Agenten und Multiagentensysteme

Seminar: Verteilte Prozesse

Wilfried Teiken
wteiken@cl-ki.uni-osnabrueck.de

28. Oct 1999

Agenten

Unter Agenten werden im folgenden Systeme verstanden, die folgende Eigentschaften erfüllen:

- *Autonomie*, d. h. die Agenten verfolgen kein vorgegebenes Programm sondern haben zur Erfüllung ihrer Aufgaben einen gewissen Spielraum. Außerdem könne die Agenten eigenständig *aktiv* werden.

- *Rationalität*, d. h. sie verfolgen ihre Ziele, indem sie versuchen die zu lösenden Aufgaben zielgerichtet abzuarbeiten.

- *Reaktivität*, d. h. der Agent muß auf äußere Einflüsse reagieren können. Dies betrifft sowohl die Notwendigkeit einer *Sensorik* als auch der *Echtzeitfähigkeit*.

Agenten nach Russel und Norvig

„An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors.“

„For each possible percept sequence, an ideal rational agent should do whatever action is expected to maximize its performance measure, on a basis of the evidence provided by the percept sequence and whatever built-in knowledge the agent has.“

Verteilte KI vs. Multiagenten

Ein *Verteiltes KI-System* ist ein System, das in seiner Gesamtheit geplant wird um eine bestimmte Aufgabe zu lösen.

$\Rightarrow$ klare starr modellierte Aufgabenverteilung
effizienter Ablauf

Ein *Multiagentensystem* ist ein System aus (einzeln geplanten) Komponenten, die in ihrer Gesamtheit in der Lage sind, ein Problem gemeinsam zu lösen.

$\Rightarrow$ dynamische Aufgabenverteilung
wechselnde Kompetenzen
evtl. ineffizienter Ablauf

Vorteile von Agentensystemen

- *Robustheit* gegenüber äusseren Einflüssen, d. h. Agenten müssen auf eine große Bandbreite von möglichen *Störquellen*, also nichtantizipierten Ereignissen, reagieren können;

- *Flexibilität*, d. h. ein System kann durch die Kombination verschiedener bestehender Agenten oder die Integration neuer Agenten an eine Aufgabe angepaßt werden.

Die Robustheit eines Agentensystems wird durch die klare Trennung der verschiedenen Agenten mit strikt definierten Schnittstellen erreicht. *Lokale Ziele* garantieren dabei die Einhaltung der Schnittstellen.

Das BDI-Modell[1]

Beliefs sind das *Weltmodell* des Agenten, d. h. sie stellen die Annahmen (bzw. das Wissen) des Agenten dar;

→ Auf den *Beliefs* arbeiten i. d. R. Inferenzprozesse, die das Verhalten des Agenten bestimmen.

Desires sind die *Ziele* des Agenten, d. h. die Ausgaben, die der Agent versucht zu lösen;

→ Die *Desires* leiten den Inferenzprozess des Agenten an, sie bilden die (momentanen) Top-Level-Ziele des Agenten.

[1] Intention, Plans, and Practical Reason, Bratman 1987

Intentions

- *Intentions* sind die (partiellen) *Pläne*, die der Agent im Moment hat, d. h. sie repräsentieren die konkreten Ziele des Agenten.

- Die *Intentions* bilden das Konzept der Handlungsabsicht, d. h. sie haben folgende Eigenschaften:
  - sie sind oft unvollständig und werden hierarchisch verfeinert;
  - sie sind konsistent zu den *Beliefs*;
  - sie besitzen eine *Trägheit*, die die Neuplanung einschränkt;
  - sie schränken den Alternativenraum ein;

  ⇒ *Intentions* erlauben sowohl eine Suchraumbegrenzung als auch die Möglichkeit zur *verläßlichen* Planung.

Rationalität bei BDI-Agenten

- Ein *BDI*-Agent trifft aus interner Sicht rationale Entscheidungen, d. h.:
  - sie sind aus externer Sicht rational, d. h. sie stellen eine optimale Handlung bzgl. der aktuellen *Goals* und der *Beliefs* dar, oder
  - es gibt zwar aus externer Sicht bessere Optionen, aber diese sind inkompatibel zu den *Intentions* des Agenten. Dies kann durch den Aufwand eine Intention aufzugeben durchaus rational Begründbar sein.

- *Intentions* sind *pro-attitudes*, d. h. sie haben Einfluß auf die Motivation des Agenten und haben damit auch Einfluß auf das Verständnis von *Rationalität*

Agenten als Planer

- *klassische* Planer ($STRIPS$, $NOAH$) planen eine Handlung komplett, bevor sie sie ausführen
  $\Rightarrow$ inadäquat bei *reellen* Szenarios, da eine *reale* Welt nicht statisch ist

- *reaktive* Syteme ($PENGI$) versuchen eine Schichtenarchitektur, sind aber in ihren Plänen oft nicht optimal

- *hybride* Syteme ($IRMA$) kombinieren klassische Planung mit reaktiven Ansätzen. Pläne könne unvollständig sein und korrigiert werden.

Ein einfacher Planer

Beliefs $\neg$Painted$(Door)$, $\neg$Painted$(Ladder)$, Dry$(Ladder)$, Handempty

Goals: Painted$(Door)$, Painted$(Ladder)$

Operator	Precondition	Effect	Cost
`get-brush`	Handempty, Dry(b)	Havebrush(b), $\neg$Handempty	1.0
`paint-ceiling`	Havebrush(b), Dry$(Ladder)$, Havepaint$(Paint)$	$\neg$Dry$(Ceiling)$, $\neg$Dry(b), Painted$(Ceiling)$	2.0
`paint-ladder`	Havebrush(b), Havepaint$(Paint)$	$\neg$Dry$(Ladder)$, $\neg$Dry(b), Painted$(Ladder)$	3.0
`get-brush`	Havebrush(b)	$\neg$Havebrush(b), Handempty	1.0

Ein hierarchischer Planer

Operator	Precondition	Effect	Cost
`movelarge`(x,y)	Ispeg(x), Ispeg(y), Onlarge(x), $\neg$Onmed(x), $\neg$Onmed(y) $\neg$Onsmall(x), $\neg$Onsmall(y)	Onlarge(y), $\neg$Onlarge(x)	1.0
`movemed`(x,y)	Ispeg(x), Ispeg(y), Onmed(x), $\neg$Onsmall(x), $\neg$Onsmall(y)	Onmed(y), $\neg$Onmed(x)	1.0
`movesmall`(x,y)	Ispeg(x), Ispeg(y), Onsmall(x),	Onsmall(y), $\neg$Onsmall(x)	1.0
`Abs`(2, `movelarge`(x,y))	Ispeg(x), Ispeg(y),	Onlarge(y), $\neg$Onlarge(x)	1.0
`Abs`(2, `movemed`(x,y))	Ispeg(x), Ispeg(y), Onmed(x),	Onmed(y), $\neg$Onmed(x)	1.0
`Abs`(2, `movesmall`(x,y))	Ispeg(x), Ispeg(y), Onsmall(x),		

Ein hierarchischer Planer II

$$\begin{aligned}
\pi &= \mathsf{S}_{init} \rightarrow \mathsf{movelarge}(P_1, P_3) \rightarrow \mathsf{S}_{finish} \\
&\Downarrow \\
\pi &= \mathsf{S}_{init} \rightarrow \mathsf{moveM}(P_1, P_2) \rightarrow \mathsf{movelarge}(P_1, P_3) \rightarrow \mathsf{moveM}(P_2, P_3) \rightarrow \mathsf{S}_{finish} \\
&\Downarrow \\
&\vdots \\
&\Downarrow \\
\pi &= \mathsf{S}_{init} \rightarrow \mathsf{moveS}(P_1, P_3) \rightarrow \mathsf{moveM}(P_1, P_2) \rightarrow \mathsf{moveS}(P_3, P_2) \rightarrow \\
& \quad \mathsf{moveL}(P_1, P_3) \rightarrow \\
& \quad \mathsf{moveS}(P_2, P_1) \rightarrow \mathsf{moveM}(P_2, P_3) \rightarrow \mathsf{moveS}(P_1, P_3) \rightarrow \mathsf{S}_{finish}
\end{aligned}$$

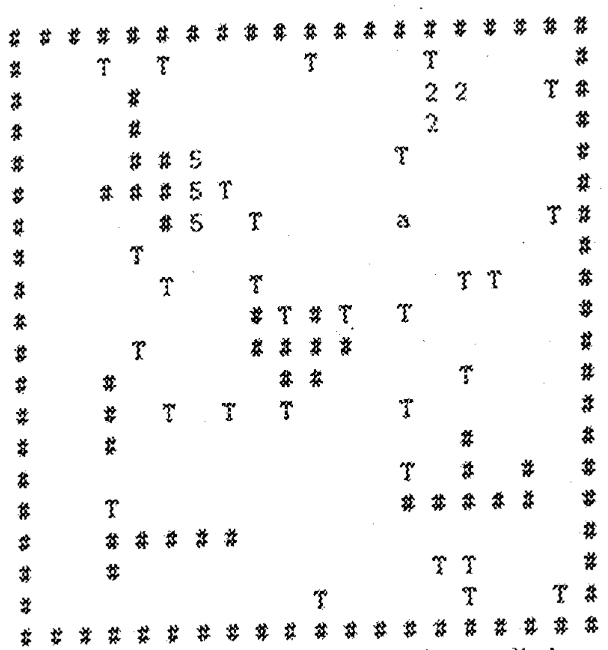
IRMA

- *IRMA* ist die Intelligent Resource-Bounded Machine Architecture, ein generischer Versuch eine *BDI*-Architektur aufzustellen.

- *IRMA*-Agenten wurden u. a. von *Martha E. Pollack* und *Marc Ringuette* tatsächlich implementiert.

- Die *IRMA*-Agenten sollen überprüfen, inwieweit sich die Annahmen Bratmans und anderer, die das *BDI*-Modell postuliert, in experimenten nachweisen lassen.

Die Tileworld



a = agent, # = obstacle, T = tile, < digits > = hole

- dynamische Welt aus Blöcken, Hindernissen, Löchern und Agenten
- Agenten können Blöcke schieben, Hindernisse sind unbeweglich, Löcher nehmen Bllocke auf
- gefüllte Blöcke geben dem Agenten Punkte, je grösser das Loch, desto mehr Punkte
- Objekte erscheinen und verschwinden in gewissen Abständen.

Eigenschaften der Tileworld

- durch die Dynamik ist statische Planung unmöglich
- durch die zumindest zeitweilige Konstanz ist rein reaktives Verhalten ineffektiv
- längerfristige Planungen und reaktives Verhalten müssen gegeneinander abgewogen werden, Parameter zur Evaluation sind:
  - die Frequenz der Weltänderungen und die Verteilungsparameter
  - die Art der Veränderung (stetig oder abrupt)
  - den Modus zur Berechnung des Filter-Schwellwertes
  - durch die Auswahlstrategie des zugrundeliegenden Planers

Die IRMA-Architektur

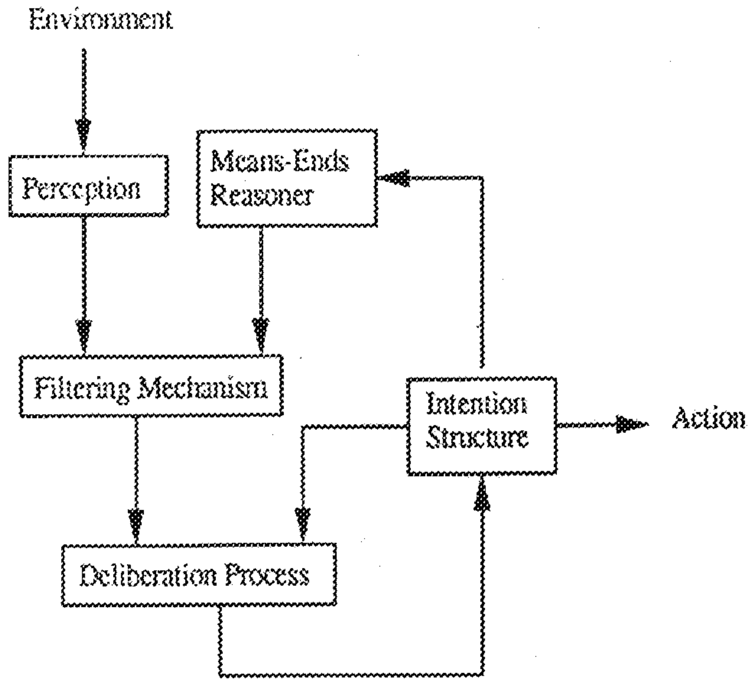


Figure 2: Tileworld Agent Architecture

- *act cycle*:
  - arbeitet auf *Intention Structure*
  - übernimmt die Wahrnehmung
- *reasoning cylce*:
  - fertigt Pläne an
  - integriert eventuell neue Optionen
- *Deliberation Process*
  - Erzeugt die *Intention Structure*
  - erhält gerfilterte Wahrnehmung

Auswertungsstrategien

Es gilt: Jedes Loch $h \in \mathcal{H}$ mit $\mathsf{score}(h) - \mathsf{score}(h_c) \geq v$, wobei h_c das gerade gewählte Loch ist, passiert den Filter. v ist der Filterschwellwert.

Simple deliberation nimmt aus der Menge aller Optionen die mit der höchsten erreichbaren Punktzahl.

LV bildet den Quotienten aus Nutzen und Aufwand, um die lohnenswerteste Option zu finden:

$$\mathsf{LV}(h) = \frac{\mathsf{score}(h)}{\mathsf{dist}(a, h) + \sum_{i=1}^{n} 2\mathsf{dist}(h, t_i)}$$

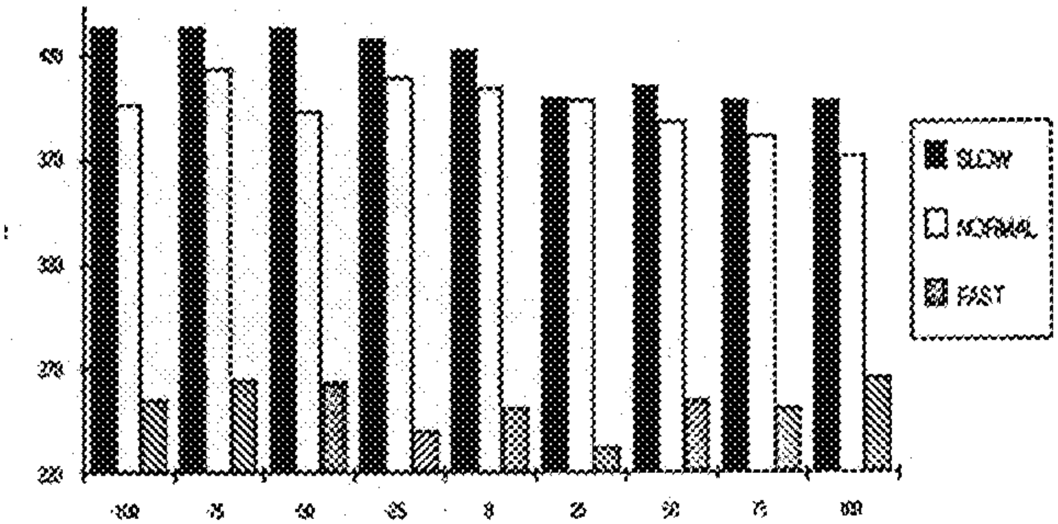
IRMA-Experiment 1

Einfacher Test, in der der Agent in der Welt Punkte sammeln soll:

- keine Filterung
- 5000 Zyklen
- normale oder 10$\times$ Geschwindigkeit bei konstanter Panungsgeschwindigkeit

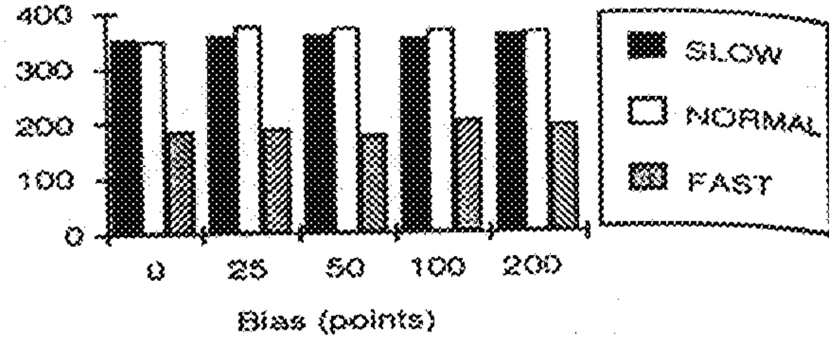
	Score for Agent				
Speed	LV	LV/Serial	Simple	Simple/Serial	Human
Normal	396	353	347	291	468
Fast	256	234	183	152	3

IRMA-Experiment 2

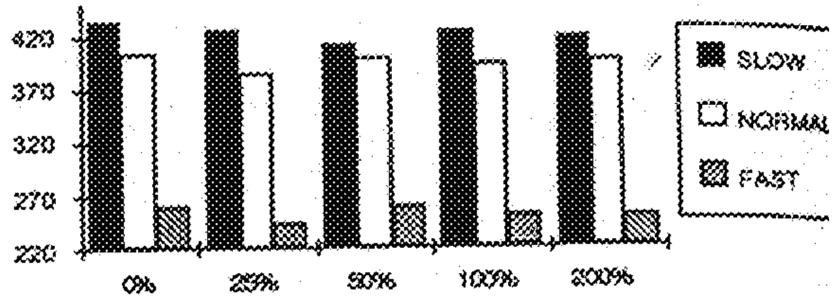


- Nur LV-Evaluator
- Filterschwellwert v von -100 bis $+100$
- Geschwindigkeiten $\frac{1}{100}$, normal und $10\times$

IRMA-Experiment 3



Experiment #3: Simple Evaluator



- Simple- und LV-Evaluator
- kein Filtering
- *Bias* bei der Zielwahl:

  Simple: existierende Pläne erhalten Bonus

  LV: neue Pläne haben Nachteil in Prozent des aktuellen LV(h_c)

Ergebnisse ?

Keine klaren Ergebisse bzgl. der Relevanz der *Intentions*, die Ursachen sind:

- falsches Verhältnis von Planungszeit und Ausführungszeit
- schwierigkeiten bei der Analyse der Rechenzeiten
- mangende Varianz der Parameter
- falsches *Modell* der Wahrnehmung

Multiagentensysteme

Die Kombination mehrerer Agenten zu einem Gesamtsystem ergibt ein Multiagentensystem. Dies ergibt mehrere Probleme:

- die Agenten müssen aufeinander reagieren können;
- die Agenten müssen kommunizieren können;
- die Agenten müssen sich koordinieren können.

Kommunikation und Koordination?

Zur Kommunikation und Koordination brauchen Agenten

- eine gemeinsame Sprache:
  - Syntax
  - Semantik
- Annahmen über den internen Zustand der *Mitagenten*:
  - Motivation
  - Ziele
  - Verhaltensmuster

Gemeinsame Sprache - KIF?

- Einfache formale Syntax (ASCII-Basiertes Format).
- Formale, modelltheoretische Syntax.
- Untersprache für Objekte, n-äre Funktionen und Relationen zur Beschreibung von Domänenontologien.
- Beschreibung der Eigenschaften von Funktionen und Relationen.
- Möglichkeit Metawissen auszudrücken.
- Erweiterbar.

$\Rightarrow$ Interlingua ?

Probleme bei KIF

- Bidirektionale Übersetzbarkeit der Anwendungssprachen nicht immer möglich.
- Probleme bei unerkannter logischer Äquivalenz verschiedener Ausdrücke.
- Unverständlichkeit der übersetzten Strukturen.
- Keine vollautomatische Übersetzung.
- Kein Protokoll zur Klärung der Terminologie.
- Keine *Transportschicht* definiert (KQML ist aber nutzbar).

Ontologien und ihre Probleme

Ontologien sind *Beschreibungen der bekannten Konzepte und ihrer gundlegenden Relationen zueinander*.

- Ontologien sind meist Anwendungsspezifisch.
- Die formale Deklaration der Konzepte ist meist unterpezifiziert.
- Konzeptausprägungen können nur unzureichend behandelt werden.
- Es gibt keine Verfahren zum Aufbau von verteilten Ontologien.
- Es ist nicht klar, wie sinnvoll über Ontologien kommuniziert werden kann.

Annahmen über die *Mitagenten*

- Agenten können interne Zustände nicht immer austauschen:
  - Beschränkungen der Kommunikationsfreiheit (Kosten, Dauer).
  - Inkompatible Architekturen.
- Agenten wollen die internen Ziele nicht immer Preisgeben:
  - Beschränkte Kooperationsbereitschaft.
  - Erhaltung der Entscheidungsfreiheit/Modularität.

$\Rightarrow$ Agenten müssen vom Verhalten auf die Ziele anderer Agenten schliessen.

Beobachtungen und Schlussfolgerungen

- *Top-Level-Goals* der anderen Agenten sind oft bekannt.
- Simulation der Planung anderer Agenten.
- Abgleich der simulierten Planung mit den tatsächlichen Beobachtungen.

Vorraussetzungen:

- Operatoren/Planungsmechanismen der anderen Agenten sind hinreichend bekannt.
- Fremde Agenten folgen dem Rationalitätsprinzip.

Konfliktvermeidung und Konfliktlösung

- Simulierte Planung der Agenten im Umfeld.
- Auswahl von (eigenen) Planungsalternativen, die nicht mit den (vermuteten) Planungen der anderen Agenten kollidieren.
- Explizite Koordination bei interferierenden Planungen:
  - Autausch von Teilaufgaben.
  - Explizite Absprache lokal suboptimaler aber global optimaler Pläne.

Koordination

- Problemanalyse
- Kompetenzabschätzung
- Teambildung
- Aufgabenverteilung
- Ressourcenzuteilung
- Konfliktresolution