Learning Evaluation Functions for Global Optimization

Justin Andrew Boyan

August 1, 1998

CMU-CS-98-152

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Thesis committee:
Andrew W. Moore (co-chair)
Scott E. Fahlman (co-chair)
Tom M. Mitchell
Thomas G. Dietterich, Oregon State University

Support for this work has come from the National Defense Science and Engineering Graduate fellowship; National Science Foundation grant IRI-9214873; the Engineering Design Research Center (EDRC), an NSF Engineering Research Center; the Pennsylvania Space Grant fellowship; and the NASA Graduate Student Researchers Program fellowship. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of ARPA, NSF, NASA, or the U.S. government.

Keywords: Machine learning, combinatorial optimization, reinforcement learning, temporal difference learning, evaluation functions, local search, heuristic search, simulated annealing, value function approximation, neuro-dynamic programming, Boolean satisfiability, radiotherapy treatment planning, channel routing, bin-packing, Bayes network learning, production scheduling

Abstract

In complex sequential decision problems such as scheduling factory production, planning medical treatments, and playing backgammon, optimal decision policies are in general unknown, and it is often difficult, even for human domain experts, to manually encode good decision policies in software. The reinforcement-learning methodology of "value function approximation" (VFA) offers an alternative: systems can learn effective decision policies autonomously, simply by simulating the task and keeping statistics on which decisions lead to good ultimate performance and which do not. This thesis advances the state of the art in VFA in two ways.

First, it presents three new VFA algorithms, which apply to three different restricted classes of sequential decision problems: Grow-Support for deterministic problems, ROUT for acyclic stochastic problems, and Least-Squares TD(λ) for fixed-policy prediction problems. Each is designed to gain robustness and efficiency over current approaches by exploiting the restricted problem structure to which it applies.

Second, it introduces STAGE, a new search algorithm for general combinatorial optimization tasks. STAGE learns a problem-specific heuristic evaluation function as it searches. The heuristic is trained by supervised linear regression or Least-Squares TD(λ) to predict, from features of states along the search trajectory, how well a fast local search method such as hillclimbing will perform starting from each state. Search proceeds by alternating between two stages: performing the fast search to gather new training data, and following the learned heuristic to identify a promising new start state.

STAGE has produced good results (in some cases, the best results known) on a variety of combinatorial optimization domains, including VLSI channel routing, Bayes net structure-finding, bin-packing, Boolean satisfiability, radiotherapy treatment planning, and geographic cartogram design. This thesis describes the results in detail, analyzes the reasons for and conditions of STAGE's success, and places STAGE in the context of four decades of research in local search and evaluation function learning. It provides strong evidence that reinforcement learning methods can be efficient and effective on large-scale decision problems.

Dedication

To my grandfather,
Ara Boyan, 1906–93

Acknowledgments

This dissertation represents the culmination of my six years of graduate study at CMU and of my formal education as a whole. Many, many people helped me achieve this goal, and I would like to thank a few of them in writing. Readers who disdain sentimentality may wish to turn the page now!

First, I thank my primary advisor, Andrew Moore. Since his arrival at CMU one year after mine, Andrew has been a mentor, role model, collaborator, and friend. No one else I know has such a clear grasp of what really matters in machine learning and in computer science as a whole. His long-term vision has helped keep my research focused, and his technical brilliance has helped keep it sound. Moreover, his patience, ability to listen, and innate personal generosity will always inspire me. Unlike the stereotypical professor who has his graduate students do the work and then helps himself to the credit, Andrew has a way of listening to your half-baked brainstorming, reformulating it into a sensible idea, and then convincing you the idea was yours all along. I feel very lucky indeed to have been his first Ph.D. student!

My other three committee members have also been extremely helpful. Scott Fahlman, my co-advisor, helped teach me how to cut to the heart of an idea and how to present it clearly. When my progress was slow, Scott always encouraged me with the right blend of carrots and sticks. Tom Mitchell enthusiastically supported my ideas while at the same time asking good, tough questions. Finally, Tom Dietterich has been a wonderful external committee member—from the early stages, when his doubts about my approach served as a great motivating challenge, until the very end, when his comments on this document were especially thorough and helpful.

Three friends at CMU—Michael Littman, Marc Ringuette, and Jeff Schneider—also contributed immeasurably to my graduate education by being my regular partners for brainstorming and arguing about research ideas. I credit Michael with introducing me to the culture of computer science research, and also thank him for his extensive comments on a thesis draft.

Only in the past year or two have I come to appreciate how amazing a place CMU is to do research. I would like to thank all my friends in the Reinforcement Learning group here, especially Leemon Baird, Shumeet Baluja, Rich Caruana, Scott Davies, Frank Dellaert, Kan Deng, Geoff Gordon, Thorsten Joachims, Andrew McCallum, Peter Stone, Astro Teller, and Belinda Thom. Special thanks also go to my officemates over the years, including Lin Chase, Fabio Cozman, John Hancock, and Jennie Kay.

The CMU faculty, too, could not have been more supportive. Besides my committee members, the following professors all made time to discuss my thesis and assist my research: Avrim Blum, Jon Cagan, Paul Heckbert, Jack Mostow, Steven Rudich, Rob Rutenbar, Sebastian Thrun, Mike Trick, and Manuela Veloso. I would also like to thank the entire Robotics Institute for allowing me access to over one hundred of their workstations for my 40,000+ computational experiments!

The third major part of what makes the CMU environment so amazing is the excellence of the department administration. I would particularly like to thank Sharon Burks, Catherine Copetas, Jean Harpley, and Karen Olack, who made every piece of paperwork a pleasure (well, nearly), and who manage to run a big department with a personal touch.

Outside CMU, I would like to thank the greater reinforcement-learning and machine-learning communities—especially Chris Atkeson, Andy Barto, Wray Buntine, Tom Dietterich, Leslie Kaelbling, Pat Langley, Sridhar Mahadevan, Satinder Singh, Rich Sutton, and Gerry Tesauro. These are leaders of the field, yet they are all wonderfully down-to-earth and treat even greenhorns like myself as colleagues.

Finally, I thank the following institutions that provided financial support for my graduate studies: the Winston Churchill Foundation of the United States (1991–92); the National Defense Science and Engineering Graduate fellowship (1992–95); National Science Foundation grant IRI-9214873 (1992–94); the Engineering Design Research Center (EDRC), an NSF Engineering Research Center (1995–96); the Pennsylvania Space Grant fellowship (1995–96); and the NASA Graduate Student Researchers Program (GSRP) fellowship (1996–98). I thank Steve Chien at JPL for sponsoring my GSRP fellowship.

* * *

On a more personal level, I would like to thank my closest friends during graduate school, who got me through all manner of trials and tribulations: Marc Ringuette, Michael Littman, Kelly Amienne, Zoran Popović, Lisa Falenski, and Jeff Schneider. Peter Stone deserves special thanks for making possible my trips to Brazil and Japan in 1997. I also thank Erik Winfree and all my CTY friends in the Mattababy Group for their constant friendship.

Second, I would like to thank the teachers and professors who have guided and inspired me throughout my education, including Barbara Jewett, Paul Sally, Stuart Kurtz, John MacAloon, and Charles Martin.

Finally, I would like to thank the two best teachers I know: my parents, Steve and Kitty Boyan. They taught me the values of education and thoughtfulness, of moderation and balance, and of love for all people and life. To thank them properly—for all the opportunities and freedoms they provided for me, and for all their wisdom and love—would fill the pages of another book as long as this one.

Table of Contents