Reinforcement Learning: Part 3
Evolution

Chris Watkins

Department of Computer Science
Royal Holloway, University of London

July 27, 2015

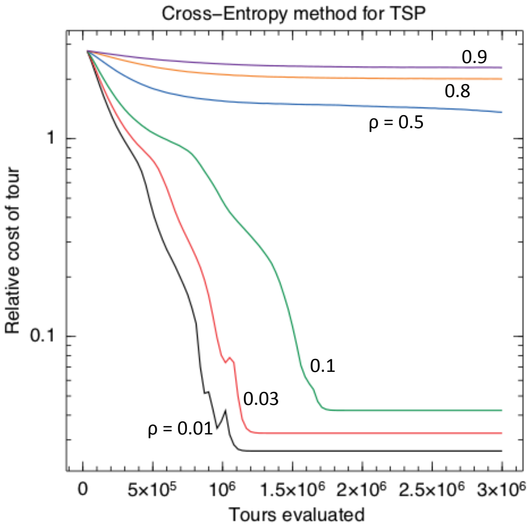
Cross-entropy method for TSP

Simple ‘genetic style’ methods can be very effective, such as the following algorithm for ‘learning’ a near-optimal tour in a TSP.

Initialise: Given: asymmetric distances between C cities
Initialise P, a matrix of transition probabilities between cities (to uniform distributions)

Repeat until convergence:

1. Generate a large number ($\sim 10C^2$) of random tours according to P, starting from city 1, under the constraint that no city may be visited twice.
2. Select the shortest 1% of these tours, and tabulate the transition counts for each city.
3. Adjust P towards the transition counts

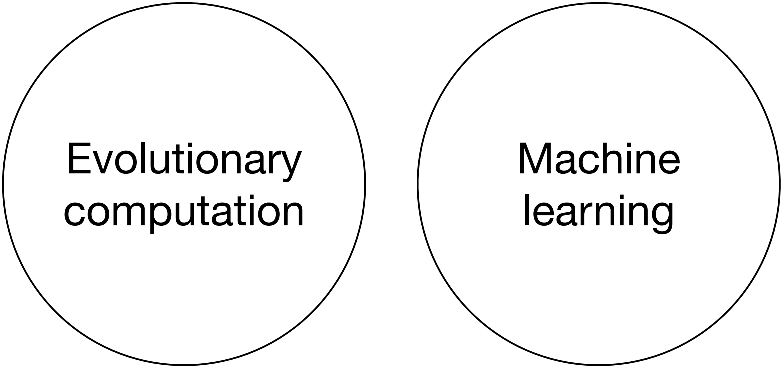


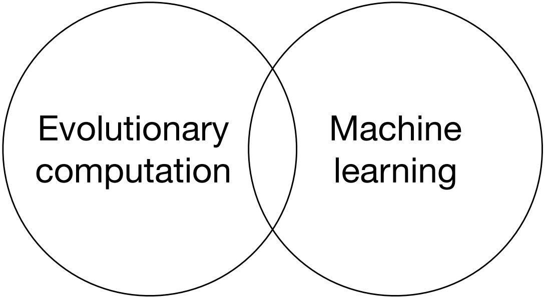
Weak selection

Fraction selected in each generation is ρ

Strong selection

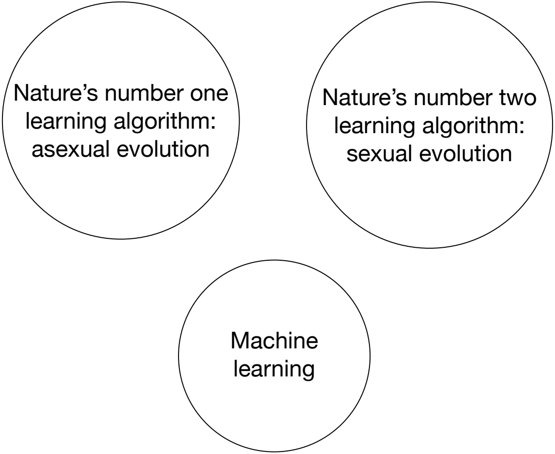
Venn diagram





- Estimation of distribution algorithms (EDA) (Pelikan and many others)
- Cross-entropy method (CEM) (Rubinstein, Kroese, Mannor 2000 - 2005)
- Evolution of reward functions (Niv, Singh)
- Simple model of evolution and learning (Hinton and Nowlan 1987)

Perhaps we should take evolution seriously?



Nature's number one learning algorithm: asexual evolution

Nature's number two learning algorithm: sexual evolution

Machine learning

Needed: an ‘enabling model’ of evolution

> *The best model of a cat is a cat.*
>
> — *Norbert Wiener*

Wiener is wryly pointing out that a model should be as simple as possible, and should contain only what is essential for the problem at hand.

We should discard as many biological details as possible, while keeping the computational essence.

If someone says “But biological detail X is important !”, then we can try putting it back in and see if it makes any difference.

Aim:

Can we construct a ‘machine learning’ style model of evolution, which includes *both* genetic evolution *and* individual learning?

Genetic mechanisms of sexual reproduction

At DNA level, the mechanisms of sexual reproduction are well known and tantalisingly simple!

Genetic mechanisms of sexual reproduction were fixed $> 10^9$ years ago, in single-celled protozoa;
since then, spontaneous evolution of advanced representational systems for multicellular anatomy and complex instincts.

Evolution is so robust you can't stop it...

Hypothesis: fair copying is the essence of sexual reproduction

Each child – each new member of the population – is a combination of genetic material copied (with errors) from other members of the population.

The copying is *fair* : all genetic material from all members of the population has an equal chance of being copied into the 'child'

A Markov chain of populations

Suppose we ‘breed’ individuals under constant conditions. Procedures for breeding, recombination, mutation and for selection are constant.

Then each population depends only on the previous population, so the sequence of populations is a Markov chain.

We will consider a sequence of populations where at each transition just one individual is removed, and just one is added. (In genetic language, this is the Moran model of overlapping populations)

Markov Chains: unique stationary distribution

A Markov chain is specified by its *transition probabilities*

$$T_{ji} = T(i \to j) = P(X_{t+1} = j | X_t = i)$$

There is a *unique*[1] stationary distribution π over the states such that

$$T\pi = \pi$$

In genetic language, the stationary distribution of the Markov chain of populations is the *mutation-selection equilibrium* .

[1] All the Markov chains we consider will be irreducible, aperiodic, and positive recurrent: such chains have a unique *stationary distribution* over states. (Any reasonable model of mutation allows some probability of any genome changing to any other.)

Markov Chains: detailed balance and reversibility

Reversibility

A Markov chain T is *reversible* iff $T\pi = \pi = T^T\pi$

Detailed balance condition

For any two states X, X',

$$\pi(X)T(X \to X') = \pi(X')T(X' \to X)$$

Kolmogorov cycle condition

For all $n > 2$, $X_1, X_2, \ldots, X_n$,

$$\begin{aligned} T(X_1 \to X_2)T(X_2 \to X_3)\cdots T(X_n \to X_1) = \\ T(X_n \to X_{n-1})T(X_{n-1} \to X_{n-2})\cdots T(X_1 \to X_n) \end{aligned}$$

Reversibility $\iff$ Detailed balance $\iff$ Kolmogorov cycle

Markov Chain Monte-Carlo (MCMC) in a nutshell

MCMC is a computational approach for sampling from a known, (typically complicated) probability distribution π.

Given a probability distribution π, *construct* a Markov chain T with stationary distribution π, and run the Markov chain to (eventually) get samples from π.

But how to construct a suitable T ?
Easiest way: define T that satisfies detailed balance for π.

Most common techniques for constructing reversible chains for a specified stationary distribution are:

- Metropolis-Hastings algorithm
- Gibbs sampling

Note that only a 'small' class of Markov chains are reversible: there is no advantage in reversible chains except that it may be easier to characterise the stationary distribution.

Evolution as MCMC

Idea: can we find a computational model of evolution that satisfies detailed balance (and so is reversible) ?

A nice stationary distribution would be:

$$\pi(g_1, \ldots, g_N) = p_B(g_1, \ldots, g_N) f(g_1) \cdots f(g_N)$$

where

- π is an (unnormalised) stationary distribution of a population of N genomes $g_1, \ldots, g_N$
- $p_B(g_1, \ldots, g_N)$ is the 'breeding distribution', the stationary probability of the population with no selection
- $f(g)$ is the 'fitness' of genome g. We require $f(g) > 0$.

Reversibility: a problem

When a reversible chain has reached its stationary distribution, an observer cannot tell if it is running forwards or backwards, because every cycle of state transitions has the same probability in both directions (as in physics).

Any model where a child must have two parents is not reversible

Given parents g_1, g_2, and their child c, the child is more similar to each parent, than the parents are similar to each other. e.g. for genomes of binary values,

$$\mathrm{HammingDistance}(g_1,c) \approx \frac{1}{2}\,\mathrm{HammingDistance}(g_1,g_2)$$

Hence in small populations with large genomes, we can identify parents and children, and so identify the direction of time

So natural evolution (and Holland's genetic algorithms) are not reversible Markov chains.

Key idea: Exchangeable Breeding

Generative probability model for a sequence of ‘genomes’ $g_1, g_2, \ldots$ Breeding only, no selection.

g_1 consists entirely of mutations

g_2 consists of elements copied from g_1, with some mutations

g_3 is elements copied from g_1 or g_2, with fewer mutations

$\ldots$

Each element of g_N is copied from one of $g_1, \ldots, g_{N-1}$, with few additional mutations...

Generating the sequence $g_1, g_2, \ldots$ does not seem biologically realistic – but *conditionally sampling* g_{N+1} given $g_1, \ldots, g_N$ can be much more plausible.

The sequence $g_1, g_2, \ldots$ needs to be exchangeable

Exchangeable sequences

Definition of exchangeability

A sequence of random variables $g_1, g_2, \dots$ is *infinitely exchangeable* iff for any N and any permutation σ of $\{1, \dots, N\}$,

$$p(g_1, \dots, g_N) = p(g_{\sigma_1}, \dots, g_{\sigma_N})$$

Which sequences are exchangeable?

de Finetti's theorem

$g_1, g_2, \dots$ is exchangeable off there is some prior distribution Θ, and a 'hidden parameter' $\theta \sim \Theta$, such that, given knowledge of θ $g_1, g_2, \dots$ are all i.i.d. samples, distributed according to θ

But de Finetti's theorem does not help us much at this point – we want an example of a plausible generative breeding distribution.

Example: Blackwell-MacQueen urn model

Pick a 'mutation distribution' H. A new genetic element can be sampled from H, independently of other mutations.

Pick a 'concentration parameter' $\alpha > 0$; this will determine the mutation rate

Generate a sequence $\theta_1, \theta_2, \ldots$ by:

- $\theta_1 \sim H$
- With prob. $\frac{1}{1+\alpha}$, $\theta_2 = \theta_1$, else w.p. $\frac{\alpha}{1+\alpha}$, $\theta_2 \sim H$
- $\ldots$
- With prob. $\frac{n-1}{n-1+\alpha}$, θ_n is randomly chosen (copied) from
  a uniform random choice of $\theta_1, \ldots, \theta_{n-1}$,
  else with prob. $\frac{\alpha}{n-1+\alpha}$, there is a new mutation $\theta_n \sim H$

This exchangeable sequence is well known in machine learning: it samples from the predictive distribution of a Dirichlet process $DP(H, \alpha)$

Example: Cartesian product of DPs

Each genome $g_i = (\theta_{i1}, \ldots, \theta_{iL})$

Generated with L independent Blackwell-MacQueen urn processes:

For each j, $1 \leq j \leq L$,

the sequence $\theta_{1j}, \theta_{2j}, \ldots, \theta_{Nj}$ is a sample from the jth urn process.

Sequence of genomes $g_1, g_2, \ldots, g_N$ is exchangeable, and is a sample from a Cartesian product of L Dirichlet processes.

This is the *simplest* plausible exchangeable breeding model for sexual reproduction.
Many elaborations possible...

Examples of possible breeding distributions

There is a rich and growing family of highly expressive non-parametric distributions, which achieve exchangeablility through random copying. Dirichlet process is the simplest.

Can construct elegant exchangeable distributions of networks, and exchangeable sequences of 'machines' with shared components...

Exchangeable breeding with tournaments (EBT)

Fitness function f, s.t. for all genomes g, $f(g) > 0$.
Current population of N genomes $g_1, \ldots, g_N$, with fitnesses $f_1, \ldots, f_N$.

Repeat forever:

1. Breed g_{N+1} by sampling from p_B conditional on the current population.

$$g_{N+1} \sim p_B(\cdot \mid g_1, \ldots, g_N)$$

2. $f_{N+1} \leftarrow f(g_{N+1})$. Add g_{N+1} into the population.
3. Select one genome i to remove from the population

$$Pr(\text{remove } g_i) = \frac{\frac{1}{f_i}}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}}$$

The genomes $\{g_1, \ldots, g_{N+1}\} \setminus \{g_i\}$ become the next population of N genomes.

EBT satisfies detailed balance

Let $G = \{g_1, g_2, \ldots, g_{N+1}\}$

To prove: $\pi(G_{\setminus N+1})T(G_{\setminus N+1} \to G_{\setminus i}) = \pi(G_{\setminus i})T(G_{\setminus i} \to G_{\setminus N+1})$

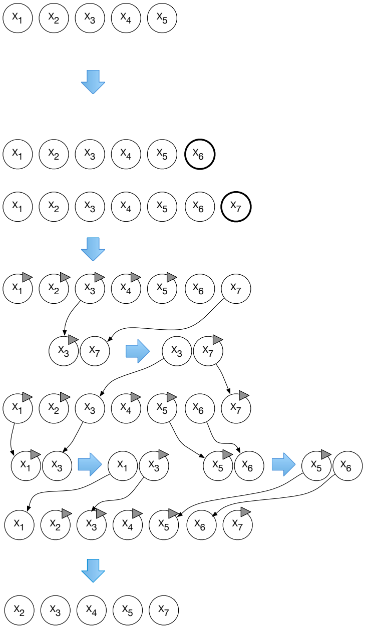
Proof:

$$\pi(G_{\setminus N+1})T(G_{\setminus N+1} \to G_{\setminus i}) =$$

$$p_B(G_{\setminus N+1})\ f_1 \cdots f_N \quad p_B(g_{N+1} \mid G_{\setminus N+1})\ \frac{\frac{1}{f_i}}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}}$$

$$= p_B(G) \quad \frac{f_1 \cdots f_{N+1}}{f_i f_{N+1}} \quad \frac{1}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}}$$

which is symmetric between g_i and g_{N+1}. *QED*.



Current generation

Breeding phase:

Any number of ‘children’ exchangeably sampled in sequence.

$$X_6 \sim p_B(\cdot \mid X_1, \ldots, X_5)$$

$$X_7 \sim p_B(\cdot \mid X_1, \ldots, X_6)$$

Selection phase:

Current generation given survival ‘tickets’ ▷

Any number of ‘tournaments’ between randomly chosen elements with and without tickets

$$Pr(X_i \text{ wins the ticket against } X_j) = \frac{f(X_i)}{f(X_i) + f(X_j)}$$

Tournaments are shown between X_3 and X_7, then between X_1, X_3, and between X_5, X_6

Next generation are the ticket owners at the end of the selection phase

Role of the population

EBT reduces to Metropolis-Hastings with a population size of 1. Why have a population larger than 1?

Two reasons:

1. Larger population concentrates the conditional breeding distribution $p_B(\cdot \mid g_1, \ldots, g_N)$.
2. Can scale log fitness as $1/N$, thus improving acceptance rate for newly generated individuals.

Stochastically generated environments

If environments are generated from an exchangeable distribution that is independent of the genome, then the environment plays a formally similar role to the genome, in that the stationary distribution for EBT with environments $v_1, \ldots, v_N$ exchangeably sampled from a distribution $P_V(\cdot)$ is:

$$\pi(g_1, \ldots, g_N, v_1, \ldots, v_N) \propto \\ p_B(g_1, \ldots, g_N) p_V(v_1, \ldots, v_N) f(g_1, v_1) \cdots f(g_N, v_N) \quad (1)$$

An intuitive interpretation of this is that each genome is ‘born’ into its own randomly generated circumstances, or ‘environment’; the current population will consist of individuals that are particularly suited to the environments into which they happen to have been born. Indeed, each ‘gene’ is part of the environment of other genes.

Individual learning

An abstract model of individual learning in different environments is that a genome g in an environment v

1. performs an experiment, observing a result x. The experiment that is performed, and the results obtained, depend on both g and v: let us write $x = \mathit{experiment}(g, v)$.
2. given g and x, the organism develops a post-learning phenotype $\mathit{learn}(g, x)$.
3. the fitness of g in v is then evaluated as

$$f(g, v) = f(\mathit{learn}(g, \mathit{experiment}(g, v)), v)$$

Individual learning

This model captures the notion that an organism:

- obtains experience of its environment;
- the experience depends both on the environment and on its own innate decision making, which depends on the interaction between its genome and the environment;
- the organism then uses this experience to develop its extended phenotype in some way;
- the organism's fitness depends on the interaction of its extended phenotype with its environment.

Minimal assumptions about 'learning'

We do not assume:

- learning is rational in any sense
- organism has subjective utilities or reinforcement