Approximate anytime inference:
Half an answer on time is better than a perfect answer too late

Frank van Harmelen
Vrije Universiteit Amsterdam

Who am I (advance warning)

- I’m not a Theoretical Computer Scientist
- I do Knowledge Representation & Reasoning
- often dealing with intractable problems
- user of theory
Robust Knowledge Representation

Reliance on logic is a **strength**
- Strong theoretical basis
- Well known properties
- Well known implementation techniques

Reliance on logic is a **weakness**
- **Crisp** (no approximate answers)
- **Abrupt** (no intermediate answers)
- **Inefficient** (no time/quality trade-off)
Why do we want \( \uparrow \rightarrow \) ?

time pressure, hard deadlines

perfect answer not needed

Another way of stating the problem:

Logic = a model of a perfect reasoner in idealised circumstances

- no wrong steps
- no missing knowledge
- no incorrect knowledge
- **unlimited time**
**Approach: Anytime Inference**

- **Quality** is function of some varying resource
  - reasoning time,
  - inference accuracy,
  - representational precision

- This function is
  - monotonic
  - *diminishing returns* (usually)
  - optimal at some time $T$
  - this $T$ not $\gg$ best abrupt algorithm
  - characterised by a performance profile

---

**How to use this?**

- many problems can be stated in terms of $\vdash$:
  - planning
  - learning
  - classifying
  - **diagnosing**
  - ...

- General idea: replace $\vdash$ with $\sim$
Two example studies

For “reasoning” take diagnosis
- replace \( \vdash \) by \( \models \)
- \( \models = 1,3-S \) (Cadoli & Schaerf)
- \( \models = BCP-k \) (Dalal)

Characterise the effects of changing \( S \) and \( k \)

\( \models^1, \models^3 \) (Cadoli & Schaerf)

Take subset \( S \) of propositional letters
Non-classical behaviour for letters \( x \) outside \( S \):

1-S assignment:
- \( x \) and \( \neg x \) are both false
- only 1 assignment possible for \( x \notin S \)
- intuition on clausal form: remove part of clause

\[
\begin{align*}
x \lor b, \neg b \lor c & \Rightarrow b, \neg b \lor c
\end{align*}
\]

- Note: theory might become \( \bot \)

(NB: For letters in \( S \): classical behaviour)
$\sim^1, \sim^3$ (Cadoli & Schaerf)

Take subset $S$ of propositional letters
Non-classical behaviour for letters $x$ outside $S$:

- 3-S assignment:
  - $x$ and $\neg x$ are not both false
  - 3 assignments possible for $x \notin S$
  - intuition on clausal form: remove entire clause

\[ a \lor b, \neg b \lor c \Rightarrow \neg b \lor c \]

- Note: theory might become $T$

$\sim^1, \sim^3$ (Cadoli & Schaerf) (ctd)

- Properties:
  - $\sim^1$ is unsound but complete
  - $\sim^3$ is incomplete but sound
  - $S$ determines “degree of approximation”:

\[
\emptyset \Rightarrow \vdash_S \Rightarrow \vdash_{S'} \Rightarrow \vdash_2 \Rightarrow \vdash_{S'} \Rightarrow \vdash_S \Rightarrow \vdash_1 \Rightarrow \vdash_{\emptyset}
\]

\[
\vdash_2 \iff \vdash_{S'} \iff \vdash_S \iff \vdash_1
\]

- Efficient anytime algorithms
- total iterative cost $\leq$ cost of $\vdash_2$
Example of $\sim^3$

$T \cup \{H_0, H_5\} \not\vdash O_3$

if $H_5 \not\in S$ then

$T \cup \{H_0, H_5\} \not\models_3 O_3$

incomplete

Example of $\sim^1$

$T \cup \{H_0\} \not\models O_3$

if $H_5 \not\in S$ then

$T \cup \{H_0\} \not\models_1 O_3$

unsound
Summary

- Anytime behaviour when $S_i$ is increased
- Previous steps can be reused

Standard def. of diagnosis

Given:
- Behaviour model: $T$
- Observations to be explained: $O^+$
- Observations to be consistent: $O^-$

Find: Explanation: $E$ such that:

\[
\begin{align*}
T \cup E \vdash O^+ \\
T \cup E \cup O^+ & \not\vdash \perp \\
T \cup E \cup O^- & \not\vdash \perp
\end{align*}
\]

\[\text{ABD} \quad \text{CBD}\]
Theorems using $\sim^{1,3}$:

- **ABD$_3$ shrinks** when $S$ increases
- **CBD$_1$ grows** when $S$ increases
- Any **CBD$_1$ is contained in** a classical diagnosis
- Any **ABD$_3$ contains** classical diagnoses

\[
\phi = \{ABD_1^0\} \subseteq \{ABD_1^S\} \circ \neq \{ABD_2^2\} \\
\{ABD_2^2\} \supset \{ABD_3^S\} \not\subset \{ABD_3^0\} = \emptyset
\]

\[
\emptyset = |\{ABD_1^0\}| \leq |\{ABD_1^S\}| \leq |\{ABD_2\}| \\
|\{ABD_2\}| \geq |\{ABD_3^S\}| \geq |\{ABD_3^0\}| = 0
\]

Approximate Entailment

- Semantically well-founded
- Computationally attractive
- Dual

- Parameter $S$ is crucial for appropriate behaviour
- Almost no quantitative analysis
How to choose S?

- No good general strategy exists
- Problem-specific heuristics:
  - \( S = \{\text{urgent causes}\} \)
    first diagnosis is most urgent candidate
  - \( S = \{\text{unreliable components}\} \)
    first diagnosis is most likely candidate
  - \( S = \{\text{specific observations}\} \)
    first diagnosis is most specific candidate

Second case study: \( \text{BCP}_k \)

\( \text{BCP}_0: \)
- use all literals + a clause from \( T \) to derive a new literal \( \alpha \)
- add \( \alpha \) to \( T \)

Example:

\[
(P \lor Q), (P \lor \neg Q), (\neg P \lor Q), (\neg P \lor \neg Q) \vdash \perp \\
(P \lor Q), (P \lor \neg Q), (\neg P \lor Q), (\neg P \lor \neg Q) \not\vdash \perp
\]
Basic notions: $\text{BCP}_0$ & $\text{BCP}_k$ (ctd)

- Example:
  
  $\text{Th} = (P \lor Q), (P \lor \neg Q), (\neg P \lor S \lor T), (\neg P \lor S \lor \neg T)$

  1. $\text{Th} \vdash_0 P$
  2. $\text{Th}, P \vdash_0 S$  **but**
  3. **not** $\text{Th} \vdash_0 S$
  4. $\vdash_0$ **cannot chain**

Basic notions $\text{BCP}_0$ & $\text{BCP}_k$ (ctd)

- $\text{BCP}_k$ = allow chaining on length $\leq k$

\[
\begin{array}{c}
T \vdash_0 \phi \\
\hline
T \vdash_k \psi ; T, \psi \vdash_k \phi \\
\vdash_k \phi
\end{array}
\]

- **if** $|\psi| \leq k$
- When $k$ **grows**, $\text{BCP}_k$ is **more complete**
- For some $k$, $\text{BCP}_k$ is complete
Examples of BCP<sub>k</sub> diagnosis

Simple theorems

Small <i>k</i>: fewer diagnostic problems

- if ... ⊨ <i>k</i>⊥ then ... ⊨ <i>k</i>⊥

Small <i>k</i>: more diagnostic solutions

- \{E \mid E \text{ is a } BCP_{k+1} \text{ solution}\} ⊆ \{E \mid E \text{ is a } BCP_k \text{ solution}\}
Thm: Upperbound on $k$ (surprise)

- For a given component $c$:
  - $n = \#$ of unknown inputs & outputs of $c$
- For inferences on $c$: $\text{BCP}_{n-1} = \text{classical}$

Corrollary: take $k' = \max \ n-1$ over all $c$
- Then: $\{\text{BCP}_{k'} \ E\} = \{\text{classical } E\}$
- NB: independent of circuit size, independent of component complexity only dependent on component size!

Thm: Sharper upperbound on $k$

- Def: ($mcs = \text{minimal conflict set}$)
  - $k(mcs) = \text{minimal } k \text{ needed to prove}$
    $$... \cup mcs \vdash_{k} \bot$$
  - $k^* = \max k(mcs)$ over all $mcs$
- Then: $\{\text{BCP}_{k^*} \ E\} = \{\text{classical } E\}$

Of course:
- $k^* \leq k'$
- $k^*$ not practical
**Low $k$ can do a lot (surprise)**

Thm: for problem diagnosis

$k=1$ suffices for all binary circuits

Thm: for problem recognition

$k=0$ suffices for all known-input circuits

---

**Open questions**

- How tight is $k'$ ($k' \geq k^*$) ?
- Is $k$ a good complexity indicator ?

- Both are empirical questions ?
Analysis of performance profiles

- Qualitative:

```
\begin{align*}
\text{recall} & \rightarrow \text{time} \\
\text{precision} & \rightarrow \text{time} \\
\text{assignments} & \rightarrow \text{time}
\end{align*}
```

Take home message:

- Approximate forms of deduction
- Anytime algorithm for deduction
- Realistic use of these algorithms

\[ Q \uparrow \rightarrow \text{T} \]

\[ Q \uparrow \rightarrow \text{T} \]