Notes on State Abstractions

Nan Jiang

September 28, 2018

【强化学习理论 61】StatisticalRL 5

张楚珩
清华大学 交叉信息院博士在读

这是UIUC姜楠老师开设的CS598统计强化学习（理论）课程的第四讲的第一部分，主要讲的内容是state abstraction。

原文传送门

CS598 Note4
nanjiang.cs.illinois.edu

导言

前一讲里面我们看到，要学习到一个足够好的策略，需要的样本数是和状态空间大小 $|S|$ 有呈多项式关系的。当 $|S|$ 很大的时候，就需要更多的样本。现在考虑把类似的状态聚合在一起，这样能够有效地缩小状态空间，减少所需要的样本数目。这种方法我们称作state abstraction/compression/aggregation.

具体地，考虑一个把原状态空间（original/primitive/raw state space） S 映射到聚合状态空间（abstracted state space） $\phi(S)$ 的映射 ϕ 。即可能有原状态空间里面不同的两个状态 $s^{(1)}$ ， $s^{(2)}$ 被映射到同一个状态 $\phi(s^{(1)})=\phi(s^{(2)})$ 。

还是考虑前一讲里面讨论的certainty-equivalence设定，这样的做法能够有效减小所需要的样本，即减小estimation errors；但是由于可能存在把实际不同的状态聚合到同一个状态，这样会导致更大的approximation errors。

一、Exact abstractions

1.1. 定义

Definition 1 (Abstraction hierarchy [2]). Given MDP $M = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$ and state abstraction ϕ that operates on $\mathcal{S}$, define the following types of abstractions:

1. ϕ is $\pi^\star$-irrelevant if there exists an optimal policy $\pi^\star$, such that $\forall s^{(1)}, s^{(2)} \in \mathcal{S}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$, $\pi^\star_M(s^{(1)}) = \pi^\star_M(s^{(2)})$.
2. ϕ is $Q^\star$-irrelevant if $\forall s^{(1)}, s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$, $\forall a \in \mathcal{A}$, $Q^\star_M(s^{(1)}, a) = Q^\star_M(s^{(2)}, a)$.
3. ϕ is model-irrelevant if $\forall s^{(1)}, s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$, $\forall a \in \mathcal{A}, x' \in \phi(\mathcal{S})$,
$$R(s^{(1)}, a) = R(s^{(2)}, a), \quad \sum_{s' \in \phi^{-1}(x')} P(s'|s^{(1)}, a) = \sum_{s' \in \phi^{-1}(x')} P(s'|s^{(2)}, a). \tag{1}$$
Note that the condition on transition dynamics is essentially $P(x'|s^{(1)}, a) = P(x'|s^{(2)}, a)$. It will also be convenient to define a $|\phi(\mathcal{S})| \times |\mathcal{S}|$ matrix Φ, where
$$\Phi(x, s) = \mathbb{I}[\phi(s) = x].$$
So $\Phi P(s, a)$ collapses the transition distribution over $\mathcal{S}$ to a distribution over $\phi(\mathcal{S})$, and the condition on transition dynamics can be rewritten as: $\Phi P(s^{(1)}, a) = \Phi P(s^{(2)}, a)$.

这里定义了三种abstraction，三种定义由松到紧。

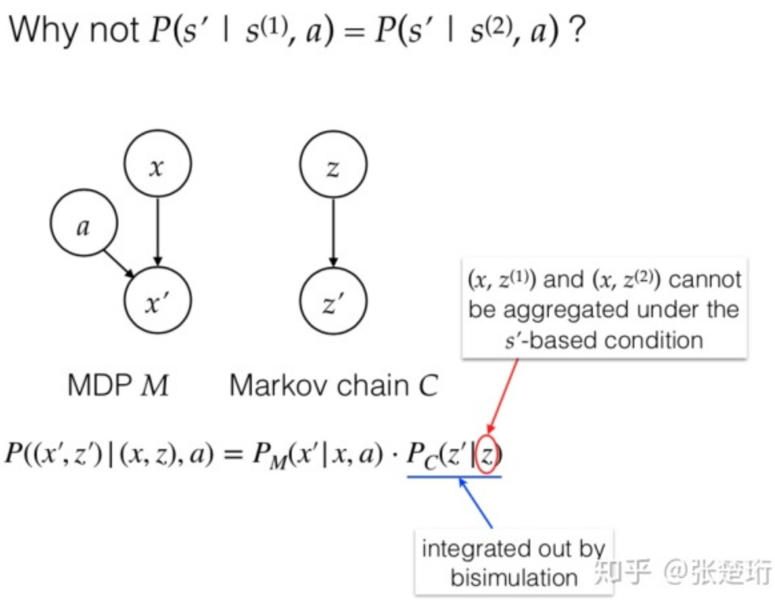
- 第一种是说只要抽象过后最优策略仍然可以被表示出来即可，极端的情况就是把最优策略选择同一种行动的状态分成一类即可，即就划分出来 $|\mathcal{A}|$ 个abstracted states。第一种抽象下，model-based和value-based方法可能不再能够适用了，只有一些直接去优化return的policy search方法还能使用。
- 第二种说的是抽象过后最优价值函数仍然可以被表示出来，极端的情况就是把最优价值函数数值相同的状态分为一类，不难看出，如果满足第二种的也能够满足第一种。这种抽象下，有一些tabular的算法还能够使用（比如Q-learning）。
- 第三种最为严格，它要求被聚合到同一个状态的不同状态不仅要具有相同的奖励，还要具有相同的dynamics。可以证明满足第三种的也能够满足第二种。这种情况下是完全等价，即任何在原问题上能用的算法，这里也能用。

最后注意到model-irrelevance对于dynamics相同的要求，只是要求转移到抽象之后状态空间 $\phi(\mathcal{S})$（这里简称x-space吧）的概率相同，并不是要求转移到原状态空间 $\mathcal{S}$（这里简称s-space吧）的概率相同，即

- Model-irrelevant: $\forall a \in A,$ $\quad R(s^{(1)}, a) = R(s^{(2)}, a)$
  (bisimulation) $\forall a \in A, x' \in \phi(S),$ $\quad \underbrace{P(x' \mid s^{(1)}, a)}_{\sum_{s' \in \phi^{-1}(x')} P(s' | s^{(1)}, a)} = P(x' \mid s^{(2)}, a)$

其实这样的要求更宽泛一些，具有更高的抽象能力。为了理解这件事情讲义举了个例子，考虑一个

MDP但是和另外一个无关的马可夫链耦合到一起，形成联合的状态空间 ，其中 x 是原来MDP中的状态空间， z 是这个无关的马可夫链的状态空间。显然，我们的目标是通过抽象能够完全剔除掉这个无关的马可夫链的。分析可知，只有这上面规定的这样的做法才能把这个无关的马可夫链剔除。



1.2. 相互关系

Theorem 1 (Theorem 2 of [2][1]). *Model-irrelevance implies $Q^\star$-irrelevance, which further implies $\pi^\star$-irrelevance.*

The proof is deferred to Section 2.

等会我们会证明这个定理的一个加强版本。

1.3. 性质

Uniqueness of coarsest bisimulation

这里讲的bisimulation就是前面讲的第三种model-irrelevant。结论是对于一个MDP存在一个唯一的abstraction使得抽象出来的状态数目最少。证明的思路就是如果存在任意的两个bisimulation ϕ_1 和 ϕ_2 那么我们就可以把它们合并成一个新的 ϕ_{12} ，合并的方法如下。只要任意两个状态 $s^{(1)}$ ， $s^{(2)}$ 被原来两个bisimulation中的某一个映射到同一个abstracted state，它们就会被 ϕ_{12} 映射到同一个

abstracted state。反复这样只要有不一样的bisimulation就合并，最后总能合并到一个最coarse（形成x-space中状态数目最少）并且唯一的bisimulation。

下面证明这件事情。要证明 ϕ_{12} 是一个bisimulation就需要证明对于任意两个满足 $\phi_{12}(s^{(1)})=\phi_{12}(s^{(2)})$ 的状态 $s^{(1)}$ 和 $s^{(2)}$ ，它们满足bisimulation的两个条件。关于reward的条件容易证明，即对于任意的 $\phi_{12}(s^{(1)})=\phi_{12}(s^{(2)})$ ，一定有 $r(s^{(1)},a)=r(s^{(2)},a)$ ，因为这个等号要么通过 ϕ_1 、要么通过 ϕ_2 的reward条件连结。下面证明的是关于dynamics的条件，即对于任意 有 $P(y'|s^{(1)},a)=P(y'|s^{(2)},a)$ 。

Proof. We prove by showing that for any bisimulations ϕ_1 and ϕ_2 of M, their *common coarsening* is also a bisimulation, denoted as ϕ_{12}. We define ϕ_{12} by giving its equivalence criterion: for any $s^{(1)}$ and $s^{(2)}$, $\phi_{12}(s^{(1)}) = \phi_{12}(s^{(1)})$ if and only if the two states are equivalent under either ϕ_1 or ϕ_2. Now we verify that ϕ_{12} is bisimulation. The reward condition is obviously satisfied, so it remains to check the transition condition.

Due to symmetry we consider any two states such that $\phi_1(s^{(1)}) = \phi_1(s^{(2)})$. For any $y' \in \phi_{12}(\mathcal{S})$,

$$\begin{aligned}
P(y'|s^{(1)},a) &= \sum_{s' \in \phi_{12}^{-1}(y')} P(s'|s^{(1)},a) \\
&= \sum_{x' \in \phi_1(\phi_{12}^{-1}(y'))} \sum_{s' \in \phi_1^{-1}(x') \bigcap \phi_{12}^{-1}(y')} P(s'|s^{(1)},a) \\
&= \sum_{x' \in \phi_1(\phi_{12}^{-1}(y'))} \sum_{s' \in \phi_1^{-1}(x')} P(s'|s^{(1)},a) \qquad (\phi_1^{-1}(x') \text{ is always entirely inside } \phi_{12}^{-1}(y')) \\
&= \sum_{x' \in \phi_1(\phi_{12}^{-1}(y'))} \sum_{s' \in \phi_1^{-1}(x')} P(s'|s^{(2)},a) \qquad (\phi_1(s^{(1)}) = \phi_1(s^{(2)})) \\
&= P(y'|s^{(2)},a).
\end{aligned}$$

On the second line, $\phi_1(\phi_{12}^{-1}(y'))$ is a set of abstract states in $\phi_1(\mathcal{S})$, formed by mapping each element of $\phi_{12}^{-1}(y')$ with ϕ_1 (recall that a set does not contain duplicate elements). The next step follows from the fact that any equivalence class in $\mathcal{S}$ induced by ϕ_{12} can always be partitioned into disjoint subsets, where each subset is a *complete* equivalence class under ϕ_1. Therefore, when we calculate $P(y'|s^{(1)},a)$, we can first sum over each smaller equivalence class under ϕ_1, and those probabilities will be the same for $s^{(1)}$ and $s^{(2)}$ as these two states are equivalent under ϕ_1 and the smaller equivalence classes are complete. As a consequence, the outer sum is also equal, and the result follows.

考虑 y' 是一个大桌面放在地上，地上互不重叠地铺满大饼（ ϕ_1 ）、同时也互不重叠地铺满披萨（ ϕ_2 ）。第一行到第二行说的是，对于桌面上的点求和可以看做，对于和桌面重合的每一张大饼，求和大饼和桌面重合的部分；第二行到第三行说的是由于不存在大饼压过桌面边缘的情况，因此对于桌面上的点求和可以看做，对于和桌面重合的每一张大饼求和；第三行到第四行说的是每个大饼内部，都可以把式子中的 $s^{(1)}$ 换成 $s^{(2)}$ ；最后一行说的是同理收回来，可以得到桌面内部都可以把式子中的 $s^{(1)}$ 换成 $s^{(2)}$ 。

二、Approximate abstractions

前面的定义都是各种准确的abstraction定义，它们很难找到也很难验证，实际中会用到近似的abstraction。我们这里先仿照前面定义三种近似的abtraction，并且给出这三种近似abstraction之间的关系，它们的关系是对于前面Theorem1的加强。

2.1. 定义

Definition 3 (Approximate abstractions). Given MDP $M = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$ and state abstraction ϕ that operates on $\mathcal{S}$, define the following types of abstractions:

1. ϕ is an $\epsilon_{\pi^\star}$-approximate $\pi^\star$-irrelevant abstraction, if there exists an abstract policy $\pi : \phi(\mathcal{S}) \to \mathcal{A}$, such that $\|V_M^\star - V_M^{[\pi]_M}\|_\infty \le \epsilon_{\pi^\star}$.
2. ϕ is an $\epsilon_{Q^\star}$-approximate $Q^\star$-irrelevant abstraction if there exists an abstract Q-value function $f : \phi(\mathcal{S}) \times \mathcal{A} \to \mathbb{R}$, such that $\|[f]_M - Q_M^\star\|_\infty \le \epsilon_{Q^\star}$.
3. ϕ is an (ϵ_R, ϵ_P)-approximate model-irrelevant abstraction if for any $s^{(1)}$ and $s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$, $\forall a \in \mathcal{A}$,

$$|R(s^{(1)}, a) - R(s^{(2)}, a)| \le \epsilon_R, \quad \left\|\Phi P(s^{(1)}, a) - \Phi P(s^{(2)}, a)\right\|_1 \le \epsilon_P. \tag{3}$$

Note that Definition 1 is recovered when all approximation errors are set to 0.

为了方便分析，定义了lifting，它表示了如何把x-space上定义的函数投影到s-space上。注意到x-space上的任意函数都可以投影到s-space上，形成piece-wise constant function。但是反过来不行，以为s-space上的函数不能保证每一个piece都是一样的数值。

Definition 2 (*lifting*). For any function f that operates on $\phi(\mathcal{S})$, let $[f]_M$ denote its *lifted* version, which is a function over $\mathcal{S}$, defined as $[f]_M(s) := f(\phi(s))$. Similarly we can also lift a state-action value function. Lifting a real-valued function f over states can also be expressed in vector form: $[f]_M = \Phi^\top f$.

2.2. 相互关系

The following theorem characterizes the relationship between the 3 types of approximate abstractions, with Theorem 1 as a direct corollary.

Theorem 2. *(1) If ϕ is an (ϵ_R, ϵ_P)-approximate model-irrelevant abstraction, then ϕ is also an approximate $Q^\star$-irrelevant abstraction with approximation error $\epsilon_{Q^\star} = \frac{\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{2(1-\gamma)^2}$.*
(2) If ϕ is an $\epsilon_{Q^\star}$-approximate $Q^\star$-irrelevant abstraction, then ϕ is also an approximate $\pi^\star$-irrelevant abstraction with approximation error $\epsilon_{\pi^\star} = 2\epsilon_{Q^\star}/(1-\gamma)$.

即第三种近似可以用来bound第二种近似，第二种近似可以用来bound第一种近似。我们可以连用上面两个结论用以使用第三种近似来bound第一种近似，但是我们后面可以看到，如果不经过第二步，可以得到一个更紧的bound。同时注意到，第一种近似表明了x-space上求解到的最优策略相比于直接在s-space上求解到的最优策略的性能损失上界，这是我们很关心的事情。

又卡了，下一讲将分别证明：

- 用第三种近似bound第二种近似
- 用第三种近似bound第一种近似（loss of abstract model）
- 用第二种近似bound 第一种近似（loss of abstract model）

发布于 2019-05-24

强化学习 (Reinforcement Learning)