Was bisher geschah

Wissensrepräsentation und -verarbeitung durch

- Künstliche Neuronale Netze (insbes. auch CNN)
- Zustandsübergangssysteme
- Klassische Logiken
- Regelsysteme in klassischer Aussagen- und Prädikatenlogik
- Logische Programme (Prolog, Datalog)
- Nichtmonotonen Schließens bei unvollständigem Wissen (closed world assumption, schwache Negation)
- Beispiele zum Planen
- Answer Set Programming
- Mehrwertige Logiken,
  z.B. dreiwertige Łukasiewicz-Logik

Fuzzy-Logiken

Fuzzy-Logik: Sammelbegriff für verschiedene Logiken,
meist mit

- Wahrheitswertbereich $[0, 1] \subseteq \mathbb{R}$ (oder Teilmenge davon)
- $\neg x \mapsto 1 - x$
- Wahrheitswertkonstanten 0 und 1

und definiert über Wahrheitswertfunktionen für andere Junktoren,
oft $\wedge$, $\vee$ oder $\rightarrow$

Beispiele:

- Standard-Fuzzy-Logik: $\wedge \mapsto \min$, $\vee \mapsto \max$
- Produkt-Logik: $x \wedge y \mapsto xy$
- Łukasiewicz-Logik: $x \rightarrow y \mapsto \min(1, 1 - x + y)$

Vierwertige Logik

sinnvoll z.B. zum Umgang mit widersprüchlichen Informationen

Beispiel: parakonsistente Logik von Belnap (1977)

Wahrheitswertbereich $\{0, \perp, \top, 1\}$ (auch $\{0,1\}^2$)

mit zwei Ordnungen (beide partiell):

- Wahrheits-Ordnung: $0 <_W \perp <_W 1$ und $0 <_W \top <_W 1$
- Informations-Ordnung: $\perp <_I 0 <_I \top$ und $\perp <_I 1 <_I \top$

Wahrheitswertfunktionen:

$\neg$	$\perp$	0	1	$\top$
	$\perp$	1	0	$\top$

$\wedge$	$\perp$	0	1	$\top$
$\perp$	$\perp$	0	$\perp$	0
0	0	0	0	0
1	$\perp$	0	1	$\top$
$\top$	0	0	$\top$	$\top$

$\vee$	$\perp$	0	1	$\top$
$\perp$	$\perp$	$\perp$	1	1
0	$\perp$	0	1	$\top$
1	1	1	1	1
$\top$	1	$\top$	1	$\top$

Mehrwertige Mengen und Relationen

Übergang von zweiwertigen zu mehrwertigen Mengen:

- Menge $M \subseteq U$ mit
  charakteristischer Funktion $\chi_M : U \to \{0, 1\}$
- Mehrwertige Menge über Wahrheitswertbereich W
  Funktion $M : U \to W$
  ordnet jedem $x \in U$ einen Wahrheitswert
  (Zugehörigkeitsgrad) zu

Relation: Menge $R \subseteq A \times B$ von Paaren (Tupeln)
Übergang von zweiwertigen zu mehrwertigen Prädikaten (Relationen):

- Relation $R \subseteq A \times B$ mit
  charakteristischer Funktion $\chi_R : A \times B \to \{0, 1\}$
- Mehrwertige Relation über Wahrheitswertbereich W
  Funktion $R : A \times B \to W$
  ordnet jedem Paar $(a, b) \in A \times B$ einen Wahrheitswert zu

Erinnerung: Eigenschaften sind einstellige Relationen (Mengen).

Unsichere Regelsysteme

Ansätze:

1. Unsicherheit in den Daten (Wahrheitswerte an Fakten)
   Regel $l_1 \wedge \ldots \wedge l_n \rightarrow h$
   angewendet auf Voraussetzungen l_i, je mit Wahrheitswert w_i
   ordnet h den Wert $f(w_1, \ldots, w_n)$ zu
2. Unsicherheit in den Regeln (Wahrheitswerte an Regeln)
   Regel $l_1 \wedge \ldots \wedge l_n \rightarrow_w h$
   angewendet auf Voraussetzungen l_i
   ordnet h den Wert w zu, falls alls l_i erfüllt sind
3. Kombination aus beiden

Beispiel

ProbLog
(`https://dtai.cs.kuleuven.be/problog/index.html`)

Fakten mit Wert:

```
0.5::heads1.
0.6::heads2.
```

Regeln:

```
twoHeads :- heads1, heads2.
```

Anfragen:

- `heads1 = 0.5`
- `heads2 = 0.6`
- `twoHeads = 0.3`

Beispiel

Fakten mit Wert:

```
0.5::heads1.
0.6::heads2.
```

Regeln:

```
someHead :- heads1.
someHead :- heads2.
```

Anfragen:

- `heads1 = 0.5`
- `heads2 = 0.6`
- `someHead = 0.8`

Beispiel

Fakten und Regeln mit Wert:

```
0.3::stress(X) :- person(X).
0.2::influences(X,Y) :- person(X), person(Y).

smokes(X) :- stress(X).
smokes(X) :- friend(X,Y), influences(Y,X), smokes(Y).

0.4::asthma(X) :- smokes(X).
```

WH: Wahrscheinlichkeiten

Begriffe:

- Zufalls-Experiment
- (endlicher) Wahrscheinlichkeitsraum $(\Omega, 2^\Omega, P)$
- Elementar-Ereignis
- zufälliges Ereignis

Beispiele:

- Experiment: dreimal würfeln,
- Ereignis V: Augenzahlen sind paarweise verschieden,
- Elementar-Ereignisse: $\{(x, y, z) \mid x, y, z \in \{1, \dots, 6\}\}$
- $P(V)$ bei Gleichverteilung?

WH: Bedingte Wahrscheinlichkeiten

Definition:
Bedingte Wahrscheinlichkeit von Ereignis A unter Ereignis B:

$$P(A \mid B) = P(A \cap B)/P(B)$$

Beispiele:

- zwei Würfel, $A =$ Augensumme ist > 7,
  $B =$ beide Zahlen sind ungerade.
- B eine Ursache (für Fehler, Krankheit, usw.),
  A eine Auswirkung (Symptom) (leichter zu beobachten)

Unterschied zu bisher betrachteten Regelsystemen:

- bisher: Aussagen über Wahrheit (von Aussagen)
- jetzt: Aussagen über Wahrscheinlichkeit (von Ereignissen)

Satz von Bayes

Satz von Bayes (einfache Form):

$$P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

Beweis: Def. von $P(X \mid Y)$ einsetzen, vereinfachen.

Anwendung: Rechnen mit bedingten Wahrscheinlichkeiten

- 1/3 aller Studenten haben ein Notebook.
- 1/10 aller Studenten studieren Informatik.
- 9/10 aller Informatik-Studenten haben ein Notebook.
- Sie sehen einen Studenten mit einem Notebook.
- Mit welcher Wahrscheinlichkeit studiert er Informatik?

Das ist ein Beispiel für probabilistische Inferenz.
wird verallgemeinert auf längere Ketten von
Ursache-Wirkung-Beziehungen

WH: Unabhängige Ereignisse

Def: Ereignisse A, B heißen (stochastisch) unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$.

Satz: $P(B) > 0 \Rightarrow (A$ und B unabh. $\iff P(A \mid B) = P(A))$.

Bsp:
zwei Würfel, $A =$ Augensumme > 7, $B =$ beide Zahlen ungerade.
A und B sind *nicht* unabhängig.

Def: Nicht unabhängige A, B heißen *korreliert*.

Vorsicht: das bedeutet nicht,
dass A die Ursache für B ist, oder B die für A.
Es könnte z.B. eine gemeinsame Ursache C für A und B geben.
(correlation does not imply causation)

Beispiele:

- $A =$ schweres Fahrzeug, $B =$ hoher Verbrauch, $C =$ unwegsames Gelände
- $A =$ geringes Geburtsgewicht, $B =$ hohe Säuglingssterblichkeit, $C =$ starkes Rauchen

Diskrete Zufallsgrößen

- Def: Zufallsgröße ist Funktion $X : \Omega \to$ endl. Menge ($\subseteq \mathbb{R}$)
- einfachster Fall: $\Omega = \{0,1\}^k$

  $X_k = (\vec{x} \mapsto \vec{x}_k)$ (die k-te Komponente)
- dann Wsk-Raum bestimmt durch Wsk der Elementar-E.,

  Bsp: $P(0,0) = 1/3, P(0,1) = 1/6, P(1,0) = 0, P(1,1) = 1/2$
- (Motivation für Bayes-Netz: beschreibt solchen Wsk-Raum durch deutlich weniger als 2^k Parameter)
- zu Zufallsgröße X betrachte Ereignis $X = e$,

  Bsp (Fortsetzung): $P(X_1 = 0 \cap X_2 = 1) = 1/6$.

  $P(X_2 = 1) = 1/6 + 1/2 = 2/3$, $P(X_1 = 0) = \dots$
- Def. Zufallsgrößen X, Y sind unabhängig:
  jedes $X = e$ ist unabhängig von jedem $Y = f$

Kausal-Diagramme

Kausal-Diagramm: DAG

- Knoten: Sachverhalte
- Kanten: (vermutete) kausale (ursächliche) Beziehungen

Beispiel:

- Knoten: Winter, glatt, Tom betrunken, Unfall Tom / Jerry

Verbindungsmuster:

- seriell: $W \to G \to U$
- teilend: $G \to T$, $G \to J$
- zusammenführend: $B \to T$, $G \to T$

Bayes-Netze: Motivation, Definition

- Bayes-Netz (alternativ: *believe network*) ist DAG
  - Knoten: Zufallsvariablen
  - Kanten: (vermutete) kausale (ursächliche) Beziehungen
- Anwendung: probabilistisches Schließen,
  Bestimmung wahrscheinlicher Ursachen für Symptome
- BN erfunden von Judea Pearl, erhielt (u.a.) dafür den *ACM Turing Award* 2011,

  `https://amturing.acm.org/award_winners/pearl_2658896.cfm`
- benannt nach Thomas Bayes (1701–1761),
  Satz von Bayes über bedingte Wahrscheinlichkeiten

Definition Bayes-Netz

- Syntax: ein Bayes-Netz N ist ein Paar (G, Θ) mit
  - G ist DAG, Knoten sind Zufallsgrößen
  - Θ: für jeden Knoten X mit Eltern $X_1, \ldots, X_k$:
    Wahrscheinlichkeiten $P(X = e \mid X_1 = e_1 \cap \cdots \cap X_k = e_k)$
    für alle $[e, e_1, \ldots, e_k] \in W^{k+1}$
- Semantik: N beschreibt Wahrscheinlichkeitsraum durch
  $P(X = e) = P(X = e \mid \ldots X_k = e_k \ldots) \cdot \prod_k P(X_k = e_k)$

  induktive Definition:
  IA: Quellen des DAG (ohne Vorgänger, d.h., ohne Bedingungen, d.h., $\prod \emptyset = 1$)

Beispiel Bayes-Netz

(nach Judea Pearl)

- Knoten: Einbruch R, Erdbeben E, Alarmanlage A (zuhause), John ruft (auf Arbeit) an J, Mary ruft an M.
- Kanten mit Parametern (Bsp)
  - $P(R = 1) = 0.001$, $P(E = 1) = 0.002$
  - $P(A = 1 \mid R = 0, E = 1) = 0.29, \dots$

Graphische Darstellung: Tafel

Bedingte Unabhängigkeit und BN

- (Wdhlg.) Def A und B unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$.
- Def: A und B bedingt unabhängig bezüglich C: $P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C)$.
  (Vorstellung: wir schränken den Wsk-Raum ein auf die Elementar-Ereignisse aus C, verwenden dort die Standard-Def. der Unabh.)
- Def: bedingte Unabh. von (diskreten) Zufallsgrößen entsprechend
- Satz: für jedes BN N, für alle $X, Y \in N$ mit $X \not\to_N^* Y$: X und Y sind bedingt unabh. bezüglich der Eltern von X.

Inferenz mit BN

- die Diagnose-Aufgabe: gegeben ein BN, gesucht sind bedingte Wahrscheinlichkeiten der Ursache(n),
  unter der Bedingung von Beobachtungen
- Bsp: $P(\text{Einbruch} = 1 \mid \text{John} = 1 \cap \text{Mary} = 1)$
- Bsp: $P(\text{Einbruch} = 1 \mid \text{John} = 1 \cup \text{Mary} = 1)$
- kann exakt bestimmt werden, dauert jedoch $2^{|N|}$
  kann nicht besser gehen, weil aussagenlogische Erfüllbarkeit auf dieses Inferenzproblem reduziert werden kann
- die Alternative sind schnellere (Simulations)Verfahren,
  die einen Näherungswert liefern

Reading Group KW

Judea Pearl (2018):
The Seven Tools of Causal Inference with Reflections on Machine Learning

https://ftp.cs.ucla.edu/pub/stat_ser/r481.pdf