Anytime Approximate Modal Reasoning

Fabio Massacci
Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza"
via Salaria 113, I-00198 Roma
email:massacci@dis.uniroma1.it

Abstract

Propositional modal logics have two independent sources of complexity: unbounded logical omniscience and unbounded logical introspection. This paper discusses an approximation method to tame both of them, by merging propositional approximations with a new technique tailored for multi-modal logics. It provides both skeptical and credulous approximations (or approximation that are neither of the two).

On this semantics we build an anytime proof procedure with a simple modification to classical modal tableaux. The procedure yields approximate proofs whose precision increases as we have more resources (time, space etc.) and we analyze its semantical and computational "quality guarantees". *

Introduction

The study of tractability of logical inference has received a major attention in AI. For propositional logic the obstacle to tractability is *unbounded logical omniscience*: given a knowledge base we can always derive all its logical consequences. Modal logics have another source of intractability, *unbounded logical introspection*: we can reason about our beliefs about somebody's belief about...

Logical omniscience has received most attentions and, as noted out in (Fagin, Halpern, & Vardi 1995), it has been tamed essentially by weakening the "omniscience" or by weakening the "logical".

The first approach is based on *inference procedures which are tractable but incomplete* such as unit resolution (del Val 1994) or boolean constraint propagation (Dalal 1996a). We can also restrict inferences to a subset of formulae as in (Dalal & Etherington 1992). In this way we have a complete procedure for a subset of the language and a sound but incomplete (yet tractable) system for the whole language.

Completeness for the whole language can be recovered by lemmaizing off-line (*knowledge compilation*) or on-line (*analytic cut*). We compute all answers of the tractable inference mechanism and store them for future retrieval.

The technique of (Selman & Kautz 1996) only included Horn lemmas and was complete only for horn relaxations of the original problem. New systems use general clauses as lemmas (del Val 1994; Dalal 1996a). Then we can start an incremental knowledge compilation by imposing incremental bounds on the size of the clauses used as lemmas and construct an anytime reasoning procedure following (Dean & Boddy 1988; Russel & Zilberstein 1991). Since the number of lemmas grows exponentially, the result may be a blow-up of the knowledge base (Selman & Kautz 1996).

An incremental construction can also be done by lemmaizing on-line (Crawford & Kuipers 1989; Dalal 1996a) i.e. by using cut (also called "Socratic completeness").

The dual approach is weakening the "logical" and keeping the "omniscience" in a logic weaker than classical logic. Typically one uses a variant of Belnap's four valued logic as done by Levesque (1984) and further studied, for example, in (Fagin, Halpern, & Vardi 1995; Lakemeyer 1994,1996; Schaerf & Cadoli 1995).

We can construct *incremental approximations* with the techniques of (Fagin & Halpern 1988) and (Schaerf & Cadoli 1995): we consider a subset of the propositional variables as interesting, i.e. deserving a classical interpretation, while the rest of propositions is given a three-valued interpretation. By increasing the set of interesting variables we can regain completeness in an incremental fashion. In (Schaerf & Cadoli 1995) we find sound & incomplete and complete & unsound reasoning while (Fagin & Halpern 1988) is limited to sound & incomplete reasoning.

We still have a problem of computational complexity since tractability results are restricted to special normal forms. In the general case, "approximate" logical consequence in classical logic corresponds to the "exact" logical consequence in relevance logic, which is CO-NP-complete (Fagin, Halpern, & Vardi 1995; Schaerf & Cadoli 1995).

Most works cited above use modal operators for explicit (tractable) beliefs and implicit (intractable) beliefs. Yet tractability results are usually restricted to sentences of the form $\mathbf{K}_a A \supset \mathbf{K}_b B$ where A and B are propositional formulae. When full introspection is allowed by Lakemeyer (1994,1996), it is only for modal logic K45 which is in NP (Halpern & Moses 1992). In a nutshell, the expressivity of modal operators is not used. It is not a case that the problem of unbounded logical introspection is hardly mentioned.

The development of tractable inference systems for $\mathcal{ALC}$ (a variant of modal logic K) has been done by (Patel-Schneider 1989) where a four-valued semantics is adopted for tractability. The study of incremental approximations has been made in (Schaerf & Cadoli 1995) where approximation is essentially an incremental rewriting procedure.

In the rest of the paper, we present the intuitions behind our proposal and recall some preliminaries. Then we develop an approximation for the propositional part of the reasoning, an approximation for the modal part and combine the two. Finally we show how to transform a traditional tableau calculus due to Fitting (1983) into an anytime deduction procedures for which we discuss the semantical and computational quality of the approximation.

Semantical Intuitions

The main intuition behind our approximate semantics stems from the observation, following (Russel & Zilberstein 1991), that the difficulty behind unbounded logical omniscience and introspection is not the "logical", nor the "omniscience" nor the "introspection" but the "unbounded".

This difficulty is caused by the egalitarian attitude of classical modal logic: all propositional atoms are equal, all possible worlds are equal. So, we must consider a potentially exponential number of propositional assignments (unbounded logical omniscience) for a potentially exponential number of worlds (unbounded logical introspection).

Therefore, logical inference has an "all-or-nothing" behavior: either we have considered (implicitly or explicitly) all worlds and all propositions and then we can answer "A is a theorem" or we don't know what to say.

The idea behind *anytime reasoning* is that it should be possible to interrupt our reasoning process at any time and get *approximate answers* whose quality improves with time and which eventually converge to the exact answer. For instance, suppose that the exact computation requires time 2^{cn} in the worst case. An anytime reasoner gives a first approximate answer in time 2^{c1} and then improves the the answer in time 2^{c2} and so on until the exact answer is obtained or the process is interrupted. In the worst case n steps are required for completion (although we may be luckier). Anyhow, if we interrupt the process at time t we obtain the approximate answer provided at the kth checkpoint such that $2^{ck} \leq t < 2^{c(k+1)}$. See (Dean & Boddy 1988; Russel & Zilberstein 1991) for further details.

What is difficult to grasp is the meaning of approximate answer for logical inference: what is an approximate proof or an approximate counter model?

Consider the propositional case; suppose that we have coded a planning problem in propositional logic (Kautz & Selman 1992) and that we are looking for a plan (a model for the KB). To approximate logical omniscience, we start by considering just few atoms as describing *interesting* properties and therefore deserving a classical interpretation. Other variables stand for *unknown* properties and so are the formula composed with them. They don't specify properties of "interesting models". Other atoms corresponds to *incoherent* properties and do not contribute to "interesting proofs".

When we start the search for an approximate model, we process interesting formulae as usual. Each time we find an unknown formula we conclude the branch of the search with failure: at this level of approximation, interesting models (plans) do not require to give truth values to unknown propositions. When we meet incoherent formulae we simply ignore them, we assume that we can fix them later. Thus we do not consider incoherent formulae in the (approximate) proof that a plan does not exist; only contradictions between interesting formulae matter.

From the viewpoint of inference, the presence of incoherent propositions leads to more models (we assume these formulae are approximately satisfied) and thus fewer proofs than classical logic (sound approximations). Unknown formulae yield fewer models (these formulae cannot be approximately satisfied) and thus more proofs (complete approximation). By suitably combining interesting, unknown and incoherent propositions we can reconstruct the tractable cases of (Levesque 1984; Schaerf & Cadoli 1995; Fagin, Halpern, & Vardi 1995) and many more.

This technique is not sufficient for modal logics because of the (possibly exponential) number of possible worlds. The key observation is that interesting worlds should not be "too remote" from the real world i.e. they should be reachable with few introspection steps. We curtail the "unbounded" logical introspection by imposing that "too many" steps of introspection always lead outside interesting worlds.

Again, we divide worlds into three sets: incoherent, unknown and interesting worlds. In incoherent worlds we enforce any kind of statement, including contradictory statements, while at unknown worlds we do not enforce any statement whatsoever.

So we can define the notion of k-skeptical and k-credulous agents: a k-skeptical agent is logical omniscient and able of introspective reasoning but only up to k levels of introspection. Afterwards, she will give up and assume that an incoherent world is reached. The dual attitude will be taken by a credulous agent.

Note that the notion of incoherent and unknown worlds is substantially different from the notion of "impossible" world used in the literature (Levesque 1990). The notion of impossible world refers to "inaccessible" worlds, i.e. worlds which are not reachable via the accessibility relation $\mathcal{R}_a$. Our notion is closer to the "non-normal" worlds introduced by Kripke (see (Fitting 1983)) and the "incoherent" and "incomplete" worlds of (Fagin, Halpern, & Vardi 1995)[page 212], although our use and formalization is different.

We have defined two independent approximations: one for the propositional part of the logic and an approximation for the modal part. These approximations can be combined in many ways. For example, we known that incoherent formulae are approximately satisfied in interesting worlds and interesting formulae never hold in unknown worlds. So what happens to incoherent formulae in unknown worlds?

One choice is to make the propositional part predominant: incoherent proposition always hold in a world, no matter which world. The opposite choice gives more importance to worlds: unknown worlds never satisfies a formula, no matter which formula. In the latter case the propositional approximation is a secondary approximation and the modal one is a primary approximation. This is our choice, mainly because it makes computations simpler.

We are only left with the last step: an anytime algorithm.

We decided to pick a traditional system off-the-shelves and transform it into an anytime procedure based on our semantics. The idea is to use a refutation procedure and change the termination conditions: devise complete approximations by accepting more (unknown) formulae as approximately contradictory, devise sound approximation by accepting more (incoherent) formulae as approximately satisfiable.

Preliminaries on Modal Logics

An introduction can be found in (Fitting 1983; Halpern & Moses 1992). We construct *modal formulae* A, B from atomic propositions $p \in \mathcal{P}$ and agent names $a \in \mathcal{A}$:

$$A, B \quad ::= \quad p \mid \neg A \mid A \wedge B \mid \mathbf{K}_a A$$

We use signed formulae so that the α, β, π, ν notation of Fitting (1983) can be used for other connectives. A *signed formula* φ, ψ is a pair $\mathtt{t}.A$ or $\mathtt{f}.A$. The intuitive interpretation is that A is respectively true and false.

The semantics is based on *Kripke models* i.e. triples $\left\langle W, \{\mathcal{R}_a\}_{a\in\mathcal{A}}, \mathcal{I}\right\rangle$ where W is a non empty set, whose elements are called *worlds*, $\{\mathcal{R}_a\}_{a\in\mathcal{A}}$ is a family of binary *accessibility relations* over W, and $()^{\mathcal{I}}$ is function, called *valuation* from propositional letter to subset of worlds. Intuitively the worlds in $(p)^{\mathcal{I}} \subseteq W$ are those where p is true.

Let $\mathcal{M}$ be a model, w a world in $\mathcal{M}$, and φ a signed formula, the forcing relation $w \Vdash \varphi$ is defined below:

$w \Vdash \mathtt{t}.p$	if	$w \in (p)^{\mathcal{I}}$
$w \Vdash \mathtt{t}.\neg A$	if	$w \Vdash \mathtt{f}.A$
$w \Vdash \mathtt{t}.A \wedge B$	if	$w \Vdash \mathtt{t}.A$ and $w \Vdash \mathtt{t}.B$
$w \Vdash \mathtt{t}.\mathbf{K}_a A$	if	$\forall v \in W : w\mathcal{R}_a v$ implies $v \Vdash \mathtt{t}.A$
$w \Vdash \mathtt{f}.p$	if	$w \notin (p)^{\mathcal{I}}$
$w \Vdash \mathtt{f}.\neg A$	if	$w \Vdash \mathtt{t}.A$
$w \Vdash \mathtt{f}.A \wedge B$	if	$w \Vdash \mathtt{f}.A$ or $w \Vdash \mathtt{f}.B$
$w \Vdash \mathtt{f}.\mathbf{K}_a A$	if	$\exists v \in W : w\mathcal{R}_a v$ and $v \Vdash \mathtt{f}.A$

When $w \Vdash \varphi$ holds we say that *"w satisfies φ in $\mathcal{M}$"*. We say that $\left\langle W, \{\mathcal{R}_a\}_{a\in\mathcal{A}}, \mathcal{I}\right\rangle$ is *model for A* iff every world in W satisfies $\mathtt{t}.A$; it is a *counter model for A* if there is world $w \in W$ which satisfies $\mathtt{f}.A$.

Our definition of logical consequence is a modal extension of the propositional definition in (Dalal 1996b).

Definition 1 *A formula A is a* logical consequence *of KB (in symbols $KB \models A$) iff every model for KB is not a counter model for A. If $\emptyset \models A$ we say that A is* valid.

Just Propositional Approximation

For the propositional approximation assume an *approximate partition* (for short ap) of the set of propositional atoms $\mathcal{P}$ into three sets: *unknown* propositions $unk(\mathcal{P})$, *interesting* propositions $int(\mathcal{P})$, and *incoherent* propositions $inc(\mathcal{P})$.

We only change the forcing relation $\Vdash$ for propositions:

$w \Vdash \mathtt{t}.p$	if	$p \in inc(\mathcal{P})$
	elseif	$p \in int(\mathcal{P})$ & $w \in (p)^{\mathcal{I}}$
$w \Vdash \mathtt{f}.p$	if	$p \in inc(\mathcal{P})$
	elseif	$p \in int(\mathcal{P})$ & $w \notin (p)^{\mathcal{I}}$

The corresponding logical consequence is $\models^{ap}$.

Combined with our key definition of logical consequence, this is enough to limit logical omniscience to interesting formulae. We can use incoherent formulae for sound but incomplete (i.e. skeptical) reasoning and unknown formulae for complete but unsound (i.e. credulous) reasoning.

For instance with incoherent concepts we may be incomplete: if $q \in int(\mathcal{P})$ and $p \in inc(\mathcal{P})$ we have

$\not\models^{ap} \mathbf{K}_a(p \supset q) \wedge \mathbf{K}_a p \supset \mathbf{K}_a q$.

With unknown concepts we may draw unsound conclusions. If $q \in int(\mathcal{P})$ and $p \in unk(\mathcal{P})$ then

$\models^{ap} \mathbf{K}_a(p \vee q) \supset \mathbf{K}_a q$.

What is interesting is the presence of formulae which mixes interesting, unknown, and incoherent propositions. It turns out that we can evaluate the overall nature of a formula in polynomial time (see the section on tableaux).

Just Modal Approximation

For the modal approximation we partition the worlds of a Kripke model into interesting worlds $int(W)$, incoherent worlds $inc(W)$, and unknown worlds $unk(W)$.

At first there is a constraint on the accessibility relation: *from an incoherent (unknown) world we only access incoherent (unknown) worlds*. Formally, $w\mathcal{R}_a v$ and $w \in inc(W)$ implies that $v \in inc(W)$. Dually $w\mathcal{R}_a v$ and $w \in unk(W)$ implies $v \in unk(W)$.

The next step is a revision of the forcing relation for propositional atoms (the rest is unchanged):

$w \Vdash \mathtt{t}.p$	if	$w \in inc(W)$
	elseif	$w \in int(W)$ & $w \in (p)^{\mathcal{I}}$
$w \Vdash \mathtt{f}.p$	if	$w \in inc(W)$
	elseif	$w \in int(W)$ & $w \notin (p)^{\mathcal{I}}$

Now, we can prove that if one only see incoherent worlds then everything is believed (the world satisfies $\mathtt{t}.\mathbf{K}_a A$ for all A) and everything unbelieved(the world also satisfies $\mathtt{f}.\mathbf{K}_a A$). If all worlds nearby are unknown then nothing is believed and nothing is unbelieved.

To approximate the introspection capabilities of an agent, we consider first the single agent case.

A *bounded agent* as a triple $\langle a, k_a, type\rangle$ where k_a is an integer and $type$ is either c (for credulous) or s (for skeptical). We say that agent a is k_a*-skeptical* in the former case and k_a*-credulous* in the latter.

A simple method for bounding introspection is to say that if a is k_a-skeptical (credulous) then every simple $\mathcal{R}_a$ path with more than $k_a + 1$ worlds terminates into an incoherent (unknown) world. Unfortunately, this definition is too strong: as in (Fagin, Halpern, & Vardi 1995), skeptical agents would believe too few valid formulae and credulous agents would believe too many invalid formulae because we have only specified the "maximal" distance between the "real world" and incoherent or unknown worlds. Without a minimal distance, the "real" world may be too close to uninteresting worlds.

To avoid this problem we introduce *realistic worlds*:

Definition 2 *Let $\langle a, k_a, type\rangle$ be a bounded agent and $\left\langle W, \{\mathcal{R}_a\}_{a\in\mathcal{A}}, \mathcal{I}\right\rangle$ be an approximate model with a partition of worlds. A world $w \in W$ is k_a-*realistic *if it is an interesting world and if*

- $k_a = 0$ *and a is* skeptical *then $w\mathcal{R}_a v$ implies $v \in inc(W)$;*

- k_a=0 *and* a *is* credulous *then* $w\mathcal{R}_a v$ *implies* $v \in unk(W)$*;*
- $k_a > 1$ *then* $w\mathcal{R}_a v$ *implies that* v *is* $k_a - 1$-realistic.

So, in the definition of forcing for π formulae

$$w \Vdash \mathtt{f}.\mathbf{K}_a A \text{ iff } \exists v \in W : w\mathcal{R}_a v \ \& \ v \Vdash \mathtt{f}.A \qquad (1)$$

if w is a k_a-realistic world then v is a $k_a - 1$-realistic world. Each time we satisfy a false belief $\mathtt{f}.\mathbf{K}_a A$ we get closer to unknown and incoherent worlds.

We upgrade the notion of counter model: given a bounded agent $\langle a, k_a, type\rangle$, a model $\langle W, \{\mathcal{R}_a\}_{a\in\mathcal{A}}, \mathcal{I}\rangle$ is an *approximate counter model* for A if there is a k_a-realistic world $w \in W$ which satisfies $\mathtt{f}.A$. The rest is unaltered and we denote the corresponding logical consequence as $\models^{k_a - type}$.

Some examples can be helpful to clarify the semantics. For 1-skeptical agents plain logical omniscience holds:

$$\models^{1-s} \mathbf{K}_a(p \supset q) \wedge \mathbf{K}_a p \supset \mathbf{K}_a q \qquad (2)$$

although unbounded logical omniscience may not hold:
$\not\models^{1-s} \mathbf{K}_a(\mathbf{K}_a(p \supset q) \wedge \mathbf{K}_a p) \supset \mathbf{K}_a\mathbf{K}_a q$

The situation is reversed for 1-credulous agents: (2) holds but unsound formulae may become approximately valid.
$\models_L^{1-c} \mathbf{K}_a(\mathbf{K}_a p \vee \mathbf{K}_a q) \supset \mathbf{K}_a\mathbf{K}_a q$

Still, 1-credulous agents remain sound if the reasoning does not require introspection over 1 level of modal nesting:
$\not\models_L^{1-c} (\mathbf{K}_a\mathbf{K}_a p \vee \mathbf{K}_a q) \supset \mathbf{K}_a q$

With many agents around reasoning may remain hard. Using the same techniques of (Halpern & Moses 1992) we can prove that *deciding approximate validity remains* PSPACE-*hard if we only bound self-introspection*. The trick is to use $\mathbf{K}_a\mathbf{K}_b A$ formulae to reason about mutual beliefs.

For bounding mutual introspection, we introduce the notion of *family of bounded agents* as the set $\mathcal{F} = \{\langle a, k_a, type_a\rangle\}_{a\in\mathcal{A}}$. We assume that there is only one triple in $\mathcal{F}$ for every agent a and we say that agent a is k_a*-skeptical* if $\langle a, k_a, s\rangle \in \mathcal{F}$ and it is k_a-credulous if $\langle a, k_a, c\rangle \in \mathcal{F}$.

Definition 3 *Let* $\mathcal{F}$ *be a family of bounded agents and* $\langle W, \{\mathcal{R}_a\}_{a\in\mathcal{A}}, \mathcal{I}\rangle$ *be an approximate model with a partition of worlds. A world* $w \in W$ *is* $\mathcal{F}$-realistic *if it is an interesting world and for every agent* a *in the family*

- $k_a = 0$ *and* a *is* skeptical *then* $w\mathcal{R}_a v$ *implies* $v \in inc(W)$*;*
- k_a=0 *and* a *is* credulous *then* $w\mathcal{R}_a v$ *implies* $v \in unk(W)$*;*
- *if* $k_a > 0$ *then* $w\mathcal{R}_a v$ *implies that* v *is* $\mathcal{F}/a$-realistic.

where the family $\mathcal{F}/a$ *is obtained by* $\mathcal{F}$ *by setting* for all agents *the bound to* $k_a - 1$ *and the type to* $type_a$.

Hence, in (1) we converge towards incoherent or unknown worlds no matter the path we follow. This definition has a intuitive motivation: when passing from w to v we are performing an introspection step for the agent a. Loosely speaking it is a who is now reasoning about the beliefs of other agents in his own head. If a is skeptical so he will reason skeptically when reasoning about other agents.

We can prove that if A has counter model for KB then it has a counter model with at most $2^{k_m} + 3$ worlds where k_m is the maximum bound for every agent in family $\mathcal{F}$.

Combined Approximation

We choose the modal approximation as the primary approximation. This requires a revision of the forcing relation:

$$\begin{array}{lll}
w \Vdash \mathtt{t}.p & \text{if} & w \in inc(W) \\
 & \text{elseif} & w \in int(W) \ \& \ p \in inc(\mathcal{P}) \\
 & \text{elseif} & w \in int(W) \ \& \ p \in int(\mathcal{P}) \ \& \ w \in (p)^{\mathcal{I}} \\
w \Vdash \mathtt{f}.p & \text{if} & w \in inc(W) \\
 & \text{elseif} & w \in int(W) \ \& \ p \in inc(\mathcal{P}) \\
 & \text{elseif} & w \in int(W) \ \& \ p \in int(\mathcal{P}) \ \& \ w \notin (p)^{\mathcal{I}}
\end{array}$$

Approximate Tableaux

We use *prefixed tableaux* by Fitting (1983) and present only the aspects we need for the anytime reasoner.

The basic component is a prefixed signed formula$\sigma : \varphi$ where σ is an alternating sequence of integers and agent names called *prefix* and defined as $\sigma ::= 1 \mid \sigma.a.n$.

The basic intuition is that σ is a "name" for a world where φ is satisfied. The key advantage for this representation is that the name explicitly codes the introspection steps from the "real" world denoted by 1 to the world σ.

The rest of the machinery is pretty standard: a *tableau* for proving that A is a logical consequence of KB is a tree where the root is labelled by $1 : \mathtt{f}.A$ and where nodes are added and labelled by prefixed signed formulae according tableau rules (Fitting 1983). A *branch* $\mathcal{B}$ is a path from the root to the leaf. Intuitively, each branch is a tentative model and the tableau construction is an attempt to construct a counter model for A. If we fail (because we find a contradiction in each branch) then A is a logical consequence; if we succeed (because we have reduced all formulae in a branch without finding contradictions) then we have the required counter model.

The simplest way to plug our approximation semantics onto this machinery is to change the rules that terminates the search by accepting branches which are only approximately contradictory or approximately reduced.

We need to lift the interesting, unknown and incoherent meaning that we have given to atoms and worlds to general prefixed formulae in a quick way.

So, we construct a *superficial valuation* of a formula, i.e. a function $\mathrm{V}(a, k, \varphi)$ with values in the set $\{inc, unk, int\}$. The first argument is the agent a for which the formula is evaluated, the second the current bound of introspection and the last is the signed formula. If the agent is not specified we use the wild card $*$.

For propositional letters we have

$$\begin{array}{l} \mathrm{V}(\_,\_,\mathtt{t}.p) \\ \mathrm{V}(\_,\_,\mathtt{f}.p) \end{array} = \begin{cases} int & \text{if } \in int(\mathcal{P}) \\ unk & \text{if } p \in unk(\mathcal{P}) \\ inc & \text{if } p \in inc(\mathcal{P}) \end{cases}$$

We use $\_$ as a don't care condition.

The function is defined inductively on the propositional and modal connectives. For negation we have that $\mathrm{V}(a, k, \mathtt{t}.\neg A) = \mathrm{V}(a, k, \mathtt{f}.A)$ and similarly for $\mathtt{f}.\neg A$. The other cases are shown in Table 1 and Fig. 1. We use the variable X to represent the table in compact form so in the third line of the $\mathtt{t}.A \wedge B$-table we have that if $\mathrm{V}(\_,\_,\mathtt{t}.B) = X$ then $\mathrm{V}(\_,\_,\mathtt{t}.A \wedge B) = X$.

t.A	t.B	t.$A \wedge B$
unk	X	unk
X	unk	unk
inc	X	X
int	int	int

f.A	f.B	f.$A \wedge B$
unk	X	X
inc	X	inc
X	inc	inc
int	int	int

Table 1: Superficial Valuation

$$\mathrm{V}(a, 0, \mathtt{f}.\mathbf{K}_a A) = \begin{cases} inc & \text{if } a \text{ is skeptical} \\ unk & \text{if } a \text{ is credulous} \end{cases}$$
$$\mathrm{V}(a, k+1, \mathtt{f}.\mathbf{K}_a A) = \mathrm{V}(a, k, \mathtt{f}.A)$$

$$\mathrm{V}(a, 0, \mathtt{t}.\mathbf{K}_a A) = \begin{cases} inc & \text{if } a \text{ is skeptical} \\ int & \text{if } a \text{ is credulous} \end{cases}$$
$$\mathrm{V}(a, k+1, \mathtt{t}.\mathbf{K}_a A) = \begin{cases} inc & \text{if } \mathrm{V}(a, k, \mathtt{t}.A) = inc \\ int & \text{otherwise} \end{cases}$$

Figure 1: Meaning of Formulae for Bounded Agents

The intuition behind the superficial evaluation is simple. Consider negation: if a formula is interesting, then also its negation is interesting. On the other case if a formula represents a unknown concept also its negation is an unknown concept. In the case of disjunction, if one disjunct of a formula is unknown then the overall status of the formula will be given by the remaining disjunct.

If the bound or the agent is not yet specified, we use the wild card $*$ as a symbol and the equations below, where $\langle a, k_a, type \rangle \in \mathcal{F}$:

$$\mathrm{V}(*, *, \mathtt{f}.\mathbf{K}_a A) = \mathrm{V}(a, k_a, \mathtt{f}.\mathbf{K}_a A)$$
$$\mathrm{V}(*, *, \mathtt{t}.\mathbf{K}_a A) = \mathrm{V}(a, k_a, \mathtt{t}.\mathbf{K}_a A)$$

We define the *first step of introspection* as first$(1) = *$ and first$(1.a_1.n_1.\dots a_k.n_k) = a_1$ and say that σ *belongs to an agent* a whenever first$(\sigma) = a$. Intuitively, this means that a is the "main" agent responsible for the chain of introspective reasoning which lead to the world σ. The *length* of a prefix, denoted by $|\sigma|$ is the number of introspection steps, i.e. $|1| = 0$ and $|1.a_1.n_1.\dots a_k.n_k| = k$.

Intuitively, each time we pass from σ to $\sigma.b.n$ we are performing an introspection step. So the length of σ tells how many introspection steps that we have already done. If σ belongs to an agent a with bound k_a then we are only left with $k_a - |\sigma|$ introspection steps.

Once V() is defined we simply need to compute $\mathrm{V}(\text{first}(\sigma), k_{\text{first}(\sigma)} - |\sigma|, \varphi)$ and reduce it only if it is interesting. Moreover we consider it classically contradictory with $\sigma : \overline{\varphi}$ only if both formulae are interesting. In all other cases an approximate decision (without search) is taken.

Formally we revise the definition of reduced formulae and contradictory branches from (Fitting 1983):

Definition 4 *Let $\mathcal{F}$ be a family of bounded agents and ap an approximate partition. The prefixed formula $\sigma : \varphi$ is* approximately reduced *iff*

- *it is interesting and reduced according the classical definition*
- *or $\sigma = 1$ and $V(*, *, \varphi) = inc$;*
- *or σ belongs to a and $V(a, k_a - |\sigma|, \varphi) = inc$.*

Definition 5 *Let $\mathcal{F}$ be a family of bounded agents and ap an approximate partition. A* branch *$\mathcal{B}$ is* approximately contradictory *iff*

- *$\mathcal{B}$ is contradictory for interesting formulae according the classical definition,*
- *or $\sigma = 1$ and $V(*, *, \varphi) = unk$,*
- *or σ belongs to a and $V(a, k_a - |\sigma|, \varphi) = unk$.*

If all tableaux branches are contradictory according the classical definition we have a *classical proof*. In a similar way a branch yields a *classical countermodel* if it is not contradictory and fulfils the classical definition.

We have now the formal machinery to construct an approximating sequence of proofs of increasing precision with the anytime algorithm in Fig. 2.

Input The modal formulae KB and A, and a time limit

Output A tableau $\mathcal{T}_k$

1. **choose** an approximate partition so that the set of interesting variables $I_0 = int(\mathcal{P})$ is small (e.g. logarithmic in the size of KB and A);
2. **choose** a family of bounded agents $\mathcal{F}_0 = \{ \langle a, k_a, type_a \rangle \}_{a \in \mathcal{A}}$ where k_a is a small constant, and decide whether $type_a$ is *c*redulous or *s*keptical;
3. **compute** an approximate tableau $\mathcal{T}_k$;
4. **if** the approximate tableau $\mathcal{T}_k$ is a classical proof or a classical counter model **then** stop;
5. **else if** not timeout
   (a) **find** the part of $\mathcal{T}_k$ where non classical reduction or closure rules have been applied
   (b) **increase** the precision so that one approximately reduced or contradictory formula change status by
      - **either** enlarging the set of interesting variables from I_k to I_{k+1} generating a new partition ap_{k+1};
      - **or** increasing the bound k_a for an agent a in the family $\mathcal{F}_k$ generating a new family $\mathcal{F}_{k+1}$;
   (c) **return to** (3) for a new $\mathcal{T}_{k+1}$;

Figure 2: Anytime Deduction Algorithm

If we time out after $\mathcal{T}_k$, we have available either an approximate proof (all branches of $\mathcal{T}_k$ are approximately closed) or an approximate counter model (an open branch of $\mathcal{T}_k$). With enough resources we can arrive at a stage where every agents in $\mathcal{F}_k$ has a bound k_a sufficiently large to guarantee that a classical proof (or counter model) is found.

For what regard the *semantical direction* of the approximation, we can increase the soundness of the procedure by increasing the bound of credulous agents or by decreasing the number of unknown proposition. Increasing the bound of skeptical agents or decreasing the number of incoherent proposition increase the completeness.

Quality of Approximation

Now we analize the quality of the approximation:

Theorem 1 (Approx. Soundness and Completeness) *Given an approximate partition and a family of bounded agents, the formula A is an approximate logical consequence of KB iff it has a approximate tableau proof.*

The standard soundness proof would proceed by induction showing that if there is a contradictory branch then we can unwind the contradiction down to atomic level. The only difficulty is to rule out "contradictory" incoherent formulae.

For the completeness the hard part is construction of the model from a reduced but not contradictory branch. It is similar to the traditional construction (Fitting 1983) with two tweaks: use two extra worlds (one subsuming all unknown worlds and one for all incoherent worlds) and consider only interesting propositions for the valuation $()^{\mathcal{I}}$.

The next result shows that, given enough resources we are guaranteed to find the correct (classical) answer:

Theorem 2 (Convergence Guarantee) *For every A and KB there is a k_0 such that for all $k \geq k_0$, $\mathcal{T}_k$ is a classical tableau proof for A iff $KB \models A$.*

The last step is proving the anytime complexity guarantee. At first we note that the verification of $\mathrm{V}(a, k, \varphi)$ takes polynomial time in the size of A and KB.

Without restrictions, we cannot guarantee that the timing of the computation of $\mathcal{T}_k$ and $\mathcal{T}_{k+1}$ respects the anytime guarantees from (Dean & Boddy 1988; Russel & Zilberstein 1991). We need the following assumptions:

- for generating ap_{k+1} and $\mathcal{F}_{k+1}$, apply a polynomial time algorithm to not interesting parts of $\mathcal{T}_k$;
- apply rules for constraint propagation (Massacci 1998) before other rules and apply rule for reducing $\mathtt{t}.A \wedge B$ formulae before other rules.

Then we can prove an *anytime guarantee* on the performance of the algorithm as follows: every branch has 2^{k_b} prefixes where k_b is the maximum bound of the family of bounded agents and for each prefix we can generate a number of branches which is exponential in the number of occurrences of interesting propositions.

This is not a good bound since occurrences of interesting formulae may be many. To improve the bound we consider *modal atoms* i.e. signed formulae $\mathtt{f}.A$ (or $\mathtt{t}.A$) where A where either $A = \mathbf{K}_a B$ or A is an atomic proposition.

If A is a modal atom such that both $\sigma : \mathtt{t}.A$ and $\sigma : \mathtt{f}.A$ are interesting, then we apply the cut rule to $\sigma : \mathtt{t}.A$ and $\sigma : \mathtt{f}.A$. Then, the systematic application of constraint propagation before other rules eliminate the atom from the branch. So an induction on the number of modal atoms provide the required anytime guarantee using the number (not the number of occurrences) of interesting modal atoms.

Conclusions

We have proposed an approximation technique for propositional modal logic which encompasses both propositional and modal aspects. The approximations makes it possible to tame unbounded logical omniscience and unbounded logical introspection both in the single and multi-agent case.

With this approximation it is possible to define skeptical (sound) approximations, credulous (complete) approximations and mixed approximations where some agents reason in a skeptical way and other agents in a credulous way.

We designed an anytime procedure which uses an *existing* deduction procedure (prefixed tableau) and we have discussed the "quality guarantees" of this approximate anytime deduction mechanism.

There are a number of open problems such as the extension to first order logic, the definition of a measure theory for the approximation, the developement of heuristics for the incremental selection of interesting propositions.

References

Crawford, J., & Kuipers, B. 1989. Towards a theory of access-limited logic for knowledge representation. In *Proc. of KR-89*, 67–78.

Dalal, M., & Etherington, D. 1992. Tractable approximate deduction using limited vocabularies. In *Proc. of the 9th Canadian Conf. on AI (AI-92)*, 206–221.

Dalal, M. 1996a. Anytime families of tractable propositional reasoners. In *Proc. of AI and Math. (AI/MATH-96)*, 42–45.

Dalal, M. 1996b. Semantics of an anytime family of reasoners. In *Proc. ECAI-96*, 360–364.

Dean, T., and Boddy, M. 1988. An analysis of time-dependent planning. In *Proc. AAAI-88*, 49–54.

del Val, A. 1994. Tractable databases: How to make propositional unit resolution complete through compilation. In *Proc. of KR-94*, 551–561.

Fagin, R., & Halpern, J. 1988. Belief, awareness, and limited reasoning. *AIJ* 34:39–76.

Fagin, R.; Halpern, J.; and Vardi, M. 1995. A nonstandard approach to the logical omniscience problem. *AIJ* 79:203–240.

Fitting, M. 1983. *Proof Methods for Modal and Intuitionistic Logics*. Reidel.

Halpern, J., and Moses, Y. 1992. A guide to completeness and complexity for modal logics of knowledge and belief. *AIJ* 54:319–379.

Kautz, H., & Selman, B. 1992. Planning as satisfiability. In *Proc. ECAI-92*, 359–363.

Lakemeyer, G. 1994, 1996. Limited reasoning in first-order knowledge bases. *AIJ* 71:1994 and *AIJ* 84:209–255, 1996.

Levesque, H. 1984. A logic of implicit and explicit belief. In *Proc. of AAAI-84*, 198–202.

Levesque, H. 1988. Logic and the complexity of reasoning. *J. of Philosophical Logic* 17:355–389.

Levesque, H. 1990. All I know, a study in autoepistemic logic. *AIJ* 42.

Massacci, F. 1998. Constraint propagation in propositional and modal tableaux. In *Proc. TABLEAUX-98*.

Patel-Schneider, P. 1989. A four-valued semantics for terminological logic. *AIJ* 38:319–351.

Russel, S., & Zilberstein, S. 1991. Composing real-time systems. In *Proc. IJCAI-91*, 212–217.

Schaerf, M., and Cadoli, M. 1995. Tractable reasoning via approximation. *AIJ* 74:1–62.

Selman, B., & Kautz, H. 1996. Knowlege compilation and theory approximation. *JACM* 43:193–224.