

Künstliche Intelligenz

Einführung

Dr.-Ing. Stefan Lüdtkke

Universität Leipzig

Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI)

Leipzig



Ziel der Lehrveranstaltung

- Künstliche Intelligenz: Teilbereich der Informatik, der sich mit der Erstellung *intelligenter Agenten* beschäftigt
- Angrenzende Fachgebiete: Neurowissenschaften, Psychologie, Linguistik, ...
 - Diese beschäftigen sich mit der Erforschung menschlicher Intelligenz
 - Uns geht es stattdessen um *rationale Agenten*

Was ist Intelligenz?

An agent is intelligent to the extent that what it does is likely to achieve what it wants, given what it has perceived”

(Stuart Russell, Human Compatible: AI and the Problem of Control)

Literatur und Inhalte



Stuart Russel und Peter Norvig

Artificial Intelligence – A Modern Approach (AIMA), 3. Ausgabe

Prentice Hall, 2009

aima.cs.berkeley.edu

Inhalte (Plan)

- I Einführung – Intelligente Agenten
- II Problemlösung durch Suchen
- III Schlussfolgern unter Unsicherheit
- IV Maschinelles Lernen
- V Decision Theory

Organisatorisches

- Lehrveranstaltungen vom 27.02. bis 14.03.
- Voraussetzung für die Teilnahme an der Klausur: Erfolgreiche Teilnahme am *Projekt*
 - Weitere Infos dazu nächste Woche
- Vorlesungsmaterialien: stefanluedtke.github.io/teaching

Zeitplan, voraussichtlich

	9 - 11	11 - 13	13 - 15	15 - 17
Mo 27.02.	V1 Einführung V2 Agenten	U1 Python	U1 Python	
Di 28.02.	V3 Suche uninformiert	V4 Suche	U2 Agenten	V5 Suche: CSPs
Mi 01.03.	U3 Suche, CSPs	V6 Game Playing	U3 Suche, CSPs	
Do 02.03.	V7 Planung	U4 Planung	V8 Wahrscheinlichkeitsrechnung	U5.1 Wahrscheinlichkeitsrechnung
Fr 03.03.	V9 Bayesian Networks	V10 Prob. Inferenz	U5.2 Inferenz	
Mo 06.03.	V11 HMMs	U5.3 HMMs	V12 ML Intro, Clustering	
Di 07.03.	V13 Entscheidungsbäume	U6.1 Entscheidungsbäume	V14 ANNs	
Mi 08.03.	V15 ML: Backprop	V16 ML Evaluation	U6.2 ML Evaluation	Projekt
Do 09.03.	V17 MDPs	U7 MDPs	Projekt	Projekt
Fr 10.03.	V18 Q-Learning	V19 Spieltheorie	Projekt	
Mo 13.03.	Projekt	Projekt	Projekt	Projekt
Di 14.03.	Projekt	Projekt	Projekt	Projekt

Was ist Intelligenz?

Herangehensweisen, Intelligenz zu definieren:

- äußerlich sichtbares Verhalten vs. innere Denkprozesse
- Menschliche Intelligenz vs. rationales Verhalten

Was ist künstliche Intelligenz?

Welches Systemverhalten erfüllt das Kriterium „Intelligent“?

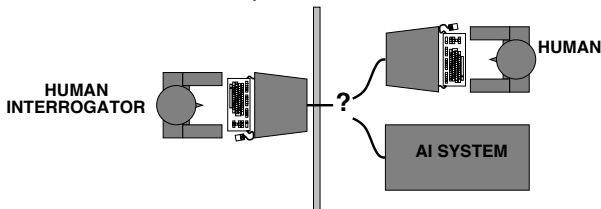
Menschlich Denken	Rational Denken
„The exciting new effort to make computers think ... <i>machines with minds</i> , in the full and literal sense.“ (Haugeland 1984) „[The automation of] activities that we associate with human thinking, activities such as decision making, problem solving, learning ...“ (Bellman, 1978)	„The study of mental faculties through the use of computational models.“ (Charniak & McDermott, 1985) „The study of computations that make it possible to perceive, reason, and act.“ (Winston, 1992)
Menschlich Handeln	Rational Handeln
„The art of creating machines that perform functions that require intelligence when performed by people.“ (Kurzweil, 1990) „The study of how to make computers do things at which, at the moment, people are better.“ (Rich & Knight, 1991)	„Computational intelligence is the study of the design of intelligent agents.“ (Poole et al., 1998) „AI ... is concerned with intelligent behavior in artifacts.“ (Nilsson, 1998)

- Als nächstes: Kurzer Einblick in jede der 4 Sichtweisen

Menschlich Handeln

Turing (1950) „Computing machinery and intelligence“

- „Können Maschinen denken?“ → „Können sich Maschinen intelligent verhalten?“
- Praktischer Test: Imitationsspiel



- Vorhersage, dass in 2000 eine Maschine eine 30% Chance haben könnte, einen Laien für 5 Minuten von ihrer Menschlichkeit zu überzeugen
- Diskutiert bereits alle wesentlicher Argumente gegen KI, die in den folgenden 50 Jahren aufgebracht werden
- Identifiziert bereits wesentliche Komponenten von KI-Systemen: Wissen, Schließen, Sprachverständnis, Lernfähigkeit

Menschlich Handeln

Turing (1950) „Computing machinery and intelligence“

Problematik des Turing-Tests:

- Nicht reproduzierbar
- Nicht konstruktiv
- Nicht mathematisch analysierbar

Menschlich Denken

Kognitions- und Neurowissenschaften

- 1960er: „Kognitive Revolution“: Psychologie der Informationsverarbeitung ersetzt den vorherrschenden orthodoxen Behaviorismus
- Erfordert wissenschaftliche Theorien der internen Prozesse im Gehirn
 - Welches Abstraktionsniveau? – „Wissen“? „Schaltkreise“?
 - Wie validieren? – Verhaltenstests (top down)? Direkte Analyse neurologischer Daten (bottom up)?
- Beide Ansätze (Kognitionswissenschaften und Neurowissenschaften) jetzt von KI getrennt
- Siehe auch: *Human Brain Project* – Ziel unter anderem: Nachbildung (von Teilen) des Gehirns durch Computermodelle. Finanziert durch die Europäische Union (1,19 Mrd. Euro)

Rationales Denken

Grundregeln des Schließens

- Schlussssysteme, die Regeln für korrekte Inferenz festlegen
- Durch syntaktische Schlussregeln und formal präzise Definition wird maschinelles Schließen möglich
- Normativ statt beschreibend
- Seit Aristoteles: Was sind „korrekte“ Argumentations- und Denkprozesse?
- von hier direkte Linie über Mathematik und Philosophie zur KI
- Probleme:
 - Formalisierung von Alltagswissen („Common Sense“)
 - Behandlung von Unsicherheit
 - Nicht jede „intelligente Handlung“ erfordert einen mathematischen Beweis

Rationales Handeln

Das Richtige tun

- Rationales Verhalten = das Richtige tun
- Das Richtige = das, was die erwartete Zielerreichung maximiert, gegeben die verfügbaren Informationen
- Dabei irrelevant, ob man „viel denken muss“, um das Richtige herauszufinden
- Charakterisierung von „Intelligenz“ auf Basis der *Wirkung* in Bezug auf ein objektives Kriterium (Rationalität)
- Perspektive dieser Vorlesung

Rationale Agenten

- Ein *Agent* ist eine Entität, die ihre Umgebung wahrnimmt und in dieser Umgebung handelt
- In dieser Vorlesung diskutieren wir den Entwurf von *rationalen Agenten*
- Abstrakt ist ein Agent eine Funktion, die Beobachtungssequenzen auf Aktionen abbildet:

Agent: [Percept] \rightarrow Action

- Für beliebige Umgebungen und Aufgaben suchen wir dann diejenigen Agenten, die die beste Leistung erreichen
- Problem: Aufgrund von Einschränkungen in Speicherplatz und Rechenleistung wird „perfekte Rationalität“ nicht immer erreichbar sein
 - Wir suchen somit das beste Programm für die gegebenen Rechnerressourcen – „bounded rationality“

Starke vs. Schwache KI

- Schwache KI: Systeme, die in bestimmten Teilbereichen menschliche Leistungsfähigkeit erreichen, z.B. Schach
 - KI-Forschung beschäftigt sich fast ausschließlich mit Erstellung von schwacher KI
- Starke KI (Artificial General Intelligence, AGI): KI, die alle intellektuellen Aufgaben erlernen oder verstehen kann, die ein Mensch auführen kann
 - Fernziel der KI-Forschung
 - Unpräzise Definition

Aktuelle AGI-Debatte

- AGI unpräzise definiert (können ja nicht mal Intelligenz definieren...)
- Ergebnisorientierte Definition von Francois Chollet: “We will know we have AGI when the majority of the world’s GDP is being produced by autonomous AI agents.”
- Aktuelle Debatte, wie lange wir noch von einer AGI entfernt sind
- Eure Prognose zu AGI?

Aktuelle AGI-Debatte

- AGI unpräzise definiert (können ja nicht mal Intelligenz definieren...)
- Ergebnisorientierte Definition von Francois Chollet: “We will know we have AGI when the majority of the world’s GDP is being produced by autonomous AI agents.”
- Aktuelle Debatte, wie lange wir noch von einer AGI entfernt sind
- Eure Prognose zu AGI?
- Prognosen reichen von “AGI in 8 Jahren” bis zu “nicht in den nächsten 50 Jahren”
- Meine Prognose
 - aktuelle KI-Modelle (Language Models, Vision Models) als nützliche Tools, aber Skalierung nicht ausreichend für AGI
 - 10% Wahrscheinlichkeit, für AGI in 15 Jahren
 - 30% Wahrscheinlichkeit für AGI in 30 Jahren

Zusammenfassung

- Ziel des Forschungsbereichs Künstliche Intelligenz ist die Erstellung rationaler Agenten
- Dafür kommen unter anderem Methoden der Logik, Suchverfahren, Statistik, Machine Learning und Decision Theory zum Einsatz, die wir uns in dieser Vorlesung anschauen werden