Model: representation of normal or abnormal behaviour and, possibly, internal structure

Formalisation of model-based diagnosis:
- *consistency-based diagnosis* (normal behaviour), and
- *abductive diagnosis* (abnormal behaviour)
Abductive diagnosis

Correspondence between predicted *abnormal* behaviour and observed behaviour $\Rightarrow$ defect!

Originators:
- D. Poole, “Explanation and prediction: an architecture for default and abductive reasoning”, *Computational Intelligence*, vol. 5, nr. 2, 97–110, 1989
Causal models

Causality: combination of causes gives rise to effects
- flu causes fever
- fever causes chills
- fever causes thirst
- sport also causes myalgia
- flu causes myalgia

Using logic: \((\text{Cause}_1 \land \cdots \land \text{Cause}_n) \rightarrow \text{Effect}\)

Example:
- \(\text{fever} \rightarrow \text{chills}\)
- \(\text{fever} \rightarrow \text{thirst}\)
- \(\text{sport} \rightarrow \text{myalgia} \cdots\)
Causality and implication

\( x \text{ causes } y: \text{ Causes}(x, y) \)

Axiomatisation:

- **transitivity:**
  \[ \forall x \forall y \forall z ((\text{Causes}(x, z) \land \text{Causes}(z, y)) \rightarrow \text{Causes}(x, y)) \]

- **antisymmetry:**
  \[ \forall x \forall y (\text{Causes}(x, y) \rightarrow \neg \text{Causes}(y, x)) \]

- **reflexivity:**
  \[ \forall x \text{ Causes}(x, x) \text{ (by definition this excludes antisymmetry)} \]

With implication: \( x \text{ causes } y \equiv x \rightarrow y \)

- **transitivity ✓:** \( \{P \rightarrow Q, Q \rightarrow R\} \models P \rightarrow R \)

- **no antisymmetry but contraposition:** \( \{P \rightarrow Q, \neg Q\} \models \neg P \)
  \( (P \rightarrow Q \equiv \neg Q \rightarrow \neg P) \)

- **reflexivity ✓:** \( \models P \rightarrow P \)
Weak and strong causality

- **Strong causality:** $C \rightarrow E$
  "If $C$ present, then $E$ **must** also be present"

- **Weak causality:** $(C \land \alpha) \rightarrow E$ or simplified $C \land \alpha \rightarrow E$
  "If $C$ present, then $E$ **may** be present" ($\alpha$ is incompleteness assumption)
Weak and strong causality

**Weak causality** ("may cause"): $C \land \alpha \rightarrow E$
($\alpha$ is incompleteness assumption)

**Example:**

- $\text{fever} \land \alpha_1 \rightarrow \text{chills}$
- $\text{fever} \rightarrow \text{thirst}$
- $\text{sport} \rightarrow \text{myalgia}$
Generalisation

- Representation for classes of cases
- \( \forall x (S(x) \rightarrow S'(f(x))) \), with \( S \) and \( S' \) states

\[
\begin{array}{c|c}
  x & f(x) \\
  \vdots & \vdots \\
\end{array}
\]

- \( \forall x (\text{Interest}(x) \rightarrow \text{Rate}(\text{Stocks}(x))) \)

\[
\begin{array}{c|c}
  x & \text{Stocks}(x) \\
  \text{high} & \text{low} \\
  \text{low} & \text{high} \\
\end{array}
\]
**Causal specification:** $\Sigma = (\Delta, \Phi, \mathcal{R})$, with:
- $\Delta$: potential causes and incompleteness assumptions
- $\Phi$: facts that can be observed
- $\mathcal{R}$: causal model

**Prediction** $V \subseteq \Delta$: $\mathcal{R} \cup V \vDash E$, with $E \subseteq \Phi$ ($E$ can be observed)
Example prediction

Causal specification: $\Sigma = (\Delta, \Phi, R)$

- **Example 1:** $R \cup \{flu, \alpha_1\} \models \{chills, thirst\}$
- **Example 2:**
  $$R \cup \{flu, \alpha_1, \alpha_2\} \models \{chills, thirst, myalgia\}$$
- **Example 3:** $R \cup \{sport\} \models myalgia$
Diagnostic problem

Causal specification: $\Sigma = (\Delta, \Phi, \mathcal{R})$

Observed facts: $F = \{\text{myalgia, thirst}\}$

Diagnosis $D$?

1. Prediction that explains $F$, formal: $\mathcal{R} \cup D \models F$
2. ··· but that does not explain too much

Example diagnoses: $D = \{\text{flu, } \alpha_2\}$, $D' = \{\text{sport, flu}\}$ and $D'' = \{\text{flu, } \alpha_1\}$?
Don’t explain too much!

Causal specification: $\Sigma = (\Delta, \Phi, \mathcal{R})$

Observed facts: $F = \{\text{myalgia, thirst}\}$

Facts that should \textit{not} be explained: $C = \{\neg\text{chills}\}$

Formal: $D \subseteq \Delta$ is a \textit{diagnosis} if:

1. $\mathcal{R} \cup D \models F$ (covering condition)
2. $\mathcal{R} \cup D \cup C \not\models \bot$ (consistency condition)
Causal specification: \( \Sigma = (\Delta, \Phi, \mathcal{R}) \)

Observed facts: \( F = \{ \text{myalgia, thirst} \} \)

Facts that should not be explained:

\[ C = \{ \neg \text{chills} \} \]

\[ \mathcal{R} \cup \{ \text{flu, } \alpha_1, \alpha_2 \} \cup \{ \neg \text{chills} \} \models \bot \]

\[ \Rightarrow D = \{ \text{flu, } \alpha_1, \alpha_2 \} \text{ no diagnosis} \]
Abduction = anticausal reasoning

Abduction:

\[
\begin{array}{c}
\text{Effect, Cause} 
\rightarrow 
\text{Effect} \\
\text{Cause}
\end{array}
\]

Idea: *reversal of the causal relation*

Example:

\[\text{fever} \rightarrow \text{thirst results in thirst} \rightarrow \text{fever}\]

Thus:

\[\{\text{thirst} \rightarrow \text{fever, thirst}\} \models \text{fever}\]

Conclusion:

Abduction = deduction with implication reversal
Abduction and deduction

Reversal of the causal relations in $\mathcal{R}$ and addition to $\mathcal{R}$ is called the completion of $\mathcal{R}$

Basic idea:

$d_1 \to f$
$d_2 \to f$

indicates that $d_1$ and $d_2$ are possible explanations for $f$;

$$f \to (d_1 \lor d_2)$$

makes this explicit

Together:

$$\{d_1 \to f, d_2 \to f, f \to (d_1 \lor d_2)\} \equiv \{f \leftrightarrow (d_1 \lor d_2)\}$$
Abducibles

\[ \Delta: \text{defects, some of them derivable from other defects, some not:} \]

- abducible: defects \( d \) not derivable
- non-abducible: each defect that can be derived

\[ \Phi: \text{findings, also non-abducible} \]

\( A \) are abducibles; \( N \) are non-abducibles

Example:

\[
\begin{align*}
  d_1 & \rightarrow d_2 \\
  d_3 & \rightarrow f_1 \\
  d_2 & \rightarrow f_2
\end{align*}
\]

\( A = \{d_1, d_3\} \); \( N = \{d_2, f_1, f_2\} \)
Completion

\[ \mathcal{R} = \{ \varphi_{1,1} \rightarrow n_1, \ldots, \varphi_{1,n_1} \rightarrow n_1, \]
\[ \quad \vdots \]
\[ \varphi_{m,1} \rightarrow n_m, \ldots, \varphi_{m,n_m} \rightarrow n_m \} \]

\[ N = \{ n_i \mid 1 \leq i \leq m \} \] is the set of non-abducible literals, and

each \( \varphi_{i,j} \) denotes a conjunction of defect literals, possibly including an assumption literal

Predicate completion of \( \mathcal{R} \) with respect to \( N \):

\[ \text{COMP}[\mathcal{R}; N] = \mathcal{R} \cup \{ n_1 \rightarrow \varphi_{1,1} \lor \cdots \lor \varphi_{1,n_1}, \]
\[ \quad \vdots \]
\[ n_m \rightarrow \varphi_{m,1} \lor \cdots \lor \varphi_{m,n_m} \} \]
Example:

Example:

\[ \mathcal{R} = \{ P \land Q \rightarrow V, \]
\[ T \rightarrow V, \]
\[ T \rightarrow U \} \]

with \( N = \{ V, U \} \) results in

\[ \text{COMP}[\mathcal{R}; N] = \{ V \leftrightarrow ((P \land Q) \lor T), U \leftrightarrow T \} \]

Let \( V \) be observed: \( \text{COMP}[\mathcal{R}; N] \cup \{ V \} \models ((P \land Q) \lor T) \)

i.e. two alternative diagnoses: \( (P \land Q) \) and \( T \)

Conclusion: Abduction = deduction in a completed theory
Deduction of the solutions

- $\mathcal{P} = (\Sigma, F)$ is an abductive diagnostic problem
- $\text{COMP}[\mathcal{R}; N]$ is the predicate completion of $\mathcal{R}$ with respect to $N$, the set of non-abducible literals in $\mathcal{P}$

A solution formula $S$ for $\mathcal{P}$ is defined as the most specific formula consisting only of abducible literals, such that

$$\text{COMP}[\mathcal{R}; N] \cup F \cup C \models S$$

where $C$ is defined as:

$$C = \{ \neg f \in \Phi \mid f \in \Phi, f \notin F, f \text{ is a positive literal} \}$$
Solution formula

**Theorem.** Let $\mathcal{P} = (\Sigma, F)$ be an abductive diagnostic problem. Let $C$ be constraints and $S$ be a solution formula for $\mathcal{P}$. Let $H \subseteq \Delta$ be a set of abducible literals, and let $I$ be an interpretation of $\mathcal{P}$, such that for each abducible literal $a \in A$: $\models_I a$ iff $a \in H$. Then, $H$ is a solution to $\mathcal{P}$ iff $\models_I S$.

**Proof.**
Conjuncts in $S$ are equivalent to observed findings $f \in F$, that are logically entailed by $\mathcal{R} \cup H$, or to non-observed findings $\neg f \in C$ that are consistent with $\mathcal{R} \cup H$. Hence, an interpretation $I$ for which $\models_I H$, that falsifies each abducible in $\Delta \setminus H$, satisfying every $f \in F$ and each $\neg f \in C$ that has been rewritten, must satisfy this collection of conjuncts, i.e. $S$. 
Example

\[ R:\]

\begin{align*}
\text{fever} \land \alpha_1 & \rightarrow \text{chills} \\
\text{flu} & \rightarrow \text{fever} \\
\text{fever} & \rightarrow \text{thirst} \\
\text{flu} \land \alpha_2 & \rightarrow \text{myalgia} \\
\text{sport} & \rightarrow \text{myalgia}
\end{align*}
Example

\[ \text{COMP}[\mathcal{R}; \{ \text{chills, thirst, myalgia, fever} \}] \]

\[ = \mathcal{R} \cup \{ \text{chills} \to \text{fever} \land \alpha_1, \]
\[ \text{fever} \to \text{flu}, \text{thirst} \to \text{fever}, \]
\[ \text{myalgia} \to (\text{flu} \land \alpha_2) \lor \text{sport} \} \]

\[ = \{ \text{chills} \leftrightarrow \text{fever} \land \alpha_1, \]
\[ \text{fever} \leftrightarrow \text{flu}, \]
\[ \text{thirst} \leftrightarrow \text{fever}, \]
\[ \text{myalgia} \leftrightarrow (\text{flu} \land \alpha_2) \lor \text{sport} \} \]

Note that

\[ \text{COMP}[\mathcal{R}; \{ \text{chills, thirst, myalgia, fever} \}] \cup F \cup C \models \]
\[ S \equiv (\text{flu} \land \alpha_2) \lor (\text{flu} \land \text{sport}) \]

given that \( F = \{ \text{thirst, myalgia} \} \) and \( C = \{ \neg \text{chills} \} \)
Example

\[ \text{COMP}[R; \{ \text{chills, thirst, myalgia, fever} \}] \]

\[ = R \cup \{ \text{chills} \rightarrow \text{fever} \land \alpha_1, \quad \text{fever} \rightarrow \text{flu}, \text{thirst} \rightarrow \text{fever}, \quad \text{myalgia} \rightarrow (\text{flu} \land \alpha_2) \lor \text{sport} \} \]

\[ = \{ \text{chills} \leftrightarrow \text{fever} \land \alpha_1, \quad \text{fever} \leftrightarrow \text{flu}, \quad \text{thirst} \leftrightarrow \text{fever}, \quad \text{myalgia} \leftrightarrow (\text{flu} \land \alpha_2) \lor \text{sport} \} \]

\[ \text{COMP}[R; \{ \text{chills, thirst, myalgia, fever} \}] \cup F \cup C \models \neg (\text{fever} \land \alpha_1) \]

because \( \{ \neg \text{chills}, \text{chills} \leftrightarrow (\text{fever} \land \alpha_1) \} \models \neg (\text{fever} \land \alpha_1); \quad \neg (\text{fever} \land \alpha_1) \text{ is not part of } S, \text{ because fever is non-abducible} \]
Set covering diagnosis

\[ \mathcal{N} = (\Delta, \Phi, C) \] is called a causal net, where:
- \( \Delta \) is a set of possible defects,
- \( \Phi \) is a set of elements called observable findings, and
- \( C \) is a binary relation

\[ C \subseteq \Delta \times \Phi \]

called the causation relation

A diagnostic problem in the set-covering theory of diagnosis: \( \mathcal{D} = (\mathcal{N}, F) \), where \( F \subseteq \Phi \) is a set of observed findings
Further notions

From defects to causes and vice versa:

- **effects function** \( e : \wp(\Delta) \to \wp(\Phi) \) is defined as follows: for each \( D \subseteq \Delta \):

\[
e(D) = \bigcup_{d \in D} e(\{d\})
\]

where \( e(\{d\}) = \{f \mid (d, f) \in C\} \)

- **causes function** \( c : \wp(\Phi) \to \wp(\Delta) \) is defined as follows: for each \( E \subseteq \Phi \):

\[
c(E) = \bigcup_{f \in E} c(\{f\})
\]

where \( c(\{f\}) = \{d \mid (d, f) \in C\} \)
Example

\[ e(D) = \bigcup_{d \in D} e(\{d\}) \]

where

\[ e(\{d\}) = \begin{cases} 
\{\text{cough, fever, sneezing}\} & \text{if } d = \text{flu} \\
\{\text{cough, sneezing}\} & \text{if } d = \text{common cold} \\
\{\text{fever, dyspnoea}\} & \text{if } d = \text{pneumonia} 
\end{cases} \]
Set-covering diagnosis

Let $\mathcal{D} = (\mathcal{N}, F)$ be a diagnostic problem, where $F$ denotes a set of observed findings. Then, a set-covering diagnosis of $\mathcal{D}$ is a set of defects $D \subseteq \Delta$, such that:

$$e(D) \supseteq F$$

Let $F = \{cough, fever\}$ then

$$D_1 = \{flu\}$$

is a diagnosis, but

$$D_2 = \{flu, common\ cold\}$$
$$D_3 = \{common\ cold, pneumonia\}$$

and $D_4 = \{flu, common\ cold, pneumonia\}$ are also diagnoses for $F$
Mapping to abductive diagnosis

Define

\[ e(\{d\}) = \{f_1, \ldots, f_n\} \]

and construct for each \( d \in \Delta \):

\[ d \land \alpha_{f_1} \rightarrow f_1 \]
\[ d \land \alpha_{f_2} \rightarrow f_2 \]
\[ \vdots \]
\[ d \land \alpha_{f_n} \rightarrow f_n \]

Note no interactions between defects!

\[ e(D) \supseteq F \iff \mathcal{R} \cup H \models F \text{ and } \mathcal{R} \cup H \not\models \bot, \text{ with } D \text{ defects in } H \]
Alternative diagnostic definitions

- **Minimal cardinality:** a diagnosis $D$ of $F$ is an explanation of $D$ iff it contains the minimum number of elements among all diagnoses of $F$.

- **Irredundancy:** a diagnosis $D$ of $F$ is an explanation of $D$ iff no proper subset of $D$ is a diagnosis of $F$.

- **Relevance:** a diagnosis $D$ of $F$ is an explanation of $D$ iff $D \subseteq c(F)$.

- **Most probable diagnosis:** a diagnosis $D$ of $F$ is an explanation of $D$ iff $P(D|F) \geq P(D'|F)$ for any diagnosis $D'$ of $F$.

- A diagnosis $D$ is called a **minimal-cost explanation** of $D$ iff $\sum_{d \in D} \text{cost}(d) \leq \sum_{d \in D'} \text{cost}(d)$.
Example

Diagnoses $D_i$, $i = 1, \ldots, 4$, are relevant diagnoses, because $c(\{cough, fever\}) \supseteq D_i$

Irredundant diagnoses of $F$ are $D_1$ and $D_3$

There is only one minimal cardinality diagnosis: $D_1$
Diagnostic problem in AILog (developed by David Poole):

- facts, denoted by $\text{FACTS}$
- a set of hypotheses, denoted by $\text{HYP}$, and
- a set of constraints, denoted by $C$

$\text{FACTS}$ and constraints $C$ are formulae in first-order logic; hypotheses act as abducibles = assumables in AILog

A set $\text{FACTS} \cup H$ is called an explanation of a closed formula $g$, where $H$ is a set of ground instances of hypothesis elements in HYP, iff:

(1) $\text{FACTS} \cup H \models g$, and

(2) $\text{FACTS} \cup H \cup C \not\models \bot$. 
Example

assumable a1.
assumable a2.
assumable fever.
assumable flu.
assumable sport.

chills <- fever & a1.
fever <- flu.
thirst <- fever.
myalgia <- flu & a2.
myalgia <- sport.

false <- chills. % constraint
Calling AILog

ailog: create_nogoods. % enforce consistency
ailog: ask thirst & myalgia.

Yields the following results:
Answer: thirst & myalgia.
Assuming: [a2, fever, flu].
  [more,ok,how,help]: more.
Answer: thirst & myalgia.
Assuming: [fever, sport].
  [more,ok,how,help]: more.
Answer: thirst & myalgia.
Assuming: [a2, flu].
  [more,ok,how,help]: more.
Answer: thirst & myalgia.
Assuming: [flu, sport].
  [more,ok,how,help]: more.
No more more answers.
Conclusions

- Model-based (consistency-based and abductive) reasoning is suitable if there are domain models available

- Consistency-based diagnosis: no or only scarce knowledge of problems in domain

- Abductive diagnosis: causal models of abnormal behaviour available

- Integration of consistency-based and abductive diagnosis is possible

- Relationship with Bayesian networks