A delta-rule approximation to Bayesian inference in change-point problems

Robert C. Wilson, Matthew R. Nassar, and Joshua I. Gold

Whether it is a stock market crash, falling in love, or task switching in controlled behavioral experiment, life is full of "change-points" that cause the rules of the game to change abruptly, often without warning. To thrive in these circumstances, individuals must recognize these events quickly and adapt their behavior accordingly – for example by pulling out their money, putting on cologne, or adopting the appropriate task rules.

As with other inference problems, ideal-observer models can help shed light on the computational demands of, and solutions to, change-point problems. Online Bayesian solutions to this problem have been reported [Adams & MacKay, Technical report Cambridge University, 2007; Fearnhead & Liu J. R. Statist. Soc. B 69(4):589, 2007], which can capture many aspects of human behavior on a simple change-point task [Nassar et al., in preparation]. However, despite this success, the computations implied by the Bayesian models seem biologically implausible. These computations require the brain to maintain and update a constantly growing probability distribution (the "run-length distribution") over all possible locations of the last change-point.

In this work we address this shortcoming by systematically reducing the full Bayesian solution to show that it can be well approximated by a form of delta rule. This reduced Bayesian algorithm updates its current beliefs about the world based on the difference between the model's current predictions about the world and the observed reality. Given the substantial evidence for this kind of prediction-error signal in the brain [e.g. Schultz W., et al., Science 275:1593, 1997], the model is more plausible than the full Bayesian solution. This approach also yields a massive reduction in computational cost, going from $O(t)$ computations per time step at time t, to $O(1)$.

Our approach rests on drastically simplifying the representation of the run-length distribution. In particular, we introduce a reduced distribution that is not maintained over all possible run-lengths but instead on just two run-lengths: the first a run-length of zero, which assumes that a change-point just occurred, and the other a non-zero run-length, which represents the expected time since the last change-point. Using this reduced distribution, we derive update equations for the parameters of the model that, for the mean of the predictive distribution, take the form a delta rule. We find that the learning rate depends on both average run-length and the probability of a change-point at the current time, given the latest data. We also show that an extension of this work to deal with hierarchical change-point models [Wilson et al., submitted] leads to a hierarchy of delta-rules. This hierarchical model can adjust to increasingly complex change-point dynamics, including cases not considered by the reported online Bayesian solutions, in which the frequency of change-points is unknown.

This work establishes a formal relationship between a biologically plausible algorithm based on the delta rule and ideal-observer Bayesian models. The results allow us to characterize in detail the strengths and limitations of both kinds of model in performing on-line inference for a range of complex change-point problems.