Today

- de Kleer's assumption based truth maintenance
- Belief Revision

Reason maintenance

Recall we look at reasoning maintenance systems:

- to keep track of dependencies in output of a (different) reasoning system
- so that we can quickly repair when things go wrong, and make changes safely and quickly

de Kleer's TMS

de Kleer aims to record the reasoning dependencies, rather than the IN/OUT of Doyle. Instead, the system records only how deductions would depend on any assumptions. That is, the job of ATMS is to record under what sets of assumptions any conclusion holds. In ATMS, each node has, in addition to the tag showing what it stands for (what you would normally call its label), something else which de Kleer confusingly terms a label; this is a set of sets of assumptions. The node follows from any one of such sets of assumptions if every assumption in the set holds.

If a node turns out to be contradictory, according to the reasoning system, then the ATMS notes that all the sets of assumptions labelling it lead to a contradiction.

Noting contradictions

The convenient way to do this noting is to have a node which stands for 'false', and to add any such contradictory sets to its label. For each node, a check is made such that none of its sets of assumptions is a superset of any other; if so, it should be deleted. The ATMS does this to ensure that no superfluous assumption appears anywhere in a set within any label.

Thus, if one argument uses only a subset of another argument's assumptions, then we forget about the more extravagant argument.

How it works

To have a look at the idea of how ATMS functions, suppose that early on in the deductive process (performed by a reasoning engine) there have been nodes for assumptions $a_1 \ldots a_5$, and that there are already two nodes A and B, with labels as follows

A	$\{ a_1, a_2 \}\{ a_2, a_5 \}$
B	$\{ a_1 \}\{ a_2, a_3, \}\{ a_4 \}$

Suppose the false node contains $\{ a_4, a_5 \}$.

Suppose we want to add a justification that A and B imply C (C a new node).

Computing new label

First, find the pairwise sets of assumptions, one set for pair from each label:

1 $\{ a_1, a_2 \}$
2 $\{ a_1, a_2, a_3 \}$
3 $\{ a_1, a_2, a_4 \}$
4 $\{ a_1, a_2, a_5 \}$
5 $\{ a_2, a_3, a_5 \}$
6 $\{ a_2, a_4, a_5 \}$

Now remove any that has another as a subset (here, 2,3,4).

Now remove any with contradictory assumptions as a subset (here, 6).

So get new label (blend with old label, if C already there).

Blending labels

Take the union of the two labels, delete any set in this union which has another member in the union as a subset of it.

Since the new and old labels were already consistent there is no need to go looking for sets which have contradictory subsets, this time. If the whole process changes the label for C, and recorded justifications show that other nodes depended on C, then the label changes have to be propagated forward to those other nodes, and on from them, and so on.

Propagating information about inconsistency

Suppose, at a certain stage, C is found to be a nogood by the reasoning system (that is, every set of assumptions labelling it indicates a contradiction).

Then each member of its label is added to the label of the specific node standing for 'false', and all such members, and their supersets, are removed from the label of every other node.

Any set of (inconsistent) assumptions being a superset of another within the 'false' node's label itself is also removed.

Advantages

First, all the consequences of a set of assumptions can be explored together; no backtracking is involved, and the system is not striving to maintain one mutually consistent set of assumptions as in Doyle's TMS. This suits certain kinds of application.

Second, it can be very efficiently implemented, since the main operations involved are set operations such as union and subset checking. If sets are represented as bit strings these kinds of operations can be performed directly by hardware.

Problems?

The obvious disadvantage is that the system is driven by forward-chaining reasoning, so there is no natural progression towards a particular desired goal.

There is considerably more to ATMS than this; we have only covered the basic idea. For example, in some applications it might be desirable to constrain the ATMS to considering only those sets of assumptions which contain at least one, or perhaps exactly one, of a given set (of particular assumptions).

Justified or coherent?

So far we have looked at belief revision from a logical and algorithmic point of view. There are other ways to approach the topic, taking more cognitive aspects into consideration.

Two possibilities: base the revision on

- *foundations*: propositions are believed if they have justifications (e.g. derivations)
- *coherence*: propositions fit together coherently (e.g. they are not contradictory); as many beliefs as possible should be retained.

TMSs are foundational; we look now at the coherence approach.

Choosing between revisions

An important element from coherence theories is the basis for choice in situations when logic and the structure of justifications alone allow more than one way of revising the system of beliefs, represented as a belief set.

If, by the acquisition of new information, a statement f and its negation $\neg f$ both become believed, there are two ways of reaching a new consistent belief set. Either f and beliefs supporting it or $\neg f$ and beliefs supporting it could be dropped.

One way to approach this is to use a preference ordering over beliefs in order to decide in such conflicts. This ordering is called *epistemic entrenchment*.

Entrenched beliefs

Deciding a conflict (of believing both f and $\neg f$) requires going back in a chain of reasoning and deciding on the relative merits of beliefs that underlie the conflicting beliefs.

These are very likely to impinge not only on the present problem, but on a larger part of the reasoner's system of belief.

And it seems that among these, *some beliefs* prove to be **more tenacious** than others.

The more tenacious beliefs will usually be those that are *more central to the reasoner's whole system of beliefs*, that are *more useful for his general style of argumentation*.

This phenomenon is well documented for scientific as well as common-sense reasoning.

Example

Suppose we believe

1. All European swans are white
2. The bird in the trap is a swan
3. The bird in the trap comes from Sweden
4. Sweden is in Europe

And we look and see that the bird in the trap is black.

A reasoning system (that knows that black and white are incompatible properties) will spot a contradiction.

Representation of beliefs and inference

A logic-based formalism is used to represent belief. Beliefs are modeled as statements in (a restricted) first-order logic. The deductive process in this logic is modeled in an appropriate meta-theory.

The reasoner starts from a set of assumptions, which are taken to be self-evident beliefs that need no justification. Only formulae derivable from these are believed. The assumptions are distinguished from other beliefs by the meta-theory. The agent's beliefs are assumed to be closed under the inference system used.

If the agent's theory includes some formula f and its negation $\neg f$, this contradiction is resolved by choosing a theory that only includes one of them.

Representing entrenchment

This choice is made from a meta level viewpoint, because properties outside the object level logic of beliefs (the assumptions' degrees of epistemic entrenchment) serve as decision criteria.

In order to allow this, reasoning about what is believed (on the object level) is effected on the meta level, via the predicate BEL(ieved). This meta level is consistent. If certain sentences are derivable in it (namely that there is an f such that both f and $\neg f$ are BELieved), a decision rule is invoked to determine which subset of object level beliefs to choose.

Object level

This is mostly standard, looking at a subset of FOL.

1. *Terms:* only constants are taken as terms, of which there are only finitely many.

2. *Basic formulae:* are of the form $p(t_1, t_2, \ldots, t_n)$, where p is an n-ary predicate symbol and the t_i are terms.

3. *Well-formed formulae:* are basic formulae, negations of basic formulae, or of the form
$$A_1 \wedge A_2 \wedge \ldots \wedge A_n \rightarrow (\neg)B,$$
for basic formulae $A_1 \ldots A_n, B$.

We assume the standard notion of logical consequence

meta-level

The meta language is a first-order language with equality.

1. *Logical constants* and *logical variables* denote well-formed formulae of the object language and rational numbers between 0 and 1. Every propositional formula of the object language (say f) is assigned a logical constant (say $\hat{f}$), and every rational number between 0 and 1 is assigned a logical constant (itself).

2. *Terms* are logical constants and logical variables and complex terms constructed from formulae of the object language in the following way:
   - $f \wedge g$ is denoted by $and(\hat{f}, \hat{g})$
   - $f \rightarrow g$ is denoted by $implies(\hat{f}, \hat{g})$
   - $\neg f$ is denoted by $not(\hat{f})$

We can simplify the notation: For example (IMP), the modus ponens for beliefs,

$$\forall \hat{f}\ \ \forall \hat{g}[BEL(implies(\hat{f}, \hat{g})) \rightarrow (BEL(\hat{f}) \rightarrow BEL(\hat{g}))]$$

will be written as

$$\forall f\ \ \forall g[BEL(f \rightarrow g) \rightarrow (BEL(f) \rightarrow BEL(g))].$$

But this is still first-order logic !

Meta-theory ctd

The meta language has a number of special predicates.

- **Assumptions:**

  $ASSUMPTION(f)$ is true if formula f is one of the assumptions of the agent, i.e. a belief that is justified by the empty set.

- **Tautologies:**

  $TAUTOLOGY(f)$ is true if formula f is a constructive tautology.

- **Belief:**

  $BEL(f)$ is true if formula f is in the agent's belief set.

meta-theory ctd

BEL derives belief in possibly non-atomic formulae of the object language from belief in other such formula. Inferring new beliefs proceeds via the following rules:

All assumptions are believed (ASS):

$$\forall f[ASSUMPTION(f) \rightarrow BEL(f)]$$

Conjunction (CON):

$$\forall f\ \forall g[(BEL(f) \wedge BEL(g)) \equiv BEL(f \wedge g)]$$

Implication (IMP):

$$\forall f\ \forall g[BEL(f \rightarrow g) \rightarrow (BEL(f) \rightarrow BEL(g))]$$

That's all . . .

Only beliefs that are justified by these axioms are believed. For any ϕ, the following holds (MINBEL):

$$\begin{array}{l}\forall f[ASSUMPTION(f) \rightarrow \phi(f)] \wedge \\ \forall f\ \forall g[(\phi(f) \wedge \phi(g)) \equiv \phi(f \wedge g)]\ \wedge \\ \forall f\ \forall g[\phi(f \rightarrow g) \rightarrow (\phi(f) \rightarrow \phi(g))] \\ \quad \rightarrow\ \ \forall f[BEL(f) \rightarrow \phi(f)]\end{array}$$

- **Epistemic entrenchment:**

  Degrees of epistemic entrenchment are assigned to object language formulae in the meta language. $EE(f, e)$ means that f has the degree of epistemic entrenchment (a rational number between 0 and 1) of e. No believed assumption has 0, and only believed tautologies have 1.

Summary

- de Kleer's assumption based TMS
- Belief Revision via epistemic entrenchment