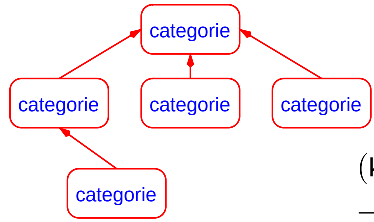
Model-gebaseerd Redeneren

- ‘Traditionele’ kennissystemen:
  - regel gebaseerd (ervaringsregels)
    if conditions then actions/conclusions fi
    ⋮
  - gebruikte redeneermethoden:
    - top-down inferentie: redeneer van doelen subdoelen naar feiten
    - bottom-up inferentie: redeneer van feiten naar eindconclusies/acties
  - recent: business rules
- ‘Moderne’ kennissystemen: model gebaseerd

Ervaringsregels

Probleemoplossen op grond van ervaringsregels:

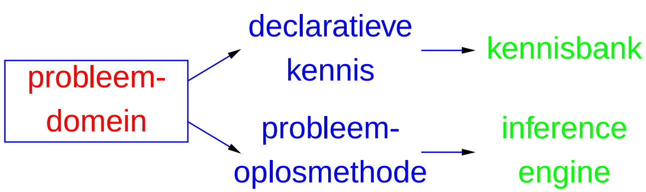


$(\text{kenmerk}_1 \wedge \cdots \wedge \text{kenmerk}_n)$
$\longrightarrow$ categorie

Nadelen:

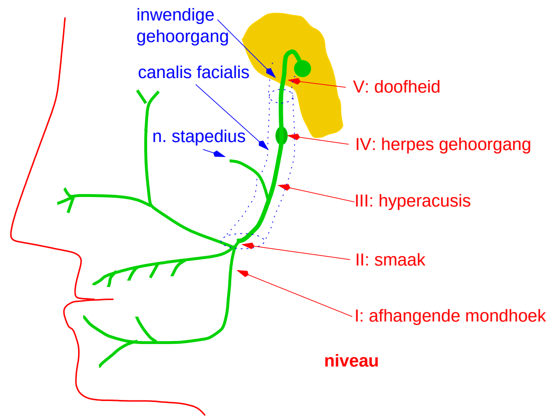
- kennis over structuur en werking niet gebruikt
- kennis moeilijk onderhoudbaar (aanpasbaar)

Gebruik van modellen



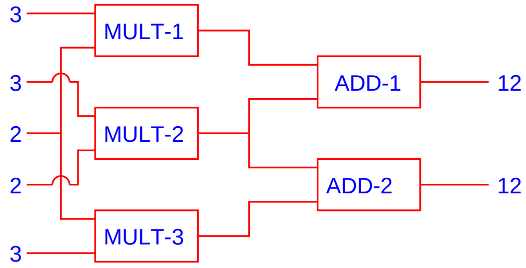
- Modellen:
  - zijn meestal niet toegespitst op oplossen van één probleem $\Rightarrow$ herbruikbaarheid
  - gaan over momentaan gedrag, ontwikkeling van gedrag in de tijd, structuur
- Methoden: diagnose, beslissen, voorspellen, planning

Medische diagnostiek: facialisparese



Diagnose:
$\text{Symptomen-niveau}_{i+1} =$
$\quad \text{Symptomen-niveau}_i \cup \text{Nieuwe-symptomen}$

Structuur èn gedrag



- Structuur:
  - componenten: MULT-1, MULT-2, ...
  - bedrading
- Gedrag:
  - gedrag van individuele componenten
  - combineren van gedrag

Methode: model-gebaseerde diagnose

- Model: representatie van normaal of afwijkend gedrag en eventueel van de interne structuur
- Formalisatie:
  - *consistentie-gebaseerde diagnose*, en
  - *abductieve diagnose*

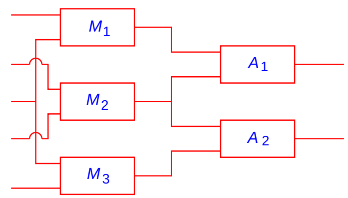
Consistentie-gebaseerde diagnose

Verschil tussen voorspeld gedrag en geobserveerd gedrag $\Rightarrow$ defect!

Geestelijke vaders:

- R. Reiter, "A Theory of diagnosis from first principles", *Artificial Intelligence*, vol. 32, 57–95, 1987.
- J. de Kleer, A.K. Macworth, and R. Reiter, "Characterising diagnoses and systems", *Artificial Intelligence*, vol. 52, 197–222, 1992.

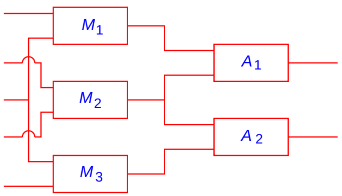
Normaal gedrag



SYSteemspecificatie $\mathrm{SYS} = (\mathrm{SB}, \mathrm{COMPS})$:

- Componenten die *defect* kunnen zijn: $\mathrm{COMPS} = \{M_1, M_2, M_3, A_1, A_2\}$
- SB (SysteemBeschrijving):
  - *generieke* beschrijving van functies van componenten
  - declaratie van componenten: $\mathrm{MUL}(M_1)$, $\mathrm{ADD}(A_1)$
  - verbindingen tussen componenten

Normaal gedrag: formalisering



SYSteemspecificatie $\mathrm{SYS} = (\mathrm{SB}, \mathrm{COMPS})$:

- SB (SysteemBeschrijving):

  $\forall x(\mathsf{MUL}(x) \to \mathrm{in}_1(x) \times \mathrm{in}_2(x) = \mathrm{out}(x))$
  $\forall x(\mathsf{ADD}(x) \to \mathrm{in}_1(x) + \mathrm{in}_2(x) = \mathrm{out}(x))$

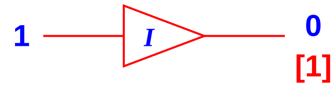
  $\mathrm{MUL}(M_1), \mathrm{MUL}(M_2), \mathrm{MUL}(M_3), \mathrm{ADD}(A_1), \mathrm{ADD}(A_2)$
  $\mathrm{in}_1(A_1) = \mathrm{out}(M_1), \mathrm{in}_2(A_1) = \mathrm{out}(M_2)$
  $\mathrm{in}_1(A_2) = \mathrm{out}(M_2), \mathrm{in}_2(A_2) = \mathrm{out}(M_3)$

- $\mathrm{COMPS} = \{M_1, M_2, M_3, A_1, A_2\}$

Ab predicaat

- $\mathsf{Ab}(c)$: component c is *abnormaal*
- $\neg\mathsf{Ab}(c)$: component c is niet abnormaal, i.e. *normaal*

Voorbeeld (Inverter I):



$\mathrm{SB} = \{\forall x((\mathrm{INV}(x) \wedge \neg\mathrm{Ab}(x)) \to \neg(\mathrm{out}(x) = \mathrm{in}(x))),\ \mathrm{INV}(I)\}$

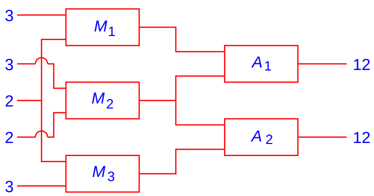
- Input: $\mathrm{in}(I) = 1$; geobserveerde output: $\mathrm{out}(I) = 1$

$$\mathrm{SB} \cup \{\mathrm{in}(I) = 1, \mathrm{out}(I) = 1\} \cup \{\neg\mathrm{Ab}(I)\} \vDash \bot$$

$$\mathrm{SB} \cup \{\mathrm{in}(I) = 1, \mathrm{out}(I) = 1\} \cup \{\mathrm{Ab}(I)\} \nvDash \bot$$

(assumptie dat I is (ab)normaal is (in)consistent

Normaal gedrag: formalisering



SYSteemspecificatie SYS = (SB, COMPS):

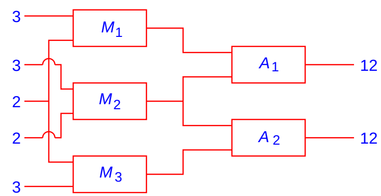
- SB (Systeem Beschrijving):

  $\forall x((\mathsf{MUL}(x) \wedge \neg\mathrm{Ab}(x)) \to \mathrm{in}_1(x) \times \mathrm{in}_2(x) = \mathrm{out}(x))$
  $\forall x((\mathsf{ADD}(x) \wedge \neg\mathrm{Ab}(x)) \to \mathrm{in}_1(x) + \mathrm{in}_2(x) = \mathrm{out}(x))$

  $\mathrm{MUL}(M_1), \mathrm{MUL}(M_2), \mathrm{MUL}(M_3), \mathrm{ADD}(A_1), \mathrm{ADD}(A_2)$
  $\mathrm{in}_1(A_1) = \mathrm{out}(M_1), \mathrm{in}_2(A_1) = \mathrm{out}(M_2)$
  $\mathrm{in}_1(A_2) = \mathrm{out}(M_2), \mathrm{in}_2(A_2) = \mathrm{out}(M_3)$

- $\mathrm{COMPS} = \{M_1, M_2, M_3, A_1, A_2\}$

Voorspellen van normaal gedrag

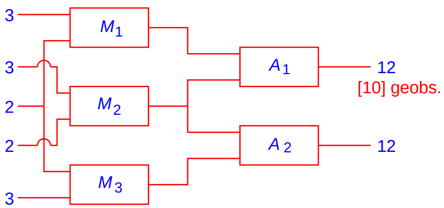


- Systeemspecificatie SYS = (SB, COMPS), met SB systeembeschrijving, en COMPS componenten
- Een voorspelling:

  $\mathrm{SB} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS}$ is niet defect$\} \cup\ \mathrm{Inputs} \vDash \mathrm{Gedrag}$
- Voorbeeld:

  $\mathrm{SB} \cup \{\neg\mathrm{Ab}(M_1), \neg\mathrm{Ab}(M_2), \neg\mathrm{Ab}(A_1)\} \cup \{\mathrm{in}_1(M_1) = 3,$
  $\mathrm{in}_2(M_1) = 2, \mathrm{in}_1(M_2) = 3, \mathrm{in}_2(M_2) = 2\} \vDash \mathrm{out}(A_1) = 12$

Er is een defect!



Stel $(\mathrm{out}(A_1) = 10) \in \mathrm{Inputs}$, dan:

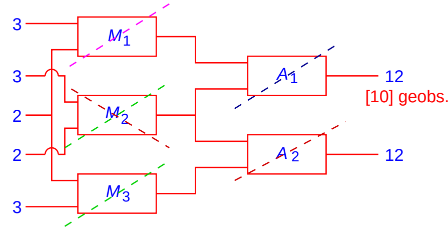
$$\mathrm{SB} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS}\} \cup \mathrm{Inputs} \vDash \bot$$

want:

- $\mathrm{SB} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS}\} \cup \mathrm{Inputs} \vDash \mathrm{out}(A_1) = 12$, en
- $\mathrm{SB} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS}\} \cup \mathrm{Inputs} \vDash \mathrm{out}(A_1) = 10$

$\Rightarrow$ defecte component

Welke componenten zijn defect?



Mogelijke diagnosen (defecte componenten) D:

- $D = \{A_1\}, \{M_1\}, \{M_2, M_3\}, \{A_2, M_2\}$, want
  $\mathrm{SB} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS} - D\}$
  $\cup \{\mathrm{Ab}(c) \mid c \in D\} \cup \mathrm{Inputs} \nvDash \bot$
- D is de *kleinste* verzameling, want $D = \mathrm{COMPS}$ zou anders ook altijd een diagnose zijn

$\Rightarrow$ multipele diagnosen

Algoritmisch

- Enumeratie van alle diagnosen: #P complete (NP hard voor opsomming)

$$\mathrm{Ab}(A_1) \wedge \neg\mathrm{Ab}(A_2) \wedge \neg\mathrm{Ab}(M_1) \wedge \neg\mathrm{Ab}(M_2) \wedge \neg\mathrm{Ab}(M_3)$$
$$\mathrm{Ab}(A_1) \wedge \mathrm{Ab}(A_2) \wedge \neg\mathrm{Ab}(M_1) \wedge \neg\mathrm{Ab}(M_2) \wedge \neg\mathrm{Ab}(M_3)$$
$$\mathrm{Ab}(A_1) \wedge \mathrm{Ab}(A_2) \wedge \mathrm{Ab}(M_1) \wedge \neg\mathrm{Ab}(M_2) \wedge \neg\mathrm{Ab}(M_3)$$
$$\vdots$$

- Heuristische methoden:
  - hitting set algoritme (Reiter)
  - assumption-based truth maintenance system (ATMS, De Kleer)
- Inperkingen: bijvoorbeeld maximaal 2 defecten, dan bovengrens op complexiteit

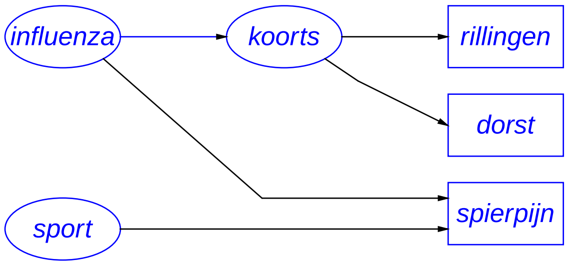
Abductieve diagnose

Overeenstemming tussen voorspeld *afwijkend* gedrag en geobserveerd gedrag $\Rightarrow$ defect!

Geestelijke vaders:

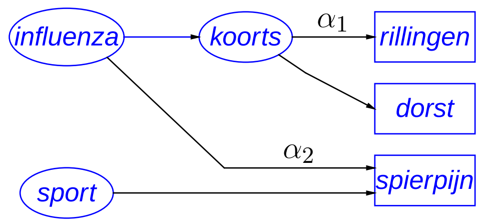
- L. Console, D. Theseider Dupré and P. Torasso, “A theory of diagnosis for incomplete causal models”, In: *IJCAI'89*, 1311–1317, 1989
- D. Poole, “Explanation and prediction: an architecture for default and abductive reasoning”, *Computational Intelligence*, vol. 5, nr. 2, 97–110, 1989

Causale modellen



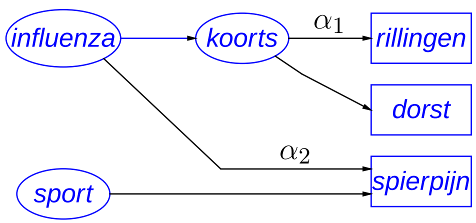
- Causaliteit: combinatie van Oorzaken geeft aanleiding tot bepaalde Gevolgen
- Logische representatie:
  Oorzaak$_1 \wedge \cdots \wedge$ Oorzaak$_n \rightarrow$ Gevolg
- Voorbeeld:
  koorts $\longrightarrow$ *rillingen*
  koorts $\longrightarrow$ *dorst*
  sport $\longrightarrow$ *spierpijn*

Zwakke en sterke causaliteit



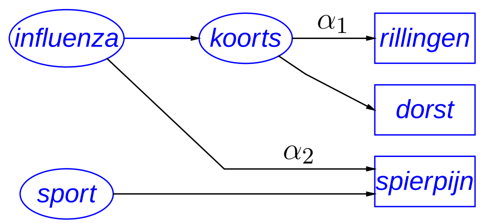
- Sterke causaliteit: $O \rightarrow G$
  "Als O aanwezig is, dan *moet* G aanwezig zijn"
- Zwakke causaliteit: $O \wedge \alpha \rightarrow G$
  "Als O aanwezig is, dan *kan* G aanwezig zijn" (α is onvolledigheidsassumptie)
- Voorbeeld:
  koorts $\wedge\ \alpha_1 \longrightarrow$ *rillingen*
  koorts $\longrightarrow$ *dorst*
  sport $\longrightarrow$ *spierpijn*

Voorspelling



- Causale specificatie: $\Sigma = (\Delta, \Phi, \mathcal{R})$, met:
  - Δ: mogelijke oorzaken en onvolledigheidsassumpties
  - Φ: feiten die geobserveerd kunnen worden
  - $\mathcal{R}$: causaal model
- Voorspelling $V \subseteq \Delta$: $\boxed{\mathcal{R} \cup V \vDash E}$,
  met $E \subseteq \Phi$ (E is observeerbaar)

Voorbeelden voorspelling

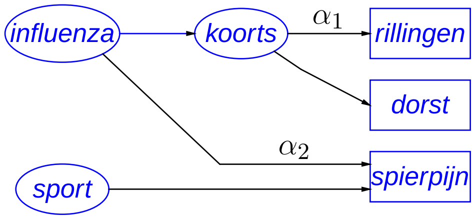


Causale specificatie: $\Sigma = (\Delta, \Phi, \mathcal{R})$

- Voorbeeld 1: $\mathcal{R} \cup \{$*influenza*$, \alpha_1\} \vDash \{$*rillingen*, *dorst*$\}$
- Voorbeeld 2:

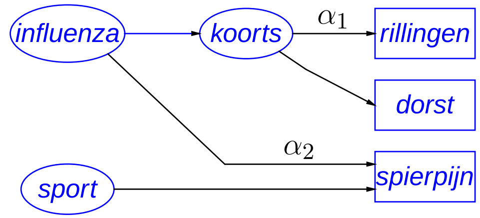
  $\mathcal{R} \cup \{$*influenza*$, \alpha_1, \alpha_2\} \vDash$
  $\{$*rillingen*, *dorst*, *spierpijn*$\}$
- Voorbeeld 3: $\mathcal{R} \cup \{$*sport*$\} \vDash$ *spierpijn*

Diagnostisch probleem



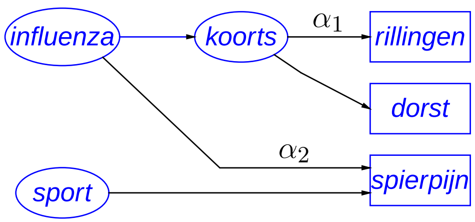
- Causale specificatie: $\Sigma = (\Delta, \Phi, \mathcal{R})$
- Geobserveerde feiten: $F = \{\textit{spierpijn}, \textit{dorst}\}$
- Diagnose D?
  (1) Voorspelling die F *verklaart*: $\mathcal{R} \cup D \vDash F$
  (2) $\cdots$ maar niet te veel verklaart
- Voorbeelden van diagnosen: $D = \{\textit{influenza}, \alpha_2\}$, $D' = \{\textit{sport}, \textit{influenza}\}$ en $D'' = \{\textit{influenza}, \alpha_1\}$?

Niet te veel verklaren!



- Causale specificatie: $\Sigma = (\Delta, \Phi, \mathcal{R})$
- Geobserveerde feiten: $F = \{\textit{spierpijn}, \textit{dorst}\}$
- Feiten die *niet* verklaard mogen worden: $C = \{\neg\textit{rillingen}\}$
- Formeel: $D \subseteq \Delta$ is een *diagnose*, als:
  (1) $\mathcal{R} \cup D \vDash F$ (coveringconditie)
  (2) $\mathcal{R} \cup D \cup C \nvDash \bot$ (consistentieconditie)

Consistentieconditie



- Causale specificatie: $\Sigma = (\Delta, \Phi, \mathcal{R})$
- Geobserveerde feiten: $F = \{\textit{spierpijn}, \textit{dorst}\}$
- Feiten die *niet* verklaard mogen worden:

$$C = \{\neg\textit{rillingen}\}$$

$\mathcal{R} \cup \{\textit{influenza}, \alpha_1, \alpha_2\} \cup \{\neg\textit{rillingen}\} \vDash \bot$

$\Rightarrow D = \{\textit{influenza}, \alpha_1, \alpha_2\}$ *geen* diagnose

Abductie = anticausaal redeneren

Abductie:

$$\frac{\text{Gevolg},\ \text{Oorzaak} \rightarrow \text{Gevolg}}{\text{Oorzaak}}$$

Idee: *omkering van causale relatie*

Voorbeeld:

koorts $\rightarrow$ *dorst* resulteert in *dorst* $\rightarrow$ *koorts*

Dus:

$$\{\textit{dorst} \rightarrow \textit{koorts}, \textit{dorst}\} \vDash \textit{koorts}$$

Conclusie:

Abductie = deductie met implicatie omkering

Abductie en deductie

Omkering van causale relaties in $\mathcal{R}$ en toevoegen aan $\mathcal{R}$ heet de completion van $\mathcal{R}$; notatie: $\mathrm{COMP}(\mathcal{R})$

Voorbeeld:

$$\mathcal{R} = \{P \wedge Q \rightarrow V, \\ T \rightarrow V, \\ T \rightarrow U\}$$

resulteert in $\mathrm{COMP}(\mathcal{R}) = \{V \leftrightarrow ((P \wedge Q) \vee T), U \leftrightarrow T\}$

Zij V geobserveerd: $\mathrm{COMP}(\mathcal{R}) \cup \{V\} \vDash ((P \wedge Q) \vee T)$

i.e. twee alternatieve diagnosen: $(P \wedge Q)$ and T

Conclusie: Abductie = deductie in a *completed theorie*

Conclusies

- Model-gebaseerd redeneren bruikbaar als in domein modellen beschikbaar zijn
- Consistentie-gebaseerde diagnose: geen of geringe ervaring aanwezig met problemen in domein
- Abductieve diagnose: causale modellen van afwijkend gedrag beschikbaar
- Integratie van consistentie-gebaseerde en abductieve diagnose is mogelijk
- Relatie tussen abductieve en consistentie-gebaseerde diagnose enerzijds, en probabilistische diagnose met Bayesiaanse netwerken anderzijds