Making better decisions when outcomes are uncertain

Posted by ashley | 04 May, 2017

Markov decision processes are mathematical models used to determine the best courses of action when both current circumstances and future consequences are uncertain. Theyve had a huge range of applications in natural-resource management, manufacturing, operations management, robot control, finance, epidemiology, scientific-experiment design, and tennis strategy, just to name a few.

But analyses involving Markov decision processes (MDPs) usually make some simplifying assumptions. In an MDP, a given decision doesnt always yield a predictable result; it could yield a range of possible results. And each of those results has a different value, meaning the chance that it will lead, ultimately, to a desirable outcome.

Characterizing the value of given decision requires collection of empirical data, which can be prohibitively time consuming, so analysts usually just make educated guesses. That means, however, that the MDP analysis doesnt guarantee the best decision in all cases.

In theProceedings of the Conference on Neural Information Processing Systems, published last month, researchers from MIT and Duke University took a step toward putting MDP analysis on more secure footing. They show that, by adopting a simple trick long known in statistics but little applied in machine learning, its possible to accurately characterize the value of a given decision while collecting much less empirical data than had previously seemed necessary.

In their paper, the researchers described a simple example in which the standard approach to characterizing probabilities would require the same decision to be performed almost 4 million times in order to yield a reliable value estimate.

With the researchers approach, it would need to be run 167,000 times. Thats still a big number except, perhaps, in the context of a server farm processing millions of web clicks per second, where MDP analysis could help allocate computational resources. In other contexts, the work at least represents a big step in the right direction.

People are not going to start using something that is so sample-intensive right now, says Jason Pazis, a postdoc at the MIT Laboratory for Information and Decision Systems and first author on the new paper. Weve shown one way to bring the sample complexity down. And hopefully, its orthogonal to many other ways, so we can combine them.

Unpredictable outcomes

In their paper, the researchers also report running simulations of a robot exploring its environment, in which their approach yielded consistently better results than the existing approach, even with more reasonable sample sizes nine and 105. Pazis emphasizes, however, that the papers theoretical results bear only on the number of samples required to estimate values; they dont prove anything about the relative performance of different algorithms at low sample sizes.

Pazis is joined on the paper by Jonathan How, the Richard Cockburn Maclaurin Professor of Aeronautics and Astronautics at MIT, and by Ronald Parr, a professor of computer science at Duke.

Although the possible outcomes of a decision may be described according to a probability distribution, the expected value of the decision is just the mean, or average, value of all outcomes. In the familiar bell curve of the so-called normal distribution, the mean defines the highest point of the bell.

The trick the researchers algorithm employs is called the median of means. If you have a bunch of random values, and youre asked to estimate the mean of the probability distribution theyre drawn from, the natural way to do it is to average them. But if your sample happens to include some rare but extreme outliers, averaging can give a distorted picture of the true distribution. For instance, if you have a sample of the heights of 10 American men, nine of whom cluster around the true mean of 5 feet 10 inches, but one of whom is a 7-foot-2-inch NBA center, straight averaging will yield a mean thats off by about an inch and a half.

With the median of means, you instead divide your sample into subgroups, take the mean (average) of each of those, and then take the median of the results. The median is the value that falls in the middle, if you arrange your values from lowest to highest.

Value proposition

The goal of MDP analysis is to determine a set of policies or actions under particular circumstances that maximize the value of some reward function. In a manufacturing setting, the reward function might measure operational costs against production volume; in robot control, it might measure progress toward the completion of a task.

But a given decision is evaluated according to a much more complex measure called a value function, which is a probabilistic estimate of the expected reward from not just that decision but every possible decision that could follow.

The researchers showed that, with straight averaging, the number of samples required to estimate the mean value of a decision is proportional to the square of the range of values that the value function can take on. Since that range can be quite large, so is the number of samples. But with the median of means, the number of samples is proportional to the range of a different value, called the Bellman operator, which is usually much narrower. The researchers also showed how to calculate the optimal size of the subsamples in the median-of-means estimate.

The results in the paper, as with most results of this type, still reflect a large degree of pessimism because they deal with a worst-case analysis, where we give a proof of correctness for the hardest possible environment, says Marc Bellemare, a research scientist at the Google-owned artificial-intelligence company Google DeepMind. But that kind of analysis doesn't need to carry over to applications. I think Jason's approach, where we allow ourselves to be a little optimistic and say, Let's hope the world out there isn't all terrible, is almost certainly the right way to think about this problem. Im expecting this kind of approach to be highly useful in practice.

The work was supported by the Boeing Company, the U.S. Office of Naval Research, and the National Science Foundation.