Understanding Learning Trajectories With Infinite Hidden Markov Models

Sebastian A. Bruijns (sebastian.bruijns@tuebingen.mpg.de)
Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics
Tübingen, Germany

The International Brain Laboratory (info@internationalbrainlab.org)

Peter Dayan (dayan@tue.mpg.de)
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
University of Tübingen, Tübingen, Germany

Abstract

Learning the contingencies of a complex experiment is not easy. Individuals learn in an idiosyncratic manner, revising their strategies multiple times as they are shaped, or shape themselves. They may even end up with different asymptotic strategies. This long-run learning is therefore a tantalizing target for the sort of quantitatively individualized characterization that descriptive models can provide. However, any such model requires a flexible and extensible structure which can capture the rapid introduction of radically new behaviours as well as slow changes in existing ones. We suggest a dynamic input-output infinite hidden semi-Markov model whose latent states are associated with specific behavioural patterns. This model encompasses a countably infinite number of potential states, and so can capture new behaviours by introducing states; equally, dynamical evolution of the behavioural pattern specified by a single state allows tracking of slow adaptations in existing behaviours. We fit this model to around 10,000 trials per mouse as they learned to perform a contrast detection task over multiple stages. We quantify different stages of learning via the number and psychometric characteristics of behavioural states, providing comprehensive insight into the highly individualised learning trajectories of animals.

Keywords: learning; behavioural modelling; nonparametric Bayesian; hidden Markov model

Experimental Data

We fit data from 24 mice learning a contrast detection task (Laboratory et al., 2021) over an average of 18 sessions, each of around 600 trials. On each trial, a Gabor patch of a controlled contrast appears equiprobably (except in special cases, see below) on the left or right side of the screen. The mouse turns a wheel to indicate the side, and is rewarded for correct choices (on 0% contrast trials, where both sides are empty, one side is randomly rewarded). Initial training involves only the easiest inputs (100% and 50% contrasts); more difficult inputs (25%, 12.5%, 6.125% and lastly 0% contrast) are introduced as the subject improves. If the mouse makes a mistake on a 50% or 100% contrast trial, the stimulus is repeated on the same side, to encourage unbiased behaviour. Because of this, strongly biased policies can lead to less than 50% reward rate.

Model

We use a dynamic input-output infinite hidden semi-Markov model (ioiHSMM) to characterise the relationship between inputs (contrast on a trial, recent past choices and feedback), and output (the animal's current choice). At this model's heart is an infinite hidden Markov model, which is a nonparametric Bayesian model (Johnson & Willsky, 2013). It has all the basic components of a hidden Markov model (initial state distribution, transition matrix, observation distributions), but also performs inference over the number of realized hidden states (subject to an inbuilt Occam's razor). This flexibility is critical to the model's ability to address the different numbers of stages through which different animals transition during learning. Every state also has a duration distribution (making the model semi-Markov) in the form of a negative-binomial distribution, as the implicit geometric dwell times of conventional hidden Markov models were insufficient. The model employs logistic regression to represent an input-output relationship that maps input features to response probabilities via a set of weights. This allows a comprehensive treatment of all factors that might influence choice. The weights together define an extended psychometric function, and (in the generative model) are dynamic, changing between sessions by an amount drawn from a zero-mean Gaussian distribution. The model combines and expands on aspects of recent models that successfully describe animal behaviour, namely the generalized linear model hidden Markov model for finding discrete behavioural states (Ashwood et al., 2022), and non-state based logistic regression combined with a random walk prior on psychophysical weights to track the trajectory of subjects' decision-making strategies (Roy, Bak, Akrami, Brody, & Pillow, 2021).

The states of the ioiHSMM are collectively able to describe the stages of understanding of the animal (giving us insight into which factors currently influence its choices). States can change slowly across sessions, tracking gradual learning in the behaviour of a mouse; but they can also be replaced abruptly by new states, e.g., if a sudden insight into the task causes a drastic change in behaviour. Using appropriate priors, we use Gibbs sampling to fit this generative model to all sessions of individual mice. We verify the validity of our fitting procedure via recovery analysis and show cross-validation performance on par with existing models for quantifying behavioural changes in mouse behaviour (Roy et al., 2021).

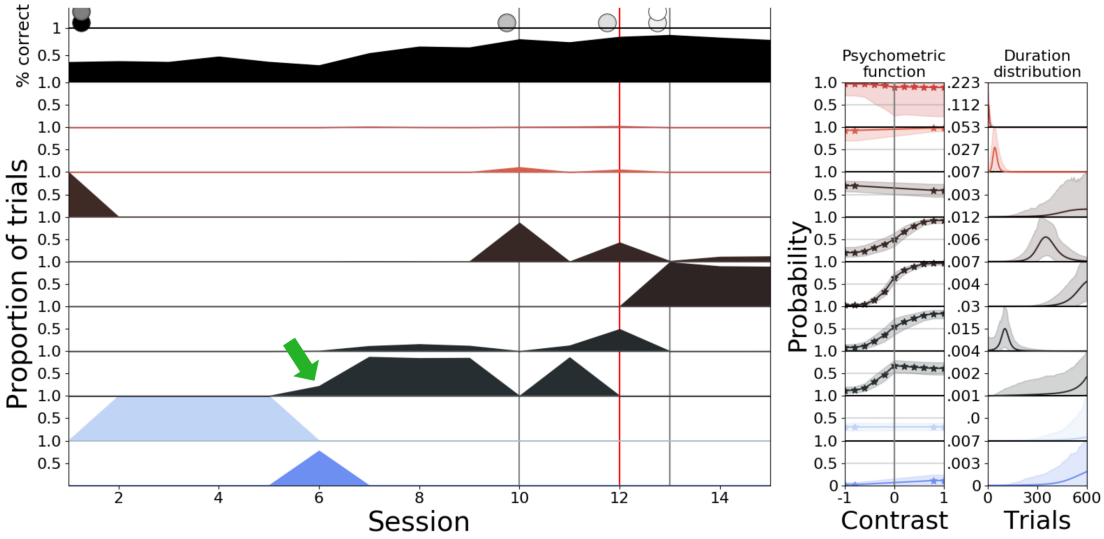

Figure 1: Posterior mass assigned to states over sessions for an example mouse. The model postulates 9 relevant behavioural states, coloured and sorted by their degree of bias. The top row shows the subject's overall accuracy over sessions. Vertical lines indicate when new contrasts are introduced (grey circles). On the right we plot mean psychometric function (PMF) with 95% credible interval for each state (contrasts on the left encoded as negative numbers). As some of the states never encountered specific contrasts (e.g. state 1 is only active with 100% and 50% contrasts), we show a reduced PMF, the points at which the PMF is defined are marked by stars. For better visibility we plot all points on the PMF as equidistant.

Results

Figure 1 shows a summary of the posterior sampling from the ioiHSMM for one mouse. The mouse goes through a number of stages in the first 6 sessions, which have different overall biases, but are all indifferent to the stimulus side. In session 6, the animal first uses a state (state 3, green arrow) which distinguishes between the two sides (though answers are still at chance for contrasts on the right), signifying an important step in learning. This state is used again throughout the next session, and is always replaced by a state with an even better PMF (state 4) towards the end of a session. It is unclear why this more competent state only appears towards the end. At the end of the training protocol the mouse's behaviour is dominated by one state with an unbiased PMF and small lapse rates. **Figure 2** shows the posterior over states within one session of the same mouse. The 3 different states which explain a number of consecutive trials in this part of the session tease apart subtle differences in the behaviour of the animal: It starts with a reasonable PMF, but a higher lapse rate for left contrasts (negative numbers). This is followed by a brief state that is extremely biased towards rightwards answers. Lastly another state with a reasonable PMF takes over, but this time with a higher lapse rate for contrasts on the right side.

While this description of behaviour naturally lends itself to analyses on the level of the individual, we have also begun to search for commonalities and trends across the population of animals. By warping the varying numbers of training sessions onto a common timeline, we can plot overall trends of state usage (**figure 3**). Initial sessions are usually dominated by just one state, but behaviour then quickly differentiates across multiple states, as the animal apparently tries out different strategies. Towards the end we see a slight pruning of the state numbers again; in particular the last sessions tend to be made up out of one very good state, and one state that has been described as being 'disengaged' (higher lapse rates or higher bias).

In sum, we created a highly flexible and comprehensive model for quantifying the entire learning trajectory of animals. We plan to use this model to study differences in the manner of learning of populations, such as a population of mice which do not succeed in learning the task, or mice with a genetic predisposition to autism spectrum disorder. We will also perform neural recordings during these training sessions, and, by using the states we found as predictors for interpreting neural data, will verify more directly their validity.

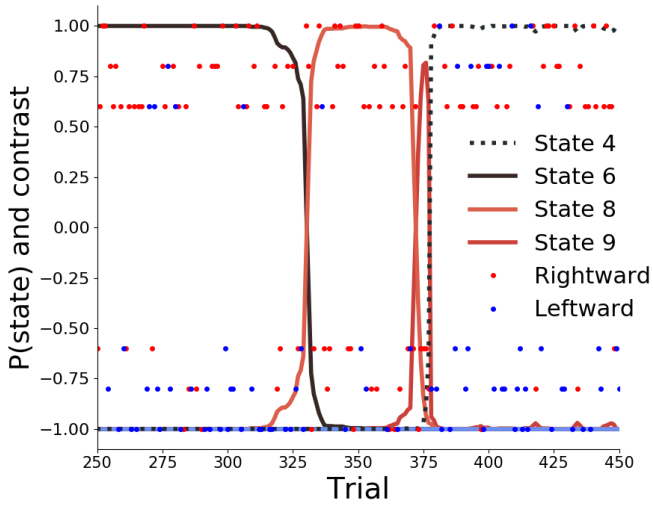

Figure 2: Excerpt of the posterior from the session of the mouse from **Figure 1** marked by the red line. Lines show the posterior probability of any given state being active on a trial, coloured dots indicate the contrast and the animal's response.

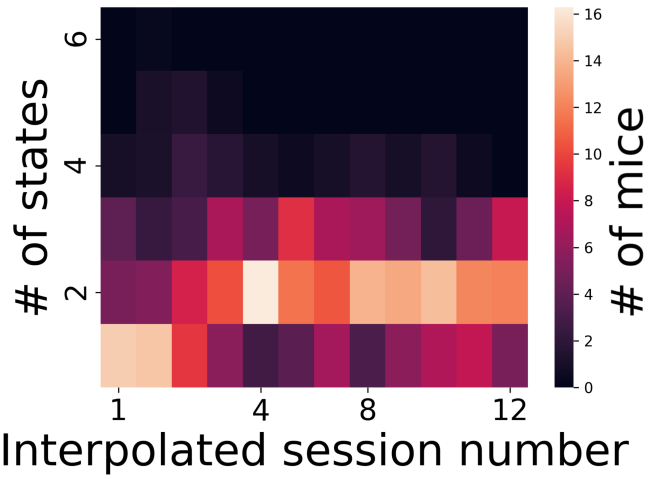

Figure 3: By interpolating the different numbers of sessions it took each mouse to complete training onto a single timeline, we can plot how many states are used across time as a heatmap.

Acknowledgments

We thank Jonathan Pillow, Alexandre Pouget, Zoe Ashwood, Nicholas Roy and the Behavioural Analysis and Theory working groups of the International Brain Lab for discussions. This work was funded by the Simons Foundation (SAB, PD), the Max Planck Society (SAB, PD) and the Alexander von Humboldt Foundation (PD).

References

Ashwood, Z. C., Roy, N. A., Stone, I. R., Urai, A. E., Churchland, A. K., Pouget, A., ... Laboratory, T. I. B. (2022, Feb 01). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, *25*(2), 201-212. doi: 10.1038/s41593-021-01007-z

Johnson, M. J., & Willsky, A. S. (2013, feb). Bayesian nonparametric hidden semi-markov models. *J. Mach. Learn. Res.*, *14*(1), 673–701.

Laboratory, T. I. B., Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., ... Zador, A. M. (2021, may). Standardized and reproducible measurement of decision-making in mice. *eLife*, *10*, e63711. doi: 10.7554/eLife.63711

Roy, N. A., Bak, J. H., Akrami, A., Brody, C. D., & Pillow, J. W. (2021). Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron*, *109*(4), 597-610.e6. doi: https://doi.org/10.1016/j.neuron.2020.12.004