Research 1969 - 2024

Tom Westerdale

May 1, 2024

Main Result:

Strings of actions in learning systems and strings of genes in evolving systems change their probabilities according to the same simple formula, at least for a large class of learning adaptive plans.

1969 - 2006:

I believe it was in 1962 that I first heard John Holland outline the basic similarity of learning and evolution.[1] But crucial steps in the formalization of that similarity were missing, and after 1969 I directed my research toward filling in those steps.

The problem was that the probability change formulae seemed to differ depending on whether the adaptation was evolution or learning. We demonstrated that if the learning system uses a trace method to estimate values of actions, then there is no fundamental difference in the formulae.[3] The problem appeared when a temporal difference method was used.

2007 - 2013:

In 2007 I became free of any requirement to lecture or publish, and I was able to concentrate full time on the problem. I thought that the heart of the problem was in the biases inherent in temporal difference methods, biases that disappear when the Markov property holds in the system. I was wrong. I only gradually came to realize that I did not even understand what was going on in a simple adaptive Markov chain. The adaptive plan gives us the time derivatives of the *conditional* transition probabilities. But I could not produce a usable formula for the time derivative of an *unconditional* transition frequency in even the simplest adaptive plan. All approaches to the problem failed.

2014 - 2015:

Then in 2014 I suddenly realized that if I treat time symmetrically it does not double the problem, it solves it. Over the next six months I worked out the details and found that the required formulae were surprisingly simple. Further work extended the analysis to situations where the Markov property does not hold and biases occur.

2016 - 2024:

Ever since the seminal paper by Sutton and Barto [2] I've realized that once I solved my problem, their reinforcement learning approach should extend the analysis to learning systems we really care about. So my work since 2016 has been a combination of extending the analysis and simplifying its presentation. It took two years to prove a needed extension to a convergence proof reported by Sutton and Barto, and about as long to prove bucket brigade convergence in probability. These convergence proofs allowed a proper generalization. It turns out that generalizing also shortens, producing for the first time a write up of reasonable length.[4][5] (2024)

References

[1] J. H. Holland. *Adaptation in Natural and Artificial Systems.* Univ. of Michigan Press, Ann Arbor, 1975.

[2] R. S. Sutton and A. G. Barto. Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88(2):135–170, 1981.

[3] T. H. Westerdale. A Reward Scheme for Production Systems with Overlapping Conflict Sets. *IEEE Trans. Syst., Man, Cybern.*, SMC-16(3):369–383, 1986.

[4] Tom Westerdale. Learning Resembles Evolution – the Markov Case. pages 1–12, 2024. unpublished. URL `https://www.dcs.bbk.ac.uk/~tom/briefsymmetric.pdf`.

[5] Tom Westerdale. Learning resembles Evolution Even when using Temporal Diffference. pages 1–15, 2024. unpublished. URL `https://www.dcs.bbk.ac.uk/~tom/briefgeneral.pdf`.