EPISTEMISCHE LOGIK

Epistemisch relevante Handlungen

Manuel Bremer
University of Düsseldorf, Germany
www.mbph.de

Handlungen

- Handlungen verändern die Wirklichkeit, so dass gegeben eine Handlung auch Wissen bezüglich der Wirklichkeit neu gewonnen bzw. revidiert werden muss.
- Algorithmische Vorgänge und Schritte in Programmen können als ‚Handlungen' (in einem über intentionale Akteure hinaus erweiterten Sinne) aufgefasst werden. Deshalb sollte die Epistemische Logik in der KI Modelle epistemischer Zustände mit Modellen von Programmschritten kombinieren.
- Handlungen überführen Zustände in Zustände.

Verlautbarungen

- Eine besonders epistemisch relevante Gruppe von Handlungen sind Verlautbarungen. In der epistemischen Logik wird (zunächst) angenommen, dass Verlautbarungen nur bezüglich des Wahren stattfinden. In diesem Falle sind Verlautbarungen eine Quelle des Wissens. Sie erzeugen zugleich wechselseitiges Wissen.
- Ein ‚broadcast' in einem Netzwerk wäre ein typisches Beispiel für eine Verlautbarung (dass z.B. eine Netzwerk-ressource frei oder besetzt ist).
- Eine Verlautbarung wird notiert: $!\varphi$

Zustandsübergänge

- Zur Modellierung von Handlungen tritt neben die epistemische Zugänglichkeitsrelation eine Übergangsrelation $H_\alpha(w,w^*)$ zwischen Zuständen/möglichen Welten, verstanden als die von der Handlung/dem Ereignis α verursachte Zustandsänderung. α führt von *w* zu *w**.
- Als Notation wird eine inhaltlich gefüllte Box verwendet: [α], wobei α für eine Handlung steht. Es wird also eine eigene syntaktische Kategorie für Handlungen eingeführt.
- Wenn eine Handlung α, ausgeführt in Zustand w, immer die Konsequenz ψ hat (d.h. ψ wird/bleibt wahr): $w \vDash [\alpha]\psi$

Semantik

- Semantisch: $w \models [\alpha]\psi$ gdw. $\forall w'(H_\alpha(w,w') \supset w' \models \psi)$
- Die Übergänge werden i.d.R. – und weniger problematisch bei technischen Systemen – als deterministisch angenommen: $H\alpha(w,w') \wedge H\alpha(w,w'') \supset w' = w''$.

Die Übergangsrelation H ist also eine *Funktion*.

Dual

- Es lässt sich wie in der Modallogik ein Dual $<\alpha>\psi$ einführen:

$$<\alpha>\psi \stackrel{\text{def}}{=} \neg[\alpha]\neg\psi$$

mit der Wahrheitsbedingung:

$$w \vDash <\alpha>\psi \text{ gdw. } \exists w'(H_\alpha(w,w') \wedge w' \vDash \psi)$$

- Doch insofern H *funktional* ist (bei „[ ]“ also nicht mehrere Welten mit Konsequenzen betrachtet werden) und für unanwendbare Handlungen α in einer Welt w, festgelegt werden kann $H_\alpha(w,w)$, erweitert dies *nicht* die Sprache.

- Theorem:

$$\models [\alpha]\varphi \equiv \langle\alpha\rangle\varphi$$

- Insofern *nur wahrheitsgemäße* Verlautbarungen betrachtet werden (s.u.), ist $\langle !\varphi\rangle\top$ kein Theorem:

$$\not\models \langle !\varphi\rangle\top$$

denn ist φ falsch, ist nach der zugrunde gelegten Semantik auch $\varphi \wedge \top$ falsch, also auch $\langle !\varphi\rangle\top$.

Handlungen in Programmen

- Die Menge der Handlungen in Programmen wird oft angegeben wie folgt:

  act (als Metavariable, wie α) umfasst:

  - elementare Handlungen a_1, a_2 …
  - skip (nichts tun, Fortsetzen mit nächster Handlung)
  - fail (abbrechen)
  - act_1; act_2 (sequentielle Verkettung)
  - $?\varphi$ (Test auf φ)
  - $!\varphi$ (Broadcast von φ)
  - if φ then act_1 else act_2 fi

- while φ do act_1 od

“fi” und “do” sind Endmarken.

• Diese ‚Handlungen‘ kann man direkt als Ereignisse mit der entsprechenden Übergangsinterpretation verstehen.

• Um Handlungen von verschiedenen Subjekten zu unterscheiden, bietet sich ein Operator „$do_x\alpha$“ an: Agent *x* vollzieht Handlung α. (Führt ein Vorgang unabhängig von einem bestimmten Agenten zu einer Konsequenz, kann man den Index und „do“ auch unterdrücken.)

• Die direkten epistemischen Handlungen Test und Broadcast bedürfen auch eigener Axiome (s.u.); zunächst seien die nicht-epistemischen Handlungen betrachtet.

Wissen von Handlungseffekten

- In der epistemischen Logik lässt sich mit diesen Ausdrücken eine zweisortige Logik (als Erweiterung von S4 z.B.) einführen (eine Sorte für Handlungen, eine für Sätze).

- Ein Satz wie „$W_a([do_b\alpha]\varphi)$“ besagt dann: Agent *a* weiß, dass ein Vollziehen der Handlung α durch Agent *b* zum Zustand φ führt.

- Formal $w \vDash W_a([do_b\alpha]\varphi$ gdw. $\forall w‘(R_a(w,w‘) \supset \forall w‘‘(H_{b\alpha}(w’,w’’) \supset w‘‘ \vDash \varphi))$

Semantik der Programmschritte

• Damit die nicht-epistemischen Handlungen die intendierte Interpretation erhalten, müssen an H Bedingungen gestellt werden:

(i) $w \vDash \varphi \rightarrow H_{(\text{if } \varphi \text{ then } \alpha \text{ else } \beta \text{ fi})}(w) = H_{\alpha}(w)$

$w \nvDash \varphi \rightarrow H_{(\text{if } \varphi \text{ then } \alpha \text{ else } \beta \text{ fi})}(w) = H_{\beta}(w)$

(ii) $H_{(\text{while } \varphi \text{ do } \alpha \text{ od})}(w) = H_{(\text{if } \neg\varphi \text{ then skip else fail fi})}((H_{(\text{if } \varphi \text{ then } \alpha \text{ else skip fi})})^{*}(w))$

(iii) $H_{(\alpha;\beta)}(w) = H_{\beta}(H_{\alpha}(w))$

(iv) $H_{\text{skip}}(w) = w$

(v) $(\forall w,w')(\neg H_{\text{fail}}(w,w'))$

Plangemäß

- Eine Handlung verläuft ‘plangemäß’ für einen Agenten, wenn dieser weiß, dass ein Zustand eingetreten ist nach der Handlung, welcher das Resultat ist, das die Handlung seinem Wissen gemäß haben soll:

(W[ ]) $W_i[\alpha]\varphi \supset [\alpha]W_i\varphi$

- Semantisch: $\models W_i[\alpha]\varphi \supset [\alpha]W_i\varphi$ gdw. $\forall w_1,w_2(H_\alpha(w_1,w_2) \supset \forall w_3(R_i(w_2,w_3) \supset \exists w_4(R_i(w_1,w_4) \wedge H_\alpha(w_4,w_3))))$

d.h. kennt Agent *i* ein Resultat der Handlung α, dann kennt *i* einen Zustand, der als Ausgangszustand α‘s diente.

Informativ

- Eine Handlung α ist ‚informativ' für einen Agenten, wenn dieser durch Ausführen von α eine Information φ gewinnt:

$$w \models [do_i \alpha](W_i \varphi \vee W_i \neg\varphi)$$

- Semantisch: $\forall w_1, w_2 (H_{do(i,\alpha)}(w_1, w_2) \supset \forall w_3 (R_i(w_2, w_3) \supset (w_2 \models \varphi \equiv w_3 \models \varphi))$

- Eine plangemäße Handlung, deren Ausführung zu φ führt, ist informativ bezüglich φ.

Echt informativ

- Jede Handlung ist irrelevant ‚informativ‘ bezüglich von logischen wahren Sätzen, da diese (in S4) schon gewusst wurden! Obige Bedingung gilt trivialerweise, da

$$\forall w_2,w_3(R_i(w_2,w_3) \supset (w_2 \vDash \varphi \equiv w_3 \vDash \varphi))$$

wahr ist.

- Eine Handlung ist ‘echt informativ‘ genau dann, wenn sie informativ ist und folgende Zusatzbedingung gilt:

$$\exists w_4(R_i(w_1,w_4) \wedge w_4 \vDash \varphi) \wedge \exists w_5(R_i(w_1,w_5) \wedge w_5 \nvDash \varphi)$$

d.h. φ war vor dem Ausführen von α nicht bekannt.

Theoreme bezüglich von Handlungen

- Auch ohne epistemische Begriffe gelten schon einige rein handlungslogische Theoreme, etwa:

$$\vDash [?\varphi]\psi \equiv (\varphi \supset \psi)$$

oder

$$\vDash [\alpha;\beta]\varphi \equiv [\alpha][\beta]\varphi$$

Öffentliche Ankündigungen

- Ein broadcast ist eine öffentliche Ankündigung, nach der wechselseitiges Wissen bezüglich des Angekündigten bestehen sollte. Dies zeigte sich z.B. im Szenario der ‚muddy children‘.
- Vorausgesetzt wird i.d.R., dass diese broadcasts wahrheitsgemäß sind. Es treten keine Täuschungen auf!
- Die Semantik sieht demgemäß vor, dass *nach* einem broadcast nur solche Welten zu betrachten sind, in denen das Verkündete wahr ist. In dieser Restmenge muss die Nachbedingung gelten.

- Bei [!φ] wird also in einem Modell *M* die Menge der möglichen Welten *W* bezüglich φ filtriert. Das Modell mit der entsprechenden Restmenge von Welten ist M_{φ}.
- Formal semantisch:

$$M,w \vDash [!\varphi]\psi \quad \text{gdw.}$$

$$M,w \nvDash \varphi \;\vee\; H_{!\varphi}(w,w') \wedge M_{\varphi},w' \vDash \psi$$

d.h. die Behauptung der Konsequenzen ψ wird als trivial wahr angesehen, wenn $M,w \nvDash \varphi$, insofern ja vorausgesetzt wird, dass in einer solchen Situation *kein broadcast erfolgt*. Ein broadcast hat ein relevantes Resultat ψ, wenn nach dem broadcast ψ der Fall ist (also unter der Wahrheit der verkündeten Bedingung φ).

- $[!\varphi]\psi$: Nach dem Verkünden von φ ist ψ der Fall.
- Da nur wahre Verlautbarungen vollzogen werden, werden auf der anderen Seite (d.h. der Seite der formalen Semantik) Verlautbarungen, deren Gegenstand nicht vorliegt, nicht für diesen Fall relevant sein, d.h. wiederum, dass einfach gilt:

$$w \nvDash \varphi \rightarrow w \vDash [!\varphi]\psi$$

- Alternativ könnte man *partielle* Handlungsrelationen wählen: Wenn die Handlungsrelation für $[!\varphi]$ in w nicht definiert ist, gibt es auch *keine* erreichbare Welt, in der ψ falsch ist: $\neg\exists w_2\ H_{!\varphi}(w_1,w_2) \supset w \vDash [!\varphi]\psi$

Epistemische Konsequenzen

• Die Abhängigkeit, dass nach dem Verkünden von φ ein Resultat ψ der Fall ist, macht Sinn vor allem für epistemische Konsequenzen. ψ betrifft epistemische Zustände.

• Beispiel:

$$\vDash [!\varphi] W_i \varphi$$

bzw. sogar

$$\vDash [!\varphi] W_{Ak} \varphi \quad \text{[kollektives Wissen per broadcast]}$$

• Wenn „$W_{Aw}\varphi$“ für wechselseitiges Wissen in der Gruppe A steht, sollte sogar gelten: $\vDash [!\varphi] W_{Aw} \varphi$

Mooresche Probleme

• Einige scheinbar einfache Zusammenhänge gelten jedoch nicht, aus Gründen, die an Mooresche Sätze erinnern:

$$\not\models [!\varphi]\varphi$$

Gegenbeispiel $\not\models [!\neg W_i\varphi \wedge \varphi]\neg W_i\varphi \wedge \varphi$

Dieser Satz hat *keine* Modelle, denn in $M_{\neg W i\varphi \wedge \varphi}$ müsste $W_i\varphi$ gelten, da dort nur φ-Welten enthalten sind: ↯

Hier gilt sogar:

$$\models [!\neg W_i\varphi \wedge \varphi]\neg(\neg W_i\varphi \wedge \varphi)$$

d.h. die Äußerung eines Umstandes hebt diesen auf.

• Analog, wenn der Vater der ‚muddy children‘ sagte: „Kein Kind kennt den Fakt, dass genau *n* Kinder schmutzig sind“.

• Allgemein kann in einer intersubjektiven Situation Belehrung eines Akteurs über dessen Nichtwissen stattfinden, wobei zugleich entsprechendes Wissen entsteht:

$$\exists w\ w \vDash [!W_a(\varphi \wedge \neg W_b\varphi)]\neg W_a(\varphi \wedge \neg W_b\varphi)$$

Nach der Verkündung von „$\varphi \wedge \neg W_b\varphi$“ weiß *b* Bescheid.

Axiome für Öffentliche Verlautbarungen

- Für „!φ“ wurden eine Reihe von Axiomen vorgeschlagen, von denen einige im Folgenden erläutert werden sollen.
- Die volle Stärke der Logik der öffentlichen Verlautbarungen zeigt sich, wenn die Sprache einen Operator für wechselseitiges Wissen enthält.
- Die Nummerierung bezieht sich natürlich nur auf die Erweiterungen eines schon vorhanden Basissystems der Epistemischen Logik wie KTS4 mit den entsprechenden anderen Axiomen und Regeln.
- Die Gesamtliste der Axiome ist redundant.

Normalität

- (Axiom 1) $[!\lambda](\varphi \supset \psi) \supset ([!\lambda]\varphi \supset [!\lambda]\psi)$

Zur Erinnerung: die Welten sind maximal und konsistent. Das Axiom zeigt im Übrigen, dass eine Handlung nie nur eine Konsequenz haben kann, sondern immer mindestens den logischen Abschluss der Konsequenz(en).

- (Regel 1) $\vdash\varphi \rightarrow \vdash[!\psi]\varphi$

Logische Wahrheiten bleiben erhalten – egal, was die Leute sagen.

Faktenerhaltung

- (Axiom 2) $[!\varphi]p \equiv \varphi \supset p$

Die Richtung „$\supset$“ drückt den Zusammenhang aus.
„$\subset$“ drückt aus, dass logisch resultierende *Fakten* nicht verloren gehen können (durch die Verkündigung eines wahren Umstandes).

- Letzteres gilt ja gerade nicht für epistemische Umstände (s.o.), die mit Verlautbarungen (insbesondere von epistemischen Zuständen) variieren, *selbst* wenn sie von diesen impliziert werden.

Funktionalität

- (Axiom 3) $[!\varphi]\neg\psi \equiv (\varphi \supset \neg[!\varphi]\psi)$

Wir gehen – wie immer – von wahren Verlautbarungen aus. Lässt sich die Wahrheit φ nicht verkünden mit dem Resultat ψ (rechte Seite), dann muss das (Maximalität der Welten) daran liegen, dass $\neg\psi$ das Resultat der Verlautbarung ist. Konsistenz fordert die andere Richtung.

- In Axiomen dieser Form drückt φ im Antecedenz der Konditionale auf der rechten Seite aus, dass wir über wahre Verlautbarungen sprechen. (Wenn φ falsch ist, gilt $[!\varphi]\psi$ ja in der formalen Semantik als irrelevanterweise wahr, s.o.)

Wissen von Verlautbarungen

- (Axiom 4) $[!\varphi]W_i\psi \equiv (\varphi \supset W_i(\varphi \supset [!\varphi]\psi))$

Wenn bezüglich einer eventuellen wahren Verlautbarung jemand *weiß*, dass sie ψ mit sich bringt (rechte Seite), so bringt diese Verlautbarung Wissen bezüglich ψ mit sich.

Die andere Richtung gilt, insofern das durch die Verlautbarung von φ bedingte Wissen stabil sein muss bezüglich des Umstandes, dass φ zu Recht verlautbart wurde. Der Agent *weiß nun* von den Effekten einer solchen Verlautbarung.

- Es gilt nicht einfach:

$$[!\varphi]W_i\psi \equiv W_i[!\varphi]\psi$$

denn semantisch wird „$[!\varphi]W_i\psi$“ wahr, falls φ falsch ist (s.o.), ohne dass deswegen Wissen bezüglich der Fälle vorliegt, in denen φ wahr ist.

- Epistemische Operatoren können im Allgemeinen nicht über Handlungsoperatoren transportiert werden.

Kompositionalität von Verlautbarungen

- (Axiom 5) $[!\varphi][!\psi]\lambda \equiv [!\varphi \wedge [!\varphi]\psi]\lambda$

Bringt die Verlautbarung von φ dem Umstand hervor, dass eine (weitere) Verlautbarung von ψ in λ resultiert (linke Seite), dann muss eine Verlautbarung von φ *und* des Umstandes, dass diese Verlautbarung (alleine schon) ψ bedingt, auch das Resultat λ der Verlautbarung von ψ im Anschluss an eine von φ bedingen.

Die andere Richtung gilt trivialerweise.

Konjunktivität der Konsequenzen

- (Axiom 6) $[!\varphi](\psi \wedge \lambda) \equiv [!\varphi]\psi \wedge [!\varphi]\lambda$

Wie schon erwähnt, kann eine Verlautbarung mehrere Konsequenzen haben, die miteinander bestehen und als solche auch zusammengefasst werden können (Richtung von rechts nach links).

- Nur auf diese Weise lässt sich die Anzahl der Operatoren „[ ]“ in einem Satz reduzieren, durch Zusammenfassen der Konsequenzen.

Von Verlautbarungen zu wechselseitigem Wissen

- (Regel 2) $\vdash \lambda \supset [!\varphi]\psi,\ \vdash (\lambda \wedge \varphi) \supset W_i\psi \ \rightarrow$

  $\vdash \lambda \supset [!\varphi]W_{Aw}\psi$

Wenn unter einer Bedingung λ die Verlautbarung von φ das Resultat ψ hat und im Fall der bloßen Wahrheit von φ (d.h. *ohne* Verlautbarung) ein beliebiger Agent ψ weiß, dann wissen unter dieser Bedingung alle *nach der Verlautbarung* von φ voneinander, dass sie die Konsequenz ψ kennen.

- Die Regel ähnelt dem ‚Induktionsaxiom' in Systemen mit kollektivem und wechselseitigen Wissen.
- Jeder Agent schließt hier von sich auf andere, und da die Verlautbarung öffentlich ist, wissen dies alle voneinander, schreiben sich also das entsprechende wechselseitige Wissen zu.
- Nach Voraussetzung der Bedeutung von ‚öffentlich' bekommen alle Agenten eine entsprechende Verlautbarung also solche mit, d.h. die Verlautbarung selbst ist immer wechselseitiges Wissen.

Erfolgreiche Verlautbarung

- Für wechselseitiges Wissen gilt ein Prinzip der erfolgreichen Verlautbarung:

$$\vDash [!W_{Aw}\varphi]W_{Aw}\varphi$$

- Auf einem Prinzip der erfolgreichen Verlautbarung basieren auch einige performative Sprechakte (wie Bekennen).

Einige Theoreme

• Die folgenden Theoreme gelten im Gesamtsystem (d.h. bei Verwendung von „W“ und „W_{Aw}“).

(i) $\models \varphi \supset [!\varphi]\varphi$

(ii) $\models [!p]W_i p$

(für *elementare* Fakten, folgt aus Axiom 2)

(iii) $\models [!\neg p]W_{Aw}\neg p$

(iv) $\models W_{Aw}(\varphi \supset W_{Ak}\varphi) \supset (\varphi \supset W_{Aw}\varphi)$

(das ‚Induktionsaxiom“ ist jetzt herleitbar)

(v) $\models [!W_{Aw}\varphi]W_{Aw}\varphi$

(vi) $\models [!\varphi]\psi$ gdw. $\models [!\varphi]W_{Aw}\psi$

Literatur

Diese Folien basieren auf:

- Meyer/van der Hoek, *Epistemic Logic for AI and Computer Science*, S. 96 – 102;

- Ditmarsch et al., *Handbook of Epistemic Logic*, Kap. 6

- Ditmarsch/van der Hoek/Kooi, *Dynamic Epistemic Logic*, Kap. 4