Cornell University
Center for Advanced Computing

Red Cloud with MATLAB case study

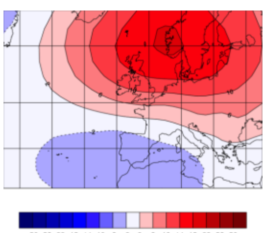
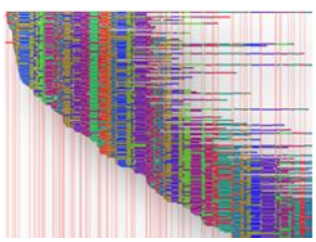
CAC
we enable your success

Peter Frazier – Bayesian Optimization via Simulation with Correlated Sampling and Correlated Prior Beliefs

How do you build algorithms that find the optimum in a system that is represented by a simulation rather than equations?

Finding the Answer

Peter Frazier of Cornell's School of Operations Research and Information Engineering is interested in finding the most efficient way to collect information and make decisions. He applies mathematical methods from Bayesian statistics and sequential decision-making under uncertainty to a broad array of theoretical and practical applications which he calls "optimal learning" or "information collection" problems. To solve these problems, decisions must be made about what type of information and how much information to collect and at what cost. With the right information, learning occurs quickly, resulting in more efficient solutions or behaviors. Some examples of information collection problems are calibrating a model of climate change, accurately reconstructing a collection of whole genomes from fragmented genetic data, or setting the right schedule or staffing level for a large organization.


For example, in finding the best staffing level for a large hospital, when the value of one alternative is learned (the quality of care at a particular staffing level), this suggests that similar alternatives (staffing levels with similar number of staff in each unit) are likely to have similar values. A correlated prior distribution allows this information to be used by an optimization algorithm and thereby improves efficiency.

"I study these questions within a mathematical framework that uses decision theory to pose the problem of collecting information in an optimal way as a Markov decision process (MDP)," explains Frazier. "I then study the solutions of these MDPs using dynamic programming."

This set of mathematical problems and tools is most frequently called sequential design of experiments, but is also related to reinforcement learning, decision theory, active learning, budgeted learning, and multi-armed bandits (a statistical approach that takes its name from slot machines). Frazier often refers to the design of sequential experiments as optimal learning because he is asking how we can learn optimally. Most frequently, he and his colleagues work on problems in simulation optimization and the global optimization of expensive functions.

Their goal is to find new algorithms for calibrating and optimizing simulations of large scale systems that occur in nature, engineering, and individual and collective human behavior. They have worked on problems in computer vision, logistics, information retrieval, neuroscience, inventory control, and drug discovery.

Improved Research

Research Metrics

- Reduce Time to Insight: Accelerate the design and testing of algorithms by using a compute/analysis resource that "bursts" on demand.
- Scale to an Optimal Execution Environment: Run Parallel Computing Toolbox codes on an optimal number of cores in the cloud (using MATLAB Distributed Computing Server) rather than procure dedicated hardware/software for only periodic use.

Research Challenge

Bayesian analyses require sophisticated computations, including the use of simulation methods. Typically, simulation techniques are employed when it is difficult or impossible to represent the system with equations that can then be solved to predict and achieve the best outcomes. Among other sources of difficulty is inherent randomness in the system, which leads to outcomes expressed in probabilistic terms. While simulation techniques can take this randomness into account, many simulation runs are required just to predict the outcome of each possible decision, much less to find the best decision. Hence Frazier and his colleagues face a particular challenge in building algorithms that find the optimum in a system that is represented by a simulation rather than equations.

Solution

Frazier and his colleagues have developed the first set of knowledge-gradient methods to take advantage of both correlated beliefs and correlated sampling. Frazier sees simulation optimization as a decision-theory problem in sequential experimental design, with the objective of finding solutions with minimal error and computational effort. "Red Cloud with MATLAB is enabling us to test our theories faster," said Frazier. "Proofs of these conjectures would constitute an important theoretical advance, and would lead to new high-performance algorithms." The research team looks forward to taking the next steps toward the development of efficient value of information-based optimization via simulation for large-scale problems that take advantage of correlated sampling.

The Clients

Peter Frazier, Operations Research and Information Engineering, Cornell University

- Air Force Office of Scientific Research Outstanding Young Investigator award
- NSF EAGER award; NSF Review Panel – Service Enterprise Systems
- Associate Editor, *Operations Research;* reviewer for more than ten journals

Jing Xie, Ph.D. student, Operations Research and Information Engineering

- Chinese National Honor Program for Fundamental Sciences

- Finalist in the INFORMS Junior Faculty Interest Group Paper Competition, 2011.

The Collaborative Relationship

"The on demand convenience of Red Cloud with MATLAB is ideal for our work in sequential decision-making and optimal methods for collecting information. It provides the software we need when we need it, enabling us to develop simulation optimization and feasibility determination algorithms faster and more efficiently. We are not burdened with procuring and maintaining our own computational resources and can share the cloud resource with other researchers, providing economies of scale for all. We look forward to continuing to improve our use of Bayesian statistics and dynamic programming using Red Cloud with MATLAB."

Peter Frazier
Assistant Professor
School of Operational Research and Information Engineering

Acknowledgement

This research is supported by Air Force Office of Scientific Research grant AFOSR FA9950-11-1-0083: "Decision-theoretic methods in simulation optimization."