DOI:10.1145/1735223.1735246

Technical Perspective
Learning to Act in Uncertain Environments

By Peter L. Bartlett

THE PROBLEM OF decision making in an uncertain environment arises in many diverse contexts: deciding whether to keep a hard drive spinning in a netbook; choosing which advertisement to post to a Web site visitor; choosing how many newspapers to order so as to maximize profits; or choosing a route to recommend to a driver given limited and possibly out-of-date information about traffic conditions. All are sequential decision problems, since earlier decisions affect subsequent performance; all require adaptive approaches, since they involve significant uncertainty. The key issue in effectively solving problems like these is known as the *exploration/exploitation trade-off*: If I am at a crossroads, when should I go in the most advantageous direction among those that I have already explored, and when should I strike out in a new direction, in the hopes I will discover something better?

The following paper by Ganchev, Kearns, Nevmyvaka, and Vaughan considers a sequential decision problem from the financial domain: how to allocate stock orders across a variety of marketplaces, each with an unknown volume available, so as to maximize the number of orders that are filled. These marketplaces are known as *dark pools* because they allow traders to keep their transactions hidden. The popularity of these dark pools has grown enormously, as traders making large transactions hope to reduce their market impact, that is, the tendency for the price to move in an unfavorable direction. Because transactions are hidden, the characteristics of the various dark pools are uncertain, and can only be discovered by active exploration.

The broad approach followed by the authors is based on an intuitive heuristic that is reminiscent of a title you might encounter in the self-help section of a bookstore: "optimism in the face of uncertainty." The idea is to treat uncertain outcomes as optimistically as the data allows: pick the alternative that, in the best possible world, is consistent with our experiences so far, and leads to the best outcome. One alternative might be chosen either because it is clearly superior to all others or because there is not enough data to rule out that possibility. If it turns out this alternative is a poor choice, at least it leads to a reduction in uncertainty, so that in the future it can be confidently avoided. This approach naturally leads to a balance between the desire to exploit information that has already been gathered, and the need to explore uncertain alternatives.

The authors illustrate how the optimism heuristic can be successfully applied to the dark pools problem. One striking feature of this result is that their approach is successful despite the fact that the number of distinct states in this problem is enormous: it is exponential in the number of venues. They exploit the favorable structure of the problem, and in particular the way it decomposes neatly across the distinct venues. Their approach involves a modification of a conventional nonparametric statistical estimator for censored data—the Kaplan-Meier estimator, which is used in survival analysis. They use one of these estimators for each venue in order to decide on the allocation. Their modification to the Kaplan-Meier estimator incorporates the optimism heuristic by encouraging exploration near the boundary of the region of state space that has already been adequately explored. The key result is that this approach can successfully adapt to unknown markets: under the assumption that the volumes available in the various venues are independent random variables, they prove that their strategy rapidly performs almost as well as the optimal allocation.

The following paper considers a sequential decision problem from the financial domain: how to allocate stock orders across a variety of marketplaces, each with an unknown volume available, so as to maximize the number of orders that are filled.

The heuristic of optimism in the face of uncertainty is known to perform well in other sequential decision problems. For instance, in small Markov decision problems, it leads to learning algorithms that have small regret: the amount of utility gathered per time step rapidly approaches the best possible value. The key challenge in this area is the development of methods that can deal with very large problems: in a wide range of applications, the state space is enormous; the dark pools problem is typical in this regard. For good performance in a case like this, it seems essential that the problem exhibits some kind of helpful structure. The work detailed in the following paper shows how the generic approach of optimism in the face of uncertainty can be applied to exploit the structure of a very large sequential decision problem to give an adaptive strategy that allows automatic performance optimization. This is an approach that will certainly see wider application.

Peter L. Bartlett is a professor in the Computer Science Division and Department of Statistics at the University of California, Berkeley.