

Grundlagen der Künstlichen Intelligenz

26. Aussagenlogik: Grundlagen

Malte Helmert

Universität Basel

28. April 2014

Einordnung

Einordnung:

Aussagenlogik

Umgebung:

- **statisch** vs. dynamisch
- **deterministisch** vs. nicht-deterministisch vs. stochastisch
- **vollständig** vs. partiell vs. nicht **beobachtbar**
- **diskret** vs. stetig
- **ein Agent** vs. mehrere Agenten

Lösungsansatz:

- problemspezifisch vs. **allgemein** vs. lernend

(Aber Anwendungen auch in reichhaltigeren Umgebungen.)

Aussagenlogik: Überblick

Kapitelüberblick Aussagenlogik:

- 26. Grundlagen
- 27. Logisches Schliessen und Resolution
- 28. DPLL-Algorithmus
- 29. Lokale Suche und Ausblick

Motivation

Aussagenlogik: Motivation

Diese Woche: Aussagenlogik

- Modellierung und Repräsentation von Problemen und Wissen
- Grundlage **allgemeiner** Problembeschreibungen und Lösungsstrategien
(↷ Handlungsplanung, nächste Woche)
- erlaubt automatisches **logisches Schliessen** (Thema heute)

Zusammenhang zu CSPs

- letzte Woche: Constraint-Satisfaction-Probleme
- Erfüllbarkeitsproblem der Aussagenlogik kann als **nicht-binäres CSP** über $\{\mathbf{F}, \mathbf{T}\}$ aufgefasst werden
- Formel kodiert die Constraints
- Gesucht: erfüllende Zuweisung von Werten zu Variablen
- SAT-Algorithmen für diese Probleme: \rightsquigarrow DPLL (**Freitag**)

Aussagenlogik: Beschreibung von Zustandsräumen

Aussagevariablen für Missionare und Kannibalen:

`two-missionaries-on-left-shore`

`one-cannibal-on-left-shore`

`boat-on-left-shore`

...

- Problembeschreibung verständlich für allgemeine Problemlöser
- Zustände repräsentiert als Wahrheitswerte der atomaren **Propositionen** (Aussagevariablen)

Aussagenlogik: Intuition

Propositionen: atomare (nicht weiter teilbare)
Aussagen über die Welt.

Propositionen bilden mit **logischen Verknüpfungen**
wie „und“, „oder“ und „nicht“ die aussagenlogischen Formeln.

Syntax

Syntax

Σ Alphabet von Propositionen

(z. B. $\{P, Q, R, \dots\}$ oder $\{X_1, X_2, X_3, \dots\}$).

Definition (aussagenlogische Formel)

- \top und \perp sind Formeln.
- Jede Proposition aus Σ ist eine (atomare) Formel.
- Wenn φ eine Formel ist, dann auch $\neg\varphi$ (**Negation**).
- Wenn φ und ψ Formeln sind, dann auch
 - $\varphi \wedge \psi$ (**Konjunktion**)
 - $\varphi \vee \psi$ (**Disjunktion**)
 - $\varphi \rightarrow \psi$ (**Implikation**)

Bindungsstärke: $(\neg) > (\wedge) > (\vee) > (\rightarrow)$

(bei Bedarf Klammern setzen)

Semantik

Semantik

Eine Formel ist **wahr** oder **falsch**;
je nach dem, wie ihre Propositionen **interpretiert** werden.

Semantik: Intuition

- Propositionen sind entweder wahr oder falsch. Dies wird durch eine **Interpretation** bestimmt.
- Wahrheitswert einer Formel folgt aus Wahrheitswerten ihrer Propositionen.

Beispiel

$$\varphi = (P \vee Q) \wedge R$$

Wahrheitswert für beispielhafte Interpretationen:

- Wenn P und Q falsch sind und R wahr ist, dann ist φ falsch.
- Wenn P und R wahr sind, dann ist auch φ wahr (unabhängig vom Wahrheitswert von Q).

Semantik: formal

- definiert über **Interpretation** $I : \Sigma \rightarrow \{\mathbf{T}, \mathbf{F}\}$
- Interpretation I ist **Belegung** der Propositionen in Σ .
- Wann ist eine Formel φ unter einer Interpretation I wahr?
Symbolisch: Wann gilt $I \models \varphi$?

Definition ($I \models \varphi$)

- $I \models \top$ und $I \not\models \perp$
- $I \models P$ gdw. $I(P) = \mathbf{T}$ für $P \in \Sigma$
- $I \models \neg\varphi$ gdw. $I \not\models \varphi$
- $I \models \varphi \wedge \psi$ gdw. $I \models \varphi$ und $I \models \psi$
- $I \models \varphi \vee \psi$ gdw. $I \models \varphi$ oder $I \models \psi$
- $I \models \varphi \rightarrow \psi$ gdw. $I \not\models \varphi$ oder $I \models \psi$
- $I \models \Phi$ für eine Formelmenge Φ gdw. $I \models \varphi$ für alle $\varphi \in \Phi$

Beispiele

Beispiel (Interpretation I)

$$I = \{P \mapsto \mathbf{T}, Q \mapsto \mathbf{T}, R \mapsto \mathbf{F}, S \mapsto \mathbf{F}\}$$

Welche Formeln sind unter I wahr?

- $\varphi_1 = \neg(P \wedge Q) \wedge (R \wedge \neg S)$. Gilt $I \models \varphi_1$?
- $\varphi_2 = (P \wedge Q) \wedge \neg(R \wedge \neg S)$. Gilt $I \models \varphi_2$?
- $\varphi_3 = (R \rightarrow P)$. Gilt $I \models \varphi_3$?

Terminologie

Definition (Modell)

Eine Interpretation I heisst **Modell** von φ wenn $I \models \varphi$.

Definition (erfüllbar etc.)

Eine Formel φ heisst

- **erfüllbar**, wenn es eine Interpretation I mit $I \models \varphi$ gibt.
- **unerfüllbar**, wenn φ nicht erfüllbar ist.
- **falsifizierbar**, wenn es eine Interpretation I mit $I \not\models \varphi$ gibt.
- **allgemeingültig** (= **gültig** = eine **Tautologie**),
wenn $I \models \varphi$ für alle Interpretationen I gilt.

Definition (logische Äquivalenz)

Formeln φ und ψ heissen **logisch äquivalent** ($\varphi \equiv \psi$),
wenn für alle Interpretation I gilt: $I \models \varphi$ gdw. $I \models \psi$.

Wahrheitstabelle

Wahrheitstabelle

Wie kann (Un-) Erfüllbarkeit, Falsifizierbarkeit, Allgemeingültigkeit automatisch für eine Formel bestimmt werden?

↪ einfache Methode: **Wahrheitstabelle**

Beispiel: Ist $\varphi = ((P \vee H) \wedge \neg H) \rightarrow P$ allgemeingültig?

P	H	$P \vee H$	$((P \vee H) \wedge \neg H)$	$((P \vee H) \wedge \neg H) \rightarrow P$
F	F	F	F	T
F	T	T	F	T
T	F	T	T	T
T	T	T	F	T

$I \models \varphi$ für alle Interpretationen I $\rightsquigarrow \varphi$ ist allgemeingültig.

- Erfüllbarkeit, Falsifizierbarkeit, Unerfüllbarkeit?

Normalformen

Normalformen: Begriffe

Definition (Literal)

Wenn $P \in \Sigma$, dann heissen die Formeln P sowie $\neg P$ **Literale**.

P heisst **positives Literal**, $\neg P$ heisst **negatives Literal**.

Das **komplementäre Literal** zu P ist $\neg P$ und umgekehrt.

Wenn ℓ ein Literal ist, bezeichnet $\bar{\ell}$ das komplementäre Literal zu ℓ .

Normalformen: Begriffe

Definition (Literal)

Wenn $P \in \Sigma$, dann heissen die Formeln P sowie $\neg P$ **Literale**.

P heisst **positives Literal**, $\neg P$ heisst **negatives Literal**.

Das **komplementäre Literal** zu P ist $\neg P$ und umgekehrt.

Wenn ℓ ein Literal ist, bezeichnet $\bar{\ell}$ das komplementäre Literal zu ℓ .

Definition (Klausel)

Eine Disjunktion von 0 oder mehr Literalen heisst **Klausel**.

Die **leere Klausel** \perp wird auch als \square geschrieben.

Klauseln aus nur einem Literal heissen **Einheitsklauseln**.

Normalformen: Begriffe

Definition (Literal)

Wenn $P \in \Sigma$, dann heißen die Formeln P sowie $\neg P$ **Literale**.

P heisst **positives Literal**, $\neg P$ heisst **negatives Literal**.

Das **komplementäre Literal** zu P ist $\neg P$ und umgekehrt.

Wenn ℓ ein Literal ist, bezeichnet $\bar{\ell}$ das komplementäre Literal zu ℓ .

Definition (Klausel)

Eine Disjunktion von 0 oder mehr Literalen heisst **Klausel**.

Die **leere Klausel** \perp wird auch als \square geschrieben.

Klauseln aus nur einem Literal heißen **Einheitsklauseln**.

Definition (Monom)

Eine Konjunktion von 0 oder mehr Literalen heisst **Monom**.

Normalformen

Definition (Normalformen)

Formel φ ist in **konjunktiver Normalform** (KNF, Klauselform), wenn sie eine Konjunktion von 0 oder mehr Klauseln ist:

$$\varphi = \bigwedge_{i=1}^n \left(\bigvee_{j=1}^{m_i} \ell_{i,j} \right)$$

Formel φ ist in **disjunktiver Normalform** (DNF), wenn sie eine Disjunktion von 0 oder mehr Monomen ist:

$$\varphi = \bigvee_{i=1}^n \left(\bigwedge_{j=1}^{m_i} \ell_{i,j} \right)$$

Normalformen

Zu jeder Formel gibt es logisch äquivalente Formeln in KNF und in DNF.

Umwandlung in KNF

Wichtige Umformungsregeln für Umwandlung in KNF:

- $\varphi \rightarrow \psi \equiv \neg\varphi \vee \psi$ ((\rightarrow)-Elimination)
- $\neg(\varphi \wedge \psi) \equiv \neg\varphi \vee \neg\psi$ (De Morgan)
- $\neg(\varphi \vee \psi) \equiv \neg\varphi \wedge \neg\psi$ (De Morgan)
- $\neg\neg\varphi \equiv \varphi$ (Doppelnegation)
- $(\varphi \wedge \psi) \vee \eta \equiv (\varphi \vee \eta) \wedge (\psi \vee \eta)$ (Distributivität)

Es gibt Formeln, bei denen jede äquivalente KNF- bzw. DNF-Formel exponentiell länger ist.

Zusammenfassung

Zusammenfassung (1)

- **Aussagenlogik** bildet die Basis für allgemeine Repräsentation von Problemen und von Wissen.
- **Propositionen** (atomare Formeln) sind nicht weiter teilbare Aussagen über die Welt.
- **Aussagenlogische Formeln** kombinieren atomare Formeln mit \neg , \wedge , \vee , \rightarrow zu komplexeren Aussagen.
- **Interpretationen** legen fest, welche atomaren Formeln wahr sind und welche falsch.

Zusammenfassung (2)

- wichtige Begriffe:
 - Modell
 - erfüllbar, unerfüllbar, falsifizierbar, allgemeingültig
 - logisch äquivalent
- bestimmte Arten von Formeln:
 - atomare Formeln und Literale
 - Klauseln und Monome
 - konjunktive Normalform und disjunktive Normalform