Today

- Epistemic Entrenchment

Reminder of Epistemic Entrenchment

- Assign to statements in KB a degree of desire to hang on to them
- Use restricted syntax in KB
- If contradiction is found, analyse support for contradictory statements
- Throw away least entrenched statement in KB that solves the problem

References

For current work on belief revision, see

- "Belief Revision", P. Gärdenfors (ed), Cambridge UP, Tracts in Theoretical Computer Science
- Forthcoming workshop on Belief Revision and Dynamic Logic `www.irit.fr/~Andreas.Herzig/Esslli05`

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.

Epistemic entrenchment

Degrees of epistemic entrenchment are assigned to object language formulae in the meta language. $EE(f, e)$ is true if formula 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.

Preference between assumptions

Because of the foundational design of the system, any evaluation of a derived belief in these terms must ultimately depend on the evaluation of its underlying assumptions. The choice of a new belief set is based on formulae derived on this meta-level. These formulae must have assumptions as arguments. This is expressed by the special predicate $PREFER(a_1, a_2)$, where a_1, a_2 are assumptions, one of which supports f and the other of which supports $\neg f$.

This expression is equivalent to a conjunction of various $BEL(.)$, $EE(., .)$ and possibly other relations between beliefs (equality and arithmetical relations).

Decision criterion

Decision criterion *maximin*:

Let the beliefs f and $\neg f$ have the sets of assumption bases M_1 and M_2.

Let the elements of each assumption base $m_{ij} \in M_i, i = 1, 2, j = 1, ..., J_i$ be $a_{ijk}, k = 1, ..., K_{ij}$ with degrees of epistemic entrenchment given as $EE(a_{ijk}, e_{ijk})$.

Find $\hat{a_{ij}}$ such that $EE(\hat{a_{ij}}, \hat{e_{ij}})$ and $\hat{e_{ij}} \leq e_{ijk}, k = 1, ..., K_{ij}$. The assumption $\hat{a_{ij}}$ thus has the minimum degree of epistemic entrenchment – the 'bottleneck'.

Now find $\bar{a}_i$ such that $EE(\bar{a}_i, \bar{e}_i)$ and $\bar{e}_i \geq \hat{e_{ij}}, j = 1, ..., J_i$. The assumption $\hat{a}_i$ thus has the maximum epistemic entrenchment among all the bottlenecks in the assumption bases in M_i. The decision criterion *maximin* says that the element of the conflicting pair $f, \neg f$ with the higher maximal bottleneck degree of entrenchment is kept.

Representing logical consequence

The logical consequence operator can be defined as

$$Cn_{BEL}(S) \quad = \quad \{f | S \vdash_{BEL} f\}$$

where $\vdash_{BEL}$ represent deduction in the object logic.

The representation of belief sets in the present system will usually start from a **belief base** $BASE(K)$ for belief set K. This base is the set of assumptions of the object language.

We can use these ideas to prove properties of the revision criterion.

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.

Formalising

In our restricted language:

$$inTrap(s)$$
$$euro(s) \wedge swan(s) \rightarrow white(s)$$
$$inTrap(s) \rightarrow swan(s)$$
$$inTrap(s) \rightarrow swedish(s)$$
$$swedish(s) \rightarrow euro(s)$$
$$inTrap(s) \rightarrow \neg white(s)$$

Assign entrenchment degrees

$$inTrap(s) \quad 0.99 \tag{1}$$
$$euro(s) \wedge swan(s) \rightarrow white(s) \quad 0.7 \tag{2}$$
$$inTrap(s) \rightarrow swan(s) \quad 0.8 \tag{3}$$
$$inTrap(s) \rightarrow swedish(s) \quad 0.6 \tag{4}$$
$$swedish(s) \rightarrow euro(s) \quad 0.95 \tag{5}$$
$$inTrap(s) \rightarrow \neg white(s) \quad 0.92 \tag{6}$$

Here it is easy to see the expected outcome (reject second statement).

Note that we do *not* treat the numbers as probabilities – if we did, then we should combine then, e.g. by Bayesian methods.

Combining arguments

Suppose we have some additional information:

the bird in the trap is tagged
the tag classifies it as white
tag classifications are correct

Translate:

$$inTrap(s) \rightarrow tagged(s,t) \quad 0.95 \tag{7}$$
$$tagged(s,t) \wedge correct(t) \rightarrow white(s) \quad 0.99 \tag{8}$$
$$correct(t) \quad 0.96 \tag{9}$$

Now there is another argument for the swan being white.

Combining entrenchments

White or not white?

For each argument, find associated entrenchments, and take min in each:

For white:

$\{\ 0.99, 0.7, 0.8, 0.6, 0.95\ \}$, min 0.6
$\{\ 0.99, 0.95, 0.99, 0.96\ \}$, min 0.95

For black:

$\{\ 0.92\ \}$, min 0.92.

The maximum value for each side decides whether the black or white conclusion is kept (here, white).

Revising the KB

When we have worked out which conclusion to hang on to, we need to revise the KB so that it becomes logically consistent. In this case, just drop one statement (that the swan is not white). In general, need to do more.

An algorithm, to keep positive case:

```
UNTIL KB consistent DO
   for each A set of support assumptions for not F
       find least entrenched A
       delete A
```

NB this is non-deterministic, depending on order the arguments are treated in. Other algorithms are possible, some may keep more or less of the original beliefs.

Properties of belief revision

Technically, a belief set can be changed by adding or removing beliefs. Since the system is foundational and beliefs are closed under logical consequence, belief sets can only be expanded or contracted by assumptions.

Expansion Expanding belief set K by assumption A results in the belief set

$$K_A^+ = Cn_{BEL}(BASE(K) \cup \{A\}),$$

where $BASE(K)$ is K's belief base.

Contraction Contracting belief set K by assumption A results in the belief set

$$K_A^- = Cn_{BEL}(BASE(K) \setminus \{A\}),$$

Properties of belief Revision ctd

There are conditions that any belief revision should satisfy (called AGM postulates).

For expansion of a belief set with a new sentence A, get e.g.

(K$^+$ 1) For any sentence A and any belief set K, K_A^+ is a belief set.

(K$^+$ 2) $A \in K_A^+$

(K$^+$ 3) $K \subseteq K_A^+$

(K$^+$ 4) If $A \in K$, then $K_A^+ = K$.

(K$^+$ 5) If $K \subseteq H$, then $K_A^+ \subseteq H_A^+$.

What this gives us

- A belief revision system using a meta-logic which ensures that every belief is either an assumption or justified by other beliefs and that satisfies the AGM postulates.
- an algorithm has which provides a policy of dealing with incoming information, given initial epistemic entrenchment

Summary

- Belief revision
- Foundational and Coherence approaches
- Epistemic entrenchment
- Meta-theory to compute preferred belief revision