

Einführung in die Methoden der Künstlichen Intelligenz

Einführung

PD Dr. David Sabel

SoSe 2014

Womit beschäftigt sich die Künstliche Intelligenz?

Ziel der Künstlichen Intelligenz:

Herstellung eines **intelligenten Agenten**

Auch: Herstellung eines möglichst guten

- autonomen
- lernenden
- intelligenten
- automatischen

Informationssysteme

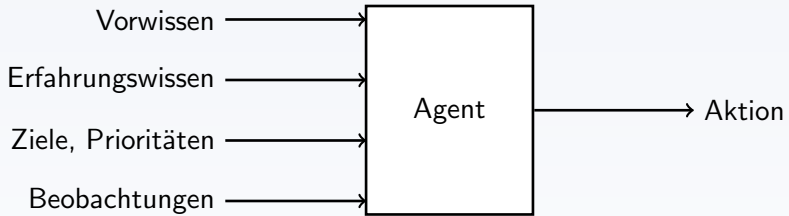
Beispiele

- Taschenrechner
- Schachspielende Computer,
z.B. Deep Blue, Deep Thought und Deep Fritz
- Sprachübersetzer wie z.B. GoogleTranslate, Babelfish, etc.
- Wissensbasierte Systeme in vielen Varianten
- Texterkennungssysteme in Kombinationen mit
Hintergrundwissen wie z.B. IBM Watson (Jeopardy-Gewinner)
- Roboter, z.B. Haushaltsroboter wie Staubsaugerroboter,
Industrieroboter, etc.
- intelligente Informationssysteme

Roboter und weitere Begriffe

- Roboter** = Intelligenter Agent, Computer im Kern, agiert mit der **physikalischen** Umwelt
- Softbot** = Software-Roboter, Umwelt ist i.A. nicht physikalisch, hat Wissensbasis und gibt Antworten und Ratschläge
- Webbot** = Web-Roboter, (inter-)agiert im WWW, z.B. um Suchdatenbanken zu erstellen
- Chatbot** = Chat-Roboter, interagiert in einem Chat

Agent, allgemein



- **Vorwissen**: Wissen über die Umgebung z.B. Karte
- **Erfahrungswissen**: erlerntes Wissen, Testfälle
- **Ziele**: üblicherweise mit Prioritäten und Wichtigkeiten versehen.
- **Beobachtungen**: über die Umgebung und über sich selbst

Nächste **Aktion** als Ausgabe (aufgrund von Schlussfolgern, Lernen,...)

Fragestellungen

- Was ist Künstliche Intelligenz?
- Was zeichnet eine Methode als **KI-Methode** aus?
- Was ist ein **intelligenter** Agent?
- Können Computer **denken**?

Es gibt keine allgemein richtige Antwort auf diese Fragen!

Klassifizierung der Ansätze nach Russel & Norvig

	menschlich	rational
Handeln	menschliches Handeln	rationales Handeln
Denken	menschliches Denken	rationales Denken

Menschliches Handeln

	menschlich	rational
Handeln	menschliches Handeln	rationales Handeln
Denken	menschliches Denken	rationales Denken

Fernziel: Systeme erschaffen, die **analog zu Menschen handeln**

Entwickle Methoden, sodass der Computer Dinge tun kann,

- die momentan nur der Mensch kann
- in denen der Mensch noch den Computern überlegen ist

Nachweis, ob Ziel erreicht: Vergleiche Maschine und Mensch
(z.B. Turing-Test, folgt noch)

Menschliches Denken (1)

	menschlich	rational
Handeln	menschliches Handeln	rationales Handeln
Denken	menschliches Denken	rationales Denken

Ziel: Computer „denkt **wie** ein Mensch“

- Forschung: Wie denkt der Mensch, entwickle Modelle dafür
- Z.B. durch psychologische Experimente, Hirntomografie . . .
- Stelle Theorie auf, setze danach in ein System um
- Ziel erreicht: Wenn Ein-/Ausgaben dem menschlichen Verhalten gleichen

Menschliches Denken (2)

- Ansatz ist eher **interdisziplinär: Kognitionswissenschaft**
- Beispiele: Wie erkennen wir Gesichter, Verständnis des Spracherwerbs, ...
- System ist **kognitiv adäquat**, wenn:
 - arbeitet strukturell und methodisch wie ein Mensch
 - erzielt entsprechende Leistungen
- Achtung: Taschenrechner ist **nicht kognitiv adäquat**, da er **anders addiert** als der Mensch
- Daher eher:

kognitive Simulation \neq künstliche Intelligenz
- Zusehr **exakte Nachahmung** des menschlichen Denkens

Rationales Denken

	menschlich	rational
Handeln	menschliches Handeln	rationales Handeln
Denken	menschliches Denken	rationales Denken

- Formalisierung von Denken durch **Axiome** und **korrekte** Schlussregeln
- Üblich: Verwendung einer **Logik** („logischer Ansatz“)
- Ziel: Implementiere **Deduktionssystem**, das sich **intelligent** verhält
- Vorteil: Mathematisch eindeutig
- Hürden: Formalisierung des Problems & Wissens in einer Logik

Rationales Handeln (1)

	menschlich	rational
Handeln	menschliches Handeln	rationales Handeln
Denken	menschliches Denken	rationales Denken

- Agenten-Ansatz
- Agent = System, dass auf seine **Umgebung** (Eingaben) eine **Reaktion** (Ausgaben) durchführt.
- Agent sollte **autonom** agieren, sich auf Änderungen **anpassen** und ein **Ziel** verfolgen

Rationales Handeln (2)

- **Rationaler** Agent: Agent **maximiert** Ergebnis, bestmögliches Ergebnis
- Allgemeiner Ansatz, kann auch die anderen Ansätze miteinbeziehen:
 - z.B. Verwenden einer Logik
 - kann menschlich Denken / Handeln, wenn dies auch rational ist, aber keine Grundbedingung (daher mehr Freiheit in der Methodik)
- Rationalität kann mathematisch “sauber” definiert werden.
- Philosophische Fragen über Intelligenz kann man dann außen vor lassen.

Ausrichtungen

Kognitionswissenschaft:

Menschliches Handeln / Denken analysieren,
modellieren, nachahmen

Ingenieurmäßig ausgerichtete KI:

Entwickle Methoden, Techniken, Werkzeuge zum
Lösen komplexer Anwendungsprobleme. Z.B.

- Deduktionstechniken
- KI-Programmiersprachen
- neuronale Netze als Mustererkenner
- Lernalgorithmen
- wissensbasierte Systeme
- Expertensysteme

Ausrichtung der Vorlesung

- ingenieurmäßige Ausrichtung
- direkt programmierbare KI-Methoden
- insbes. Logiken in verschiedenen Varianten

Gebiete der Künstlichen Intelligenz (Auswahl)

- Programmierung strategischer Spiele (Schach, ...)
- Automatisches/Interaktives Problemlösen und Beweisen
- Natürlichsprachliche Systeme (Computerlinguistik)
- Bildverarbeitung
- Robotik
- (medizinische) Expertensysteme
- Maschinelles Lernen

Philosophische Aspekte

Philosophische Richtungen

- Materialismus
- Behaviorismus
- Funktionalismus

Materialismus

- Grundgedanke: Es gibt **nichts außer Materie**
- Insbesondere auch Geist, Gedanken, Wille, ...
entsteht aus der Materie
- Prinzip: **Alles** was den Menschen ausmacht, kann durch Naturwissenschaft erforscht / analysiert werden
- Konsequenz: Alles ist grundsätzlich auch **konstruierbar**
- Insbesondere: Prinzipiell ist auch der denkende, intelligente Mensch konstruierbar

Behaviorismus

- Grundprinzip: Nur das **Beobachtbare** ist Gegenstand der Wissenschaft
- Folgerung: Nur verifizierbare Fragen sind sinnvoll
- Glauben, Ideen, Wissen nur indirekt beobachtbar
- Bewusstsein, Ideen, Furcht, ...
sind **Umschreibungen** für **Verhaltensmuster**
- Verifizieren benötigt evtl. unendlich viele Beobachtungen
- Falsifizieren: Eine Beobachtung genügt
- Äquivalenz von Systemen (z.B. Mensch vs. Maschine)
gegeben **bei gleichem Ein-/Ausgabe-Verhalten**

Funktionalismus

- Grundidee: Geistige Zustände (Ideen, Glauben, Furcht,...) sind **interne Zustände** eines komplexen Systems
- Einzig die Funktion definiert die Semantik eines Systems
- Zustand S_1 des Systems A ist **funktional äquivalent** zu Zustand S_2 des Systems B , wenn:
 - A im Zustand S_1 und B im Zustand S_2 liefern bei gleicher Eingabe die gleiche Ausgabe
 - Nachfolgezustände von A und B sind funktional äquivalent
- Z.B. Röhren-Radio und Transistor-Radio sind funktional äquivalent
- Aber: CD-Player und Radio sind nicht äquivalent.
- Konsequenz: Mensch ist im Prinzip ein endlicher Automat mit Ein-/Ausgabe.

Starke und schwache KI-Hypothese

Schwache KI-Hypothese

Maschinen (Computer, Roboter, . . .) können agieren, **alsob** sie intelligent wären

Starke KI-Hypothese

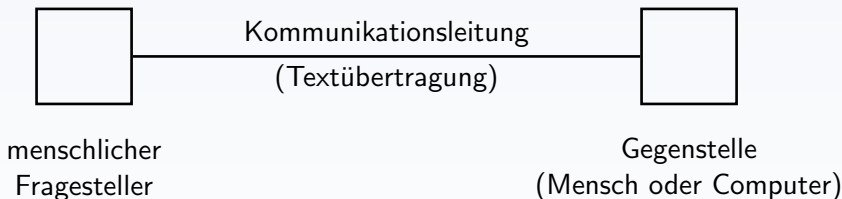
Maschinen (Computer, Roboter, . . .) können **wirklich** denken und **simulieren nicht nur** das Denken.

In der KI-Forschung

- Schwache KI-Hypothese wird als gegeben hingenommen
- Starke KI-Hypothese: Pragmatische Sichtweise: irrelevant, hauptsache das System funktioniert

Turing-Test

Test zum Nachweis der starken KI-Hypothese, vorgeschlagen von Alan Turing:



- Mensch stellt schriftliche Fragen an Computer / Mensch
- Begrenzte Zeit
- Test ist **bestanden**, wenn Fragesteller nicht unterscheiden kann, ob Gegenstelle Mensch oder Computer ist

Turing-Test (2)

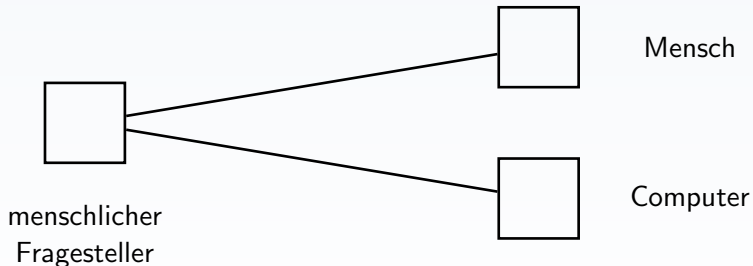
- Kritik: Nicht objektiv, da der Test von den Fähigkeiten des Fragestellers abhängt
- Auch z.B. Wissen des Fragestellers über Fähigkeiten eines Computers
- Abhilfe: Test mit mehreren Personen wiederholen

Totaler Turing-Test Unterschied zum Turing-Test:

Zusätzlich Videoübertragung und Objekterkennung

Loebner-Preis

- Jährlich ausgetragener Wettbewerb (seit 1991)
- Chatbots-Test ähnlich zum Turingtest
- Gewinner: Chatbot, der die Jury am meisten überzeugt hat (Bronze-Medaille)



Loebner-Preis (2)

Einmalige Preise:

- Silbermedaille (25.000 USD): Erster Chatbot, der die Jury überzeugend täuscht
- Goldmedaille (100.000 USD): wie Silbermedaille aber zusätzlich mit audio-visueller Kommunikation

Weder Silber noch Gold wurden bisher vergeben.

Turing-Test: Pro / Contra

Pro:

- Halbwegs einsichtiges Kriterium für “Intelligenz”

Contra:

- System als riesige Datenbank mit vorgefertigten Antworten

Ist das System **intelligent**?

ELIZA

- Von J. Weizenbaum entwickeltes Programm, das als Softbot einen **Psychotherapeuten** simuliert
- Konnte manche Menschen täuschen

Techniken:

- Vorgefertigte Phrasen, falls das System nichts versteht
“Erzählen Sie mir aus Ihrer Jugend”
- Mustererkennung: in der Eingabe wird nach Schlüsselwörtern gesucht
“Erzählen Sie mir mehr über *xyz*”

Chinesischer Raum

Gedankenexperiment von John Searle als Gegenargument zur starken KI-Hypothese:

- Jemand, der kein Chinesisch versteht, sitzt in einen Raum
- Im Raum:
 - Stapel mit Chinesischen Zetteln
 - Handbuch (in Muttersprache) mit Regeln wie aus Chinesischen Zetteln neue Chinesische Zettel erzeugt werden können
- Ein Chinesischer Zettel wird durch Schlitz reingereicht
- Person erzeugt neue Zettel auf Stapel und gibt einen Zettel nach außen

Chinesischer Raum (2)

Fragen

- **Versteht** die Person Chinesisch?
- **Versteht** das **Gesamtsystem** etwas?

J. Searle:

Kein Teil des System versteht irgendetwas

Gegenargument (Behaviorismus):

Das Gesamtsystem versteht etwas,
da das Verständnis **beobachtbar** ist

Das Prothesenexperiment

Annahme: Neuronen können künstlich nachgebaut werden
(elektronische Neuronen)

Experiment: Ersetze einzelne Neuronen durch elektronische Neuronen

Frage

Ab welcher Anzahl verwandelt sich das Prothesen-Gehirn in einen Computer, der nichts versteht?

Folgerungen:

- Entweder: Starke KI-Hypothese gilt und nichts ändert sich
- Oder: Es gibt etwas, das noch unbekannt ist (Geist,...)

Symbolverarbeitungshypothese

Physikalisches Symbolsystem

- Symbole, denen eine Bedeutung in der Realität zugeordnet werden kann
- System erzeugt aus eingegebener Symbolstruktur (z.B. String von Symbolen) weitere Symbolstrukturen

Symbolverarbeitungshypothese (Alan Newell und Herbert Simon)

Es kann ein physikalisches Symbolsystem konstruiert werden, das intelligentes Verhalten zeigt (den Turingtest besteht).

Matt Ginsberg: Ziel der Künstlichen Intelligenz: Konstruktion eines physikalischen Symbolsystems, das zuverlässig den Turingtest besteht.

Hypothese des Konnektionismus

Man benötigt subsymbolische, verteilte, parallele Verarbeitung, um intelligente Maschine zu konstruieren

- Implikation: Man benötigt künstliche neuronale Netze
- Gegenargument: Man kann künstliche neuronale Netze auch (als Software) auf normaler Hardware programmieren

KI-Paradigmen

Zwei wesentliche Paradigmen

- **Physikalisches Symbolsystem**
 - explizites Programmieren
 - verwenden von Logiken, Schlussregeln, Inferenzverfahren
 - Stärken: Ziehen von Schlüssen, strategische Spiele, Planen, . . .
- **Lernverfahren**
 - insbesondere durch künstliche neuronale Netze
 - Stärken: Bildererkennung, Musterverarbeitung, verrauschte Daten, maschinelles Lernen, adaptive Systeme

Komplexes KI-System benötigt i.A. **alle** Paradigmen

Analyse und Programmieren von Teilaspekten

Arbeitshypothese:

- Untersuche und programmiere Teilaspekte an kleinen Beispielen.
- Danach: Untersuche große Beispiele und kombiniere die Lösungen der Teilaspekte

Problematiken:

- Kombination oft nicht möglich oder ineffizient
- Lösung für kleine Beispiele (Mikrowelten) muss nicht funktionieren für realistische Beispiele (Makrowelten)

Wissensrepräsentationssysteme

Wissensrepräsentationshypothese (Brian Smith)

Die Verarbeitung von Wissen läßt sich trennen in:

- Repräsentation von Wissen, wobei dieses Wissen eine Entsprechung in der realen Welt hat
- Inferenzmechanismus, der Schlüsse daraus zieht.

⇒ Basis für Programme, deren innere Struktur als Modellierung von Fakten, Wissen, Beziehungen und als Operationen, Simulationen verstanden werden kann.

Komponenten:

- 1 **Formale Sprache:** Festlegung der gültigen syntaktischen Formen (Wissensbasis, Anfragen)
- 2 **Semantik:** Bedeutung der Sätze der formalen Sprache (i.A. modular aufgebaut)
- 3 **Inferenz-Prozedur** (operationale Semantik) Wie kann man Schlüsse ziehen?
Diese Inferenzen müssen **korrekt** bzgl. der Semantik sein.

Implementierung:

- Parser für die formale Sprache
- Implementierung der Inferenzprozedur.

Geschichte der KI (1)

1950 A. Turing: Imitationsspiel

1956 J. McCarthy Dartmouth Konferenz: Start des Gebietes
"artificial intelligence" und Formulierung der Ziele.

1957- 1962 "allgemeiner Problemlöser"

LISP (higher-order, Parser u. Interpreter zur Laufzeit)
Dameprogramm von A. Samuels (adaptierend "lernend")
Newell & Simon: GPS (General Problem Solver)

1963-1967 spezialisierte Systeme semantische Verarbeitung
(Benutzung von Spezialwissen über das Gebiet)
Common Sense Knowledge Problem
Resolutionsprinzip im automatischen Beweisen

Geschichte der KI (2)

1968- 1972 Modellierung von Mikrowelten

- MACSYMA mathematische Formel-Manipulation
- DENDRAL Expertensystem zur Bestimmung von organischen Verbindungen mittels Massenspektroskopie
- SHRDLU, ELIZA: erste Versuche zur Verarbeitung natürlicher Sprache

1972-1977 Wissensverarbeitung als Technologie

- medizinisches Expertensystem: MYCIN
- D. Lenats AM (automated mathematics)
- KI-Sprachentwicklungen: PLANNER, KRL PROLOG
KL-ONE

Geschichte der KI (3)

1977- Entwicklung von Werkzeugen Trend zur Erarbeitung von Grundlagen;

- Expertensystem-Schalen (E-MYCIN)
- Fifth generation project in Japan (beendet)
- Computerlinguistik
- künstliche neuronale Netze
- logisches Programmieren
- 1985 Doug Lenats CYC

Erzeugung eines vernünftigen Programms mit Alltagswissen durch Eingabe einer Enzyklopädie

Geschichte der KI (4)

Aktuelle Forschungsrichtungen

- Technologie zum textuellen Sprachverstehen und zur Ermöglichung von Mensch-Computer-Dialogen. (Wahlster: Verbmobil-Projekt)
- Robotik-Ansatz: (Embodied artificial intelligence, R. Pfeifer). Untersuchung und Entwicklung von Sensorik, Motorik, d.h. physikalische Gesetzmäßigkeiten des Roboters und der Steuerung mittels Computer, „intelligentes Insekt“
- Automatische Deduktion
- Logik
- Robotik
- künstliche neuronale Netze
- ...

Fazit der bisherigen Erfahrungen

- Fernziel scheint mit aktuellen Methoden, Techniken, Hardware und Software nicht erreichbar
- Motivationen und Visionen der KI sind in der Informatik verbreitet.
- Direkte Programmierung von Systemen (ohne sehr gute Lernalgorithmen) hat Grenzen in der Komplexität des einzugebenden Modells (Regeln, Fakten, ...)

Fazit der bisherigen Erfahrungen (2)

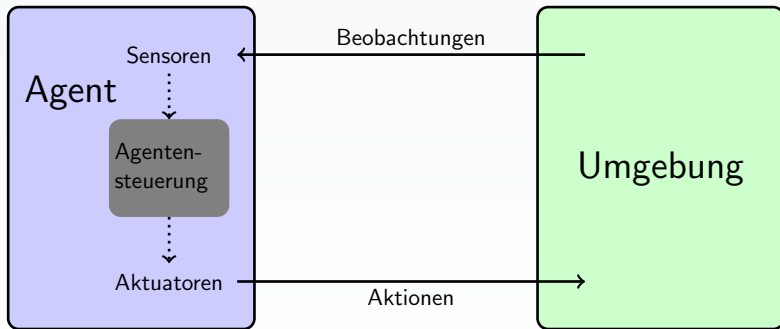
- Teilbereiche sind eigenständigen Forschungsgebiete: Z.B. Sprachverarbeitung, Robotik, Automatische Deduktion, . . .
- Aktuelle Forschungsziele sind eher spezialisierte Aspekte: Logik und Inferenzen, Mensch-Maschine-Kommunikation, Lernverfahren (adaptive Software), Repräsentationsmechanismen; eingebettete KI, nicht-Standard-Logiken und Schlussfolgerungssysteme,

Intelligente Agenten

Agent = Überbegriff für alle KI-Systeme

Ein Agent hat

- *Sensoren* zum Beobachten seiner Umgebung und
- *Aktuatoren* (Aktoren; Effektoren) um die Umgebung zu manipulieren.



Intelligente Agenten (2)

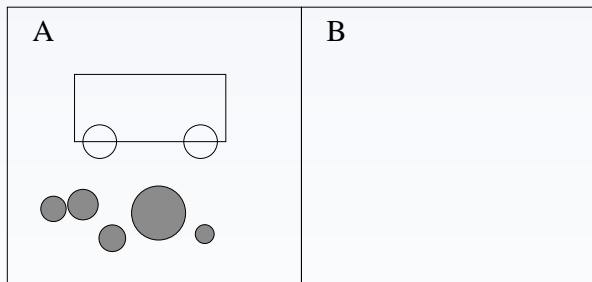
Agent

- macht Beobachtungen (Folge = Beobachtungssequenz)
- Aktion beeinflussen Umgebung und evtl. ihn selbst (z.B. Position)

Agentenfunktion: $\{\text{Beobachtungsfolgen}\} \rightarrow \{\text{Aktionen}\}$.

Agentfunktion kann durch das **Agentenprogramm** implementiert werden

Beispiel: Staubsaugerwelt (Russel & Norvig)



- Orte: *A* oder *B*
- Jeder Ort: Dreckig / Sauber
- Agent kann nur aktuellen Ort beobachten (Sauber/Dreckig)
- Aktionen: `InsAndereQuadrat`, `Saugen` und `NichtsTun`.

Beispiel: Staubsaugerwelt (Russel & Norvig) (2)

Problem:

Wann ist der Agent (das zugehörige Programm)
gut / vernünftig bzw. intelligent ?

Notwendig:

Performanzmaß, d.h. eine **Leistungsbewertung** des Agenten

Z.B.

- Alles immer maximal sauber
- Möglichst sauber, aber wenig Stromverbrauch
- Möglichst sauber, aber wenig störend
- ...

Der optimale agierende Agent ist der **intelligente** Agent.

Intelligenter Agent

Definition

Ein **vernünftiger (intelligenter, rationaler) Agent** ist derjenige, der stets die optimale Aktion bzgl des Performanzmaßes wählt, aufgrund seiner Beobachtungsfolge und seines Vorwissens über die Umgebung.

Guter Agent:

- mittels der Sensoren Wissen über die Umgebung sammeln;
- lernfähig, bzw. sich adaptiv zu verhalten, aufgrund der Beobachtungssequenz

Agent wird als *autonom* bezeichnet, wenn der Agent eher aus seinen Beobachtungen lernt und nicht auf vorprogrammierte Aktionen angewiesen ist.

Klassifikationen

- **Vollständig beobachtbar** vs. **teilweise beobachtbar**.
Der Staubsauger kann z.B. nur sein eigenes Quadrat beobachten.
- **Deterministisch** vs. **Stochastisch**
Der Dreck erscheint zufällig in den Quadraten.
- **Episodisch** vs. **sequentiell**
Episodisch: Feste unabhängige Zeitabschnitte,
in denen beobachtet agiert wird
Sequentiell. Es gibt Spätfolgen der Aktionen.

Umgebungen (2)

Klassifikationen

- **Statisch** vs. **Dynamisch**

Dynamisch: die Umgebung kann sich **während der Nachdenkzeit** des Agenten verändern.

Semi-Dynamisch: Umgebung unverändert,
aber Performanzmaß verändert sich
Z.B.: Schachspielen mit Uhr.

- **Diskret** vs. **Stetig**.

- **Ein Agent** oder **Multiagenten**.

Bei Multiagenten: Klassifizierung in:

Gegner / Wettbewerber / Kooperierenden Agenten.

Umgebungen, Beispiele (Russel & Norvig)

Arbeits- umgebung	Beobacht- bar	Deter- min- istisch	Episo- disch	Statisch	Diskret	Agenten
Kreuzwortsrätsel	vollst.	det.	seq.	statisch	diskret	1
Schach mit Uhr	vollst.	det.	seq.	semi	diskret	n
Poker	teilw.	stoch.	seq.	statisch	diskret	n
Backgammon	vollst.	stoch.	seq.	statisch	diskret	n
Taxifahren	teilw.	stoch.	seq.	dyn.	stetig	n
Medizinische Diagnose	teilw.	stoch.	seq.	dyn.	stetig	1
Bildanalyse	vollst.	det.	episod.	semi	stetig	1
Interaktiver Englischlehrer	teilw.	stoch.	seq.	dyn.	diskret	n

Struktur des Agenten

Agent besteht aus:

- **Architektur des Agenten:**
Physikalischer Aufbau inkl. Sensoren und Aktuatoren
- Programm des Agenten

Arten von Agenten (1)

- **Tabellengesteuerter Agent** (endlicher Automat) mit einer Tabelle von Einträgen:
Beobachtungsfolge1 \mapsto Aktion1
Beobachtungsfolge2 \mapsto Aktion2
- leicht abgeändert: **Agent mit Zustand**
implementiert Funktion der Form
(Zustand, Beobachtung) \mapsto Aktion, Zustand'

Arten von Agenten (2)

Einfacher Reflexagent

- Tabellengesteuert (endlicher Automat) mit Tabelle von Einträgen:
 - Beobachtung1 \mapsto Aktion1
 - Beobachtung2 \mapsto Aktion2
- Wählt Aktion aufgrund der direkten Wahrnehmung

Arten von Agenten (3)

Modellbasierte Strukturierung:

- Zustandsbasiert
- Zustand modelliert die Umgebung
- D.h. Wahrnehmung werden in Modell übertragen
- Berechnungen auf dem Modell: Z.B. Was bewirken Aktionen?

Arten von Agenten (3)

Zweckbasierte Strukturierung (goalbased, zielbasiert):

- Aktionen steuern auf das (vorgebene) Ziel zu.
- Z.B. Taxi kann anhand des Ziels entscheiden ob links oder recht abbiegen
- Agent kann erkennen, ob er das Ziel erreicht hat.

Arten von Agenten (4)

Nutzenbasierte Strukturierung (utility-based, nutzenbasiert):

- Nächste Aktion wird anhand einer **Nutzenfunktion** (internes Gütemaß) berechnet.
- benötigt vorgegebene **Bewertungsfunktion (utility function)**

Arten von Agenten (5)

Lernende Agenten:

- Lernverfahren erhöhen die Autonomie des Agenten.
- Ausführungsmodul: bisheriges Programm
- Lernmodul: verwendet Beobachtungen um Parameter im Programm anzupassen
- Lernmethoden: online, offline (mit Trainingsphase)