Consistency of Feature Markov Processes

Peter Sunehag and Marcus Hutter

THE AUSTRALIAN NATIONAL UNIVERSITY

ALT 2010

Introduction

- Feature Markov (Decision) Processes is a history based approach to Sequence Prediction and Reinforcement Learning (RL). introduced by Hutter 2009
- We also consider Sequence Prediction with side information.
- In RL we have sequences of actions, observations and rewards. Actions and observations are side information for predicting future rewards.
- The actions are chosen with the aim of maximizing total long term reward in some sense.

States and Features

- If we have a suitable set of features that define states that are Markov, i.e. the sequence is transformed into a Markov sequence of states and the RL problem into a Markov Decision Process (MDP), then there exists sucessfull methods for solving the problems.
- The Feature Markov Decision Process framework (Hutter 2009) aims at learning such a feature map based on a globally (for a class of feature maps) defined cost criterium.

States and Features

- If we have a suitable set of features that define states that are Markov, i.e. the sequence is transformed into a Markov sequence of states and the RL problem into a Markov Decision Process (MDP), then there exists sucessfull methods for solving the problems.
- The Feature Markov Decision Process framework (Hutter 2009) aims at learning such a feature map based on a globally (for a class of feature maps) defined cost criterium.
- We present a consistency theory in this article/talk.
- Some first empirical results in (Mahmoud 2010)
- Empirical investigations are ongoing by Phuong Nguyen, Mayank Daswani.

- Consider the set of all infinite sequences $y_t, t = 1, 2, \ldots$ of elements from a finite alphabet Y. We equip the set with the σ-algebra that is generated by the cylinder sets $\Gamma_{y_{1:n}} = \{x_{1:\infty} | \, x_t = y_t, t = 1, \ldots, n\}$

- Consider the set of all infinite sequences $y_t, t = 1, 2, \ldots$ of elements from a finite alphabet Y. We equip the set with the σ-algebra that is generated by the cylinder sets $\Gamma_{y_{1:n}} = \{x_{1:\infty} | \, x_t = y_t, t = 1, \ldots, n\}$

Definition

A sequence $y_{1:\infty}$ defines a probability distribution on infinite sequences if the (relative) frequency of every finite substring of $y_{1:\infty}$ converges asymptotically. The probabilities of the cylinder sets are defined to equal those limits:

$$\Gamma_{z_{1:m}} := \lim_{n\to\infty} \#\{t \le n : y_{t+1:t+m} = z_{1:m}\}/n$$

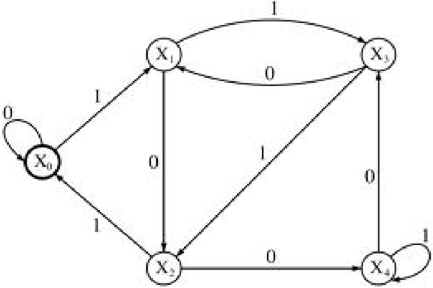
We call such sequences ergodic

Sources and Feature Maps

- Our primary class of sources and maps are defined from FSMs.

Definition

We are interested in Finite State Machines (FSM) with the property that s_{t-1} (internal state) and y_t (current element in a sequence) determine s_t. If we also have probabilities $Pr(y_{t+1} \mid s_t)$ we have defined a generating distribution.
Probabilistic-Deterministic Finite Automata (PDFA)
Used in (Mahmoud 2010)



Proposition

If we have FSM (of bounded memory) and the state transitions are ergodic it will a.s. generate an ergodic sequence

Sources and Feature Maps

Proposition

If we have FSM (of bounded memory) and the state transitions are ergodic it will a.s. generate an ergodic sequence

- Special case of maps: Trees. (Ineffecient representation of bounded memory FSM, many states)

Proposition

If we have FSM (of bounded memory) and the state transitions are ergodic it will a.s. generate an ergodic sequence

- Special case of maps: Trees. (Ineffecient representation of bounded memory FSM, many states)
- More generally, feature map $\Phi : Y^* \to S$ from histories (finite strings) to states. $\Phi(y_{1:t}) = s_t$. Feature Markov Process

Proposition

If we have FSM (of bounded memory) and the state transitions are ergodic it will a.s. generate an ergodic sequence

- Special case of maps: Trees. (Ineffecient representation of bounded memory FSM, many states)
- More generally, feature map $\Phi : Y^* \to S$ from histories (finite strings) to states. $\Phi(y_{1:t}) = s_t$. Feature Markov Process
- Task: Find Φ while observing the sequence y_t.

Proposition

If we have FSM (of bounded memory) and the state transitions are ergodic it will a.s. generate an ergodic sequence

- Special case of maps: Trees. (Ineffecient representation of bounded memory FSM, many states)
- More generally, feature map $\Phi : Y^* \to S$ from histories (finite strings) to states. $\Phi(y_{1:t}) = s_t$. Feature Markov Process
- Task: Find Φ while observing the sequence y_t.
- Given Φ, estimating probabilities $Pr(y|s)$ is simple (frequency, KT, Laplace)

Criteria for Selecting Feature Maps

- Given a feature map Φ and $y_{1:t}$, we also have a (fictitious) state sequence s_t.
- We can estimate $Pr(y_{t+1} = y|s_t = s)$ using frequency estimates $\frac{\#\{t|y_{t+1}=y, s_t=s\}}{\#\{t|s_t=s\}}$ and then we have defined a generative distribution for sequences

Criteria for Selecting Feature Maps

- Given a feature map Φ and $y_{1:t}$, we also have a (fictitious) state sequence s_t.
- We can estimate $Pr(y_{t+1} = y | s_t = s)$ using frequency estimates $\frac{\#\{t | y_{t+1}=y, s_t=s\}}{\#\{t | s_t=s\}}$ and then we have defined a generative distribution for sequences
- Alternative, estimate $Pr(s_{t+1} = s' | s_t = s)$ and $Pr(y_{t+1} = y | s_{t+1} = s)$. HMM parameters
- Let $p_{t,\Phi}$ be the distribution estimated at time t based on Φ.

Criteria for Selecting Feature Maps

- Given a feature map Φ and $y_{1:t}$, we also have a (fictitious) state sequence s_t.
- We can estimate $Pr(y_{t+1} = y | s_t = s)$ using frequency estimates $\frac{\#\{t | y_{t+1}=y, s_t=s\}}{\#\{t | s_t=s\}}$ and then we have defined a generative distribution for sequences
- Alternative, estimate $Pr(s_{t+1} = s' | s_t = s)$ and $Pr(y_{t+1} = y | s_{t+1} = s)$. HMM parameters
- Let $p_{t,\Phi}$ be the distribution estimated at time t based on Φ.

Definition

$$Cost_t(\Phi) = \log Pr(y_{1:t} | p_{t,\Phi}) + pen(m,t)$$

where m is the number of states and $pen(m,t)$ is positive, strictly increasing in m and $\lim_{t\to\infty} \frac{pen(m,t)}{t} = 0$.

- We select the Φ (in the class) that has the smallest cost.
- We assume finite classes in this talk.

Definition

If p is a distribution and the following limit is defined

$$K(p) = \lim_{t \to \infty} \frac{1}{t} \log Pr(y_{1:t} | p)$$

then we call the limit the entropy rate of p (for $y_{1:\infty}$).

- Well defined with almost sure uniform convergence on compact sets of HMM parameters for sequences generated by ergodic Hidden Markov Models.

Analysis Strategy

Definition

If p is a distribution and the following limit is defined

$$K(p) = \lim_{t\to\infty} \frac{1}{t} \log Pr(y_{1:t}|p)$$

then we call the limit the entropy rate of p (for $y_{1:\infty}$).

- Well defined with almost sure uniform convergence on compact sets of HMM parameters for sequences generated by ergodic Hidden Markov Models.

Strategy

If Φ is such that its parameter estimates (p_t defined by frequency estimates) converge (to p_∞) and if $\frac{1}{t}\log Pr(y_{1:t}|p)$ converges uniformly (to $K(p)$) then we also have convergence for $\frac{1}{t}\log Pr(y_{1:t}|p_t)$ and the convergence is to $\lim_{t\to\infty} \frac{1}{t}\log Pr(y_{1:t}|p_\infty) = K(p_\infty)$. Finally one argues that better log-likehood will defeat lower model penalty $pen(m,t)$ due to sublinear growth of the latter and linear of the first.

Classes of Sources and Maps

- We now fit our model and map class based on FSMs into the strategy.

Proposition

If we have a map based on bounded memory FSMs and an ergodic sequence we will have converging parameters $p_{t,\Phi}$ as $t \to \infty$.

- We now fit our model and map class based on FSMs into the strategy.

Proposition

If we have a map based on bounded memory FSMs and an ergodic sequence we will have converging parameters $p_{t,\Phi}$ as $t \to \infty$.

Proposition

Suppose that we have a source defined by a bounded memory FSM (with ergodic state transition parameters) then, we will almost surely generate an ergodic sequence which is such that such that $\frac{1}{t} \log Pr(y_{1:t}|p)$ converges uniformly on compact HMM parameter sets.

Conclusion for sequence prediction

- The main result for generic sequence prediction

Theorem

If we have a finite class of maps based on FSMs of bounded memory and the data is generated by one of the maps (together with parameters that makes the Markov sequence ergodic), then almost surely the entropy rates (for all the maps) exist and there is $T < \infty$ such that whenever $t \geq T$,

$$\operatorname*{argmin}_{\Phi} Cost_t(\Phi) \in \operatorname*{argmin}_{\Phi} K(p_\Phi)$$

where $p_\Phi = \lim_{t\to\infty} p_{t,\Phi}$

Side Information

- Sequence of (x_t, y_t) and we are only really interested in y_t
- For example y_t rewards and x_t observations and actions in RL

Side Information

- Sequence of (x_t, y_t) and we are only really interested in y_t
- For example y_t rewards and x_t observations and actions in RL
- Alternative 1: Model everything $((x_t, y_t) = z_t)$ as before

Side Information

- Sequence of (x_t, y_t) and we are only really interested in y_t
- For example y_t rewards and x_t observations and actions in RL
- Alternative 1: Model everything $((x_t, y_t) = z_t)$ as before
- Alternative 2: Model states s_t and y_t

$$OCost_t(\Phi) = logPr(s_{1:t}|p_{t,\Phi}) + \log Pr(y_{1:t}|s_{1:t}, p_{t,\Phi}) + pen$$

Side Information

- Sequence of (x_t, y_t) and we are only really interested in y_t
- For example y_t rewards and x_t observations and actions in RL
- Alternative 1: Model everything $((x_t, y_t) = z_t)$ as before
- Alternative 2: Model states s_t and y_t

$$OCost_t(\Phi) = logPr(s_{1:t}|p_{t,\Phi}) + \log Pr(y_{1:t}|s_{1:t}, p_{t,\Phi}) + pen$$

- If Φ is injective Alternative 1 and 2 are the same.
- Maps based on non-empty suffix trees are injective

Side Information

- Sequence of (x_t, y_t) and we are only really interested in y_t
- For example y_t rewards and x_t observations and actions in RL
- Alternative 1: Model everything ($(x_t, y_t) = z_t$) as before
- Alternative 2: Model states s_t and y_t

$$OCost_t(\Phi) = logPr(s_{1:t}|p_{t,\Phi}) + \log Pr(y_{1:t}|s_{1:t}, p_{t,\Phi}) + pen$$

- If Φ is injective Alternative 1 and 2 are the same.
- Maps based on non-empty suffix trees are injective
- Alternative 3 (ICost): Model only y_t and marginalize out s_t
- Leads to Hidden Markov Models since s_t depends on both $y_{1:t}$ and $x_{1:t}$.

$$ICost(\Phi) = logPr(y_{1:t}|p_{t,\Phi}) + pen$$

Side Information

- Sequence of (x_t, y_t) and we are only really interested in y_t
- For example y_t rewards and x_t observations and actions in RL
- Alternative 1: Model everything ($(x_t, y_t) = z_t$) as before
- Alternative 2: Model states s_t and y_t

$$OCost_t(\Phi) = logPr(s_{1:t}|p_{t,\Phi}) + \log Pr(y_{1:t}|s_{1:t}, p_{t,\Phi}) + pen$$

- If Φ is injective Alternative 1 and 2 are the same.
- Maps based on non-empty suffix trees are injective
- Alternative 3 (ICost): Model only y_t and marginalize out s_t
- Leads to Hidden Markov Models since s_t depends on both $y_{1:t}$ and $x_{1:t}$.

$$ICost(\Phi) = logPr(y_{1:t}|p_{t,\Phi}) + pen$$

- The same kind of conditions and analysis applied to *ICost* lead to the conclusion that we will after finite time only choose between maps that yield minimal entropy rate as models of y_t.

Side Information

- ICost ignores information that only gives you a bounded finite number of bits improvement, something whose influence expires after a finite horizon.
- Example, let $y_{t+k} = x_t$. Generate by letting x_t be Bernoulli$(1/2)$

Side Information

- ICost ignores information that only gives you a bounded finite number of bits improvement, something whose influence expires after a finite horizon.
- Example, let $y_{t+k} = x_t$. Generate by letting x_t be Bernoulli$(1/2)$
- Other alternatives:
- Finite Horizon cost focus on $\log Pr(y_{t:t+k}|s_t)$
- Ex: $(k = 0)$ $\log Pr(y_{1:t}|s_{1:t}) + pen$. Close to (Mahmoud 2010) who has a Bayesian version of this
- Can fail to capture the probabilistic transition struture

Side Information

- ICost ignores information that only gives you a bounded finite number of bits improvement, something whose influence expires after a finite horizon.
- Example, let $y_{t+k} = x_t$. Generate by letting x_t be Bernoulli$(1/2)$
- Other alternatives:
- Finite Horizon cost focus on $\log Pr(y_{t:t+k}|s_t)$
- Ex: $(k = 0)$ $\log Pr(y_{1:t}|s_{1:t}) + pen$. Close to (Mahmoud 2010) who has a Bayesian version of this
- Can fail to capture the probabilistic transition struture
- The various criteria with side information does not aim at recovering the full model.
- Must relate the criteria to the sense in which we want to be optimal, for example finite horizon RL.

Summary

- Consistency of Feature Markov Processes for sequence prediction
- A framework for analyzing criteria for selecting a feature map for sequence prediction with side information
- One can only hope for optimality with respect to the sense of optimality embodied by the criteria (entropy rate of what given what)
- Ongoing empirical studies where agents based Feature Markov Decision Processes tries to solve various problems will lead us forward