A Cross-Field Review of State Abstraction for Markov Decision Processes

Congeduti, E.; Oliehoek, F.A.

Publication date
2022

Document Version
Accepted author manuscript

Published in
34th Benelux Conference on Artificial Intelligence (BNAIC) and the 30th Belgian Dutch Conference on Machine Learning (Benelearn)

Citation (APA)
Congeduti, E., & Oliehoek, F. A. (2022). A Cross-Field Review of State Abstraction for Markov Decision Processes. In *34th Benelux Conference on Artificial Intelligence (BNAIC) and the 30th Belgian Dutch Conference on Machine Learning (Benelearn)*

A Cross-Field Review of State Abstraction for Markov Decision Processes

Elena Congeduti[1][0000−0002−9568−6409] and Frans A. Oliehoek[1][0000−0003−4372−5055]

Delf University of Technology, Delft, The Netherlands
{E.Congeduti,F.A.Oliehoek}@tudelft.nl

Abstract. Complex real-world systems pose a significant challenge to decision making: an agent needs to explore a large environment, deal with incomplete or noisy information, generalize the experience and learn from feedback to act optimally. These processes demand vast representation capacity, thus putting a burden on the agent's limited computational and storage resources. State abstraction enables effective solutions by forming concise representations of the agents world. As such, it has been widely investigated by several research communities which have produced a variety of different approaches. Nonetheless, relations among them still remain unseen or roughly defined. This hampers potential applications of solution methods whose scope remains limited to the specific abstraction context for which they have been designed. To this end, the goal of this paper is to organize the developed approaches and identify connections between abstraction schemes as a fundamental step towards methods generalization. As a second contribution we discuss general abstraction properties with the aim of supporting a unified perspective for state abstraction.

Keywords: State Abstraction · Bounded Parameters Markov Decision Processes · Robust Reinforcement Learning · Model Irrelevance.

1 Introduction

Intelligent agents can not reason about every details of their structured and large world. They must necessarily base their decisions on a model of the environment that includes only a limited number of features. Intuitively, abstraction refers to the fundamental process to focus on important aspects of the surroundings while ignoring irrelevant information. Through abstraction, an agent builds compressed representations of its environment retaining only the essential information for a specific task that can be used to solve more complex and large problems. As a technique to ease decision making and learning algorithms in real-world scenarios and improve their scalability, abstraction has been extensively covered in artificial intelligence [24], operations research [23], theoretical computer science [8] and game theory [14] literature.

In this work, we focus on abstraction for Markov decision processes (MDPs) [21], for which a variety of approaches have been proposed within different research fields ranging from game-based abstraction (GBA) [19], temporal abstraction [26], value function approximation (VFA) [3]. Among them, state abstraction has been studied as a way to tackle computational and storage issues when dealing with prohibitively large sizes of the state space. The key idea is to form aggregated MDPs whose *abstract states* correspond to clusters of the original *ground states*. Grouping states together necessarily entails an additional degree of nondeterminism: it introduces further uncertainty with respect to the actual behavior of the system due to ignoring the exact ground state of the environment. Secondly, in general, partitioning the state space results in failure of the Markov property to hold. Consequently, the stochastic process induced can not be modeled as an MDP straightforwardly. Several perspectives have been adopted to deal with this issue and have led to defining abstract MDPs or other structured processes potentially suited to approximate the stochastic process on the abstract space. One natural way to represent the non-Markovian uncertainty over the ground true states is by means of partial observability. In other words, the dynamics over the partitioned abstract space can be interpreted as a Partially Observable Markov Decision Process (POMDPs) [13] for a particular choice of the observation distributions.

In this survey, we intend to highlight the close relations that ties in most of these approaches showing that many of them coincide or are equivalent when looking at them from the correct perspective. The goal of this investigation is to give rise to theoretical understanding of the various approaches. Moreover, the relevance of a unified perspective lies in the possibility to leverage techniques and solution methods developed for specific abstraction context to different abstraction approaches. The unifying perspective arising from these comparisons enlighten the key role played by the history-dependence in modeling the uncertainty on the ground state space. In fact, as already remarked by [2, 17], the state abstraction model can be viewed as a POMDP. Finally, we present a general definition of state abstraction with the aim of conveying a more organic perspective and support a unifying theoretical framework for state abstraction that can generalize most of the previous efforts in this direction. Based on this formal definition, we show how each abstraction schemes can be reinterpreted under this point of view.

Despite the attention that this topic has received lately, to date no such unified theory has been provided, relationships between approaches are still lacking or barely expressed and potential limitations and advantages unrevealed. With this respect, our work serves to bridge the gap between different research areas with the aim of inspiring new methods and techniques from the cross contamination of the different abstraction perspectives.

The rest of the paper is structured as follows. We first give an overview of previous works that survey state abstraction. We introduce a high level perspective on all the abstraction approaches and suggest an organic picture to classify them in Section 3. In Section 5, we introduce the formal definitions and dis-

cuss general abstraction properties. Then we formally describe the parallelism between state abstraction and POMDPs in Section 5.1. Finally in Section 4 we compare the different approaches by establishing connections and relations and proving equivalences.

1.1 Related Surveys

The intuitive idea of abstraction as a map from one problem representation to a new one which preserves certain properties has been introduced by [8]. In [23] the authors describe aggregation techniques and corresponding error bounds to reduce the computational burden in solving large-scale optimization problems. Although state abstraction for MDPs is directly addressed, they mainly focus their attention on discussing aggregation choices. The work [6] represents the first attempt to introduce a fairly general framework for state abstraction using propositional logic formalism. This work covers only cases where the states aggregated together share the same dynamics but not the rewards, thus only partly addressing the approximate abstraction case.

The recent growing interest in reinforcement learning along with the understanding of the potential of abstraction gave rise to several new abstraction approaches. Nonetheless, relatively few articles attempt to provide a comprehensive survey on the topic. Among them, [15] introduces a classification of abstraction approaches arranged in a hierarchical fashion according to the level of information that each of them preserves. The authors also propose a unified theoretical framework which generalizes over many of the previous aggregation mechanisms, specifically those based on bisimulations [9, 5], MDP homomorphisms [22], utile distinction [17] and policy irrelevance [12]. However, several important approaches can not directly be framed in their classification. Specifically, game-based abstraction [19], bounded parameters MDPs [10] and approaches based on robust control [20] are excluded and the tight connections that bind each other are neglected. We explore them in details in Section 4.

2 Background and Notation

A Markov Decision Process (MDP) is a tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ where $\mathcal{S}$ and $\mathcal{A}$ are finite state and action spaces, $\mathcal{T}(s'|s,a)$ and $\mathcal{R}(s,a)$ are the transition and reward functions and $\gamma \in (0,1)$ the discount factor [21]. To interact with the environment, an agent employs a policy $\pi : \mathcal{S} \longrightarrow \mathcal{A}$ mapping states to actions. The objective of an agent is to maximize its expected cumulative reward obtained by executing a policy π, that is the value

$$V^{\pi}(s) = \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^t \mathcal{R}^t | s, \pi\right]$$

when the process starts in state s. The target solution of the sequential decision problem modeled by the MDP is a policy π^* which maximizes the value $\pi^* = \text{argmax}_{\pi} V^{\pi}$.

A partially observable Markov decision process (POMDP) is an MDP in which the agent is unaware of the actual state of the system. Instead, it receives partial information on the environment state through an observation. Formally, a POMDP is a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \Omega, \mathcal{R}, \gamma)$, where $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ describes an MDP and $\mathcal{O}$, $\Omega(o|a, s')$ are the observation space and probabilities [13]. Given an action-observation history $h^t = (a^0, o^1, \ldots, a^{t-1}, o^t)$, the agent can keep track of a belief over the underlying state $b(s|h^t)$, as the probability of being in state s at time t given that history h^t. A POMDP can be transformed into an MDP over the space of all the possible histories, or equivalently the belief space, with rewards and deterministic transitions defined by

$$\begin{aligned} \rho(b, a) &= \sum_{s \in S} \mathcal{R}(s, a) b(s|h^t) \\ b(s'|h^{t+1}) &= \frac{\Omega(o|a, s') \sum_{s \in S} \mathcal{T}(s'|s, a) b(s|h^t)}{P(o|a, h^t)} \end{aligned} \tag{1}$$

for $h^{t+1} = (h^t, a, o)$. See [13] for more details.

We mark all the corresponding abstract objects with an overlying bar, as for instance the abstract state space as $\bar{\mathcal{S}}$ or an abstract policy $\bar{\pi} : \bar{\mathcal{S}} \longrightarrow \mathcal{A}$.

We use $\Delta(K)$ for the probability distributions simplex over the set K.

3 Abstraction Choices Overview

Different choices of modeling have led to the wide diversity of the state abstraction literature. Here we give an overview of the process of creating abstractions, pointing out places where different modeling choices can be made and how this gives rise to different approaches. The idea is that identifying a general procedure and attaching every approach to the specific stage of the abstraction process may help to avoid frequently encountered misconceptions, inconsistent comparisons and foster terminology alignment.

In the context of MDPs, abstraction describes the process of mapping one MDP into a new representation that retains, to some extent, the Markov property. The general abstraction process can be thought as the sequential application of the following steps:

1. Selection of the aggregation space.
2. Choice of the aggregation criteria.
3. Definition of the abstract dynamics.
4. Identification of a solution concept.

Although we do not assume that every approach has been conceived following this scheme, we believe that each of them can be characterized according to the authors choices about these four stages.

3.1 Aggregation Space

The key advantage of abstraction lies in leveraging representations to reduce the size of the initial problem and thus ease the solution search. This naturally leads to aggregation as a method to define abstract spaces and raises the question: what is a suitable space to aggregate on? We distinguish between approaches where aggregation is performed at the level of action-state histories as in feature MDPs [11] or influence-based abstraction (IBA) [18]. The second possibility consists of aggregating over the state space. Clearly, the distinction is subtle: aggregation over histories turns into state aggregation when considering states as trajectories of actions and states. In our work we mainly target the second class but it is important to underline how aggregating over spaces including the time dimension have the potential to directly overcome the non-Markovianity issue. In fact, as a POMDP regainsthe Markov property when considering histories h^t as states, by enriching the state space with histories it is more likely to find aggregation schemes that preserve the Markovianity of the system.

3.2 Aggregation Criteria

Concretely, to define state abstraction we need to establish a partition over the state space. A practical way to induce a partition is as the preimage of an aggregation function.

Definition 1. *An aggregation function* $\phi_{\bar{\mathcal{S}}} : \mathcal{S} \longrightarrow \bar{\mathcal{S}}$ *for an MDP* $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ *maps each ground state* s *to an abstract state* $\bar{s} = \phi_{\bar{\mathcal{S}}}(s)$ *with generally* $|\bar{\mathcal{S}}| \ll |\mathcal{S}|$.

With a slight abuse of notation, we use $\bar{s}$ to denote the abstract class $\phi^{-1}_{\bar{\mathcal{S}}}(s)$. Moreover, in the following text we implicitly assume the context of a ground MDP $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ and an aggregation function $\phi_{\bar{\mathcal{S}}}$, unless otherwise stated.

Several aggregation criteria, that is choices of the aggregation function, have been proposed based on different measures of similarity ranging from stochastic bisimulations [9, 5], irrelevance criteria [15, 1], MDP Homomorphisms [22], factors irrelevance [4]. Despite the diversification of notions, they overlap consistently. For instance, irrelevance criteria generalizes bisimulations: they are equivalent to one of the irrelevance characterizations identified by [15] as model irrelevance. MDPs homomorphisms also correspond to model irrelevance as long as the aggregation space is augmented by the action space. Factors irrelevance targets domains whose state space can be represented by some state variables and abstracts away entire factors that are irrelevant to the model dynamics. As such, it only induces exact abstraction, i.e. aggregation functions which cluster together states with the same transition functions. We will discuss how the exact case of every approach coincides for every choice of the aggregation function and model dynamics. Also heuristic approaches have been considered as utile distinction [17] and policy irrelevance [12] and there have been introduced methods to learn the aggregation function as model reduction techniques [9]. We refer to [7] for a complete survey of metrics for state similarity for MDPs.

3.3 Abstract Dynamics and Solutions

Given an aggregation function, we can consider the stochastic process that an MDP naturally induces over the sets of aggregated states. In general, the aggregated process does not inherit the Markov property. The only exception is when states that have the same probability of reaching any abstract state are clustered together. If in addition the reward functions are preserved then the stochastic process is an MDP which we refer to as an exact abstraction.

Definition 2 (Exact Abstraction). *The aggregated process induced over $\bar{\mathcal{S}}$ satisfies the Markov property if and only if for every $\bar{s} \in \bar{\mathcal{S}}$ and $s_1, s_2 \in \bar{s}$*

$$\sum_{s' \in \bar{s}'} \mathcal{T}(s'|s_1, a) = \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s_2, a) \qquad \forall \bar{s}', \forall a \tag{2}$$

Moreover, if the reward function satisfies

$$\mathcal{R}(s_1, a) = \mathcal{R}(s_2, a) \qquad \forall a \tag{3}$$

then we call exact abstraction the MDP $\bar{\mathcal{M}} = (\bar{\mathcal{S}}, A, \bar{T}, \bar{R}, \gamma)$ with transitions and rewards defined as

$$\bar{\mathcal{T}}(\bar{s}'|\bar{s}, a) = \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s_1, a)$$
$$\bar{\mathcal{R}}(\bar{s}, a) = \mathcal{R}(s_1, a)$$

for any of the representative state $s_1 \in \bar{s}$.

Note that in general, an abstract policy $\bar{\pi} : \bar{\mathcal{S}} \longrightarrow \mathcal{A}$ can be naturally extended to a ground policy as $\bar{\pi}(s) = \bar{\pi}(\bar{s})$. In the exact case, the optimal abstract solution $\bar{\pi}^*$ for the abstract MDP $\bar{\mathcal{M}}$, as ground policy, coincides with the optimal solution for the underlying MDP [15].

However, in most of the real-world cases few or none of such identical situations between ground states occur. Therefore, looking for exact abstractions results in a small reduction of the state space which does not facilitate substantially solutions algorithms. By relaxing the assumption on similarities between grouped states, we can hope to induce a significant state space reduction. The problem now consists of searching for a model which represents well the original aggregated process endowed with a solution concept. These two stages of abstraction differentiate the abstraction strategies that we intend to discuss thoroughly and compare in the following section.

4 A Comparison of Approaches

A first distinction between representations which strive to compensate for the Markovianity loss concerns those identifying a single abstract MDP and those employing families of MDPs.

4.1 Weighting Function Abstraction

Selecting one MDP translates essentially to set probability distributions at random for each abstract class that serve as 'weights' for the underlying states.

Definition 3 (Weighting Function Abstraction [15]). *Given a weighting function $\omega : S \longrightarrow [0,1]$, i.e. a probability distribution over each abstract class $\omega|_{\bar{s}} \in \Delta(\bar{s})$, a Weighting Function Abstraction (WFA) is an MDP $\bar{\mathcal{M}}_\omega = (\bar{\mathcal{S}}, \mathcal{A}, \bar{\mathcal{T}}_\omega, \bar{\mathcal{R}}_\omega, \gamma)$ with transitions and rewards defined as*

$$\bar{\mathcal{T}}_\omega(\bar{s}'|\bar{s}, a) = \sum_{s \in \bar{s}} \omega(s) \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a)$$

$$\bar{\mathcal{R}}_\omega(\bar{s}, a) = \sum_{s \in \bar{s}} \omega(s)\, \mathcal{R}(s, a)$$

Solving a WFA corresponds to find the optimal policy $\bar{\pi}^*_\omega$ for the abstract MDP $\bar{\mathcal{M}}_\omega$.

As remarked by [2], the issue of this approach is the idea of modeling the uncertainty over the underlying ground states by means of a stationary weighting function which is independent on the histories. This implies that the abstract process defined potentially deviates completely from the aggregated process and, as a consequence, the abstract optimal policy $\bar{\pi}^*_\omega$ may be completely ineffective. In fact, it corresponds to a myopic policy for a POMDP, which can have arbitrary loss [25].

4.2 Abstract Bounded Parameters MDPs

Relaxing the assumption of approximating the belief with a constant function, other approaches deal with uncertainty over ground states by considering families of MDPs. Bounded Parameters MDPs are generalizations of MDPs where transition and reward models are replaced by real intervals defined as the sufficient ranges to include the ground transitions and rewards.

Definition 4 (Abstract Bounded Parameters Markov Decision Processes [10]). *An Abstract Bounded Parameters MDP (ABPMDP) is a tuple $(\bar{\mathcal{S}}, \mathcal{A}, \bar{\mathcal{T}}_I, \bar{\mathcal{R}}_I, \gamma)$ where the transitions and rewards intervals are defined as*

$$\bar{\mathcal{T}}_I(\bar{s}'|\bar{s}, a) = \left[\min_{s \in \bar{s}} \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a)\,,\ \max_{s \in \bar{s}} \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a) \right] \tag{4}$$

$$\bar{\mathcal{R}}_I(\bar{s}, a) = \left[\min_{s \in \bar{s}} \mathcal{R}(s, a)\,,\ \max_{s \in \bar{s}} \mathcal{R}(s, a) \right] \tag{5}$$

Let us consider the set of all the possible MDPs whose transitions and rewards lie in the intervals (4), (5).

$$\mathcal{F}_{\text{ABPMDP}} = \{\bar{\mathcal{M}} = (\bar{\mathcal{S}}, \mathcal{A}, \bar{\mathcal{T}}, \bar{\mathcal{R}}, \gamma)\ : \\ \bar{\mathcal{T}}(\bar{s}'|\bar{s}, a) \in \bar{\mathcal{T}}_I(\bar{s}'|\bar{s}, a),\ \bar{\mathcal{R}}(\bar{s}, a) \in \bar{\mathcal{R}}_I(\bar{s}, a)\ \forall \bar{s}, a, \bar{s}'\} \tag{6}$$

According to the definitions given in Section 2, the value $\bar{V}^{\bar{\pi}}_{\bar{\mathcal{M}}}$ corresponds to the expected discounted reward for using a policy $\bar{\pi}$ in the abstract MDP $\bar{\mathcal{M}} \in \mathcal{F}$. Then, two solution policies are identified:

- An optimistically optimal policy $\bar{\pi}^*_{opt}$ satisfies $\max_{\bar{\mathcal{M}} \in \mathcal{F}} \bar{V}^{\bar{\pi}}_{\bar{\mathcal{M}}} \leq \max_{\bar{\mathcal{M}} \in \mathcal{F}} \bar{V}^{\bar{\pi}^*_{opt}}_{\bar{\mathcal{M}}}$, for every abstract policy $\bar{\pi}$.
- A pessimistically optimal policy $\bar{\pi}^*_{pes}$ satisfies $\min_{\bar{\mathcal{M}} \in \mathcal{F}} \bar{V}^{\bar{\pi}}_{\bar{\mathcal{M}}} \leq \min_{\bar{\mathcal{M}} \in \mathcal{F}} \bar{V}^{\bar{\pi}^*_{pes}}_{\bar{\mathcal{M}}}$, for every abstract policy $\bar{\pi}$.

In other words, the idea is to find the polices that perform optimally in the most and least favorable MDP of the ABPMDPs family $\mathcal{F}$. Clearly, considering a general MDP belonging to the ABMDP family would not bring better results than fixing a weighting function. However, solutions in face of pessimisms and optimisms allow to set lower and upper bounds for the target solution. We refer to [10] for the proof that the solution problems are well-posed and for the details on the bounds.

4.3 Abstract Robust MDPs

Strictly connected to the idea of allowing uncertainties in transitions and rewards model, other abstraction formulations rely on games structures [19, 30]. The idea is to model abstractions as a simple stochastic two-player game where a second agent takes on the additional uncertainty introduced by aggregation. We focus our attention specifically on methods based on robust MDP framework [29]. As BPMDPs, robust MDPs originate from the idea of MDPs with imprecise transitions probabilities [28] as models to deal with uncertainty due to statistical estimations of the model dynamics from historical data.

The key idea of the game-based abstraction is to represent the component of the nondeterminism introduced through abstraction by an additional agent whose actions, every time step, specify a probability distribution over the current abstract state. Formally, we introduce a *nature agent* whose action space corresponds to the ground state space $\mathcal{S}$. Its policy is defined by a function $\xi : \bar{\mathcal{S}} \times A \longrightarrow \Delta(\mathcal{S})$, such that $\xi(\bar{s}, a) \in \Delta(\bar{s})$. Different choices can be made as to what such policies condition on. We take the most widely adopted perspective which is referred to as (s, a)-rectangularity [20]. Once set the nature policy, the problem turns into a regular MDP.

Definition 5 (Abstract Robust Markov Decision Processes [20]). *Given a nature policy ξ, an Abstract Robust MDP (ARMDP) is an MDP $\mathcal{M}_\xi = (\bar{\mathcal{S}}, \mathcal{A}, \bar{\mathcal{T}}_\xi, \bar{\mathcal{R}}_\xi, \gamma)$ where the transitions and rewards are defined as*

$$\bar{\mathcal{T}}_\xi(\bar{s}'|\bar{s}, a) = \sum_{s \in \bar{s}} \xi(\bar{s}, a)(s) \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a) \tag{7}$$

$$\bar{\mathcal{R}}_\xi(\bar{s}, a) = \sum_{s \in \bar{s}} \xi(\bar{s}, a)(s)\ \mathcal{R}(s, a) \tag{8}$$

Clearly, different choices for the nature policy lead to different MDPs. The solution $\bar{\pi}^*_\xi$ identified by this approach is defined assuming a fully adversarial behavior of the nature agent, i.e. $\bar{\pi}^*_\xi$ maximizes

$$\max_{\bar{\pi}} \min_{\xi} \bar{V}^{\bar{\pi}}_{\xi}(\bar{s}) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \bar{\mathcal{R}}^t_\xi \mid \bar{s}, \bar{\pi}\right]$$

4.4 Relationships

Each of the approaches introduced induces a set of abstract MDPs. As $\mathcal{F}_{\text{BPMDP}}$ defined by (6) for the abstract BPMDP, we consider the families

$$\mathcal{F}_{\text{WFA}} = \{\bar{\mathcal{M}}_\omega : \omega \in \Delta(S),\, \omega|_{\bar{s}} \in \Delta(\bar{s})\ \forall \bar{s}\} \tag{9}$$

$$\mathcal{F}_{\text{ARMDP}} = \{\bar{\mathcal{M}}_\xi : \xi(\bar{s}, a) \in \Delta(S),\ \xi(\bar{s}, a)|_{\bar{s}} \in \Delta(\bar{s})\ \forall \bar{s},\, a\} \tag{10}$$

where $\omega|_{\bar{s}}$ and $\xi(\bar{s}, a)|_{\bar{s}}$ represent the function restrictions to the states $s \in \bar{s}$. Equations (9) and (10) define respectively the set of all the possible weighting function abstractions and abstract MDPs for different nature policies. The relationships that bind together all these approaches follow quite straightforwardly from the definitions.

Theorem 1. *Consider the WFA, ABPMDP and ARMDP as in definitions 3, 4, 5 and the corresponding sets of abstract MDPs $\mathcal{F}_{WFA}$, $\mathcal{F}_{ABPMDP}$, $\mathcal{F}_{ARMDP}$ as defined in* (9), (6),(10)*, then*

1. *The family of weighting function abstractions is a subset of the abstract MDPs generated by a nature policy which, in turn, is a subset of the abstract BPMDP family, i.e. the following chain of inclusions holds*

$$\mathcal{F}_{WFA} \subseteq \mathcal{F}_{ARMDP} \subseteq \mathcal{F}_{ABPMDP}$$

2. *If the aggregation function defines an exact abstraction, i.e.* (2), (3) *hold, then*

$$\mathcal{F}_{WFA} = \mathcal{F}_{ARMDP} = \mathcal{F}_{ABPMDP}$$

and they degenerate to the same abstract MDP corresponding to the exact abstraction.

Proof. 1. For the first inclusion we consider an arbitrary weighting function ω and the MDP $\bar{\mathcal{M}}_\omega \in \mathcal{F}_{WFA}$. Then, according to (7)*, the nature policy $\xi(\bar{s}, a) \coloneqq \omega|_{\bar{s}}$ induces an equivalent abstract robust MDP $\bar{\mathcal{M}}_\xi = \bar{\mathcal{M}}_\omega \in \mathcal{F}_{ARMDP}$. In essence, the subset of nature policies ξ which are independent on the actions, $\xi(\bar{s}, a_1) = \xi(\bar{s}, a_2)$ for any $a_1, a_2 \in A$, span the entire family $\mathcal{F}_{WFA}$.*

In order to show the second inclusion, it suffices to observe that for any policy ξ the transitions and rewards defined by (7) *satisfy*

$$\bar{\mathcal{T}}_\xi(\bar{s}'|\bar{s}, a) \leq \sum_{s \in \bar{s}} \xi(\bar{s}, a)(s) \max_{s \in \bar{s}} \left\{ \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a) \right\} =$$

$$\max_{s \in \bar{s}} \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a) \sum_{s \in \bar{s}} \xi(\bar{s}, a)(s) = \max_{s \in \bar{s}} \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a)$$

Likewise we can show that

$$\bar{\mathcal{T}}_{\xi}(\bar{s}'|\bar{s},a) \geq \min_{s\in\bar{s}} \sum_{s'\in\bar{s}'} \mathcal{T}(s'|s,a)$$

therefore $\bar{\mathcal{T}}_{\xi}(\bar{s}'|\bar{s},a) \in \bar{\mathcal{T}}_{I}(\bar{s}'|\bar{s},a)$ *and similarly,* $\bar{\mathcal{R}}_{\xi}(\bar{s},a) \in \bar{\mathcal{R}}_{I}(\bar{s},a)$.

2. We show first that under the assumption of exact aggregation function, the intervals (4) *and* (5) *defining the abstract family* $\mathcal{F}_{BPMDP}$ *consist of one point. In fact by* (2) *and* (3)

$$\bar{\mathcal{T}}_{I}(\bar{s}'|\bar{s},a) = \left[\sum_{s'\in\bar{s}'} \mathcal{T}(s'|s_1,a)\,,\ \sum_{s'\in\bar{s}'} \mathcal{T}(s'|s_1,a)\right]$$
$$\bar{\mathcal{R}}_{I}(\bar{s},a) = [\mathcal{R}(s_1,a)\,,\ \mathcal{R}(s_1,a)]$$

for an arbitrary choice of $s_1 \in \bar{s}$. *Thus* $\mathcal{F}_{BPMDP}$ *contains only the abstract exact MDP. From the inclusions given by 1., it follows the equality.*

According to the second claim of Theorem 1, when complete equivalence between aggregated states holds, all the approaches provide an exact approximation of the aggregated process which indeed naturally inherits the Markov property. Contrarily, when Markovianity is not preserved, in general the families do not coincide and thus neither do the solutions. Nonetheless, a weighting function abstraction corresponds to a nature policy independent on the agent action. In turn, for each nature policy an abstract robust MDP can be seen as an MDP of the abstract BPMDPs family. In general, these are all proper inclusions determined by the different dependencies of the uncertainty model: a weighing function is an history-independent function; the nature policy, assuming the (s,a)-rectangularity structure, depends on the current action; abstract BPMDP introduces an additional dependence on the next abstract state and furthermore allows different believes for transition and rewards as we clarify in the next section.

Another class of state abstraction approaches rely on value function approximations (VFA) [3, 27]. The key idea is to approximate the non-Markovian model dynamics by making use of projections onto the aggregated space. More precisely, the projections account for the uncertainty over the state space: each ground state is projected onto the abstract space by a predefined projection function. According to the projection metric chosen, different abstract MDPs can be defined. As such, the metrics play the same role as the weighting functions and the two approaches are completely equivalent. Although we do not discuss explicitly this case in Theorem 1, to formalize this idea it suffices to observe that the structure of the projections used depends only on the current abstract state and coincides precisely with that of a weighting function.

Table 1 summarizes the main features of all the approaches covered in this survey. The global picture arising consists of abstraction schemes which deal with uncertainty by introducing families of abstract MDPs and deriving solution concepts by targeting a specific MDP within the family. The distinctive feature

Table 1. Schematic representation of abstraction approaches.

Approach	Aggregation Space	Abstract representation	Solution Concept	Type	Uncertainty
IBA [18]	histories	MDP	optimality	exact	-
ABPMDPs [10]	states	MDP set	pessimism /optimism	approx	$\lambda(\bar{s}, a, \bar{s}')$
ARMDPs [20]	states	Robust MDP	pessimism	approx	$\lambda(\bar{s}, a)$
WFA [15]	states	MDP	optimality	approx	$\lambda(\bar{s})$
GBA[19]	states	stochastic game	pessimism /optimism	approx	-
VFA [3]	states	MDP	optimality	approx	$\lambda(\bar{s})$
POMDPs [2]	states	POMDP	optimality	approx	$\lambda(\bar{h}^t)$

is mainly the extent to which the abstract state-action history determines the uncertainty model. In the next section, we explore the relevance of including the abstract trajectories in the uncertainty model and how under this perspective abstraction can be interpreted as partial observability.

5 A General Framework

The analysis of similarities between state abstraction approaches lead us to a universal definition which embeds them all.

We define the sets of candidate uncertainty functions as

$$\begin{aligned} \mathcal{U}_T &= \{\lambda_T(\bar{s}, a, \bar{s}') : \lambda_T(\bar{s}, a, \bar{s}') \in \Delta(\bar{s}) \ \forall \bar{s}, a, \bar{s}'\} \\ \mathcal{U}_R &= \{\lambda_R(\bar{s}, a) : \lambda_R(\bar{s}, a) \in \Delta(\bar{s}) \ \forall \bar{s}, a\} \end{aligned} \tag{11}$$

Each pair $\lambda = (\lambda_T, \lambda_R) \in \mathcal{U}_T \times \mathcal{U}_R$ defines uniquely an abstract MDP $\bar{\mathcal{M}}_\lambda = (\bar{\mathcal{S}}, \mathcal{A}, \bar{\mathcal{T}}_\lambda, \bar{\mathcal{R}}_\lambda, \gamma)$ with transitions and rewards as

$$\begin{aligned} \bar{\mathcal{T}}_\lambda(\bar{s}'|\bar{s}, a) &= \sum_{s \in \bar{s}} \lambda_T(\bar{s}, a, \bar{s}')(s) \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a) \\ \bar{\mathcal{R}}_\lambda(\bar{s}, a) &= \sum_{s \in \bar{s}} \lambda_R(\bar{s}, a)(s) \sum_{s' \in \bar{s}'} \mathcal{T}(s'|s, a) \end{aligned}$$

Definition 6 (State Abstraction). *Given an uncertainty set as $\mathcal{U} \subset \mathcal{U}_T \times \mathcal{U}_R$, a state abstraction is the collection of MDPs*

$$\mathcal{F}_\mathcal{U} = \{\bar{\mathcal{M}}_\lambda = (\bar{\mathcal{S}}, \mathcal{A}, \bar{\mathcal{T}}_\lambda, \bar{\mathcal{R}}_\lambda, \gamma) \ : \ \lambda \in \mathcal{U}\}$$

The uncertainty set $\mathcal{U}$ essentially encodes the additional nondeterminism introduced by the aggregation and its characterizations formally capture the differences between abstraction approaches. In fact, if we consider independent uncertainty with respect to actions and next abstract states $\mathcal{U} = \{(\lambda_T, \lambda_R) \ : \ \lambda_T = \lambda_R = \lambda(\bar{s}) \in \Delta(\bar{s})\}$, then the state abstraction represents the weighting function families, $\mathcal{F}_\mathcal{U} = \mathcal{F}_{\text{WFA}}$. Instead, including the dependency on the actions as

$\mathcal{U} = \{(\lambda_T, \lambda_R) : \lambda_T = \lambda_R = \lambda(\bar{s}, a) \in \Delta(\bar{s})\}$, we obtain $\mathcal{F}_{\mathcal{U}} = \mathcal{F}_{\text{ARMDP}}$. Finally, if we allow different lambda functions for transitions and rewards and dependency on next abstract states, that is $\mathcal{U} = \mathcal{U}_T \times \mathcal{U}_R$, we obtain $\mathcal{F}_{\mathcal{U}} = \mathcal{F}_{\text{ABPMDP}}$.

5.1 Abstraction as partial observability

In [2] the authors show how the stochastic process over the aggregated space turns out to be a POMDP with the original ground MDP as the underlying MDP and the abstract states as observations. The idea is that generally the non-Markovian aggregated process can not be reasonably approximated by a Markov dynamics. Instead, including histories in the dynamics allows to define the process over the abstract space as a POMDP.

Definition 7 (Abstract POMDP). *An abstract POMDP* $\bar{\mathcal{M}}_{POMDP} = (\mathcal{S}, \mathcal{A}, \bar{\mathcal{S}}, \mathcal{T}, \Omega, , \mathcal{R}, \gamma)$ *is a POMDP with* $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ *as underlying MDP. The observation space consists of the abstract space* $\bar{\mathcal{S}}$ *with deterministic observation probabilities defined as*

$$\Omega(\bar{s}'|a, s') = \mathbb{1}_{s' \in \bar{s}}$$

To point out the relation between the abstract POMDP and Definition 6, we can think to expand the notion of uncertainty as to include the entire action-abstract state history $\bar{h}^t = (\bar{s}^0, a^1, \dots, a^{t-1}, \bar{s}^t)$ by setting $\mathcal{U} = \{\lambda_T = \lambda_R = \lambda(\bar{h}^t) \in \Delta(\bar{s}^t)\}$. Then given the transitions and rewards

$$\bar{\mathcal{T}}_\lambda(\bar{s}^{t+1}|\bar{h}^t, a^t) = \sum_{s^t \in \bar{s}^t} \lambda(\bar{h}^t)(s^t) \sum_{s^{t+1} \in \bar{s}^{t+1}} \mathcal{T}(s^{t+1}|s^t, a^t)$$
$$\bar{\mathcal{R}}_\lambda(\bar{h}^t, a^t) = \sum_{s^t \in \bar{s}^t} \lambda(\bar{h}^t)(s) \mathcal{R}(s^t, a^t)$$

then the family $\mathcal{F}_{\mathcal{U}}$ includes also non-Markovian processes. If we impose the additional constraint on the non-stationary lambda functions to follow the belief update rule (1), resulting in $\mathcal{U} = \{\lambda_T = \lambda_R = \lambda(\bar{h}^t) \in \Delta(\bar{s}^t), \lambda(\bar{h}^t) = \text{belief-up}(\bar{h}^{t-1}, a^t, \bar{s}^t)\}$, then the state abstraction $\mathcal{F}_{\mathcal{U}}$ is equivalent to the abstract POMDP $\bar{\mathcal{M}}_{\text{POMDP}}$ and provides and exact description of the aggregated process.

This argument reveals the necessity of including history in modeling uncertainty for state abstraction. In the case of weighting function, for instance, we are approximating the belief $\lambda(\bar{h}^t)$ with an arbitrary history-independent function $\omega = \lambda(\bar{s}^t)$. Therefore we can expect to achieve at most the value of an optimal memory less policy which is well-known that can obtain arbitrarily bad performance [16].

On the other hand, one may think that by modeling state abstraction as a POMDP we would lose all the benefits of the problem size reduction for which abstraction has been introduced [17]. One idea may be to adopt intermediate solutions where the history is partially included and explore trade-off situations between problem size reduction and history dependency.

6 Conclusions

In this paper, we survey the main approaches for state abstraction developed within reinforcement learning, planning, operations research and game theory communities. We characterize a general abstraction scheme by identifying stages of the abstraction building process corresponding to the choice of the aggregation space and function, the model dynamics and solution concept. Then, we mainly focus on comparing techniques introduced to model the abstract dynamics and solutions as weighting function abstractions, bounded parameters MDPs, robust MDPs and value function approximations. We show how they can all be interpreted under the unified perspective of families of abstract MDPs, where the distinctions are established by the model of the uncertainty employed. Finally we introduce a unified formal framework which generalizes all the prior approaches and highlight how to embed the partial observability perspective into this general definition of state abstraction.

Acknowledgements This project had received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No.758824 —INFLUENCE).

References

1. Abel, D., Hershkowitz, D., Littman, M.: Near optimal behavior via approximate state abstraction. In: International Conference on Machine Learning (ICML). pp. 2915–2923 (2016)
2. Bai, A., Srivastava, S., Russell, S.J.: Markovian state and action abstractions for MDPs via hierarchical MCTS. In: International Joint Conference on Artificial Intelligence (IJCAI). pp. 3029–3039 (2016)
3. Bertsekas, D.P.: Approximate dynamic programming (2008)
4. Boutilier, C., Dearden, R., Goldszmidt, M.: Stochastic dynamic programming with factored representations. Artificial intelligence **121**(1-2), 49–107 (2000)
5. Dean, T., Givan, R., Leach, S.: Model reduction techniques for computing approximately optimal solutions for Markov decision processes. In: UAI. pp. 124–131 (1997)
6. Dearden, R., Boutilier, C.: Abstraction and approximate decision-theoretic planning. Artificial Intelligence **89**(1-2), 219–283 (1997)
7. Ferns, N., Castro, P.S., Precup, D., Panangaden, P.: Methods for computing state similarity in markov decision processes. In: UAI. p. 174–181 (2006)
8. Giunchiglia, F., Walsh, T.: A theory of abstraction. Artificial intelligence **57**(2-3), 323–389 (1992)

9. Givan, R., Dean, T., Greig, M.: Equivalence notions and model minimization in Markov decision processes. Artificial Intelligence **147**(1-2), 163–223 (2003)
10. Givan, R., Leach, S., Dean, T.: Bounded-parameter Markov decision processes. Artificial Intelligence **122**(1-2), 71–109 (2000)
11. Hutter, M.: Extreme state aggregation beyond MDPs. In: International Conference on Algorithmic Learning Theory. pp. 185–199. Springer (2014)
12. Jong, N.K., Stone, P.: State abstraction discovery from irrelevant state variables. In: International Joint Conference on Artificial Intelligence (IJCAI). pp. 752–757 (2005)
13. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artificial intelligence **101**(1-2), 99–134 (1998)
14. Koller, D., Pfeffer, A.: Representations and solutions for game-theoretic problems. Artificial intelligence **94**(1-2), 167–215 (1997)
15. Li, L., Walsh, T.J., Littman, M.L.: Towards a unified theory of state abstraction for mdps. ISAIM **4**(5), 9 (2006)
16. Littman, M.L.: Memoryless policies: Theoretical limitations and practical results. In: International conference on simulation of adaptive behavior. p. 238 (1994)
17. McCallum, R.: Reinforcement learning with selective perception and hidden state (1997)
18. Oliehoek, F., Witwicki, S., Kaelbling, L.: A sufficient statistic for influence in structured multiagent environments. Journal of Artificial Intelligence Research **70**, 789–870 (2021)
19. Parker, D., Norman, G., Kwiatkowska, M.: Game-based abstraction for Markov decision processes. In: International Conference on the Quantitative Evaluation of Systems. pp. 157–166 (2006)
20. Petrik, M., Subramanian, D.: Raam: The benefits of robustness in approximating aggregated MDPs in reinforcement learning. In: Advances in Neural Information Processing Systems. pp. 1979–1987 (2014)
21. Puterman, M.L.: Markov decision processes: Discrete stochastic dynamic programming (1994)
22. Ravindran, B., Barto, A.G.: SMDP homomorphisms: An algebraic approach to abstraction in semi-Markov decision processes. In: International Joint Conference on Artificial Intelligence (IJCAI). pp. 1011–1016 (2003)
23. Rogers, D.F., Plante, R.D., Wong, R.T., Evans, J.R.: Aggregation and disaggregation techniques and methodology in optimization. Operations Research **39**(4), 553–582 (1991)
24. Saitta, L., Zucker, J.D.: Abstraction in artificial intelligence and complex systems, vol. 456. Springer (2013)
25. Singh, S.P., Yee, R.C.: An upper bound on the loss from approximate optimal-value functions. Machine Learning **16**(3), 227–233 (1994)
26. Sutton, R.S., Precup, D., Singh, S.: Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. Artificial intelligence **112**(1-2), 181–211 (1999)
27. Van Roy, B.: Performance loss bounds for approximate value iteration with state aggregation. Mathematics of Operations Research **31**(2), 234–244 (2006)
28. White III, C.C., Eldeib, H.K.: Markov decision processes with imprecise transition probabilities. Mathematics of Operations Research **42**(4), 739–749 (1994)
29. Wiesemann, W., Kuhn, D., Rustem, B.: Robust Markov decision processes. Mathematics of Operations Research **38**(1), 153–183 (2013)

30. Winterer, L., Junges, S., Wimmer, R., Jansen, N., Topcu, U., Katoen, J.P., Becker, B.: Motion planning under partial observability using game-based abstraction. In: IEEE Conference on Decision and Control (CDC). pp. 2201–2208 (2017)