Counterfactual Models

Jeroen van Maanen*

June 26, 2004

Abstract

This paper presents a way to turn interaction history tree models [Maa03] into counterfactual models that can be used to choose actions for an autonomously exploring learning system.

1 Introduction

.

2 Preparations

For this paper we need to expand the definitions as presented in [Maa03] a bit. Lets do a brief recap.

The set $\mathbb{B} := \{0, 1\}$ is the set of binary digits, or bits. $\mathbb{B}^*$ is the set of all finite bit sequences. We denote the empty sequence (a sequence of zero bits) as $\varepsilon \in \mathbb{B}^*$. For $a, b \in \mathbb{B}^*$ we denote the concatenation of a and b by ab. For sets $A, B \subseteq \mathbb{B}^*$ we define $AB := \{ab \mid a \in A, b \in B\}$. Furthermore we define $A^0 := \{\varepsilon\}$, $A^{n+1} := A^n A$, and $A^* := \bigcup_{n \in \mathbb{N}} A^n$. We denote the cardinality of a set A by $|A|$, so we can write 'A is finite' as $|A| < \omega$.

Let X and Y be 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, input words are also called responses. We define the set of events $E := YX$. Note that E is prefix-free.

We will need a suffix relation that respects words or events (remember that a prefix-free set is not necessarily suffix-free). We define for a set $A \subseteq \mathbb{B}^*$ that a bit sequence s is an A-suffix of a bit sequence a if and only if there is a $p \in A^*$ such that $ps = a$. Notation $a \sqsupseteq_A s$.

As in [Maa03] we define the set of interaction histories as $H := E^*Y$. For each $h \in H$ and $g \in E^*$ we call gh an extension of h. Note that $gh \sqsupseteq_E h$. If $g \neq \varepsilon$ then we call gh a proper

* Jeroen.van.Maanen@xs4all.nl

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 of the environment are based on trees of interaction histories. The class of interaction history trees is defined as:

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

In this paper there will be a number of occasions when we want to assign weights to the members of a set in such a way that the complete set of weights has a finite description. Sometimes these weight are to be interpreted as a probability distribution over the set, at other places we need the weights to be simple natural numbers that encode how often something has happened, and sometimes we need the weights to be rational numbers that represent utilities of available alternatives. To generalize these situations we define the class of weight assignments for a countable set A as

$$\mathcal{W}_A := \left\{ f : A \rightarrow \mathbb{Q} \;\middle|\; |\{a \in A \mid f(a) > 0\}| < \omega \wedge \forall a \in A[f(a) \geq 0] \right\} \tag{2}$$

A weight function $f \in \mathcal{W}_A$ is well-defined if there is at least one $a \in A$ such that $f(a) > 0$. If $f \in \mathcal{W}_A$ is well-defined, then it corresponds with a probability distribution over A:

$$p_f(a) := \frac{f(a)}{\sum\limits_{w \in A} f(w)} \tag{3}$$

The class of interaction tree models is defined as:

$$\mathcal{M} := \{\psi : T \rightarrow \mathcal{W}_X \mid T \in \mathcal{T}\} \tag{4}$$

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

3 Policies

In this paper the emphasis is on *policies* for the learning subject rather than *models* of the environment. However, there is an interesting duality between the two. We just need to reverse the roles of X and Y. Recall that we defined the set of events E as YX. We now define the set of reactions R as XY. We also define a slightly shifted analogs of H and $\mathcal{T}$: $H' := R^*X$ and

$$\mathcal{T}' := \{T \subseteq H' \mid 1 \leq |T| < \omega \wedge \forall g \in R^*, h \in H'[gh \in T \rightarrow h \in T]\} \tag{5}$$

We can now define the class of policies as:

$$\mathcal{P} := \{\pi : T \rightarrow \mathcal{W}_Y \mid T \in \mathcal{T}'\} \tag{6}$$

4 Quick gains

We would like to let a learning subject maximize the growth of its model of the environment. In this section we look at a local utility function that will be extended in the next section. Let $h_t \in E^*$ denote the complete history of the interactions between the learning subject and its environment so far. Let $y \in Y$ be a candidate action. Determine the longest suffix of $h_t y$ that is contained in T. Formally, let $h \in T$ such that

$$h_t y \sqsupseteq_E h \wedge \forall h' \in H[(h_t y \sqsupseteq_E h' \wedge h' \sqsupseteq_E h) \rightarrow h' \notin T] \tag{7}$$

We call h the *state* of the interaction given $h_t y$.

Let $\hat{x} \in X$ such that

$$\forall x \in X[(\varphi(h))(x) \leq (\varphi(h))(\hat{x})] \tag{8}$$

Now determine $t, n \in \mathbb{N}^+$ such that **if** the environment would respond with $\hat{x}$ the first t times that the learning subject would find itself in state h, **then** the first $t-1$ times the description length of ψ would not change and the last time it would increase by n bits.

For $h' \in E^*$ with $h'y = h$ we define a local utility function $u'_h : Y \rightarrow \mathbb{Q}$ by

$$u_{h'}(y) = n/t \tag{9}$$

TODO: solve dependency of h' on y

This can be easily turned into a policy by defining for $r \in R^*$ $x, x' \in X$ and $y \in Y$

$$(\pi(xyr))(x') := u_{yr}(x') \tag{10}$$

5 Definition of counterfactual models

In the previous section we saw that simple local optimization of actions will not result in a remarkable learning speed. If we want to do better we need to look further ahead than the single next event. The immediate question that arises is: how far do we have to look? It is tempting to formulate some ideal optimality criterion based on all infinite sequences of possible future interaction, but that is not going to help us build a working system. One of us has considered to define a utility function on the stationary distribution that arises from the current model of the environment and the variable policy and then optimize this function when varying the parameters of the policy. However, this turns out to be computationally challenging too.

So we tried another angle: lets define a data structure that can be maintained incrementally, like history tree models, and that enables the learning subject to do reasonably cheap forecasts of some utility function based on the current state of the interaction with the environment: the *counterfactual model*. The term 'counterfactual' is borrowed from *Gödel, Escher, Bach* [Hof79]. Counterfactual models are about sequences of events that are not facts, but they are not completely arbitrary either. Just like the definition of interaction

history models we begin by defining sets of sequences of words that will be used to define the domain of the models. Let $\varphi : T \to \mathcal{W}_X$ be a model of the environment. The counterfactual base corresponding to T is defined as:

$$C_T := \{g \in E^* \mid \exists h \in T[gh \in T \wedge \ell(h) < \ell(g)] \tag{11}$$

Where $\ell(a)$ is the number of words in a (*i.e.*, a member of E counts as two words).

We define the counterfactual range corresponding to T as $\rho_T := C_T \to \wp(T)$ such that

$$\rho_T(g) = \{h \in T \mid gh \in T \wedge \ell(h) < \ell(g)\} \tag{12}$$

Finally, we define the counterfactual model corresponding to φ as function $\gamma_\varphi : C_T \to \mathcal{W}_T$ that satisfies

$$\forall g \in C_T \, \forall h \in T[(\gamma(g))(h) > 0 \to h \in \rho_T(g)] \tag{13}$$

For $g \in C_T$ and $h \in \rho_T$ we define

$$(\gamma(g))(h) := w(g,h) := \prod_{\substack{k \in H, x \in X \\ kx \sqsubseteq h}} (\varphi(gk))(x) \tag{14}$$

Note that $w(g,h)$ is the probability that we can reach state h from g according to model φ.

6 Proposed policy

In this section we will propose a policy to be used by a learning subject that is based on the counterfactual model derived from the models of observed and expected data as described above. We do not make the claim that this model is optimal in any sense. We do claim that it achieves the following goals. It explores an unknown environment without the need for external evaluations of its actions. If the environment is generated by an element ψ^* of $\mathcal{M}$ (the so-called $\mathcal{M}$-closed case) and ψ^* meets some reasonable conditions, then, with probability one, this policy will cause ψ to converge to an model that is equivalent to ψ^*. For each event, the necessary updates on the data structures necessary to choose the actions according to this policy can be done in a time that is sub-linear in the number of events.

Let $\varphi : T \to \mathcal{W}_X$ be the current model of observed data so far. Let $\psi : T' \to \mathcal{W}_X$ for some $T' \subseteq T$ be the current model of expected data so far. Let $\gamma_\varphi : C_T \to \mathcal{T}$ be the counterfactual model corresponding to φ. We define:

$$T_R := \{xr \in R^*X \mid x \in X, r \in E^* \mid \exists y \in Y[ry \in T]\} \tag{15}$$

Note that $T_R \in \mathcal{T}'$ (See Equation 5.)

We propose to let a learning subject use a policy $\pi : T_r \to \mathcal{W}_Y$ that is defined as follows. Let $h \in T$ determine t and n as in Section 4. Define a utility function $u \in \mathcal{W}_T$ by

$$u(h) = w(\varepsilon, h)^{t-1} \cdot n \tag{16}$$

(See Equation 14 for the definition of w.) Let $r \in T_R$ and let

$$C_r := \{\langle g, h\rangle \in E^* \times H \mid r \sqsupseteq_E g \wedge gh \in T\} \tag{17}$$

Calculate for each action y the expected reachable utility

$$f(y) := \sum_{\substack{\langle g,h\rangle \in C_r \\ y \sqsubseteq h}} u(h) \cdot (\gamma_\varphi(g))(h) \tag{18}$$

Define $\pi(r)$ as p_f, the normalization of f (See Equation 3).

References

[Hof79] Douglas R. Hofstädter. *Gödel, Escher, Bach*. Basic Books, 1979.

[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`.

[Maa03] Jeroen van Maanen. *Interaction History Tree Models*, 2003. `http://www.lexau.org/pub/TreeModel-latest.pdf`.

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