

# METHODEN DER KÜNSTLICHEN INTELLIGENZ

Ipke Wachsmuth, Universität Bielefeld

## Kurztext zu Teil 4: Spezielle Schlussverfahren

Dieser Abschnitt befasst sich mit der Frage: *Wie lassen sich auch bei unsicherem und unvollständigem Wissen Schlüsse ziehen?* Schlussfolgern in Anwendungsdomänen ist in vielen Fällen nicht auf die deduktiven Inferenzen, die im vorigen Abschnitt Thema waren, zu stützen. So kann es sein, dass aufgrund der vorliegenden Fakten nur Hypothesen über mögliche andere Fakten gewonnen werden können, oder dass das Vorliegen von Fakten über bestimmte Zusammenhänge den Anlass bildet, auf eine vermutete allgemeine Regel zu schließen. Oder regelhafte Zusammenhänge gelten mit eingeschränkter Sicherheit oder nur solange, wie nicht Ausnahmen von Regeln bekannt werden, oder es könnten schließlich zeitliche oder räumliche Abhängigkeiten zu berücksichtigen sein.

### Abduktion und Hypothesenbildung

Mit der klassischen deduktiven Inferenzregel der Logik, dem Modus Ponens, kann man aus einer bedingten Aussage und einem Faktum, d.h. einer unbedingten, also gesicherten Aussage eine gleichfalls gesicherte Konsequenz gewinnen:

Deduktives Schließen (*Modus Ponens*)

Gegeben, dass gilt: (if  $p$   $q$ ) und  $p$   
inferiere, dass gilt:  $q$

Es handelt sich um einen legalen Schluss in der Logik. Zum Beispiel, aus "Wenn es regnet, dann ist die Straße nass" und "Es regnet" folgt "Die Straße ist nass". In vielen Fällen wird jedoch der Modus Ponens auch in umgekehrter Richtung, also "rückwärts", als *abduktiver* Schluss verwendet, um eine Hypothese über ein beobachtetes Faktum zu gewinnen:

*Abduktives Schließen ("Modus Ponens rückwärts")*

Gegeben, dass gilt: (if  $p$   $q$ ) und  $q$   
inferiere, dass gilt:  $p$

Dies ist allerdings kein legaler Schluss in der Logik: Wird beispielsweise beobachtet, dass die Straße nass ist, so könnte man folgern wollen, dass es regnet. Jedoch ist das "es regnet" kein gesichertes Faktum, sondern nur eine Hypothese; es könnte das Faktum "die Straße ist nass" ja auch daraus folgen, dass es taut, oder dass ein Wasserwerfer zum Einsatz kam etc.

So ist auch das in der Diagnostik typische Zurückschließen von Beobachtungen (Symptomen) auf Zustände bzw. Objekte, die die Beobachtungen hervorrufen, eine Form der Abduktion: Wenn eine Ursache  $D$  ein Symptom  $S$  verursacht, und das Symptom  $S$  wird beobachtet, dann ist  $D$  eine mögliche Erklärung für das Symptom  $S$ . Die Abduktion ist keine logisch zwingende Schlussweise wie die Deduktion, da eine Beobachtung viele Ursachen haben kann. Die Grundlage für abduktives Schließen zur Ursachenfindung in der Diagnostik ist allerdings, dass Prämisse und Konsequenz einer Regel in einer Ursache-Folge-Beziehung stehen, was durch die logische Implikation nur unvollkommen ausgedrückt ist, da sie nicht aus sich heraus eine Kausalbeziehung modelliert.

## Induktion und Lernen

Eine andere deduktive Schlussweise ist die früher erwähnte universelle Einsetzung: Wenn etwas für alle Dinge einer betrachteten Menge gilt, dann gilt es auch für jedes einzelne Ding. D.h. vom Allgemeinen wird auf jedes Spezielle geschlossen: Alle Blätter sind grün, dann ist folglich Blatt-1 grün und Blatt-2 und Blatt-3 usw.

*Universelle Einsetzung (hier nur für eine Variable  $x$  betrachtet)*

Gegeben, dass gilt:  $(\forall x) (Px)$

inferiere, dass gilt:  $(Pa)$  für jede Konstante  $a$

Die "rückwärts"-Richtung dieser Schlussweise kann verwendet werden, um von speziellen Beobachtungen an Dingen einer Menge (Blatt-1 ist grün, Blatt-2 ist grün, Blatt-3 ist grün...) auf eine allgemeine Eigenschaft (alle Blätter sind grün) *induktiv* zu schließen:

*Induktives Schließen (einfachster Fall)*

Gegeben, dass gilt:  $(Pa), (Pb), \dots$

inferiere, dass gilt:  $(\forall x) (Px)$

Auch dies ist allerdings kein legaler Schluss in der Logik: Es könnte ja irgendwann ein Blatt vorkommen, das nicht grün ist. Induktives Schließen ist jedoch ein wichtiges Prinzip des *Lernens* – so gelangt man von empirischen Beobachtungen zu "Gesetzen", zur Vorstellung, dass etwas grundsätzlich gilt. Induktives Schließen kann dann sinnvoll eingesetzt werden, wenn man die Möglichkeit vorsieht, bei abweichenden Beobachtungen zu revidieren. Eine entwickeltere Form der Induktion ist das Konzeptlernen aus Beispielen und Gegenbeispielen, wo aus dem Vergleich zutreffender und "knapp verfehlter" Vorlagen mittels verschiedener Induktionsheuristiken eine Konzeptdefinition induziert wird (Winston's Lernprogramm).

## Probabilistisches Schließen

In vielen Anwendungsbereichen ist das Wissen über die Zusammenhänge *unsicher*. Zum Umgang mit unsicheren Wissens kann man Wissensrepräsentationen um Wahrscheinlichkeiten oder Evidenzwerte erweitern, die den Regeln zugeordnet werden, z.B. "Wenn A dann B mit 70% Wahrscheinlichkeit". Um Schlussfolgerungen abzuleiten, wird dann ein Verfahren zur Verrechnung der Evidenzwerte benötigt. Unter bestimmten Voraussetzungen sind statistische Verfahren wie das Theorem von Bayes dafür geeignet.

Der Basisalgorithmus zur Bewertung von Lösungsalternativen bei unsicherer Information besteht aus drei Schritten:

1. Starte mit der Apriori-Wahrscheinlichkeit der Lösungsalternativen;
2. Modifiziere bei jeder neuen Information die Wahrscheinlichkeit der Alternativen entsprechend den Evidenzwerten und dem Verrechnungsschema;
3. Wähle die am besten bewertete Alternative aus.

Das Theorem von Bayes eignet sich z.B. dazu, aus den Apriori-Wahrscheinlichkeiten  $P(D_i)$  einer Menge von  $n$  Diagnosen und den bedingten Wahrscheinlichkeiten  $P(S_j | D_i)$ , d.h. der statistischen Häufigkeit des Auftretens der Symptome bei gegebenen Diagnosen, die bedingte Wahrscheinlichkeit  $P(D_i | S_1 \dots S_m)$  der Diagnosen bei gegebenen Symptomen gemäß der Formel in Abb. 3 zu berechnen ("Satz von der Möglichkeit der Ursachen").

$$P(D_i / S_1 \& \dots \& S_m) = \frac{P(D_i) * P(S_1 / D_i) * \dots * P(S_m / D_i)}{\sum_{j=1}^n P(D_j) * P(S_1 / D_j) * \dots * P(S_m / D_j)}$$

**Abb. 3** Die zur Alternativenbewertung hauptsächlich benutzte Form des Theorem von Bayes

Wichtige Voraussetzungen zur Anwendung des Bayes-Theorems sind die Unabhängigkeit der Symptome untereinander, Vollständigkeit und wechselseitiger Ausschluss der Diagnosen sowie Statistiken für die Apriori- und die Symptom-Diagnose-Wahrscheinlichkeiten. Da sich diese Voraussetzungen nicht immer erfüllen lassen, kommen weitere Techniken zum Einsatz, die sich in der Praxis als brauchbar erwiesen haben. Neben dem Bayes-Theorem sind zwei verbreitete Verrechnungsschemata des probabilistischen Schließens das MYCIN- und das INTERNIST-Schema. Sie unterscheiden sich von dem Bayes-Ansatz vor allem durch die getrennte Bewertung von positiver und negativer Evidenz (also von Fakten, die für bzw. gegen eine Diagnose sprechen), durch gröbere Bewertungskategorien, durch die Einteilung der Diagnosen in Gruppen konkurrierender Diagnosen (INTERNIST) und durch die Berücksichtigung von Symptomkombinationen (MYCIN).

### Nichtmonotones Schließen

Eine Eigenschaft klassischer Logiken ist es, dass bei einer Erweiterung der zugrunde liegenden Faktenmengen alle bisher abgeleiteten Fakten (Schlussfolgerungen) erhalten bleiben, d.h. dass mit der Vergrößerung der Faktenmenge die Menge möglicher Schlussfolgerungen nicht abnimmt (Monotonieeigenschaft). Dagegen kann es in der Realität geschehen, dass das Bekanntwerden neuer Fakten die Rücknahme vorher gefolgter Fakten erfordert. Dieses Problem des *nichtmonotonen* Schließens ergibt sich aus dem inkrementellen Informationserwerb, also der Tatsache, dass das Wissen einer Anwendungsdomäne in aller Regel nicht vollständig modellierbar ist (*unvollständig* ist). Wären alle Fakten von Anfang an bekannt, entstünde kein Bedarf der Rücknahme von Schlussfolgerungen. Wenn jedoch aufgrund der vorhandenen Informationen ein plausibler Schluss gezogen wurde (z.B. Tweety ist ein Vogel, deshalb kann Tweety fliegen), der aufgrund zusätzlicher Information (Tweety ist ein Pinguin) ungültig wird, dann muss der Schluss mit allen davon abhängenden Konsequenzen revidiert werden.

Während theoretische Ansätze des nichtmonotonen Schließens in der *Nichtmonotonen Logik* formal untersucht werden, beschränken wir uns hier darauf, wie nichtmonotones Schließen mit pragmatischen Ansätzen realisiert werden kann. Am einfachsten wäre die vollständige Neuberechnung aller Inferenzen aufgrund der neuen Menge an Ausgangsdaten. Dieser Weg ist für kleinere Systeme gangbar, doch wächst der Aufwand mit der Komplexität der Schlussketten. Darum ist dieses "brute-force"-Verfahren oft nicht praktikabel. Die Kernidee zu einer Effizienzverbesserung ist das Abspeichern von Begründungen für jede Schlussfolgerung. Dann kann die Frage, ob eine Schlussfolgerung bei Änderungen weiter gültig bleibt, dadurch beantwortet werden, ob sie nach Revision noch mindestens eine gültige Begründung besitzt. Die beiden wichtigsten Rücksetztechniken, die TMS (für Truth-Maintenance-System) oder auch RMS (für Reason-Maintenance-System) genannt werden, verwalten dazu entweder

- direkte Begründungen: JTMS (Justification-based TMS), oder
- Basisannahmen, die einer Begründung zugrunde liegen: ATMS (Assumption-based TMS)

Der Unterschied lässt sich an einem einfachen Beispiel mit zwei Regeln grob verdeutlichen: Regel-1:  $A \rightarrow B$ , Regel-2:  $B \rightarrow C$ . Ist A eine Basisannahme, so wird im JTMS für C die Begründung Regel-2 abgespeichert, während im ATMS ein Kontext (d.i. eine Menge von Basisannahmen) berechnet wird, der A enthält.

Der JTMS-Basisalgorithmus ist einfach und effizient: eine Änderung eines Faktums bewirkt, dass alle damit verbundenen Begründungen überprüft werden. Falls eine Begründung ungültig ist, wird getestet, ob die zugehörige Schlussfolgerung noch weitere Begründungen hat. Falls nein, wird die Schlussfolgerung zurückgezogen und mit ihr als Input der Algorithmus rekursiv aufgerufen. Das Hauptproblem dabei ist die Behandlung von Ableitungsschleifen; z.B. würde eine Schleife  $A \rightarrow B$  und  $B \rightarrow A$  im Basisalgorithmus bewirken, dass A und B nicht mehr rücksetzbar sind. Eine Lösung dieses Problems ist das Abspeichern von nicht-zirkulären (well-founded) Begründungen für eine Schlussfolgerung.

Beim ATMS existiert für jede Schlussfolgerung ein sogenanntes Label, das aus Mengen von Basisannahmen besteht, unter denen die Schlussfolgerung gültig ist. Eine Schlussfolgerung ist gültig, wenn ihr Label mindestens eine Menge enthält, die eine Teilmenge der global gültigen Basisannahmen ist. Ein Vorteil des ATMS ist, dass die Konsequenzen verschiedener Mengen von Basisannahmen leicht miteinander verglichen werden können. Jedoch werden bei großen Wissensbasen mit sehr vielen Basisannahmen die für die Effizienz des ATMS kritischen Mengenvergleiche sehr aufwändig.

### **Temporales (und räumliches) Schließen**

Zustände und Ereignisse sind Begriffe, die implizit ein Zeitkonzept voraussetzen. Wenn ein Ereignis eintritt, dann ändern sich die Verhältnisse; ein Zustand *vor* dem Ereignis kann sich von dem *nach* dem Ereignis unterscheiden, Fakten könnten nur für begrenzte Zeit gültig sein. Die Voraussetzung zum temporalen Schließen ist die Erweiterung der Wissensrepräsentation von Fakten um Zeitangaben. Die Art der Zeitrepräsentation bestimmt, wie explizit (qualitativ/quantitativ) und effizient zeitbezogene Fragen beantwortet werden können. Am einfachsten ist die Repräsentation exakter Zeitangaben auf einer absoluten Zeitskala. Sind keine genauen Angaben verfügbar, können Zeitbezüge qualitativ (A begann vor B etc.) angegeben werden.

Zur Behandlung qualitativer Zeitbeschreibungen sind eine Reihe von Ansätzen entwickelt worden. Ein wichtiges Beispiel ist der Situationskalkül, der eine zeitlogische Beschreibung kausaler zeitlicher Zusammenhänge vornimmt, um Resultate von Ereignissen zu inferieren. Hiermit verbindet sich das berühmte-berüchtigte *Frame-Problem*: Wie lässt sich spezifizieren, welche implizit geltenden Rahmenbedingungen sich *nicht* ändern, wenn ein Ereignis eintritt? Weitere Ansätze sind das sog. *Overlap Chaining*, das den Modus Ponens auf überlappendes Gelten von Fakten und Regelzusammenhängen erweitert, und das *Time Map Management*, bei dem punkt- oder intervallbasierte Zeitkarten verwaltet werden. Diese sind im Prinzip gleich mächtig, da man ein Intervall durch Anfangs- und Endpunkt bzw. einen Zeitpunkt durch ein beliebig kleines Intervall darstellen kann, haben aber unterschiedliche Konsequenz für die Einfachheit und Eleganz der Handhabung. Am bekanntesten ist sicherlich der intervallbasierte *Allensche Zeitkalkül* mit 13 zeitvergleichenden Relationen, die eine elegante qualitative Beschreibung eigentlich aller denkbaren Zeitbeziehungen ermöglichen.

Mehr noch als temporales Schließen ist das räumliche Schließen ein intensiver Gegenstand aktueller Forschung; es kann aber in dieser Vorlesung kaum mehr als gestreift werden.