

Wissensrepräsentation

—
Christopher Habel, Özgür Özçep
Sommersemester 2005

Sitzung 20: Abduktion (2)

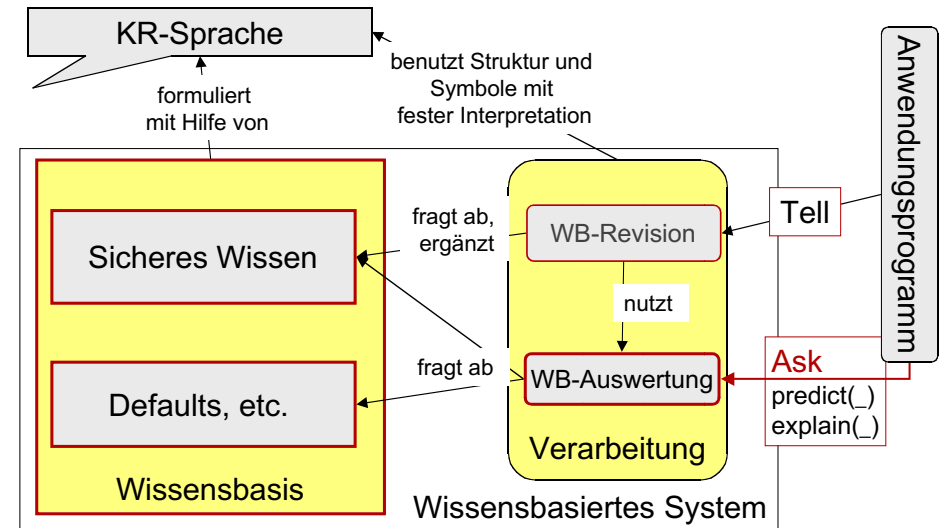
- Vorhersagen und Erklärungen
- THEORIST-Konzeption: Integration von Default-Schließen und abduktivem Schließen

Literatur

Zu Prediction & Explanation (KI)

- Poole, David (1989). Explanation and Prediction: An Architecture for Default and Abductive Reasoning. *Computational Intelligence*, 5. 97-110.
- Poole, David (1990). A methodology for using a default and abductive reasoning system. *International Journal of Intelligent Systems*, 5. 521-548.
- Poole, David; Goebel, Randy & Aleliunas, Romas (1987). Theorist: a logical reasoning system for defaults and diagnosis. In N. Cercone & G. McCalla (Eds.). *The Knowledge Frontier: Essays in the Representation of Knowledge*. (pp. 331-352). Springer Verlag: New York.
- Ronald J. Brachman & Hector J. Levesque, Knowledge Representation and Reasoning, (2004), Chapter 13, Explanation and Diagnosis.

Wissensbasiertes System Prädiktion & Erklärung



Explanations & Predictions

Architektur für Default-Schließen und Abduktives Schließen (Poole)

- F** eine Menge von Fakten, in der Domäne als sicheres Wissen angesehen
- Δ** eine Menge von Defaults: Hypothesen für Prädiktion
- Π** eine Menge von Conjectures: Hypothesen für Erklärungen
- O** eine Menge von Beobachtungen, die in der Realen Welt gemacht wurden.

THEORIST: Die Grundkonzeption

Wissen wird repräsentiert durch Formeln von $\mathcal{PL1}$.

Die Wissensbasis ist strukturiert in Formelmengen, deren Formeln unterschiedliche Rollen im Schließen und Problemlösen spielen.

- Eine Menge A geschlossener Formeln (Axiome)
- Eine Menge H Formeln (mögliche Hypothesen)

THEORIST: Definition der Basiskonzepte

Ein **Szenario** von (A, H) ist eine Menge D von Grundinstanzen von Formeln aus H , so dass $D \cup A$ konsistent ist.

Sei g eine geschlossene Formel. Ein Szenario D von (A, H) ist eine **Erklärung** von g aus (A, H) , falls $A \cup D \models g$. (g ist erklärbar.)

Ein Szenario D von (A, H) ist **maximal**, wenn es keine grössere Menge $D^* \supset D$ von Hypotheseninstanzen gibt, so dass $D^* \cup A$ konsistent ist.

Eine **Extension** von (A, H) ist die Menge der logischen Konsequenzen von $A \cup D$, wobei D ein maximales Szenario ist.

THEORIST: Eigenschaften der Basiskonzepte

Theorem

Zu g existiert eine Erklärung aus (A, H) genau dann, wenn g in einer Extension von (A, H) enthalten ist.

Beziehungen zur Defaultlogik

- $\delta \in H$ in Pooles Konzeption korrespondiert zum normalen Default $\frac{\delta(x)}{\delta(x)}$ in Reiters Defaultlogik.
- Während Reiters Defaultlogik eine Erweiterung von $\mathcal{PL1}$ um neue Inferenzregeln darstellt, behält Poole die Schlussmechanismen bei, betrachtet aber unterschiedliche Mengen von Annahmen (*hypothetical reasoning*).

Explanation & Prediction

Architektur für Default-Schließen und Abduktives Schließen (Poole)

- F** eine Menge von Fakten, in der Domäne als sicheres Wissen angesehen A
- Δ** eine Menge von Defaults: Hypothesen für Prädiktion H
- Π** eine Menge von Conjectures: Hypothesen für Erklärungen
- O** eine Menge von Beobachtungen, die in der realen Welt gemacht wurden.

Prädiktionen

- F** eine Menge von Fakten, in der Domäne als sicheres Wissen angesehen
- Δ** eine Menge von Defaults: Hypothesen für Prädiktion

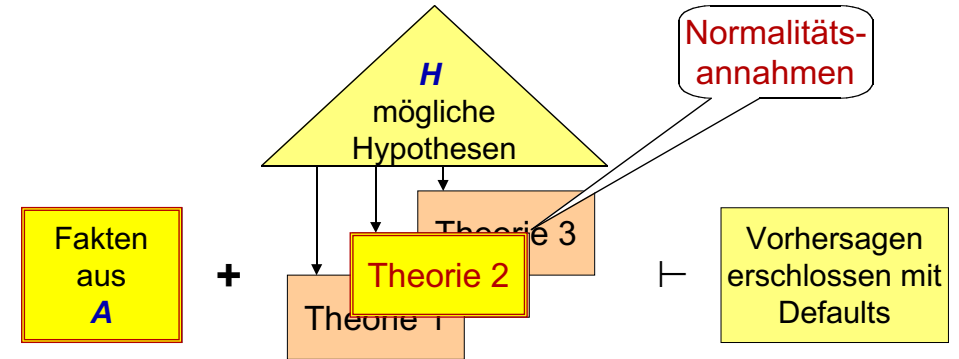
Wenn $\Delta = \emptyset$, dann liegt die konservativste Form der Prädiktion vor:

Logische Konsequenz aus der Axiomenmenge.

Wenn $\Delta \neq \emptyset$, dann liegt Default-Schließen vor, wobei Defaults als gerechtfertigte / vernünftige Annahmen angesehen werden. (Hier gibt es unterschiedlich „konservative“ Arten des Prädizierens.)

Der THEORIST-Rahmen (Vorhersage)

- Eine Menge **A** geschlossener Formeln (Axiome)
- Eine Menge **H** Formeln (mögliche Hypothesen)



Poolers Sichtweise auf Prädiktionen: Der Nixon-Diamond (Axiome und Hypothesen)

$A = \{ \forall x \neg(\text{dove}(x) \wedge \text{hawk}(x)), \text{quaker}(\text{Dick}), \text{republican}(\text{Dick}) \}$

$H = \{ \text{republican}(x) \Rightarrow \text{hawk}(x), \text{quaker}(x) \Rightarrow \text{dove}(x), \\ \text{hawk}(x) \Rightarrow \text{support_star_wars}(x), \text{quaker}(x) \Rightarrow \text{religious}(x), \\ \text{hawk}(x) \Rightarrow \text{politically_motivated}(x), \\ \text{dove}(x) \Rightarrow \text{politically_motivated}(x) \}$

Mögliche Fragen / Prädiktionen zu Dick:

dove(Dick)	hawk(Dick),
dove(Dick) \wedge hawk(Dick)	dove(Dick) \vee hawk(Dick)
support_star_wars(Dick)	
politically_motivated(Dick)	
religious(Dick)	

Default-Prädiktionen (1): Predict if explainable

Predict if explainable: Vorhersage ist auf alle Formeln zulässig, soweit sie in irgendeiner Extension sind, und solange keine Inkonsistenz auftritt.

Theorem: Es existieren nur dann mehrere Extensionen, wenn mindestens ein α existiert, derart, dass sowohl α als auch $\neg\alpha$ erklärbar sind. (α und $\neg\alpha$ liegen in unterschiedlichen Extensionen.)

Alternative Prädiktionen zu Dick (unter anderen):

$\text{hawk}(\text{Dick}) \wedge \text{politically_motivated}(\text{Dick}) \wedge \text{support_star_wars}(\text{Dick})$
 $\text{dove}(\text{Dick}) \wedge \text{politically_motivated}(\text{Dick}) \wedge \text{religious}(\text{Dick})$

Default-Prädiktionen (2): Incontestable scenarios

Ein Szenario D von (A, H) ist **unanfechtbar**, falls für alle $d \in D$ gilt, dass $\neg d$ nicht aus (A, H) erklärbar ist.

- Unanfechtbarkeit von Szenarien ist eine lokale Eigenschaft der Instanzen der Defaults, d.h. ist unabhängig von anderen Defaults in einer Erklärung.

Im Beispiel:

- Unanfechtbar ist der Default: $\text{quaker}(x) \Rightarrow \text{religious}(x)$
- Unanfechtbar erklärbar ist daher: $\text{religious}(\text{Dick})$

Fakten, Hypothesen, Beobachtungen

Fakten: die Wissensentitäten,

- die wir akzeptieren,
- die wir (in der gegenwärtigen Situation / Aufgabenstellung) nicht aufzugeben oder zu verändern beabsichtigen.

Hypothesen: die Wissensentitäten,

- die wir akzeptieren,
- die wir aber beim Vorliegen gegenteiliger Evidenz in der gegenwärtigen Situation / Aufgabenstellung aufgeben.

Beobachtungen: die Wissensentitäten,

- die wir in der gegenwärtigen Situation / Aufgabenstellung beobachten,
- deren Verbindung zu den Fakten und Hypothesen hergestellt, d.h. erklärt, werden soll.

➤ Die Einordnung als *Fakt, Hypothese oder Beobachtung* ist aufgaben- und zeitabhängig.

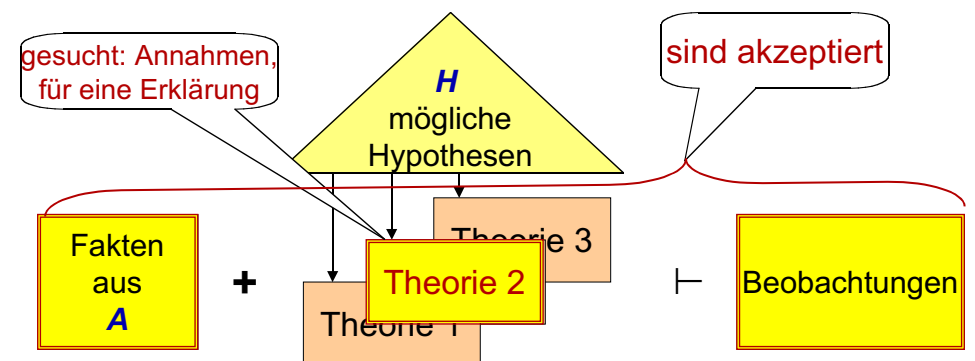
Explanation & Prediction

Architektur für Default-Schließen und Abduktives Schließen (Poole)

- F eine Menge von Fakten, in der Domäne als sicheres Wissen angesehen
 - Δ eine Menge von Defaults: Hypothesen für Prädiktion
 - Π eine Menge von Conjectures: Hypothesen für Erklärungen
 - O eine Menge von Beobachtungen, die in der Realen Welt gemacht wurden.
- A
 H

Der THEORIST-Rahmen

- Eine Menge A geschlossener Formeln (Axiome)
- Eine Menge H Formeln (mögliche Hypothesen)



Defaults & Conjectures

Defaults

- Normalitätsannahmen
- können in Prädiktion und Erklärung eingesetzt werden
- ❖ Normalerweise startet der Rechner, wenn der Einschaltknopf gedrückt wird.

Conjectures

- Abnormalitätsannahmen
- können nur in Erklärungen eingesetzt werden
- ❖ Hypothesen bei Nichtstart des Rechners
 - ❖ Gestörte Stromversorgung
 - ❖ Defekte Festplatte
 - ❖ Störung des Betriebssystems

Dies wird dann in die Überlegungen einbezogen, wenn Evidenz für das Nichtstarten vorliegt.

Erklärung für Beobachtungen

Gegeben sei F , eine Menge von Fakten, eine Menge von Defaults Δ , eine Menge von Conjectures Π und eine Menge von Beobachtungen O .

D sei eine Menge von Grundinstanzen aus Δ und P eine Menge von Grundinstanzen aus Π

$P \cup D$ ist eine **Erklärung** für O genau dann, wenn

$$F \cup P \cup D \models O \text{ und} \\ F \cup P \cup D \text{ konsistent.}$$

Wichtige Arten von Erklärungen

Minimal explanation

- Erklärungen mit den wenigsten Annahmen (insbesondere in Bezug auf Π).

Least presumptive explanation

- Eine Erklärung E_1 ist „weniger mutmaßlich“ als eine Erklärung E_2 , falls $F \cup E_2 \models E_1$.
d.h., E_1 macht weniger Annahmen als E_2

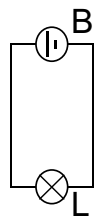
Minimal abnormal explanation

- E_1 mit conjectures P_1 und Defaults D_1 ist weniger abnormal als E_2 , falls $F \cup E_2 \models P_1$ und
 - entweder $F \cup E_1 \not\models P_2$
 - oder $F \cup E_1 \models P_2$ und $F \cup E_2 \models D_1$.

Erklärungen in der Diagnose: Ladung einer Batterie

Relationen (Das Vokabular)

battery(B)	B ist eine aufladbare Batterie
lamp(L)	L ist eine Glühbirne
connect(B,L)	B und L sind verbunden
voltage(B,V,T)	Zur Zeit T liefert B die Spannung V (und diese liegt an L an).
voltage(L,V,T)	Zur Zeit T liegt die Spannung V an L an.
battOK(B,V,T)	Zur Zeit T ist B in Ordnung und liefert V Volt Spannung
overcharged(B,V,T)	Zur Zeit T ist B überladen und liefert V Volt Spannung
flat(B,V,T)	Zur Zeit T ist B nicht hinreichend geladen und liefert V Volt Spannung
lampOK(L,T)	Zur Zeit T arbeitet L normal
dim(L,T)	Zur Zeit T glimmt L matt
lit(L,T)	Zur Zeit T leuchtet L



Beispiel: Batterie-Diagnose Spezifikationen der Batterie

Spezifikation des **Normalverhaltens** der Batterie

fact $\text{battery}(B) \wedge \text{battOK}(B,V,T) \Rightarrow \text{voltage}(B,V,T)$
fact $\text{battOK}(B,V,T) \Rightarrow 1,2 \leq V \wedge V \leq 1,6$
default $\text{battery}(B) \Rightarrow \text{battOK}(B,V,T)$

Spezifikation von Situationen des **Fehlverhaltens**

fact $\text{battery}(B) \wedge \text{overcharged}(B,V,T) \Rightarrow \text{voltage}(B,V,T)$
conjecture $\text{battery}(B) \Rightarrow \text{overcharged}(B,V,T)$
fact $\text{overcharged}(B,V,T) \Rightarrow V > 1,6$
fact $\text{battery}(B) \wedge \text{flat}(B,V,T) \Rightarrow \text{voltage}(B,V,T)$
conjecture $\text{battery}(B) \Rightarrow \text{flat}(B,V,T)$
fact $\text{flat}(B,V,T) \Rightarrow V < 1,2$

Spezifikation des Systems (Batterie – Glühbirne)

fact $\neg(\text{battery}(X) \wedge \text{lamp}(X))$
fact $\text{connect}(B, L) \wedge \text{voltage}(B,V,T) \Rightarrow \text{voltage}(L,V,T)$

Beispiel: Batterie-Diagnose Weitere Spezifikationen

Spezifikation der Spannung (Eindeutigkeit)

fact $\text{voltage}(X,V_1,T) \wedge \text{voltage}(X,V_2,T) \Rightarrow V_1 = V_2$

Spezifikation des Normalverhaltens der Lampe

fact $\text{lamp}(L) \wedge \text{lampOK}(L,T) \wedge \text{voltage}(L,V,T) \wedge V \geq 1,3 \Rightarrow \text{lit}(L,T)$
fact $\text{lamp}(L) \wedge \text{lampOK}(L,T) \wedge \text{voltage}(L,V,T) \wedge 1,0 \leq V \wedge V < 1,3 \Rightarrow \text{dim}(L,T)$
default $\text{lampOK}(L,T)$
fact $\text{lampOK}(L,T) \wedge \text{voltage}(L,V,T) \Rightarrow V \leq 1,8$
fact $\neg \text{lampOK}(L,T_0) \wedge \text{before}(T_0,T_1) \Rightarrow \text{lampOK}(L,T_1)$
fact $\neg(\text{lit}(L,T) \wedge \text{dim}(L,T))$

Beispiel: Batterie-Diagnose Erste Prädiktionen (Ohne Beobachtungen)

Was ist über die Spannung prädizierbar ?

$\text{battery}(b) \Rightarrow \exists V [1,2 \leq V \wedge V \leq 1,6 \wedge \text{voltage}(b,V,t)]$

Was ist über das Verhalten der Lampe prädizierbar ?

genauer: in einem System, in dem Lampe und Batterie verbunden sind: $\text{battery}(b) \wedge \text{lamp}(l) \wedge \text{connect}(b, l)$

- Zwei Typen von erklärenden Szenarien
 - $\text{battOK}(b, v, t)$ mit $v < 1,3$
 - $\text{battOK}(b, v, t)$ mit $1,3 \leq v \leq 1,6$

prädizierbar:

$\text{lit}(l,t) \vee \text{dim}(l,t)$

Beispiel: Batterie-Diagnose Beobachtungen und Erklärungen

Beobachtung: „Lampe und Batterie sind verbunden, die Lampe glimmt matt.“

observe $\text{battery}(b) \wedge \text{lamp}(l) \wedge \text{connect}(b, l) \Rightarrow \text{dim}(l,t)$

Erklärende Szenarien

- $\{ \text{battOK}(b, v, t), \text{lampOK}(l, t) \}$ mit $1,2 \leq v \leq 1,3$
- $\{ \text{flat}(b, v, t), \text{lampOK}(l, t) \}$ mit $1,0 \leq v < 1,2$

Vorhersagen & Erklärungen: Zwischenstand

Vorlesungen 19 & 20:

- Wissensbasierten Agenten:
Die Rolle von Vorhersagen und Erklärungen
- Abduktives Schließen und Default-Schließen: THEORIST

Vorlesungen 21 – 23

Possibilistische Erklärungen

Probabilistische Erklärungen

Erklärungen und Wissensrevision