A Stochastic Belief Management Architecture for Agent Control

Gavin Rens, Thomas Meyer, Deshendran Moodley
Centre for Artificial Intelligence Research, CSIR Meraka, South Africa
University of Cape Town, Department of Computer Science, South Africa
{grens,tmeyer,deshen}@cs.uct.ac.za

Abstract

We propose an architecture for agent control, where the agent stores its beliefs and environment models as logical sentences. Given successive observations, the agent's current state (of beliefs) is maintained by a combination of probability, POMDP and belief change theory. Two existing logics are employed for knowledge representation and reasoning: the stochastic decision logic of Rens *et al.* (2015) and p-logic of Zhuang *et al.* (2017) (a restricted version of a logic designed by Fagin *et al.* (1990)). The proposed architecture assumes two streams of observations: *active*, which correspond to agent intentions and *passive*, which is received without the agent's direct involvement. Stochastic uncertainty, and ignorance due to lack of information are both dealt with in the architecture. Planning, and learning of environment models are assumed present but are not covered in this proposal.

1 Introduction

In this paper, we propose an architecture for an agent to manage its stochastic beliefs, given streams of noisy observations and the execution of actions with uncertain effects. The architecture is not meant to compete with well-established cognitive and generally intelligent architectures. However, we see an opportunity to combine several formalisms being studied in AI in a coherent way, and to suggest some extensions of these formalisms to make the whole architecture more generally applicable and robust.

Most of the detail of the architecture will focus on the probabilistic knowledge management of the agent. This management draws mainly from probabilistic knowledge representation, probabilistic belief change and decision theory. A key aspect of the proposed agent architecture is that it can cope with incomplete information, that is, the agent can still function (albeit poorly, depending on the degree of incompleteness) with only partial knowledge about the environment and its dynamics. We shall specify how an agent's beliefs should change, given its current (possibly incomplete) beliefs, the actions it takes and the observations it receives.

In this version of the architecture, observations are of two kinds, those associated directly with the agent's performed actions, and those which are received continually (in a stream), independent of the agent's actions. When we say *independent*, we mean those observations which the agent cannot tell whether they are caused by one of its actions. Observation will also be classified as either *ontic* (caused by a physical action or event) or *epistemic* (having a purely informational source, like an announcement).

It will be seen that, given an observation, the agent's beliefs are perpetuated in particular and (usually) different ways, depending on (i) whether the observation is associated with an agent action, or extracted form the passively received stream, (ii) whether the observation is ontic or epistemic and (iii) whether sufficient knowledge about the environmental dynamics is had by the agent at the time of perceiving the observation.

The agent uses its current belief base to inform a planner which generates a policy of actions, conditioned on observations. Planning is performed online, that is, in real-time, while the agent is deployed. The architecture assumes that the agent's reward function (the value it attaches to its actions) is kept up to date (e.g., with reinforcement learning). And it is assumed that various machine learning tools and AI techniques are used to form and recognize observations, whether action-associated observations or from the passively received data steam. Language or text understanding technology is also assumed available.

Agents with reasoning and high-level control based on the proposed architecture will be better suited to deployment in environments with relatively heterogeneous information sources. A typical domain for such an agent is a robot moving around in a building, where it may be given instructions and clarifying information about its surroundings.

To summarize, this paper makes the following contributions.

1. It proposes a coherent framework for general agent knowledge management and decision-making under uncertainty and ignorance.

2. It presents a means to deal with two kinds of steams of observations: *active* and *passive*, where observations from different streams are dealt with differently.

3. It presents a technique for belief maintenance which takes into account whether beliefs should be revised (due to erroneous beliefs) or updated (due to changes in the environment).

In the next section, we define some concepts and formalisms which are used in the rest of the paper. Section 3 presents the proposed agent architecture, providing detail about some of the components. Section 4 explains the various operators the agent may use to maintain its belief base, and how it selects the appropriate operator. We summarize our research and discuss some of the issues and shortcomings of this work in Section 5.

2 Background

All information in this architecture, including environment models and the agent's current state, is represented in formal/logical languages. All observations (pieces of evidence) are interpreted as sentences of these languages. An observation classified as ontic is always a propositional logic sentence, and an observation classified as epistemic is always a *p-logic* (Sect. 2.3) sentence. The agent's state is also represented with p-logic. Models about the environment are represented with the *stochastic decision logic* (Sect. 2.2).

2.1 Notation

To assist the reader, we list some of the symbols and abbreviations used in this paper.

- SBM: stochastic belief management
- SDL: stochastic decision logic
- L_P: the propositional language
- φ, ϕ, ψ: sentences in L_P
- L_{SDL}: the language of SDL
- L_{PL}: the language of p-logic
- Φ, Ψ: sentence in L_{SDL} or L_{PL}
- Ω: set of observation names
- ω: an observation name in Ω
- BG BB: background belief-base
- OB: observation buffer
- FG OB: foreground observation-buffer (sequence of SBM observations)
- SB: state base (part of the BG BB)
- DTB: decision-theory base (part of the BG BB)
- S: a SB
- D: a DTB
- Z: a FG OB
- z: an SBM observation element in Z
- Π: the set of all probability distributions
- Π^B: the set of distributions satisfying all constraints (sentences) in B

To simplify things, the symbols for the transition function/relation (T) and the observation function / perceivability relation (O) will be overloaded. This should not cause confusion in the contexts they are used.

2.2 Stochastic Decision Logic

The *stochastic decision logic* (SDL) [Rens *et al.*, 2015] was developed for (i) representing partially observable Markov decision processes (POMDPs) [Monahan, 1982; Lovejoy, 1991] in a logical language and (ii) ascertaining the truth of a query (statements in the language of SDL), given a knowledge-base of POMDP information and any additional knowledge expressible in the language.

Syntax

The vocabulary of the language contains six sorts of objects:

1. a finite set of *fluents* $F = \{f_1, \ldots, f_\ell\}$,
2. a finite set of names of atomic *actions* $A = \{a_1, \ldots, a_m\}$,
3. a countable set of *action variables* $V_A = \{v_1^A, v_2^A, \ldots\}$,
4. a finite set of names of atomic *observations* $\Omega = \{\omega_1, \ldots, \omega_n\}$,
5. a countable set of *observation variables* $V_\Omega = \{v_1^\Omega, v_2^\Omega, \ldots\}$.
6. all *real numbers* $\mathbb{R}$,

Fluents represent particular features of the environment. Sentences formed by combining fluents using $\land$ (and), $\lor$ (or) and $\neg$ (not) are statements about the (static) environment, may be true or false, and are denoted as L_P. Sentences may be prefixed with $\forall v$ or $\exists v$ to say that sentence holds in all cases, respectively, at least one case for variable v in the sentence replaced by the appropriate action or observation name.

Intuitively, $[\![a+\omega]\!]\Phi$ means 'Φ holds in the belief state resulting from performing action a and then perceiving observation ω'. For instance, $[\![a_1+\omega_1]\!]\,[\![a_2+\omega_2]\!]\Phi$ expresses that the agent executes a_1 then perceives ω_1 then executes a_2 then perceives ω_2, after which Φ is true.

Let $\bowtie \in \{<, \leq, =, \geq, >\}$. $[a]\varphi \bowtie p$ is read 'The probability x of reaching a φ-world after executing a is such that $x \bowtie p$'. $(\omega|a) \bowtie p$ is read 'The probability x of perceiving ω, given a was performed is such that $x \bowtie p$'. **B** is a modal operator for belief. $\mathbf{B}\varphi \bowtie p$ is read 'The degree of belief x in φ is such that $x \bowtie p$'. Performing $\Lambda = [\![a_1]\!][\![a_2]\!]\cdots[\![a_k]\!]$ means that a_1 is performed, then a_2 then ... then a_k. **U** is a modal operator for utility. $\mathbf{U}\Lambda \bowtie r$ is thus read 'The utility x of performing Λ is such that $x \bowtie r$'. Evaluating some sentence Ψ after a sequence of k update operations, means that Ψ will be evaluated after the agent's belief state has been updated according to the sequence

$$\underbrace{[\![a+\omega]\!]\cdots[\![a'+\omega']\!]}_{k \text{ times}}$$

of actions and observations. $\varphi \Rightarrow \Phi$ is read 'It is a general law of the domain that Φ holds in all situations (worlds) which satisfy φ'.

The following sentence atoms are also available. $c = c'$ for testing identity between action names or between observation names, and $Reward(r)$ and $Cost(a, r)$ for statements about the immediate reward, respectively, cost.

Given a complete formalization K of the scenario sketched here, a robot, for example, may have the following queries:

- Is the degree of belief that I'll have the oil-can in my gripper greater than or equal to 0.9, after I attempt grabbing it twice in a row? That is, does $[\![\texttt{grab}+\texttt{obsNil}]\!]\,[\![\texttt{grab}+\texttt{obsNil}\,]\!]\mathbf{B}\,\texttt{holding}\geq 0.9$ follow from K?
- After grabbing the can, then perceiving that it has medium weight, is the utility of drinking the contents of the oil-can, then placing it on the floor, more than 6 units? That is, does $[\![\texttt{grab}+\texttt{obsNil}]\!]\,[\![\texttt{weigh}+\texttt{obsMedium}]\!]\,\mathbf{U}[\![\texttt{drink}]\!][\![\texttt{replace}]\!]>6$ follow from K?

The language of SDL, denoted L_{SDL}, is the set of all sentences which can be constructed from the atoms described above and the connectives $\wedge$, $\vee$, $\neg$ and $\Rightarrow$, with some restrictions [Rens *et al.*, 2015].

Semantics

The SDL is based on POMDP theory. Formally, a partially observable Markov decision process (POMDP) is a tuple $\langle\Sigma, A, T, R, \Omega, O, b^0\rangle$: a finite set of states $\Sigma=\{s_1, s_2, \ldots, s_n\}$; a finite set of actions $A=\{a_1,a_2,\ldots,a_k\}$; the *state-transition function*, where $T(s,a,s')$ is the probability of being in s' after performing action a in state s; the *reward function*, where $R(a,s)$ is the reward gained for executing a while in state s; a finite set of observations $\Omega=\{\omega_1,\omega_2,\ldots,\omega_m\}$; the *observation function*, where $O(s',a,\omega)$ is the probability of observing ω in state s' resulting from performing action a in some other state; and b^0 is the initial probability distribution over all states in Σ.

Let b be a total function from Σ into $\mathbb{R}$. Each state s is associated with a probability $b(s)=p\in\mathbb{R}$, such that b is a probability distribution over the set Σ of all states. b can be called a *belief state*.

An important function in POMDP theory is the function that updates the agent's belief state, or the *state estimation* function *SE*. $SE(a,\omega,b)=b_n$, where $b_n(s')$ is the probability of the agent being in state s' in the 'new' belief state b_n, relative to a, ω and the 'old' belief state b. Notice that $SE(\cdot)$ requires an action, an observation and a belief state as inputs to determine the new belief state.

When the states an agent can be in are *belief*-states (as opposed to objective, single states in Σ), the reward function R must be lifted to operate over belief states. The *expected* reward $\rho(a,b)$ for performing an action a in a belief state b is defined as $\sum_{s\in\Sigma}R(a,s)b(s)$.

Let $w: F\to\{0,1\}$ be a total function (aka, a *world*) assigning a truth value to each fluent. Let C be the set of $2^{|F|}$ *conceivable worlds*, that is, all possible functions w.

Definition 2.1. *An SDL structure is a tuple* $\mathcal{D}=\langle T,P,U\rangle$ *such that*

- $T: A\to\{T_a \mid a\in A\}$, *where* $T_a:(C\times C)\to[0,1]$ *is a total function from pairs of worlds into the reals. That is, T provides a transition (accessibility) relation* T_a *for each action in A. For every* $w^-\in C$, *it is required that either* $\sum_{w^+\in C}T_a(w^-,w^+)=1$ *or* $\sum_{w^+\in C}T_a(w^-,w^+)=0$.
- $O: A\to\{O_a \mid a\in A\}$, *where* $O_a:(C\times\Omega)\to[0,1]$ *is a total function from pairs in* $C\times\Omega$ *into the reals. That is, O provides a perceivability relation* O_a *for each action in A. For all* $w^+\in C$, *if there exists a* $w^-\in C$ *such that* $T_a(w^-,w^+)>0$, *then* $\sum_{\omega\in\Omega}O_a(w^+,\omega)=1$, *else* $\sum_{\omega\in\Omega}O_a(w^+,\omega)=0$;
- *U is a pair* $\langle Re, Co\rangle$, *where* $Re: C\to\mathbb{R}$ *is a reward function and Co is a mapping that provides a cost function* $Co_a: C\to\mathbb{R}$ *for each* $a\in A$.

Definition 2.2. *The probability of reaching the next belief state* b' *from the current belief state b, given a and* ω, *is* $Pr_{NB}(a,\omega,b)=\sum_{w'\in C}O_a(\omega,w')\sum_{w\in C}T_a(w,w')b(w)$.

The above definition is from standard POMDP theory.

Definition 2.3. *A* belief update *function* $BU(a,\omega,b)=b'$ *is defined such that*

$$b'(w')=\frac{O_a(w',\omega)\sum_{w\in C}T_a(w,w')b(w)}{Pr_{NB}(a,\omega,b)},$$

for $Pr_{NB}(a,\omega,b)\neq 0$.

$BU(\cdot)$ has the same intuitive meaning as the state estimation function [Kaelbling *et al.*, 1998] of POMDP theory.

A reward function over belief states is derived for the SDL in a similar fashion as in POMDP theory, however, including the notion of cost: $RC(a,b)=\sum_{w\in C}(Re(w)-Co_a(w))b(w)$.

A sentence $\Phi\in L_{SDL}$ is *satisfiable* if there exists a structure $\mathcal{D}$, a belief state b and a world w such that Φ is true when evaluated with respect to $\mathcal{D}$, b and w (denoted $\mathcal{D}bw\models\Phi$), else Φ is *unsatisfiable*. Satisfaction is defined by Rens *et al.* (2015). Let $K\subset L_{SDL}$. K is said to *entail* Φ (denoted $K\models\Phi$) if for all structures $\mathcal{D}$, all belief states b, all $w\in C$: if $\mathcal{D}bw\models\kappa$ for every $\kappa\in K$, then $\mathcal{D}bw\models\Phi$. There exists an SDL decision procedure for entailment [Rens, 2014].

Domain Specification

Rens *et al.* (2015) present a framework for domain specification. "[T]he knowledge engineer should adapt the framework as necessary for the particular domain being modeled." In the context of the SDL, the domain of interest can be divided into five parts:

Static laws (denoted as the set *SL*) have the form $\phi\Rightarrow\varphi$, where ϕ and φ are propositional sentences, and ϕ is the condition under which φ is always satisfied. They are the basic laws and facts of the domain. For instance, "A full battery allows me at most four hours of operation", "I sink in liquids" and "The charging station is in sector 14". Such static laws cannot be explicitly stated in traditional POMDPs.

Action rules (denoted as the set *AR*) must be specified.

The basic kind is the *effect axiom.* For every action a, effect axioms take the form

$$\begin{aligned}\phi_1 &\Rightarrow [a]\varphi_{11}=p_{11}\wedge\cdots\wedge[a]\varphi_{1n}=p_{1n}\\ \phi_2 &\Rightarrow [a]\varphi_{21}=p_{21}\wedge\cdots\wedge[a]\varphi_{2n}=p_{2n}\\ &\vdots\\ \phi_j &\Rightarrow [a]\varphi_{j1}=p_{j1}\wedge\cdots\wedge[a]\varphi_{jn}=p_{jn},\end{aligned}$$

where (i) for every rule i, the sum of transition probabilities $p_{i1},\ldots,p_{in}$ must lie in the range $[0,1]$ (preferably 1), (ii) for every rule i, for any pair of effects φ_{ik} and $\varphi_{ik'}$, $\varphi_{ik}\wedge\varphi_{ik'}\equiv\bot$ and (iii) for any pair of conditions ϕ_i and $\phi_{i'}$, $\phi_i\wedge\phi_{i'}\equiv\bot$.

The knowledge engineer must keep in mind that if the transition probabilities do not sum to 1, the specification is incomplete, for instance, when for some rule i, $p_{i1} + \cdots + p_{in} < 1$.

Perception rules (denoted as the set *PR*) must be specified. Let $E(a) = \{\varphi_{11}, \varphi_{12}, \ldots, \varphi_{21}, \varphi_{22}, \ldots, \varphi_{jn}\}$ be the set of all effects of action a executed under all executable conditions. For every action a, perception rules typically take the form

$$\begin{aligned}
\phi_1 &\Rightarrow (\omega_{11} \mid a) = p_{11} \wedge \cdots \wedge (\omega_{1m} \mid a) = p_{1m} \\
\phi_2 &\Rightarrow (\omega_{21} \mid a) = p_{21} \wedge \cdots \wedge (\omega_{2m} \mid a) = p_{2m} \\
&\vdots \\
\phi_k &\Rightarrow (\omega_{k1} \mid a) = p_{k1} \wedge \cdots \wedge (\omega_{km} \mid a) = p_{km},
\end{aligned}$$

where (i) the sum of perception probabilities $p_{i1}, \ldots, p_{im}$ of any rule i must lie in the range $[0, 1]$ (preferably 1), (ii) for any pair of conditions ϕ_i and $\phi_{i'}$, $\phi_i \wedge \phi_{i'} \equiv \bot$ and (iii) $\phi_1 \vee \phi_2 \vee \cdots \vee \phi_k \equiv \bigvee_{\varphi \in E(a)} \varphi$. If the sum of perception probabilities $p_{i1}, \ldots, p_{im}$ of any rule i is 1, then any observations not mentioned in rule i are automatically *unperceivable* in a ϕ_i-world.

Utility rules (denoted as the set *UR*) must be specified. Utility rules typically take the form

$$\phi_1 \Rightarrow \mathit{Reward}(r_1), \quad \ldots, \quad \phi_j \Rightarrow \mathit{Reward}(r_j),$$

meaning that in all worlds where ϕ_i is satisfied, the agent gets r_i units of reward. And for every action a,

$$\phi_1 \Rightarrow \mathit{Cost}(a, r_1), \quad \ldots, \quad \phi_j \Rightarrow \mathit{Cost}(a, r_j),$$

meaning that the cost for performing a in a world where ϕ_i is satisfied is r_i units. The conditions are disjoint as for action and perception rules.

A specification of the worlds the agent should believe it is in when it becomes active, and probabilities associated with those worlds should be provided. The agent's (partial) initial belief state (denoted *IB*) can be specified in the SDL – and will have the form

$$\mathbf{B}\varphi_1 \bowtie p_1 \quad \wedge \quad \mathbf{B}\varphi_2 \bowtie p_2 \quad \wedge \quad \cdots \quad \wedge \quad \mathbf{B}\varphi_n \bowtie p_n,$$

where the φ_i are mutually exclusive propositional sentences (i.e., for all $1 \le i, j \le n$ s.t. $i \neq j$, $\varphi_i \wedge \varphi_j \equiv \bot$). In the SBM architecture, however, the agent's state is specified with p-logic, for reasons which will become clear. Note, that any set of p-logic sentences can be translated into an SDL (partial) belief state specification at the time of reasoning.

Suppose the union of *SL*, *AR*, *PR* and *UR* is an agent's *background knowledge* and denoted K_{SDL}. In practical terms, the question to be answered in the SDL is whether $K_{SDL} \models IB \rightarrow \Phi^-$ holds, where $K_{SDL} \subset L_{SDL}$, *IB* is as described above, and $\Phi^- \in L_{SDL}$ is some sentence of interest, which excludes subformulae of the form $\varphi \Rightarrow \Phi$ (i.e., concerning laws and rules).

Several nuances about domain specification with the SDL have not been covered here. The interested reader is referred to the literature [Rens, 2014; Rens *et al.*, 2015].

2.3 P-logic and P-revision

P-logic [Zhuang *et al.*, 2017] has three kinds of atomic (p-)formulae:

- $p(\phi) \bowtie t$
- $p(\phi) \bowtie c \cdot p(\psi)$
- $p(\phi) \bowtie p(\psi) + t$

where $\phi \in F$, $p(\phi)$ is read 'the probability of ϕ, $\bowtie \in \{\le, =, \ge\}$ and t and c are rational numbers such that $0 \le t \le 1$ and $c > 0$. A p-formula is a conjunction of atomic p-formulae.

Let $L_{PL}(F)$ be the language/set of all p-formulae which can be formed by (repeated) application of conjunction, and which involves all propositions in F. From here onwards, we write L_{PL} instead of $L_{PL}(F)$, assuming F to be implicitly known.

Probability distribution b satisfies an atomic p-formula

- $p(\phi) \bowtie t$ iff $b(\phi) \bowtie t$
- $p(\phi) \bowtie c \cdot p(\psi)$ iff $b(\phi) \bowtie c \cdot b(\psi)$
- $p(\phi) \bowtie p(\psi) + t$ iff $b(\phi) \bowtie b(\psi) + t$

where the ϕ and ψ can be considered *events* in the jargon of probability theory.

b satisfies a p-formula $\Phi \wedge \Psi$ iff it satisfies Φ and Ψ. We denote the fact that b satisfies Φ as $b \Vdash \Phi$.

Zhuang *et al.* (2017) define a function $* : 2^{L_{PL}} \times L_{PL} \rightarrow 2^{L_{PL}}$, which they call a p-revision function. They prove that a p-revision function satisfies six properties which are the p-logic versions of the six basic AGM revision postulates – AGM being the dominant framework in belief revision [Alchourrón *et al.*, 1985; Gärdenfors, 1988], "which represents the agent's beliefs and input information as formulas of some background logic that subsumes classical logic."

2.4 Sets of Distributions and Entropy Optimiztion

What we have called a *fluent* and an *atomic proposition*, is called an *event* in the jargon of probability theory (assuming an event can only be true/occurred or false/not occurred). An assignment of true/false to *every* event is called an *atomic event* in probability theory. Hence, the atomic events are exactly the possible worlds W of propositional logic.

Let W be a set of possible worlds. Let $\Pi(W)$ be all possible probability distributions (aka, belief states) over the elements in W. We shall always assume that the worlds in W are all the logical truth assignments of F, and that any belief state mentioned is a distribution over a given W induced by a given F.

Let S be a subset of L_{PL} representing an agent's state, that is, let S be the agent's state base (SB). S can be thought of as a set of constraints over the distributions in $\Pi(W)$. From here onwards, we shall write Π instead of $\Pi(W)$. Then we define

$$\Pi^S := \{b \in \Pi \mid \forall \Phi \in S, b \Vdash \Phi\}$$

to be the set of belief states consistent with (satisfying all sentences in) S.

The principle of maximum entropy can be stated as follows: The true belief state is estimated to be the one consistent with known constraints, but is otherwise as unbiased as possible, or "Given no other knowledge, assume that everything is as random as possible. That is, the probabilities are distributed as uniformly as possible consistent with the available information," [Poole and Mackworth, 2010].

The Shannon entropy of a distribution b is defined as

$$H(b) := -\sum_{w \in W} b(w) \ln b(w),$$

where b is a belief state. Traditionally, given some set of distributions Π', the most entropic distribution in Π' is defined as $\arg\max_{b \in \Pi'} H(b)$. One can thus represent Π^S (and thus S) by the single 'least biased' belief state, that is, the belief state in Π^S with *maximum entropy*:

$$ME(S) := \arg\max_{b \in \Pi^S} H(b).$$

It has been extensively argued [Jaynes, 1978; Shore and Johnson, 1980; Paris and Vencovská, 1997] that maximum entropy is a reasonable inference mechanism, if not the most reasonable one (w.r.t. probability constraints).

Suppose $T_a(w, w')$ is only partially specified. That is, suppose there are two or more worlds w' for which $T_a(w, w')$ is unknown, given a and w. $O_a(w,\omega)$ may be similarly underspecified: there could be two or more observations ω for which $O_a(w,\omega)$ is unknown, given a and w.

Similarly as is done to find a representative distribution of a SB employing entropy optimization, the fully specified transition and observation functions $T_a^{full}(\cdot)$ and $O_a^{full}(\cdot)$ can be inferred from the underspecified versions as follows. Let $D \subset L_{SDL}$ be an agent's decision-theory base (DTB). We define the set of possible transition functions compatible with D as

$$\Pi^{T_a(w)} := \{b \in \Pi \mid w' \in W \text{ and}$$
$$\text{if } b(w') = p, p' \neq p, \text{ then } D \not\models \phi^w \Rightarrow [a]\varphi^{w'} = p'\},$$

where $\varphi^{w'}$ is a proposition satisfied by w' and no other world, and ϕ^w is a proposition satisfied by w and no other world. Informally, $\Pi^{T_a(w)}$ is the set of distributions representing the transition function (for action a in w) which are not disallowed by the action rules in D. Then

$$T_a^{full}(w, w') = b^{T_a(w)}(w'),$$

where

$$b^{T_a(w)} = \arg\max_{b \in \Pi^{T_a(w)}} H(b).$$

From now on, we write $T_a^{full}(w, w')$ as $T^D(w, a, w')$, where D is a DTB.

Suppose, for instance, that the propositional vocabulary (F) is $\{q, r\}$ and let $w_1 \Vdash q \wedge r$, $w_2 \Vdash q \wedge \neg r$, $w_3 \Vdash \neg q \wedge r$ and $w_4 \Vdash \neg q \wedge \neg r$. And suppose the transition information for some action a_2 executed in w_3 is captured by

$$\neg q \wedge r \Rightarrow [a_2](q \wedge r) = 0.4 \wedge [a_2](q \wedge \neg r) = 0.3 \in D$$

Then $T_{a_2}^{full}(w_3, \cdot)$ is determined/inferred using entropy maximization as the distribution

$$\{(w_1, 0.4), (w_2, 0.3), (w_3, 0.15), (w_4, 0.15)\}$$

over arrival worlds.

Let Π^Ω be all possible probability distributions over the set of observations Ω. We define the set of possible observation functions compatible with D as

$$\Pi^{O_a(w)} := \{b \in \Pi^\Omega \mid \omega \in \Omega \text{ and}$$
$$\text{if } b(w') = p, p' \neq p, \text{ then } D \not\models \phi^w \Rightarrow (\omega \mid a) = p'\}.$$

Then

$$O_a^{full}(w,\omega) = b^{O_a(w)}(\omega),$$

where

$$b^{O_a(w)} = \arg\max_{b \in \Pi^{O_a(w)}} H(b).$$

Representation of and reasoning with observation probabilities needs to be dealt with in a special way to be useful in this work. In particular, we want a way to deal with observations as propositional sentences, but to use a POMDP-style observation function to reason about observation noise. We thus need a way to go from a function defined in terms of observation names (of objects) in Ω to a function defined in terms of observations in L_P.

Events can and are viewed as *evidence*, and in belief change theory, the notion of evidence is often used interchangeably with the notion of observation. In this work, we want to be able to assign a probability to *any* proposition (received as an observation; w.r.t. a world and an action). Therefore, we shall demand that there is a POMDP observation for every possible world (atomic event).

Let there be a bijection ζ between observations in Ω and worlds in W. This implies that there are as many observation names as there are worlds. Denote the observation name that w maps to as $\zeta(w)$. Then we define

$$O^D(a,\phi,w) := \sum_{w' \in W, w' \Vdash \phi} O_a^{full}(w,\zeta(w')).$$

To illustrate the idea, suppose again that $F = \{q, r\}$ and that

$$q \wedge \neg r \Rightarrow (\omega_1 \mid a_1) = 0.2 \wedge (\omega_2 \mid a_1) = 0.6 \in D.$$

Let $\zeta = \{(w_1,\omega_1),(w_2,\omega_2),(w_3,\omega_3),(w_4,\omega_4)\}$. Now suppose we need to know the probability of an observation in w_2 given a_1 was executed to reach w_2. Given ζ, $O_{a_1}^{full}(w_2,\cdot)$ is determined using entropy maximization as the distribution

$$\{(w_1, 0.2), (w_2, 0.6), (w_3, 0.1), (w_4, 0.1)\}$$

over observations. Hence, the (inferred) probability of observing r is $O^D(a_1, r, w_2) = 0.2 + 0.1 = 0.3$ and of observing $\neg q \vee \neg r$ is $O^D(a_1, \neg q \vee \neg r, w_2) = 0.6+0.1+0.1 = 0.8$. It is difficult to see how O^D could be so generally applicable without defining ζ as it is.

3 Conceptual Model of Architecture

Figure 1 shows the conceptual model of the stochastic belief management (SBM) architecture for agent control. The following subsections discuss the architecture components in detail.

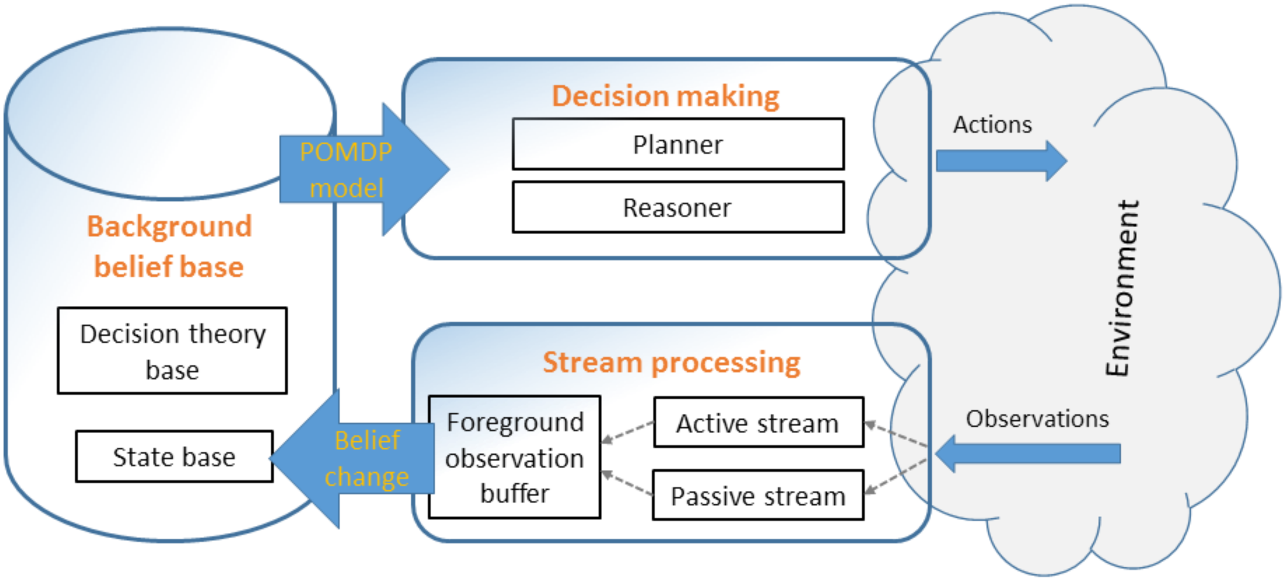


Figure 1: Conceptual model of architecture.

3.1 Observation Streams

There are two kinds of observation streams: active (atv) and passive (psv). Observations in the passive stream are assumed to be extracted from the stream periodically, for instance, every second or millisecond. Observations in the passive stream are not associated with a particular 'trigger' in the sense that the agent cannot attach a cause (action, event, request, intention) to the observation. Whereas the passive stream is periodic and continuous, the active stream is less rhythmic, due to every observation being paired with a preceding action or request for information, which is asynchronous. If a robot pushes a cup off a table and then the cup falls, then all the images and sounds associated with a cup falling on the ground can be placed in the context of the pushing action. The action thus provides an agent with contextual information helping it to update its beliefs more accurately. An action can also be a *sensing action*, for instance, `get-new-info` or `read-paragraph`. Sensing actions also create context; they help inform agents how to revise their beliefs.

For both streams, observations can be either *ontic* (physical cause) or *epistemic* (purely informational). Whether or not an observation has a physical cause can be quite philosophical. We live in a physical world, hence separating information from its physical representation can be challenging. However, we can use some common sense here: Suppose you are walking in town and a shop has a radio playing and the voice on the radio mentions the temperature in a neighboring city is 33 degrees. Considering the information of the city's temperature as an observation, what is its cause? In this case, we would say that 'neighboring city is 33 degrees' is an epistemic observation. And this observation would come through the passive stream because it was received without your solicitation.

Now suppose you know someone who lives in the neighboring city. You phone them up and ask what the temperature there is and they say it is 33 degrees. In this case, the observation is still 'neighboring city is 33 degrees' and it is still an epistemic observation. However, now it would come through the active stream because it was received due to a particular soliciting (sensing) action.

An element which is in the

- active stream and is ontic has the form (a,ϕ), where a is the action associated with observation ϕ.
- active stream and is epistemic has the form Φ, where $\Phi \in L_{PL}$.
- passive stream and is ontic has the form ϕ.
- passive stream and is epistemic has the form Φ, where $\Phi \in L_{PL}$.

Hence, all ontic elements involve a formula in L_P and all epistemic elements involve a formula in L_{PL}.

In the case of ontic observations in the active stream, a in element (a,ϕ), may loosely be thought of as the 'cause' of ϕ, but in general, ϕ is simply the dominant and meaningfully a-associated observation directly following the execution of a.

3.2 The Background Belief Base

For this version of the architecture, we divide the background belief base (BG BB) into two distinct parts. The state base (SB) tells the agent what possible situations it is in and, to some degree of completeness, the probability distribution over the possible situations. The decision-theory base (DTB) informs the agent about its decision-theory, that is, about the dynamics and utilities involved in the environment it inhabits.

SDL formulae of the form $\mathbf{B}\varphi \bowtie p$ could have been used to represent the SB, but we shall use p-logic due to its greater expressivity with respect to specifying constraints on belief states. By using p-logic, more sophisticated observations can be considered for SB maintenance and more sophisticated queries can be considered by the reasoner of the decision making component.

We propose that the the language of SDL be used to represent the DTB. It is an adequate language to specify (partial) POMDPs, using action rules, perception rules and utility rules as discussed in Section 2.2. In this study, we do not investigate how the the DTB is maintained due to interactions with the environment. But we assume that at any moment the DTB represents what the agent knows about the environmental dynamics and utilities.

Consider the SDL sentence of the form $[\![a+\omega]\!]\mathbf{B}\varphi \bowtie p$. This cannot be translated into a p-formula, yet it involves state information, that is, information about the agent's state after executing a and perceiving ω. Having knowledge about future degrees of belief (after a sequence of actions) is useful. For now, however, we shall assume that no SDL formulae involving future degrees of belief occur in the BG BB. Their inclusion is left for future work.

SDL static laws specify features of the environment which are immutable. They should thus have an influence on the agent's state beliefs, that is, on the SB. Again, in order to focus on the salient aspects of the proposed architecture, we assume that static law sentences do not occur in the BG BB.

3.3 The Foreground Observation Buffer

The *foreground observation buffer* (FG OB) keeps a record of the observations received from the passive and active streams. Observations are kept in the order which they were received. Therefore, an FG OB Z is of the form

$$(z_1, z_2, \ldots, z_n),$$

where z_i occurs before/to the left of z_j iff z_i was output temporally before z_j by the applicable stream processing module.

The SB is modified by z_1 via one of the appropriate belief change operations as soon as computational resources are available/allocated for the operation. Then z_1 is removed from Z, resulting in

$$Z = (z_2, z_3 \ldots, z_n, z_{n+1}, \ldots, z_{n+k}),$$

where k new observations have been stored in the time it takes to accommodate z_1 in the SB. Then z_2 is taken out of Z to modify the SB as soon as resources become available for the operation, and so on.

3.4 Belief Change Operations

Observations are removed from the FG OB as described in Section 3.3 and used to modify the agent's BG BB, specifically, the SB. Depending on the characteristics of the observation, an appropriate belief change operation must be selected to modify the SB. Several reference operations and a selection mechanism are discussed in Section 4.2.

3.5 Decision Making

The decision-making component comprises two subcomponents: a planner for generating behavior policies in real-time, and a reasoner for answering queries. Planning and reasoning in the SBM architecture are not the focus of this paper. Nonetheless, we can make some remarks.

Any state-of-the-art online POMDP algorithm/planner [Ross *et al.*, 2008] can be used to generate a finite policy for the agent to execute. A policy in the SBM architecture is taken to be finite length sequences of actions in a tree structure, with branching due to possible observations after actions. Such a tree structured policy is typical of online POMDP algorithms employing finite horizon forward search planning. Such algorithms require as input, a single current (root) belief state and a fully specified POMDP model. A single representative belief state b can be inferred from the SB S as $b = ME(S)$, and a (complete) transition function T^D and (complete) observation function O^D can be inferred from the (possibly incomplete) DTB D also via entropy optimization, as explained in Section 2.4.

Reasoning can be performed by asking whether some query Φ posed to the agent logically follows from the BG BB. In the case of queries about the agent's decision theory, the SDL decision procedure [Rens *et al.*, 2015] can be used – asking whether $D \models \Phi$ holds, where D is the DTB. In the case of queries about the agent's current state (of mind), the decision procedure [Fagin *et al.*, 1990] for the logic of which p-logic is a restricted version can be used – asking whether $S \models \Phi$ holds, where S is the SB. Although these two procedures exist and have been shown to be correct and terminating, they have not, to our knowledge, been implemented and optimized/analyzed with respect to efficiency.

4 General Belief Change

4.1 Appropriate Belief Representation

In this paper, we focus on the (stochastic) belief management of the state base (SB) part of the agent's background belief base (BG BB). The SB S is a set of probability constraints, not a belief state. Hence, it makes sense that p-revision is applied to S. But update as defined above operates on belief states, not sets of constraints. S represents the set of belief states consistent with it. Unfortunately, performing update of each belief state represented by S would not terminate when S represents an infinite number of belief states, which is typically the case.

One potential solution to this problem with update in this framework is to estimate a single belief state b to represent S and then apply the applicable update operation to b, producing b'. However, it is unknown how to change S, given b'. We would not want to simply keep on representing the agent's belief as a single belief state, because (i) the agent's ignorance is then ignored and (ii) p-revision is no longer applicable.

Another potential solution to this problem with update in this framework is to initially select a finite set of representative belief states $Rep(\Pi^S)$ and then never construct a new SB, but always operate on a set of belief states. This approach would mean that p-revision would not be applicable. A different kind of revision would have to be employed.

A third potential solution to this problem with update in this framework is to select a finite set of representative belief states $Rep(\Pi^S)$ and then apply the applicable update operation to every $b \in Rep(\Pi^S)$, producing $(Rep(\Pi^S))'$. An updated SB S' can then be induced from $(Rep(\Pi^S))'$. This is the method proposed by Rens *et al.* (2016) where they performed revision (via "generalized imaging") on a finite set of belief states. This approach is perhaps more acceptable, because it seems to retain some (if not all) of the ignorance captured by an SB.

In this work, we shall assume that one can always extract a finite and sufficient representative set of belief states $Rep(\Pi^S)$ from S. Denote the procedure of translating a set of belief states into a set of sentences (subset of L_{PL}) as *Sentencify*. Then we can define the update of a SB S due to observation z as

$$S \Delta z := Sentencify((\Pi^S)^{\Delta}_z),$$

where

$$(\Pi^S)^{\Delta}_z := \{b' \in \Pi \mid b' = b \Delta z,\, b \in Rep(\Pi^S)\},$$

where Δ is one of the four "Δ" operations defined below and z is an observation of the appropriate kind from the FG OB.

4.2 Operators and Operator Selection

Depending on whether an observation comes from the active or passive stream, the SB is changed differently. Furthermore, ontic observations are understood to require *belief update*, and epistemic observations are understood to require *belief revision*. Conventionally, update occurs when information is received in a dynamic environment [Katsuno and Mendelzon, 1992], while revision occurs when information is received in a static environment [Alchourrón *et al.*, 1985]. Consider the following scenario. You are new in the faculty department and you hear from someone that the door into the staff kitchen is open. Then a little while later, you hear that the kitchen-door is closed. Depending on whether you have the background knowledge that "things around here never change," you will likely process the 'kitchen-door is closed' evidence differently. Assuming all information sources are trustworthy, with the background info that things do not change, one would revise one's beliefs in a way that considers the initial belief that the door was open as a misunderstanding. On the other hand, without the background info, one would likely simply assume someone closed the door, and thus update one's belief according to the 'door-closing' context.

One can categorize observations broadly as *raw* and *marked*. A raw observation is simply the observation on its own without any information attached; the range of the degrees of belief of raw observations remain open until later processing (e.g., as an argument in a POMDP observation function). Marked observations are taken to be accompanied by information about the degree to which it can be believed, such as p-logic formulae. Another assumption made in this work is that epistemic observations are always marked. These are simplifying assumptions in order to get the project off the ground.

Regarding the dynamics of the environment, endogenous actions (of the agent) and exogenous events (in the environment; with origin outside the agent) cause transitions from one state to another. The probabilities with which these transitions occur (due to stochastic processes) my not always be known by the agent. We propose two kinds of update (when update is applicable): one method which uses state transition probabilities and one method which uses only information implicit in the structure of the available knowledge. A form of *imaging* will be used in the latter case.

Just as passive observations (not mentioning actions explicitly) can be ontic in nature, requiring update, so active observations can mention sensing actions, which trigger the reception of epistemic observations, requiring revision. Examples of sensing action are 'request information', 'activate sensor 3' and 'focus visual attention to quadrant 1'.

To summarize, we shall employ six kinds of belief change:

1. active ontic, requiring update – with state transition information
2. active ontic, requiring update – with implicit structure information
3. active epistemic, requiring revision
4. passive ontic, requiring update – with state transition information
5. passive ontic, requiring update – with implicit structure information
6. passive epistemic, requiring revision

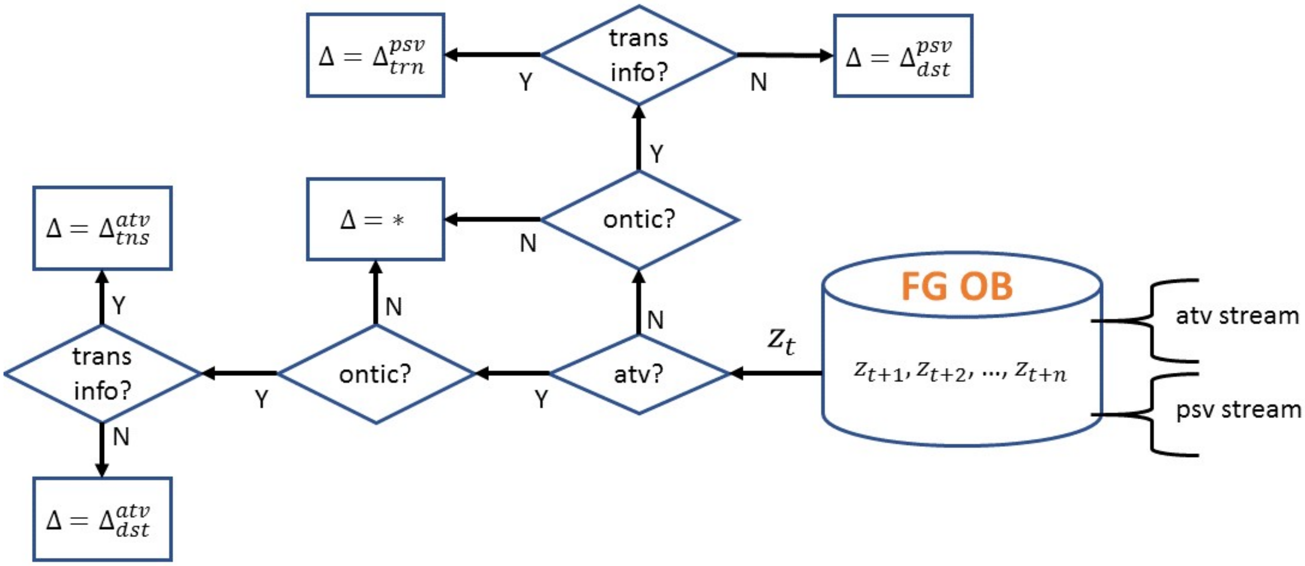
We shall assume that the agent recognizes every action as either ontic or sensing, thus allowing the agent to identify the subsequent observation as ontic, respectively, epistemic. We shall assume that observations extracted from the passive stream are marked as either ontic or epistemic. The final assumption we make is that for ontic observations, the agent can decide for every belief update, whether its transition information is sufficient for use during the update operation, or whether the less informative structural information should be used. Observation functions will always be deemed sufficient. That is, entropy optimization will always be used to infer $O^D(a, \phi, w)$, for all a, ϕ and w.

Table 1 summarizes which of the five belief change operators (to be defined below) is used in which of the six cases.

Table 1: Belief change operation per observation reception context.

	ontic		epistemic
	transition info?		
	✓	✗	
active	Δ^{atv}_{trn}	Δ^{atv}_{dst}	$*$
passive	Δ^{psv}_{trn}	Δ^{psv}_{dst}	$*$

The process for deciding on which operator to use in different contexts of observation is presented in Figure 2.



$\Delta = \Delta_{trn}^{psv}$
trans info?
$\Delta = \Delta_{dst}^{psv}$
Y
N
Y
$\Delta = \Delta_{tns}^{atv}$
$\Delta = *$
ontic?
N
FG OB
atv stream
Y
N
N
z_t
trans info?
ontic?
atv?
$z_{t+1}, z_{t+2}, \ldots, z_{t+n}$
Y
Y
psv stream
N
$\Delta = \Delta_{dst}^{atv}$

Figure 2: Process for deciding which operator to use for different observation contexts.

Definition 1 is taken almost directly from the POMDP state estimation function for update [Kaelbling *et al.*, 1998]. Definition 2 is generalized form of Lewis imaging [Lewis, 1976]. Definitions 3 is adapted from Rens (2016a).

Figure 2 is a data-flow diagram depicting the process for deciding which operator to use for different observation contexts.

Active Ontic with Transition Info

When (a,ϕ) is extracted from the active stream and ϕ is identified as being ontic and the agent deems its current model of state transitions due to a as sufficient, then belief update occurs and is defined as

$$b\,\Delta^{atv}_{trn}\,a,\phi(w) := \frac{1}{\gamma}O^D(a,\phi,w)\sum_{w'\in W} b(w')T^D(w',a,w), \quad (1)$$

where

$$\gamma := \sum_{w\in W} O^D(a,\phi,w)\sum_{w'\in W} b(w')T^D(w',a,w),$$

and where $O^D(a,\phi,w)$ and $T^D(w',a,w)$ are inferred, via entropy maximization, from the observation likelihood, respectively, transition information in D – as described in Section 2.4.

Active Ontic without Transition Info

When (a,ϕ) is extracted from the active stream and ϕ is identified as being ontic and the agent deems its current model of state transitions due to a as insufficient, then belief update occurs and is defined as

$$b\,\Delta^{atv}_{dst}\,a,\phi(w) := \frac{1}{\gamma}O^D(a,\phi,w)\sum_{w'\in W} b(w')\delta(\phi,w',w), \quad (2)$$

where

$$\gamma := \sum_{w\in W} O^D(a,\phi,w)\sum_{w'\in W} b(w')\delta(\phi,w',w),$$

and where $O^D(a,\phi,w)$ is inferred – via entropy maximization – from the observation likelihood information in D and $\delta(\phi,w',w)$ is an inverse-distance weight function, which has been designed for use with a generalized version [Rens and Meyer, 2017] of Lewis imaging.

Active Epistemic

When (a,Φ) is extracted from the active stream and Φ is identified as being epistemic, then belief revision occurs and is defined as

$$S \leftarrow S * \Phi,$$

where $*$ is a p-revision operator and S is a SB. Action a is, for instance, `get-new-info`, and Φ is a p-formula.

Passive Ontic with Transition Info

When ϕ is extracted from the passive stream and is identified as being ontic and the agent deems its current model of state transitions with respect to ϕ due to (exogenous) events as sufficient, then belief update occurs and is defined as

$$b\,\Delta^{psv}_{trn}\,\phi(w) := \frac{1}{\gamma}O^D(null,\phi,w)\sum_{w'\in W}\sum_{e\in\varepsilon} b(w')E(e,w)T^D(w',e,w), \quad (3)$$

where

$$\gamma := \sum_{w\in W} O^D(null,\phi,w)\sum_{w'\in W}\sum_{e\in\varepsilon} b(w')E(e,w)T^D(w',e,w),$$

and where $O^D(null,\phi,w)$ and $T^D(w',e,w)$ are inferred – via entropy optimization, ε is a set of (exogenous) events and E is the event likelihood function such that $E(e,w) = P(e \mid w)$ is the probability of the occurrence of event e in w. The SDL can easily be extended to represent/specify event likelihood functions; we do not expound on this here. We shall simply assume that E can be extracted from the DTB.

Passive Ontic without Transition Info

When ϕ is extracted from the passive stream and is identified as being ontic and the agent deems its current model of state transitions with respect to ϕ due to (exogenous) events as insufficient, then belief update occurs and is defined as Δ^{psv}_{dst}, which is exactly the same as Δ^{atv}_{dst}, except that the action a in $O^D(a,\phi,w)$ is taken to be *null*. The agent designer must define $O^D(null,\phi,w)$ appropriately for the domain of interest.

Passive Epistemic

When Φ is extracted from the passive stream and is identified as being epistemic, then belief revision occurs via p-revision ($*$) to change the SB, just as in the active epistemic case.

5 Conclusion and Future Work

Admittedly, this is a very preliminary proposal for an architecture, meant only to propose a basic framework and thus generate a discussion and ideas for its improvement. Unfortunately, time and other resources were unavailable to implement and evaluate a system based on the architecture.

There are many ways to sophisticate the architecture. In a sense, we have presented a skeleton framework, that is, a basic version of an agent architecture which can be expanded and elaborated upon. For instance [Van Harmelen *et al.*, 2008; Russell and Norvig, 2010; Wang and Goertzel, 2012], a generally intelligent agent should be able to represent and reason with

- the notion of defeasibility of knowledge
- independence of knowledge and conditional probabilities, e.g., employing Bayes nets
- ontologies of taxonomies
- notions of time and space
- notions of desires, intentions and goals
- notions of motivations and emotions
- knowledge abstraction and chunking

The representation and observation languages could be made richer. Belief change is a kind of learning, however, there is a need for much more learning and adaptation for an agent to be more general and autonomous. For instance, it is important that the agent can learn and maintain transition, observation and reward models. Planning and high-level control must, of course, be attended to. Attention focus is an important aspect of a sophisticated agent to be effective in the real world.

Although the SBM architecture is based on POMDPs, which assume a finite set of states, the set of states need not remain fixed over the agent's lifetime: Doshi-Velez (2009) suggests a framework called Infinite POMDP (iPOMDP) “that does not require knowledge of the size of the state space; instead, it assumes that the number of visited states will grow as the agent explores its world and only models visited states explicitly.” Moreover, environment models (represented as probability distributions) need not be completely known; as new environment dynamics are learnt, they can be incorporated as logical sentences, while the unspecified/unknown parts of the models are estimated using entropy optimization methods, as explained in the text. Also, Bayesian reinforcement learning has made headway in learning POMDP models [Ross *et al.*, 2011].

To make our architecture applicable to more general-purpose systems, where the system often runs into novel situations that have never happened before, there must be mechanisms that allow the system to compare the current situation to the past ones, and handle similarity evaluation and conflict resolution, in case that the current situation is similar to several previous ones when considering different aspects. In this regard, one could appeal to techniques of *case based reasoning* [Richter and Weber, 2013] to enhance the SMB architecture.

Although the capabilities mentioned above are necessary for a general and autonomous agent, it is unlikely that we shall be able to spend time on most of them in the foreseeable future. Where feasible, some of the capabilities could be included into the architecture as black-box modules.

The following is most likely to be our future work with the SBM architecture. Techniques for probabilistic belief change is an active area of research at this time; we would like to integrate the latest findings into the SBM architecture, as they develop. We are interested in ‘embeddedness’ of knowledge/beliefs: How should a belief base with different degrees of knowledge-embeddedness be maintained for different observation contexts and how should such a belief base be used during decision making?

With respect to planning and goal management, and for real-time planning in large POMDPs, ideas from (e.g.) the Hybrid POMDP-BDI agent architecture [Rens and Moodley, 2017], respectively, (e.g.) Partially Observable Monte-Carlo Planning (POMCP) [Silver and Veness, 2010] could be used to instantiate the planning component of the SBM architecture.

This work was inspired by Rens (2016b) who, in turn, was inspired by Voorbraak's Partial Probability Theory [Voorbraak, 1999]. And we quote Rens (2016b):

> Besides the work already cited in this paper, the following may be used as a bibliography to better place the present work in context, and to point to methods, approaches and techniques not covered in the proposed framework, which could possibly be added to it.
>
> - Probabilistic logics for reasoning with defaults and for belief change or learning [Goldszmidt and Pearl, 1996; Lukasiewicz, 2007].
> - Nonmonotonic reasoning systems with optimum entropy inference as central concept [Bourne and Parsons, 2003; Beierle and Kern-Isberner, 2008; 2009].
> - Dynamic epistemic logics for reasoning about probabilities [Van Benthem *et al.*, 2009; Sack, 2009].

Kern-Isberner and Lukasiewicz (2017) have written a useful, brief survey on reasoning under uncertainty, inconsistency, vagueness, and preferences.

References

[Alchourrón *et al.*, 1985] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

[Beierle and Kern-Isberner, 2008] C. Beierle and G. Kern-Isberner. On the modelling of an agent's epistemic state and its dynamic changes. *Electronic Communications of the European Association of Software Science and Technology*, 12, 2008.

[Beierle and Kern-Isberner, 2009] C. Beierle and G. Kern-Isberner. A conceptual agent model based on a uniform approach to various belief operations. In Bärbel Mertsching, Marcus Hund, and Zaheer Aziz, editors, *KI 2009: Advances in Artificial Intelligence: 32nd Annual German Conference on AI*, pages 273–280, Berlin, Heidelberg, September 2009. Springer Berlin Heidelberg.

[Bourne and Parsons, 2003] R. Bourne and S. Parsons. Extending the maximum entropy approach to variable strength defaults. *Ann. Math. Artif. Intell.*, 39(1–2):123–146, 2003.

[Doshi-Velez, 2009] Finale Doshi-Velez. The infinite partially observable markov decision process. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, 2009.

[Fagin *et al.*, 1990] R. Fagin, J. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87:78–128, 1990.

[Gärdenfors, 1988] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Massachusetts/England, 1988.

[Goldszmidt and Pearl, 1996] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84:57–112, 1996.

[Jaynes, 1978] E. Jaynes. Where do we stand on maximum entropy? In *The Maximum Entropy Formalism*, pages 15–118. MIT Press, 1978.

[Kaelbling *et al.*, 1998] L. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. *Artif. Intell.*, 101(1–2):99–134, 1998.

[Katsuno and Mendelzon, 1992] H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In P. Gärdenfors, editor, *Belief Revision*, pages 183–203. Cambridge University Press, 1992.

[Kern-Isberner and Lukasiewicz, 2017] G. Kern-Isberner and T. Lukasiewicz. Many facets of reasoning under uncertainty, inconsistency, vagueness, and preferences: A brief survey. *KI - Künstliche Intelligenz*, 31(1):9–13, Mar 2017.

[Lewis, 1976] D. Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85(3):297–315, 1976.

[Lovejoy, 1991] W. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28:47–66, 1991.

[Lukasiewicz, 2007] T. Lukasiewicz. Nonmonotonic probabilistic logics under variable-strength inheritance with overriding: Complexity, algorithms, and implementation. *International Journal of Approximate Reasoning*, 44(3):301–321, 2007.

[Monahan, 1982] G. Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.

[Paris and Vencovská, 1997] J. Paris and A. Vencovská. In defense of the maximum entropy inference process. *Intl. Journal of Approximate Reasoning*, 17(1):77–103, 1997.

[Poole and Mackworth, 2010] D. Poole and A. Mackworth. *Artif. Intell.: Foundations of Computational Agents.* Cambridge University Press, New York, USA, 2010.

[Rens and Meyer, 2017] G. Rens and T. Meyer. Imagining probabilistic belief change as imaging. Technical report, University of Cape Town, Cape Town, South Africa, 2017. url: http://arxiv.org/abs/1705.01172.

[Rens and Moodley, 2017] G. Rens and D. Moodley. A hybrid POMDP-BDI agent architecture with online stochastic planning and plan caching. *Cognitive Systems Research*, 43:1–20, June 2017.

[Rens *et al.*, 2015] G. Rens, T. Meyer, and G. Lakemeyer. A logic for reasoning about decision-theoretic projections. In B. Duval, J. Van den Herik, S. Loiseau, and J. Filipe, editors, *Proceedings of the Seventh Intl. Conf. on Agents and Artif. Intell. (ICAART), Revised Selected Papers*, LNAI, pages 79–99. Springer Verlaag, 2015.

[Rens *et al.*, 2016] G. Rens, T. Meyer, and G. Casini. On revision of partially specified convex probabilistic belief bases. In G. Kaminka M. Fox, editor, *Proceedings of the Twenty-second European Conference on Artificial Intelligence (ECAI-2016)*, pages 921–929. IOS Press, September 2016.

[Rens, 2014] G. Rens. *Formalisms for Agents Reasoning with Stochastic Actions and Perceptions.* PhD thesis, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, 2014.

[Rens, 2016a] G. Rens. On stochastic belief revision and update and their combination. In G. Kern-Isberner and R. Wassermann, editors, *Proceedings of the Sixteenth Intl. Workshop on Non-Monotonic Reasoning (NMR)*, pages 123–132. Technical University of Dortmund, 2016.

[Rens, 2016b] G. Rens. A stochastic belief change framework with an observation stream and defaults as expired observations. In R. Booth, G. Casini, S. Klarman,

G. Richard, and I. Varzinczak, editors, *Proceedings of the Third Intl. Workshop on Defeasible and Ampliative Reasoning (DARe-2016)*, ceur-ws.org, August 2016. CEUR Workshop Proceedings.

[Richter and Weber, 2013] M. Richter and R. Weber. *Case-Based Reasoning: A Textbook.* Springer Verlag, 2013.

[Ross *et al.*, 2008] S. Ross, J. Pineau, S. Paquet, and B. Chaib-draa. Online planning algorithms for POMDPs. *Journal of Artif. Intell. Research (JAIR)*, 32:663–704, 2008.

[Ross *et al.*, 2011] S. Ross, J. Pineau, B. Chaib-draa, and P. Kreitmann. A Bayesian approach for learning and planning in partially observable Markov decision processes. *J. Machine. Learning. Research.*, 12:1729–1770, July 2011.

[Russell and Norvig, 2010] S. Russell and P. Norvig. *Artificial. Intelligence.: A Modern Approach.* Prentice Hall, New Jersey, USA, 3nd edition, 2010.

[Sack, 2009] J. Sack. Extending probabilistic dynamic epistemic logic. *Synthese*, 169:124–257, 2009.

[Shore and Johnson, 1980] J. Shore and R. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *Information Theory, IEEE Transactions on*, 26(1):26–37, Jan 1980.

[Silver and Veness, 2010] D. Silver and J. Veness. Monte-carlo planning in large pomdps. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 2164–2172. Curran Associates, Inc., 2010.

[Van Benthem *et al.*, 2009] J. Van Benthem, J. Gerbrandy, and B. Kooi. Dynamic update with probabilities. *Studia Logica*, 93(1):67–96, 2009.

[Van Harmelen *et al.*, 2008] F. Van Harmelen, V. Lifshitz, and B. Porter. *The Handbook of Knowledge Representation.* Elsevier Science, 2008.

[Voorbraak, 1999] F. Voorbraak. Partial Probability: Theory and Applications. In *Proceedings of the First Intl. Symposium on Imprecise Probabilities and Their Applications*, pages 360–368, 1999.

[Wang and Goertzel, 2012] P. Wang and B. Goertzel, editors. *Theoretical Foundations of Artificial General Intelligence.* Atlantis Press, Paris, France, 2012.

[Zhuang *et al.*, 2017] Z. Zhuang, J. Delgrande, A. Nayak, and A. Sattar. A unifying framework for probabilistic belief revision. In F. Bacchus, editor, *Proceedings of the Twenty-fifth Intl. Joint Conf. on Artif. Intell. (IJCAI-17)*, Menlo Park, CA, 2017. AAAI Press. To appear.