Autoregressive Hidden Markov Model for Quantifying Mouse Behavior

Byron Price, Zora Che, Azar Ghahari
Boston University
Commonwealth Avenue, Boston, MA 02215, USA

Abstract

Quantifying animal behavior through time is a central question in neurobiology for understanding how genes, receptors and neural circuits interact for brain function and movement. Labeling behavior states is an open question, and unsupervised methods are preferred to garner patterns from unlabled continuous time series data. We contribute to the goal of understanding number of behavioral states in mice, using an autoregressive Hidden Markov model that is fitted through expectation maximization. We found that 20 states provided the optimal likelihood in the 5-30 states range, and that our trained model was able to identify behavioral motifs.

1 Introduction

Characterizing how naturalistic behaviors unfold over time—and how the content of behavior is altered by experimental manipulations or disease—offers a powerful lens to better understand how genes, receptors and neural circuits collaborate to enable brain function. Understanding how neural circuits create Innate behaviors requires a clear framework for characterizing how behavior is organized and evolves over time.

Through careful human observation, behavioral modules have traditionally been identified and recent technical advances have enabled more comprehensive characterization of the components of behavior. Behavioral modules and their associated transition probabilities can now be discovered systematically through automated machine vision, clustering and classification algorithms in invertebrates. however these behavioral modules have not yet been implemented in mice.

These behavioral approaches are powerful in quantifying the component of innate behaviors However, their implication requires the prior specification by human observers, perception and intuition. For example analyzing and understanding the behavior in mice requires some challenges. One of the challenges is that the most current methods track two-dimensional parameters such as the position, velocity, or 2D contour of the mice interact while they have complex three-dimensional(3D) pose dynamics. (3)

the mice behavior evolves on many timescales in parallel, identifying the correct spatio-temporal scales to modularize behavior is unclear. these modules of behaviors have variability from moment-to-moment and animal-to-animal, which causes significant challenges for identifying the content of behavioral modules

based upon recent advances in machine learning methods, identifying behavioral modules and their associated transition probabilities without human supervision can be a solution to the above challenges. This research uses 3D imaging to capture the pose dynamics of mice as they freely behave in a variety of experimental contexts.

(**?**)

(7)

Preprint. Work in progress.

(6)

1.1 Motivation

(8)

2 Methodology

2.0.1 Data Pre-Processing

The raw data are videos recorded at 30 Hz from an XBOX Kinect2 depth sensor. The sensor records the distance to objects in the environment and has an accuracy of 2mm at the recorded distance (50 cm to the floor). Mice were allowed to move freely in an open enclosure (∼ 40-by-40cm) while the depth video was recorded for 20 minutes. A static background image was first subtracted from the raw depth images and then a threshold was set to eliminate noise. Next a morphological opening procedure eliminated small objects in the image. The resulting video is of the mouse wandering in a pitch black environment. We then located the mouse within the video and found its center of mass, principal body axis, and head. This was done by converting the depth image to a set of points in the plane. The center of mass is the mean of this point cloud and the principal body axis is the first eigenvector of the covariance matrix. The head was found using a neural network classifier with a softmax output. These detected features were then verified by eye by re-watching the video with them superimposed on the mouse. With the three features, we created a new video of the mouse aligned to its principal body axis, always facing the same direction. This places the mouse in allocentric coordinates allowing us to detect deviations from a generic position such as when the body turns or rears up. Finally, we performed principal components analysis on the aligned video to reduce dimensionality from pixel space to a 10-D space representing different body positions. If the video is first mean subtracted and then converted to a matrix $\mathbf{M} \in \mathbf{R}^{T \text{by pixel}}$ where $T = 35964$ second in this case or 19.98 minutes and pixels= 6400. The singular decomposition of matrix $\mathbf{M}$ can be written as:

$$\mathbf{M} = \mathbf{USV^T}$$

The principal eigenvectors are the columns of V and the reduced-dimensionality dataset is given by

$$\mathbf{U}$$

. We therefore have for further analysis: $\mathbf{U_t}_{t=1}^{T}$ with $\mathbf{U_t} \in \mathbf{R}^{10-by-1}$ The top 10 principal components account for approximately %95 of the total variance in the aligned movie.

2.0.2 Statistical Model

To quantify behavior, we used an unsupervised learning algorithm that identifies behavioral “modules” or “syllables” and the transition probabilities between them. The model is an autoregressive Hidden Markov Model (AR-HMM), which was previously validated for this purpose in (1) Wiltschko et al. They compared the AR-HMM to a number of other models and against several non-parametric quantifications of behavior, finding that the AR-HMM provided a good fit to the data. The AR-HMM is a special case of a Hidden Markov Model:

$$\mathbf{U}_t|\mathbf{Z}_t = k \sim \mathbf{N}(\mathbf{U}_t|\mu_k + \sum_{l=1}^{L} A_k^{(l)}\mathbf{U}_{t-1}, \Sigma_k)$$
$$s.t.: \mathbf{Z}_t = k|\mathbf{Z}_{t-1} = j \sim p_{jk}$$

$$\mathbf{Z}_1 = k \sim \pi_k$$
$$\Theta = \{\mathbf{\Sigma}, \mathbf{Z}, \mathbf{A}, \pi, \mu\}$$

Each observed data point, $\mathbf{U}_t$, is modeled as a linear combination of prior observed data points, $\{\mathbf{U}_{t-1}, ..., \mathbf{U}_{t-L}\}$, through the autoregressive transition matrices $\mathbf{A}_k^{(l)}$. The indices l are lags for the autoregressive process and L represents the maximum lag (we set this hyperparameter to 3, following the original paper by Wiltschko et al.(1)). The subscript k indicates that the precise linear combination depends on the current setting of the latent variable $\mathbf{Z}_t$. The latent state, $\mathbf{Z}_t$, is a discrete random variable that can take any of K unique values corresponding to K behavioral states. The set comprises a first-order, discrete Markov chain parameterized by a transition probability matrix $\mathbf{P}$ with entries p_{jk}. Initialization of the chain is provided by $\mathbf{Z}_1$, which is governed by a categorical distribution, π. The learning goal is to infer the most probable latent behavioral states, $\mathbf{Z}_t$, for each time step from 1 to T. In order to do so, we must first learn the model parameters, Θ, using maximum likelihood estimation. In what follows, we relied heavily on the work of Rabiner and Bishop to derive an expectation maximization (EM) algorithm that learns parameters and infers the latent behavioral states (Bishop, 2006; Rabiner, 1989).

Under the AR-HMM, the complete-data log-likelihood,

$$ln[\mathscr{L}(\Theta;\mathbf{U},\mathbf{Z})] = ln[\mathbb{P}(\mathbf{U},\mathbf{Z}|\Theta)]$$

is the likelihood of the data assuming we had actually observed the latent states. If that were the case, learning model parameters would be straightforward, as we will show. The complete-data log-likelihood is:

$$\sum_{k=1}^{K}\mathbb{I}(\mathbf{Z}_1 = k)ln[\pi_k] + \sum_{t=1}^{T}\sum_{k=1}^{K}\mathbb{I}(\mathbf{Z}_t = k)ln[N(\mathbf{U}_t|\mu_k + \sum_{l=1}^{L}A_k^{(l)}\mathbf{U}_{t-1}, \mathbf{\Sigma}_k)]$$

$$\sum_{t=2}^{T}\sum_{j=1}^{K}\mathbb{I}(\mathbf{Z}_{t-1} = j)\sum_{k=1}^{K}\mathbb{I}(\mathbf{Z}_t = k)ln[p_{jk}]$$

For ease of the representation, we re-write the mean of the normal distribution as:

$$\mu_k + \sum_{l=1}^{L}A_k^{(l)}\mathbf{U}_{t-1} = \mathbf{A}_k^*\mathbf{U}_t^*$$

3 Experimental Results

We experimented with data from Datta Lab, where the mouse is already aligned and centered. We preprocessed using the methodologies described earlier, finding principal components with the top 10 components accounting for around 95% of the the total variance in the aligned frame. We seek to quantify mouse behavior by testing different numbers of states for the Autoregressive Hidden Markov Model. Since mouse behaviors differ, and that we found principal components for each mouse individually, we chose to train and validate on the continuous video of the same mice. We trained on the first 80% of the data of one preprocessed file corresponding to one continuous video of a mice, with 28771 frames and top 10 principal components. From training, we obtain parameters for a K-state Autoregressive Hidden Markov Model, optimized using the expectation maximization (EM) algorithm. For a specified number of states and a number of time lag, we obtain parameters for the model including fitted transition probability matrix P, and autoregressive transition matrix A.

Using the fitted Autoregressive Hidden Markov Model parameters, we test on the remaining 20% of the data with 7190 frames and the same top 10 principal components to calculate a log likelihood on observing testing data under the model. Since the data is a continuous time series, the testing frames follow the training frames. The higher the log likelihood on the testing frames using a set of parameters, the more likely it is for the testing data to be observed under the specified Autoregressive Hidden Markov Model. Thus, we may use it to choose the optimal number of behavioral states.

We found that we achieved the highest log-likelihood with 20 states on testing frames. Specifically we found the likelihood to be:

Figure 1: placeholder(in methods)

30 states	5.7872e04
25 states	5.7767e04
20 states	5.804e04
10 states	5.804e04
5 states	5.3487e04

This concurred with findings from prior experiments using Hidden Markov Model to quantify mouse behaviors, where researchers conjecture there are around 20 behavioral states. Moreover, the average amount of time each state lasts for the trained 20-state Autoregressive Hidden Markov model is 4.7 frames, or around 157 milliseconds. This also corresponds closely to current research on behavioral times.

3.1 Figures

(Combining the four images from methods to be the first figure here) We produced figures to illustrate the original videos, the preprocessed, aligned frames, and select principal components .
We plotted a matrix of transition probability between the states for the fitted optimal model with 20 states, and found that as expected for continuous time series data, the diagonal had the highest probabilities. Some states, such as 3 and 5, had higher than average probability of changing into one another.
To visualize the behavior states the 20-state Autoregressive Hidden Markov Model has found, we produced a figure for the first 500 frames: the upper figure consisting of rolled out frames, forming 6000-by-1 tensors for each frame; the lower figure consisting of the fitted states for each of those frames (Figure 3). We see that there are some correlation between the actual video and the designated states, as well as continuation of states for frames that have very similar in pixels.
To further observe the quantified behavioral states, we also manually watched the labeled frames. For further research, more analysis of the frames may be employed to validate with alternative methods of mouse behavior quantification.
We produced short videos for a specific state via two ways: averaging the frame sequences for all sequences of the same state, and writing several separate sequences with the same state sequentially. We found the latter provided more recognizable results. We first found all the startframe indices for a specific state, and only choosing startframe indices that proceed in the same state for more than 10 frames. We then write these short sequences of frames one after the other using MATLAB. In example video for state number 10, we can see that the mouse engage in a scratching behavior (see attached video).

??,

4 Conclusion

In this paper we experimented on quantifying mouse behavoir using Autoregressive Hidden Markov Model using real lab data. We found 20 states to be the optimal number of state for fitting the observed data to the Autoregressive Hidden Markov Model through maximizing log likelihood. Examination of labeled states showed similar behaviors in sequences of frame, as documented in video we produced. Further studies could be conducted to preprocess all mice videos collectively, finding the same principal components, then fit the Autoregressive Hidden Markov Models across different mice for generalizing behaviors, instead of focusing on one mouse.

5 Authors' contributions

BP conceptualized the research project. BP formulated and implemented the algorithm, and generated preprocessed data with princinpal components. ZC and BP designed the experiment on real data and trained and tested respectively. ZC and BP generated figures and produced videos.

(initial simulated data and performed the analysis of simulated andreal data. DH designed the extended simulation study. DH and DS performed theextended simulation study on ENVirT and other

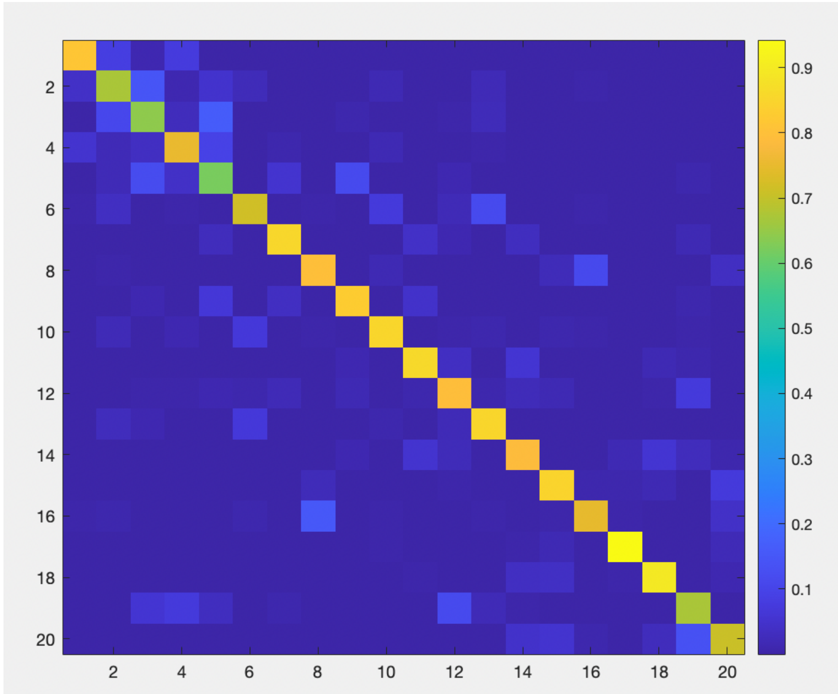

Figure 2: Matrix of transition probability between the 20 states

existing methods. DS developed the graphical user interface of ENVirT and manages the software distribution. YS and SKHcontributed in formulating the optimization algorithm. CYY, IS, BC and SLT processedand contributed in the analysis of real data. IS and DJ prepared the initial draft of themanuscript. DH, DS, SLT and SKH contributed in preparing the final version of themanuscript. All authors read and approved the final manuscript.)

References

[1] Wiltschko, A., MJ, J.,Iurilli, G., Peterson, R., Katon, J., Pashkovski, S.,, Abraira, V., Adams, R., Datta, S. “Mapping Sub-Second Structure in Mouse Behavior” Nature Methods, 2015 Dec 16;88(6):1121-1135.

[2] Buzzfeed Video. “You Won’t Believe What Obama Says In This Video! ", Youtube, 17 April 2018, https://www.youtube.com/watch?v=cQ54GDm1eL0

[3] de Chaumont F,Coura RD-S, Serreau P, Cressant A, Chabout J, Granon S, Olivo-Marin J-C. “Computerized video analysis of social interactions in mice” Nature Methods, 2012;9:410–417.

[4] Posters, Bill. “I Wish I Could...” Instagram, 13 June 2019, https://www.instagram.com/p/BypkGIvFfGZ/, Last Accessed 2 July 2020.

[5] Deepfake Detection Challenge, https://www.kaggle.com/c/deepfake-detection-challenge/overview, Last Accessed 2 July 2020.

[6] Yuezun Li, Ming-Ching Chang, Siwei Lyu “In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking”, [arXiv:1806.02877]

[7] S. Agarwal et al.,“Protecting World Leaders Against Deep Fakes”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 38-45 (2019).

[8] U. A. Ciftci et al., “FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals”, [arXiv:1901.02212]

[9] Mingxing Tan, Quoc V. Le,“EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, [arXiv:1905.11946]

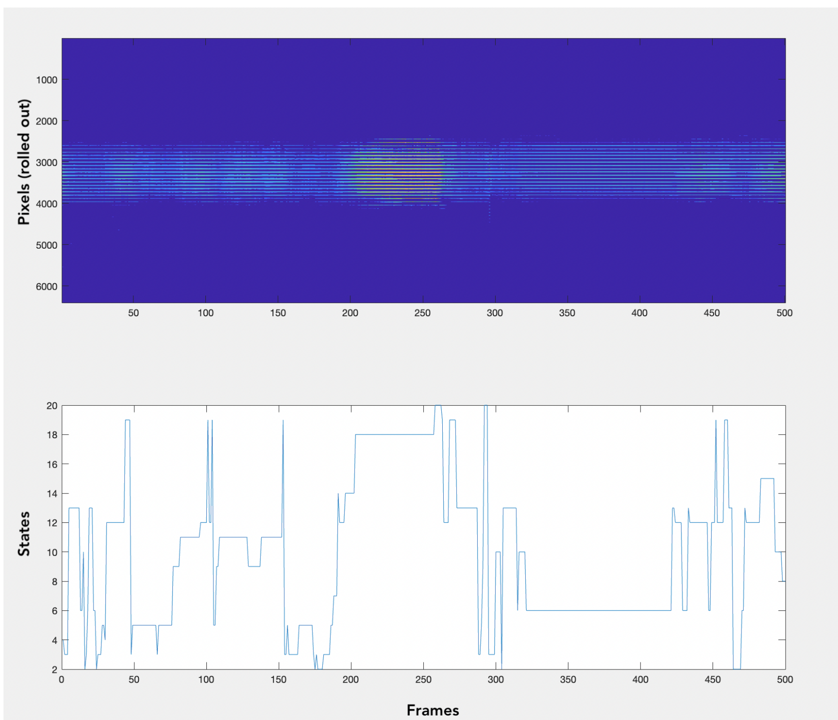

Figure 3: States and frame pixels Plotted for the First 500 Frames