Interaction History Tree Models

Jeroen van Maanen*

August 1, 2003

Abstract

This paper presents a model class that is well suited to represent Markov Decision Processes. Interaction history tree models can be easily adapted by a learning subject to model both observed behavior and expected behavior of its environment.

1 Introduction

Markov Decision Processes are powerful tools to model interactive behavior. The example that is presented in [Maa02] shows that they are powerful enough to represent environments with an interesting exploration/exploitation trade-off. This paper presents a model class that is well suited to represent Markov Decision processes: interaction history tree models, or tree models for short. These models are defined in such a way that they can be refined incrementally. As a result a learning subject that uses tree models can efficiently update its models of both observed behavior and expected behavior rather than having to search a huge space of candidate models after each atomic interaction. A measure for the size of tree models is presented that is compatible with the idea of small refinements: a small refinement leads to a small change of the size of a model. With the help of the Minimum Description Length principle [Ris89] this measure can be used by a learning subject that uses tree models to model expected behavior of the environment as well as to choose exploratory actions.

The framework to describe the interaction between the learning subject and its environment is borrowed from Hutter [Hut00]. The model class itself is based on ideas from [WRF95].

First we will define interaction history tree models. Then we will see how observed behavior can be modeled with tree models. After that we will investigate how expected behavior can be modeled. This includes the definition of a measure of the size of tree models. The next two sections discuss two technical issues: updating the structure of a tree model for observed behavior and efficient tracking of the MDL model for expected behavior. A summary concludes this paper.

* `Jeroen.van.Maanen@xs4all.nl`

2 Definition of tree models

We use the same terminology as in [MAA02]. To summarize: X and Y are prefix-free subsets of $\mathbb{B}^*$. (See [LV92] for an account of prefix-free sets.) The set X lists the input words that the learning subject can expect to receive from its environment. The set Y lists the output words that the learning subject can produce. Output words are also called actions. We also use a slightly adapted version of Hutter's notation [HUT00] for conditional probabilities: $p(y_{n-m}x_{n-m}\cdots y_n\underline{x_n})$ denotes the probability that the last input word will be x_n given that the preceding words were $y_{n-m}x_{n-m}\cdots y_n$ (Hutter requires that $m = n$).

Additionally we define for sets $A, B \subseteq \mathbb{B}^*$ the concatenation $AB := \{ab \mid a \in A, b \in B\}$. Furthermore we define $A^0 := \varepsilon$, $A^{n+1} := A^n A$, and $A^* := \cup_{n\in\mathbb{N}} A^n$. In particular we define the set of events $E := YX$, so instead of $\underline{yx} \in YX$ we will write $e \in E$. Note that E is prefix-free.

We are interested in the probability distribution over X given a state that is a function of the last words of the interaction between the learning subject and its environment up to and including the last action of the subject. Therefore we define the set of interaction histories, or histories for short, as $H := E^*Y$. For each $h \in H$ and $g \in E^*$ we call gh an extension of h. If $g \neq \varepsilon$ then we call gh a proper extension of h. If $g \in E$ we call gh an immediate extension of h. Note that we extend histories to the *left* to indicate that longer histories tell us something about the more distant *past*.

Our models will be based on trees of interaction histories. An interaction history tree is a collection T of histories that satisfies the following conditions:

- T must contain at least one element
- T must be finite
- For each history in T that can be written as a concatenation gh, where $g \in E^*$ and $h \in H$, the suffix h must also be included in T.

Formally:

$$\mathcal{T} := \{T \subseteq H \mid 1 \leq |T| < \omega \ \wedge \ \forall g \in E^*, h \in H[gh \in T \rightarrow h \in T]\} \tag{1}$$

Each interaction history tree model is a function that has an element of $\mathcal{T}$ as its domain. The range of a tree model is the set of finite distributions over X. A finite distribution over X is defined by a function from X to $\mathbb{N}$ that only assumes a non-zero value on a finite subset of its domain. If there is at least one domain element for which the defining function assumes a non-zero value, then the finite distribution is well-defined. The relation between the probabilities of the elements of X and the defining function f of a well-defined finite distribution is given by:

$$p_f(\underline{x}) := \frac{f(x)}{\sum\limits_{w\in X} f(w)} \tag{2}$$

This value is well defined because the sum in the denominator is finite and non-zero. There is exactly one finite distribution that is not well-defined. It corresponds to the defining function that is zero everywhere on its domain. We will use this trivial finite distribution, denoted $\emptyset$, as an intermediate value in Section 3.

We define the set of (defining functions of) finite distributions as:

$$\mathcal{D} := \left\{ f : X \to \mathbb{N} \;\middle|\; |\{x \in X \mid f(x) \neq 0\}| < \omega \right\} \tag{3}$$

Now we are ready to give a formal definition of the class of interaction tree models:

$$\mathcal{M} := \{\psi : T \to \mathcal{D} \mid T \in \mathcal{T}\} \tag{4}$$

A tree model $\psi \in \mathcal{M}$ is well-defined if and only if $\psi(h)$ is well defined for every $h \in \mathrm{Dom}(\psi)$.

We introduce the notation ψ_h for $\psi(h)$ because it enables us to write $\psi_h(x)$ to mean $f(x)$ where $f = \psi(h)$. We extend this notation to all $h \in H$ as follows. For each combination of $h \in H$ and $e \in E$ such that $h \in \mathrm{Dom}(\psi)$ but $eh \notin \mathrm{Dom}(\psi)$, for each $g \in E^*$, and for each $x \in X$ we define $\psi_{geh} := \psi_h$. For each $y \in Y$ such that $y \notin \mathrm{Dom}(\psi)$ and for each $g \in E^*$ we define $\psi_{gy} := f$ where $f \in \mathcal{D}$ is defined by:

$$f(x) := \sum_{y \in Y} \psi_y(x) \tag{5}$$

For a well-defined model $\psi \in \mathcal{M}$ and a history $h \in H$ we define:

$$p_\psi(h\underline{x}) := p_{\psi(h)}(\underline{x}) \tag{6}$$

Note that if $\psi \in \mathcal{M}$ is well-defined, then for each $h \in H$ the function $x \mapsto p_\psi(h\underline{x})$ is a probability distribution over X.

We will use these models in two ways: (1) to model observed behavior, and (2) to model expectations of future behavior.

3 Observed behavior

Each time when the learning subject receives a new input word $x \in X$ it revises its model of observed behavior. There are two aspects to modeling observed behavior: (1) updating the domain of the model T, and (2) updating the finite distributions $\psi(h)$. The first aspect is quite difficult to manage. We would like to gather data about longer and longer sequences of history, but we would also like to minimize the space complexity of the model. At least we have to include the last action y of the learning subject in the domain of the model before the finite distributions can be updated. It seems reasonable to place a limit on the growth of the tree and choose a rule-of-thumb to determine where to extend the tree when allowed. Possible strategies for updating the tree are discussed in Section 5. When we add a new element to the domain of the model we initially assign the finite distribution that is zero everywhere on its domain as its value. We have to make sure to update the values of the new elements to get a well-defined model again.

The second aspect is much simpler. Suppose that $\psi \in \mathcal{M}$ is our current model (after updating T) and that $d \in H$ is the data that describes the behavior between the learning system and its environment so far, up to but not including the new word $x \in X$, then our next model is defined by:

$$\psi'_h(x) := \psi_h(x) + 1 \tag{7}$$

for each $h \in \mathrm{Dom}(\psi)$ for which there is a $g \in E^*$ with $gh = d$; and:

$$\psi'_h(w) := \psi_h(w) \tag{8}$$

for each combination of $h \in \mathrm{Dom}(\psi)$ and $w \in X$ that is not covered by the previous definition.

If a learning subject starts from a model ψ with $\mathrm{Dom}(\psi) = \{y_0\}$ and it updates this model repeatedly from observed data extending the domain of the model every now and then, the resulting models will always be members of a subclass of $\mathcal{M}$:

$$\mathcal{O} := \left\{ \psi \in \mathcal{M} \;\middle|\; \forall h \in H, x \in X \left[\psi_h(x) \geq \sum_{e \in E} \psi_{eh}(x) \right] \right\} \tag{9}$$

4 Expectations of future behavior

To turn a model of observed behavior into expectations of future behavior we employ the Minimum Description Length principle (MDL) [Ris89]. To this end we define the description length $\mathbf{DL}(\psi)$ of a model $\psi \in \mathcal{M}$.

First we define how to describe $T = \mathrm{Dom}(\psi)$. Assume we have a standard enumeration $\{h_1, h_2, \ldots\} = H$ of histories, such that suffixes are listed earlier in the enumeration than histories that extend further in the past. For every element h in the enumeration that is included in T we describe the immediate extensions of h in T as follows. First we compute a threshold that is only dependent on the size of E: $a := \left\lceil \frac{|E|}{\log_2 |E|} - 1 \right\rceil$. Let n be the number of immediate extensions of h.

- If $a \leq n \leq |E| - a$ then we write a `0` and for each element of $e \in E$ a `1` if $eh \in T$ and a `0` otherwise. This is the bit vector that encodes the subset of E that defines the immediate extensions of h in T.
- If $n < a$ then we write a `1` followed by a `0`, then n using $\lceil \log a - 1 \rceil$ bits and then a run-length encoded version of the bit vector that encodes the subset of E that defines the immediate extensions of h in T using $n \lceil \log |E| \rceil$ bits.
- If $n > |E| - a$ then we write a `1` followed by a `1`, then $|E| - n$ using $\lceil \log a - 1 \rceil$ bits and then a run-length encoded version of the complement of the bit vector that encodes the subset of E that defines the immediate extensions of h in T using $(|E| - n) \lceil \log |E| \rceil$ bits.

If X is not finite, then we write a prefix-free encoded upper limit, in a standard enumeration of X, of the words in X that are mapped to a non-zero value by any ψ_h.

Then we write the description of ψ_h for every $h \in T$. If h can be written as eh', where $e \in E$ and $h' \in H$, and $\forall x \in X[\psi_h(x) = \psi_{h'}(x)]$ then we write a `0`. Otherwise write a `1` followed by the prefix-free encodings of the values $\psi_h(x)$ in the order of a standard enumeration of X (up to the uniform upper limit that was mentioned in the previous paragraph, if applicable).

Now let $\psi \in \mathcal{O}$ be a model of observed behavior and let $\varphi \in \mathcal{M}$ be a model of the expectations of behavior. The description length of the data in ψ given expectations φ is defined as:

$$\mathbf{DL}(\psi \mid \varphi) := \sum_{h \in \mathrm{Dom}(\psi)} \mathbf{DL}(\delta_h \mid \varphi_h) \tag{10}$$

where, for each $h \in \mathrm{Dom}(\psi)$, $\delta_h \in \mathcal{D}$ such that for all $x \in X$

$$\delta_h(x) := \psi_h(x) - \sum_{e \in E, eh \in \mathrm{Dom}(\psi)} \psi_{eh}(x) \tag{11}$$

and where

$$\mathbf{DL}(\delta_h \mid \varphi_h) := \sum_{x \in X} -\delta_h(x) \log_2 p_{\varphi_h}(\underline{x}) \tag{12}$$

The value $\mathbf{DL}(\psi \mid \varphi)$ can be infinity, if $\delta_h(x)$ is non-zero and $p_{\varphi_h}(\underline{x})$ is zero for one or more combinations of h and x. If $\delta_h(x)$ and $p_{\varphi_h}(\underline{x})$ are both zero for a particular combination of h and x, then $-\delta_h(x) \log_2 p_{\varphi_h}(\underline{x})$ is defined to be zero.

It can be seen that $\mathbf{DL}(\psi \mid \varphi)$ is the amount of bits necessary to describe all data d that was used to compile ψ using φ as a compression scheme.

Given observed behavior $\psi \in \mathcal{M}$, the best model of expectations of future behavior in the MDL sense is:

$$\check{\varphi} := \arg\min_{\varphi \in \mathcal{M}} \{\mathbf{DL}(\varphi) + \mathbf{DL}(\psi \mid \varphi)\} \tag{13}$$

The model $\check{\varphi}$ can be expected to perform well on a wide range of loss functions [GRÜ98].

5 Strategies for updating the tree

In Section 3 we saw that the learning subject needs a strategy for updating the tree that defines the structure of the model. One extreme strategy is just to make sure that the latest output word is included in the tree. The obvious problem with this strategy is that only direct cause/effect information is stored in the model. All information about the influence of past behavior is ignored. At the other extreme we have the strategy to always extend the tree such that the complete history of all interactions is an element of the tree. In this case no information is lost, but at a huge cost. The model of the observed behavior will be much larger than a linear description of that behavior. For the purposes of a learning subject it is not needed to keep all this data around. It follows directly from the MDL principle that the MDL model will be much smaller than a linear description of the behavior.

The author uses the following strategy in an experimental setup. For convenience, assume that X is finite. For $f \in \mathcal{D}$ we define

$$\mathbf{DL}(f) := \sum_{x \in X} \|f(x)\|_* \tag{14}$$

where $\|n\|_*$ is the length of the prefix-free encoding of n. We also reuse Equation 12 that defines $\mathbf{DL}(f \mid f')$ where $f, f' \in \mathcal{D}$.

Let $\psi \in \mathcal{M}$ be the current model of observed behavior, let $T = \mathrm{Dom}(\psi)$, let $d \in H$ describe the interaction with the environment so far, and let $g \in E^*$, $e \in E$, and $h \in H$ such that $d = geh$ with $h \in T$ and $eh \notin T$. Let

$$\check{f} := \operatorname*{arg\,min}_{f \in \mathcal{D}} \{\mathbf{DL}(f) + \mathbf{DL}(\psi_h \mid f)\} \tag{15}$$

If $\sum_{x \in X} \psi_h(x)\, \ell(x) > \vartheta \mathbf{DL}(\check{f})$ for a suitable threshold ϑ (the author uses $\vartheta = 3$) then eh is added to T.

6 Tracking the MDL model of expectations

[This section still needs work]

Let $\psi \in \mathcal{M}$ be the current model of observed behavior and let $\varphi \in \mathcal{M}$ be the current model of expected behavior. Let $S := \mathrm{Dom}(\psi)$ and let $T = \mathrm{Dom}(\varphi)$. Let $d \in H$ describe the interaction with the environment so far. Determine $g \in E^*$, $e \in E$, and $h \in T$ such that $d = geh$ and $eh \notin T$. First, we recalculate for each combination $g \in E^*$ and $h' \in H$ such that $h = gh'$ the optimal $\varphi_{h'}$. Then, for every combination of $g, g' \in E^*$ such that $d = g\,g'eh$ with $g'eh \in S$, we can calculate the effect of adding $g'eh$ to the domain of φ on $\mathbf{DL}(\varphi)$ and $\mathbf{DL}(\psi \mid \varphi)$ using just the finite distributions ψ_h, $\psi_{e'h}$ with $e' \in E$ and $\psi_{g''eh}$ with $g'' \in E^*$ such that there is a $g''' \in E^*$ with $g' = g'''\,g''$. If there is a combination for which the resulting effect is a decrease of $\mathbf{DL}(\varphi) + \mathbf{DL}(\psi \mid \varphi)$ we add the string $g'eh$ that encodes the smallest amount of events to the domain of φ. We repeat adding values to the domain of φ until there is no longer a combination g, g' for which adding $g'eh$ decreases the sum of the description length of the model of expected behavior and the description length of the data that describes the interactions so far given that model.

7 Summary and future work

Tree models as defined in Section 2 can be used efficiently by a learning subject to model observed behavior of its environment. It is unsatisfactory, however, that the last strategy that was discussed in section 5 needs an arbitrary threshold parameter. We would like to see a less arbitrary method to determine whether a finite distribution is mature enough to look for more information in longer histories that share that state.

Tree models can also be used efficiently by a learning subject to track the MDL model of expected behavior of the environment. The measurement of the description lengths is straightforward as described in Section 4. The model can be refined incrementally as described in Section 6.

The next step is to follow the reasoning of [Maa02] and define a method that produces output words for the learning subject that optimize the growth of $\mathbf{DL}(\check{\varphi})$, the description length of the MDL model for expected behavior of the environment.

References

[Grü98] Peter D. Grünwald. *The Minimum Description Length Principle and Reasoning under Uncertainty*. Institute for Logic, Language and Computation, 1998.

[Hut00] Marcus Hutter. A theory of universal artificial intelligence based on algorithmic complexity. Technical report, April 2000. 62 pages, `http://xxx.lanl.gov/abs/cs.AI/0004001`.

[LV92] Ming Li and Paul M.B. Vitányi. Inductive reasoning and Kolmogorov complexity. *Journal of Computer and System Sciences*, 44(2):343–384, April 1992.

[Maa02a] Jeroen van Maanen. *Towards a Formal Theory of Learning Systems*, 2002. `http://www.sollunae.net/zope/wiki/EnglishSummary`.

[Maa02] Jeroen van Maanen. Model growth. In *Proceedings of the Twelfth Belgian-Dutch Conference on Machine Learning*, Utrecht, 2002.

[Ris89] Jorma J. Rissanen. *Stochastic Complexity in Statistical Enquiry*. World Scientific, Singapore, 1989.

[WRF95] Marcelo J. Weinberger, Jorma J. Rissanen, and Meir Feder. A universal finite memory source. *IEEE Transactions on Information Theory*, IT-41(3):643–652, 1995.