Preference-Based Belief Revision for Rule-Based Agents

Brian Logan

School of Computer Science
University of Nottingham, UK

Outline

- belief revision
- rule-based agents
- McAllester contraction
- preferences over beliefs
- McAllester revision

Belief revision

- agents interact with other agents or their environment is to acquire new information
- in general, it is impossible to ensure that such information will be consistent with the agent's current beliefs
- when an inconsistency is discovered the agent must revise its beliefs to restore consistency
- ideally, any revision operation should be
    - rational
    - forgo less preferred beliefs, e.g., those which are less certain, credible or useful
    - computationally efficient

AGM

- AGM theory (Alchourrón, Gärdenfors, Makinson 1985) describes an idealised reasoner with a potentially infinite set of beliefs (belief set) closed under logical consequence
- inspired by work on normative reasoning
- revision is based on ideas of coherence and informational economy
- requires the changes to an agent's belief set resulting from a revision be as 'small' as possible
- if an agent has to give up a belief in A, it does not have to give up believing in things for which A was the sole justification, so long as they are consistent with its remaining beliefs

Operations on belief sets

expansion $K + A$, adds a new belief A to K, and the resulting set is closed under logical consequence

contraction $K \dot{-} A$, removes A from K and removes sufficient beliefs from K so that it no longer entails A

revision $K \dot{+} A$, if A is inconsistent with K, K is modified (as little as possible) so that it no longer entails $\neg A$, A is then added to K, and the resulting set closed under logical consequence

Basic AGM postulates for contraction

(K$\dot{-}$1) $K \dot{-} A = Cn(K \dot{-} A)$ (closure)

(K$\dot{-}$2) $K \dot{-} A \subseteq K$ (inclusion)

(K$\dot{-}$3) if $A \notin K$, then $K \dot{-} A = K$ (vacuity)

(K$\dot{-}$4) if $\nvdash A$, then $A \notin K \dot{-} A$ (success)

(K$\dot{-}$5) if $A \in K$, then $K \subseteq (K \dot{-} A) + A$ (recovery)

(K$\dot{-}$6) if $Cn(A) = Cn(B)$, then $K \dot{-} A = K \dot{-} B$ (equivalence)

Concrete operations: partial meet contraction

Definition (Partial meet contraction)

a partial meet contraction of K by A is defined as an intersection of *some* maximal subsets of K which do not imply A

Theorem (Alchourrón, Gärdenfors, Makinson)

any contraction function which satisfies postulates ($K\dot{-}1$)–($K\dot{-}6$) can be generated by a partial meet contraction function

Belief bases

- in general, belief sets are infinite, so AGM belief revision is not implementable
- one option is to focus on belief bases — finite representations of belief sets
- B_K is a belief base for a belief set K if $Cn(B_K) = K$
- expansion, contraction and revision can be defined on belief bases instead of belief sets
- if *Cn* is classical propositional consequence operator, belief base contraction and revision are still NP-hard (Nebel 1994)

Truth maintenance systems

- truth (reason) maintenance style belief revision (Doyle 1977; McAllester 1980) keeps track of dependencies between beliefs
- each belief has a set of justifications, and the reasons for holding a belief can be traced back through these justifications to a set of foundational beliefs
- when a belief *A* must be given up, sufficient foundational beliefs have to be withdrawn to render *A* underivable
- if all the justifications for *A* are withdrawn, then *A* should no longer be believed

Tractable rational revision

- AGM approaches have a well defined notion of rational belief revision, but are generally intractable
- TMS approaches are tractable (polynomial time) but logically incomplete
  - e.g., McAllester's boolean constraint propagation algorithm does not find all the classical logical consequences of boolean formulas, sacrificing completeness for efficiency
  - not clear to what extent these systems are rational
- one way to define a tractable rational revision operation is to weaken the logic of the agent so that the consistency check is no longer an expensive operation (cf Nebel 1992)

Our approach

- we assume a simple rule-based agent with a finite state and a finite program
- the agent's state or working memory (WM) is a finite set of ground literals representing the beliefs of the agent
- the agent's program consists of a set of Horn clause rules of the form

$$A_1, \ldots, A_n \rightarrow B$$

  where $A_1, \ldots, A_n, B$ are literals
- we only revise by literals — rules are not revised

Rule-based agents

the agent's language L_W contains

- literals (atomic formulas or their negations), e.g.,

$$PartOf(Bordeaux, France),\ PartOf(x, Bordeaux)$$

- rules $A_1 \wedge \ldots \wedge A_n \rightarrow B$, where $A_1, \ldots, A_n, B$ are literals, e.g.,

$$Region(x, y) \wedge PartOf(y, z) \rightarrow Region(x, z)$$

- variables in rules are assumed to be universally quantified

Agent's logic

- the agent's logic W in the language L_W contains a single inference rule, generalised modus ponens (GMP)

$$\frac{\delta(A_1), \ldots, \delta(A_n), \quad A_1 \wedge \ldots \wedge A_n \rightarrow B}{\delta(B)}$$

where δ is a substitution function

- for example

$$\frac{\begin{array}{c} Region(ChateauLafite, Pauillac),\ PartOf(Pauillac, Bordeaux), \\ Region(x, y) \wedge PartOf(y, z) \rightarrow Region(x, z) \end{array}}{Region(ChateauLafite, Bordeaux)}$$

Agent's belief set

- the agent's belief set K is the set of ground literals in the working memory (WM) and the set of its rules R
- if the agent runs its rules to quiescence, then $K = C_W(WM \cup R)$, where C_W is the closure of a set of formulas with respect to derivability in W

Example

rules

R1 $Region(x, y) \wedge PartOf(y, z) \rightarrow Region(x, z)$

R2 $Region(x, France) \rightarrow \neg Region(x, Australia)$

facts

F1 $Region(ChateauLafite, Pauillac)$

F2 $PartOf(Pauillac, Bordeaux)$

F3 $PartOf(Bordeaux, France)$

F4 $PartOf(Tasmania, Australia)$

Example continued

- the agent derives

F5 $Region(ChateauLafite, Bordeaux)$ (from F1, F2, R1)

F6 $Region(ChateauLafite, France)$ (from F5, F3, R1)

F7 $\neg Region(ChateauLafite, Australia)$ (from F6, R2)

Example continued

- suppose the agent is informed

F8 *Region*(*ChateauLafite*, *Tasmania*)

and derives

F9 *Region*(*ChateauLafite*, *Australia*) (from F8, F4 and R1)

- to restore consistency, the agent must contract by either F7 or F9

Preferred beliefs

- we assume that the agent prefers some (ground literal) beliefs to others
- the agent's preference order, $\preceq$, is a total order on *WM*
- for each non-empty $\Gamma \subseteq WM$, $w(\Gamma)$ is the minimal element of Γ with respect to $\preceq$
- for example, the agent may prefer its initial beliefs (F1-F4) to beliefs communicated by other agents (F8)

McAllester contraction

for two ground literals $\delta(A)$ and $\delta(B)$, $\delta(B)$ depends on $\delta(A)$ in K, $\delta(A) \gg_K \delta(B)$, if either

1. $\delta(A) = \delta(B)$;
2. $A_1, \ldots, A_n \rightarrow B \in K$, $\delta(A_1), \ldots, \delta(A_n) \in K$, and $\delta(A) = w(\delta(A_1), \ldots, \delta(A_n))$; or
3. $A_1, \ldots, A_n \rightarrow C \in K$, $\delta(A_1), \ldots, \delta(A_n) \in K$, $\delta(A) = w(\delta(A_1), \ldots, \delta(A_n))$, and $\delta(C) \gg_K \delta(B)$

Definition (McAllester contraction)

the McAllester contraction of K by a literal A is defined as

$$K \dot{-} A =_{df} K \setminus \{C : C \gg_K A\}$$

Characterising McAllester contractions

(K$\dot{-}$1) $K \dot{-} A = C_W(K \dot{-} A)$ (closure)

(K$\dot{-}$2) $K \dot{-} A \subseteq K$ (inclusion)

(K$\dot{-}$3) if $A \notin K$, then $K \dot{-} A = K$ (vacuity – **derivable**)

(K$\dot{-}$4) $A \notin K \dot{-} A$ (success)

(K$\dot{-}$6) if $C_W(A) = C_W(B)$, then $K \dot{-} A = K \dot{-} B$ (equivalence – **trivial**)

(K$\dot{-}$R) for each rule $A_1, \ldots, A_n \to B$, if $A_1, \ldots, A_n \to B \in K$, then $A_1, \ldots, A_n \to B \in K \dot{-} B$

(K$\dot{-}$F) If $C \in K$ and $C \notin K \dot{-} A$ then $C \gg_K A$

Representation theorem

Theorem

every McAllester contraction satisfies the postulates (K$\dot{-}$1)-(K$\dot{-}$F) and conversely, if a contraction operation satisfies (K$\dot{-}$1)-(K$\dot{-}$F), then it is a McAllester contraction

- note that the AGM recovery postulate (K$\dot{-}$5) is not satisfied

  (K$\dot{-}$5) If $A \in K$, then $K \subseteq (K \dot{-} A) + A$ (recovery)

- suppose we have a single rule $A \rightarrow B$ and $WM = \{A, B\}$
- after contraction by B, WM is empty — when we expand by B, WM contains only B

Contraction algorithm: preliminaries

- we represent WM as a directed graph consisting of two kinds of nodes: *beliefs* and *justifications*
- a justification consists of a belief and a support list containing the premises of the rule used to derive this belief: for example, $(A, [B, C])$, where A is a derived belief and the rule used to derive it is $B, C \rightarrow A$
- each justification has one outgoing edge to the belief it is a justification for, and an incoming edge from each belief in its support list
- a foundational belief D which is not derived has a justification of the form $(D, [\,])$

McAllester contraction by A

for all $j = (B, s)$ with an edge from A **do**
    remove j (and all edges to and from j) from the graph
for all $j = (A, s)$ with an edge to A **do**
    if $s == [\,]$ **then**
        remove j (and the edge from j to A)
    else
        contract by $w(s)$
remove A

- $O(kr + n)$ where r is the number of rules, k is the maximal number of premises in a rule, and n the number of ground literals in WM

Example: assigning preferences to beliefs

- define preferences in terms of the quality of justifications
- the quality of a justification is an integer in the range $0, \ldots, m$, where 0 means lowest quality and m means highest quality
- the agent associates an *a priori* quality with each non-inferential justification for its foundational beliefs
- e.g., qualities are taken from the set $\{0, \ldots, n-1\}$ where n is the number of foundational beliefs

Example: assigning preferences to beliefs

- the preference of a literal A, $p(A)$, is given by its highest quality justification

$$p(A) = \max\{qual(j) : j \in \text{justification for } A\}$$

- the quality of an inferential justification is given by the least preferred belief in its support

$$qual(j) = \min\{p(A) : A \in \text{ support of } j\}$$

- i.e., the preference of a belief is determined by the best argument for it, and the quality of an argument is as high as that of its weakest component belief
- computing preferences requires $O(n \log n + kr)$ time

Least worth and minimal change

- a McAllester contraction that uses the quality of justifications minimises the preference of the literals removed as a result of contraction
- let the worth of a set of literals be $worth(\Gamma) = \max\{p(A) : A \in \Gamma\}$

Theorem

if a McAllester contraction of K by A results in the removal of the set of literals Γ, then for any other set of literals Γ' such that $K \setminus \Gamma'$ does not imply A, $worth(\Gamma) \leq worth(\Gamma')$

- it is possible to extend the algorithm for McAllester contraction so that the set of literals removed both has least worth and is minimal — the complexity of the algorithm remains polynomial

Revision

- for a reasoner in classical logic, revision and contraction operations are definable (via the Levi and Harper identities)
- for an agent which is not a classical reasoner, contraction and revision are not inter-definable when the consistency of $K + A$ is not equivalent to $K \nvdash \neg A$
- for a rule-based agent which reasons in the logic W, applying the Levi identity to McAllester contraction results in a revision operation which does not satisfy (K$\dot{+}$5)
- instead, we define revision by A as $(K + A) \dot{-} \bot$, i.e., as expansion by A followed by elimination of all contradictions

McAllester revision

- note that we need to specify the order in which the contradictions are eliminated
- let $(B_1, \neg B_1), \ldots, (B_n, \neg B_n)$ be the list of all contradictions in $K + A = Cn(K \cup \{A\})$, ordered by preference order on $w(B_i, \neg B_i)$, and let $\sim B_i = w(\{B_i, \neg B_i\})$

Definition (McAllester revision)

the revision of K by A using ordered contraction by contradictions (OCC) is given by

$$K \dot{+} A \stackrel{df}{=} (K + A) \dot{-} \sim B_1 \dot{-} \sim B_2 \dot{-} \ldots \dot{-} \sim B_n$$

McAllester revision postulates

(K$\dot{+}$1) $K \dot{+} A = Cn(K \dot{+} A)$

(K$\dot{+}$3) $K \dot{+} A \subseteq K + A$

(K$\dot{+}$4) if $\{A\} \cup K$ is consistent, then $K + A = K \dot{+} A$

(K$\dot{+}$5) $K \dot{+} A$ is inconsistent iff A is inconsistent

(K$\dot{+}$6) if $Cn(A) = Cn(B)$, then $K \dot{+} A = K \dot{+} B$

(K$\dot{+}$R) for each rule $A_1, \ldots, A_n \rightarrow B$, $A_1, \ldots, A_n \rightarrow B \in K$ iff $A_1, \ldots, A_n \rightarrow B \in K \dot{+} B$

(K$\dot{+}$OCC) If $C \in K + A$ and $C \notin K \dot{+} A$, then for some i,

$$C \gg_{(K+A)\dot{-}\sim B_1 \dot{-} \ldots \dot{-} \sim B_{i-1}} \sim B_i$$

where $(B_1, \neg B_1), \ldots, (B_n, \neg B_n)$ are all the contradictions in $K + A$, $\sim B_i = w(B_i, \neg B_i)$, $\sim B_1 \preceq \ldots \preceq \sim B_n$, and $i \in \{1, \ldots, n\}$

Summary

- rule-based agents can be modelled as reasoners in a logic with a single inference rule of generalised modus ponens
- we can define a contraction operation for rule-based reasoners that satisfies all the basic AGM postulates for contraction (apart from the recovery postulate)
- contraction can be computed in polynomial time
- the contraction operation can be used to define a corresponding revision operation which is also polynomial time