

Wissensrepräsentationsmethoden [6.1]

- Allgemeine Unterscheidung:

  deklarative WRM
  prozedurale WRM

- Deklarative WRM:

  Datenmodelle traditioneller Datenbanksysteme
  objektorientierte WRM
  regelorientierte WRM

- Objektorientierte WRM:

  semantische Netze (KSN, SVN)
  Frame-Repräsentationen

- Regelorientierte WRM:

  Logik-orientierte WRM
  Produktionsregelsysteme

Frame-Repräsentationen [6.4]

- Frame $\approx$ Menge von Merkmal-Wert-Paaren

  Merkmal = Slot, Wert = Filler

- Vererbungshierarchie (Polyhierarchie):

  Frames erben von übergeordneten Frame deren Slots (und Filler)

  Beachte: Defaultangaben möglich!

- Vererbungstypen (Auswahl)

  SAME – identische Vererbung

  OVERRIDE – Überschreiben von Werten

  UNION – Vereinigung von Werten

Kognitive semantische Netze [6.3]

- Semantisches Netz (allg.): (markierter gerichteter) Graph

  Knoten repräsentieren Begriffe/Konzepte
  Kanten repräsentieren Beziehungen (zwischen Begriffen)

- Darstellungsmittel kognitiver semantischer Netze

  Sorten, Relationen, Schichten

  Beachte: generische Begriffe vs. Individualbegriffe

- Semantische Relationen (Auswahl)

  SUB/SUBA – Subordination von Objekten bzw. Vorgängen
  PARS – Teil-Ganzes-Beziehung
  POSS – Besitzrelation
  PROP – Eigenschaftszuordnung
  AGT – Beziehung zwischen Handlung und Handelndem
  INSTR – Beziehung zwischen Vorgang und Instrument
  DIRCL – Beziehung zwischen Vorgang und Richtung
  TEMP – Beziehung zwischen Sachverhalt und Zeitpunkt

Schlußweisen (allgemein) [7.1]

- Deduktion

  "Vorwärtsanwendung" von als gültig angenommener Regeln

  Monotones Schließen: Klassisch-logische Ableitung
  Nichtmonotones Schließen: Default Logik, CWA-Logik, ...

- Abduktion

  Beispiel: Befund & Kausalbeziehung $\rightsquigarrow$ Hypothese über Ursache

- Induktion

  Hypothetischer Schluss von Einzelfällen auf allgemeinen Zusammenhang

- Analogieschließen

- Approximatives Schließen

  Probabilistisches Schließen

  Fuzzy Reasoning

Abduktives Schließen [7.3]

- Grundschema des abduktiven Schließens:

  $A \to B$, B wahr $\rightsquigarrow$ A glaubwürdiger

- Beispiel: Diagnostisches Problem

  D Menge von Defekten
  M Menge von Manifestationen
  $C \subseteq D \times M$ Ursache-Wirkungsbeziehung

  Induzierte Ursache- und Wirkungsfunktionen:

  $\mathit{Ursache} : \wp(M) \to \wp(D),\ \mathit{Ursache}(M') = \{d \in D \mid \exists m \in M'(\langle d, m\rangle \in C)\}$
  $\mathit{Wirkung} : \wp(D) \to \wp(M),\ \mathit{Wirkung}(D') = \{m \in M \mid \exists d \in D'(\langle d, m\rangle \in C)\}$

  $E \subseteq D$ ist Erklärung von $M_0 \subseteq M$ gdw.

  $M_0 \subseteq \mathit{Wirkung}(E)$ (d.h. E ist Abdeckung von M_0)

  E ist minimal

Probabilistisches Schließen [7.4]

- Bayessches Theorem:

$$P(A|B) \;=\; \frac{P(B|A)\cdot P(A)}{P(B)}$$

- Bayessche Netze:

Endliche Menge I mit strikter partieller Ordnung $<$

Familie $(X_i)_{i\in I}$ von Zufallsvariablen mit

$$P(X_i = x_i \,|\, (X_j = x_j)_{j<i}) \;=\; P(X_i = x_i \,|\, (X_j = x_j)_{j\in v(i)})$$

wobei $v(i) =$ Menge der unmittelbaren Vorgänger von i bzgl. $<$

Es gilt:

$$\begin{aligned} P((X_i = x_i)_{i\in I}) \;&=\; \prod_{i\in I} P(X_i = x_i \,|\, (X_j = x_j)_{j<i}) \\ &=\; \prod_{i\in I} P(X_i = x_i \,|\, (X_j = x_j)_{j\in v(i)}) \end{aligned}$$

Automatisches Problemlösen [8.1]

- Problem (formal): Quadrupel $\langle P, A, Z, O \rangle$

  P – Menge der Problemzustände
  A – Anfangzustand $\in P$
  Z – Menge der Zielzustände (Lösungsmenge) $\subseteq P$
  O – Menge der Operationen ($=$ partielle Funktionen) auf P

  Lösungsweg:

  Endliche Folge $\langle o_1, \ldots, o_n \rangle$ mit $(o_n \circ o_{n-1} \circ \ldots \circ o_1)(A) \in Z$

- Problemgraph: markierter gerichteter Graph mit Wurzel

  Zustände $\rightsquigarrow$ Knoten
  Anfangszustand $\rightsquigarrow$ Wurzel
  Zielzustände $\rightsquigarrow$ (bestimmte) terminale Knoten
  (bestimmte) Pfade $\rightsquigarrow$ Lösungswege

- Problemlösen als Suche im Problemgraphen

  Strategien: Tiefe-Zuerst, Breite-Zuerst, Hill-Climbing, ...

Automatisches Problemlösen (Forts.) [8.2]

- Mehrpersonenspiele als antagonistische Probleme (Spieltheorie)

  Extensive Spiele: abwechselnde Züge (Operationen)

  Spiele mit vollständiger Information: alle zu erwartenden Züge bekannt

  Strategie: Plan der auszuführende Züge eines Spielers

  Partie: Kombination der Strategien aller Spieler (= Pfad im Spielbaum)

  Gewinnfunktion (pro Spieler): Bewertung einer Partie

  Nullsummenspiel: Bewertungen einer Partie addieren sich zu Null

- Optimale Strategie bei (extensiven) Zweipersonen-Nullsummenspielen

  MINIMAX-Verfahren (Back-up-Prozess):

  Gegeben: Spielbaum mit bewerteten Endzuständen (gemäß Spieler A)

  Sukzessive Bewertung der übrigen Zustände:

  Wenn A am Zug: Maximum der Werte der Nachfolgezustände
  Wenn B am Zug: Minimum der Werte der Nachfolgezustände

  Variante: Beschränkung der Entwicklungstiefe

Automatisches Problemlösen (Forts.) [8.2]

- ALPHA-BETA-Prinzip:

  "If you have an idea that is surely bad, do not take time to see how truly awful it is." (P. H. Winston: Artifical Intelligence)

  Selektive Entwicklung des Spielbaums:

  α = untere Schranke für Zustandsbewertung des MAX-Spielers

  β = obere Schranke für Zustandsbewertung des MIN-Spielers

  ALPHA-Abbruch bei MIN-Knoten n falls

  $\alpha(n') \geq \beta(n)$ (n' Vorgänger von n)

  BETA-Abbruch bei MAX-Knoten n falls

  $\beta(n') \leq \alpha(n)$ (n' Vorgänger von n)

LISP [4]

- Charakteristika:

  funktionale Programmiersprache, nicht statisch getypt

- Zentraler Datentyp: S-Ausdruck

  $\langle \textit{S-expr} \rangle ::= (\langle \textit{datum} \rangle \,.\, \langle \textit{datum} \rangle)$

  $\langle \textit{datum} \rangle ::= \langle \textit{atom} \rangle \mid \langle \textit{S-expr} \rangle$

  Elementarfunktionen für S-Ausdrücke: `CONS`, `CAR`, `CDR`

  $\texttt{(CONS } obj_1 \ obj_2\texttt{)} \Longrightarrow \texttt{(}obj_1 \,.\, obj_2\texttt{)}$
  $\texttt{(CAR (}obj_1 \,.\, obj_2\texttt{))} \Longrightarrow obj_1$
  $\texttt{(CDR (}obj_1 \,.\, obj_2\texttt{))} \Longrightarrow obj_2$

  Listennotation (Beispiel): `(A B C)` für `(A . (B . (C . NIL)))`

- Theoretische Basis des Funktionskonzepts: λ-Kalkül

  Funktionale Abstraktion: $f(x) \rightsquigarrow \lambda x\, f(x)$

  Funktionale Applikation: $(\lambda x\, f(x))(y) = f(y)$

LISP (Forts.) [4]

- Beispiel für rekursive Funktion: `APPND`

```
(APPND '(A B C) '(D E F)) ⟹ '(A B C D E F)
```

```
(DEFUN APPND (SEQ1 SEQ2)
   (IF (NULL SEQ1)
       SEQ2
       (CONS (CAR SEQ1) (APPND (CDR SEQ1) SEQ2))))
```

- Beispiel für Funktion höherer Ordnung: `FOLD`

`(FOLD OP X0 (X1 X2 X3))` entspricht `(OP (OP (OP X0 X1) X2) X3)`

```
(DEFUN FOLD (OP INIT SEQU)
   (IF (NULL SEQU)
       INIT
       (FOLD OP (FUNCALL OP INIT (CAR SEQU)) (CDR SEQU))))
```

```
(FOLD #'CONS 'X0 '(X1 X2 X3)) ⟹ (((X0 . X1) . X2) . X3)
```

- Viele frei verfügbare LISP- (bzw. Scheme-)Systeme

  z.B. CLISP (http://clisp.cons.org) (in SuSE-Distribution)