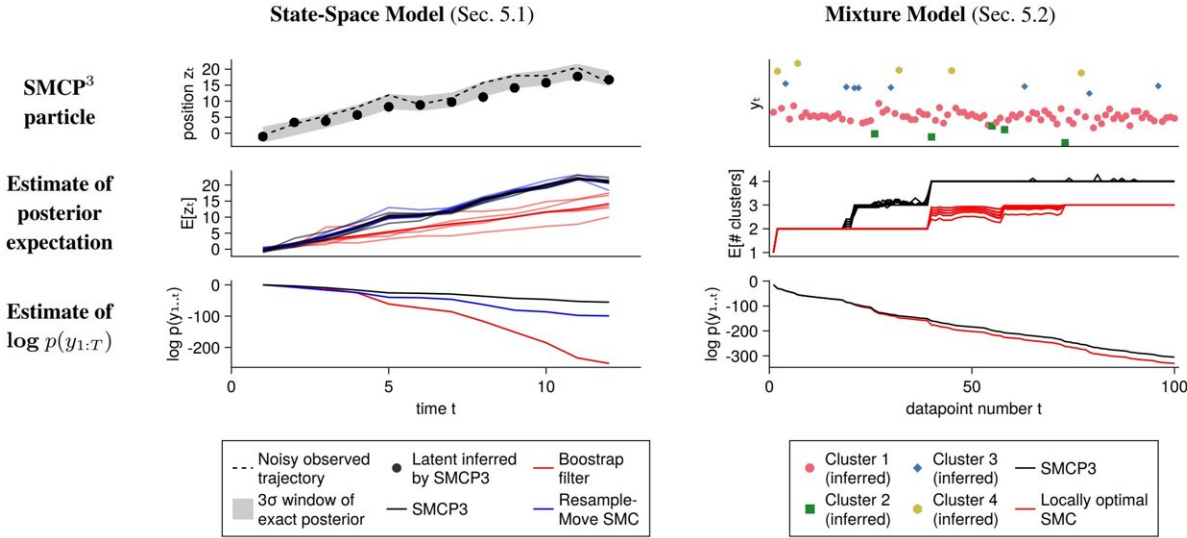# Probabilistic AI that knows how well it's working

May 25 2023, by Rachel Paiste

Illustrative results on synthetic data. Top: A single inferred posterior sample, in two models. In the state-space model, a true position is tracked from noisy observations over time. In the mixture model, clusterings are inferred for increasing subsets of the data. Middle: Estimates of posterior expectations at each step, formed from particle collections. We run each algorithm 5 times. Bottom: Log marginal likelihood estimates at each inference step; higher is better. Credit: *SMCP³: Sequential Monte Carlo with Probabilistic Program Proposals* (2023)

Despite their enormous size and power, today's artificial intelligence systems routinely fail to distinguish between hallucination and reality.

Autonomous driving systems can fail to perceive pedestrians and emergency vehicles right in front of them, with fatal consequences. Conversational AI systems confidently make up facts and, after training via reinforcement learning, often fail to give accurate estimates of their own uncertainty.

Working together, researchers from MIT and the University of California at Berkeley have developed a new method for building sophisticated AI inference algorithms that simultaneously generate collections of probable explanations for data, and accurately estimate the quality of these explanations.

The new method is based on a mathematical approach called sequential Monte Carlo (SMC). SMC algorithms are an established set of algorithms that have been widely used for uncertainty-calibrated AI, by proposing probable explanations of data and tracking how likely or unlikely the proposed explanations seem whenever given more information. But SMC is too simplistic for complex tasks.

The main issue is that one of the central steps in the [algorithm](#)—the step of actually coming up with guesses for probable explanations (before the other step of tracking how likely different hypotheses seem relative to one another)—had to be very simple. In complicated application areas, looking at data and coming up with plausible guesses of what's going on can be a challenging problem in its own right.

In self driving, for example, this requires looking at the video data from a self-driving car's cameras, identifying cars and pedestrians on the road, and guessing probable motion paths of pedestrians currently hidden from view. Making plausible guesses from raw data can require sophisticated algorithms that regular SMC can't support.

That's where the new method, SMC with probabilistic program proposals

(SMCP3), comes in. SMCP3 makes it possible to use smarter ways of guessing probable explanations of data, to update those proposed explanations in light of new information, and to estimate the quality of these explanations that were proposed in sophisticated ways. SMCP3 does this by making it possible to use any probabilistic program—any computer program that is also allowed to make random choices—as a strategy for proposing (that is, intelligently guessing) explanations of data.

Previous versions of SMC only allowed the use of very simple strategies, so simple that one could calculate the exact probability of any guess. This restriction made it difficult to use guessing procedures with multiple stages.

The researchers' SMCP3 paper shows that by using more sophisticated proposal procedures, SMCP3 can improve the accuracy of AI systems for tracking 3D objects and analyzing data, and also improve the accuracy of the algorithms' own estimates of how likely the data is. Previous research by MIT and others has shown that these estimates can be used to infer how accurately an inference algorithm is explaining data, relative to an idealized Bayesian reasoner.

George Matheos, co-first author of the paper (and an incoming MIT electrical engineering and computer science [EECS] Ph.D. student), says he's most excited by SMCP3's potential to make it practical to use well-understood, uncertainty-calibrated algorithms in complicated problem settings where older versions of SMC did not work.

"Today, we have lots of new algorithms, many based on deep neural networks, which can propose what might be going on in the world, in light of data, in all sorts of problem areas. But often, these algorithms are not really uncertainty-calibrated. They just output one idea of what might be going on in the world, and it's not clear whether that's the only

plausible explanation or if there are others—or even if that's a good explanation in the first place!"

"But with SMCP3, I think it will be possible to use many more of these smart but hard-to-trust algorithms to build algorithms that are uncertainty-calibrated. As we use 'artificial intelligence' systems to make decisions in more and more areas of life, having systems we can trust, which are aware of their uncertainty, will be crucial for reliability and safety."

Vikash Mansinghka, senior author of the paper, adds, "The first electronic computers were built to run Monte Carlo methods, and they are some of the most widely used techniques in computing and in artificial intelligence. But since the beginning, Monte Carlo methods have been difficult to design and implement: the math had to be derived by hand, and there were lots of subtle mathematical restrictions that users had to be aware of. SMCP3 simultaneously automates the hard math, and expands the space of designs."

"We've already used it to think of new AI algorithms that we couldn't have designed before."

 More information: Paper: proceedings.mlr.press/v206/lew23a/lew23a.pdf

Open-source implementation of automated SMCP3: github.com/probcomp/GenSMCP3.jl

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology