

Entropy Estimation of *P. anubis* Movement Using Bayesian Structural Inference

Katrina Brock, Animal Behavior Graduate Group



Abstract

The goal of this research is to begin to understand movement ecology from an information theoretic standpoint. Bayesian Structural Inference is used to select a family of hidden markov models to characterize the generating process of movement decisions. Entropy rate is calculated from a sample from this family of models to estimate the entropy rate and predictability of the underlying process.

Introduction

Movement ecology is a burgeoning sub-discipline of behavioral ecology that seeks to understand the causes and impact of animal movement. Development and accessibility of animal wearable tracking devices has led to a glut of movement data while theory is still being developed. One of the applications of movement ecology is predicting animal locations over time. These predictions can allow wildlife managers to make informed decisions about policies and resource allocation.

However, researchers' ability to accurately predict animal locations is constrained by the entropy in their movement. The entropy rate and predictability in human movement has already been investigated. Song et al. found that given sufficient history, it is possible to predict the cell-tower that a user's phone will communicate with at a given day an hour with 60%-93% accuracy. My project is the first step to create a product that will allow movement ecologists to use their data to estimate predictability in their study species.

Song et al. used an estimator based on the Lempel-Ziv algorithm to determine the entropy of their system. However, this method only provides a point estimate of entropy for each sequence. Instead, I have used Bayesian Structural Inference to infer a family of hidden Markov models that could have produced each observed sequence. I was then able to calculate the entropy rate of a sample of these machines. This method provides a posterior distribution of entropy rates for each sequence.

Background

The data used was from 26 individuals in a troop of olive baboons. Individuals were captured, anesthetized, and fit with tracking collars. Data was downloaded daily via radio link. After 1-34 days (depending on the individual), the collar was automatically released from the animal, and then retrieved by the researcher. GPS data was sampled at 1 Hz with one meter resolution from 6am to 6pm local time. See Strandburg-Peshkin et al. for more information.

Dynamical System

GPS locations are a discrete time, continuous space time series. Since methodologies to model this type of data as an epsilon have yet to be developed, raw data was coarse-grained into a small number of areas. Process could then be considered a discrete time-discrete alphabet epsilon machine, and the structure of that machine can be inferred using bayesian methods.

Methods

Downsampling and Coarse-graining

As a first pass, data was downsampled to one reading every 15 minutes. Then to discretize the data, k-means algorithm was run independently on each individual's downsampled location history. $K=3$ was chosen both for ease of analysis and because on visual inspection, most individuals seemed to occupy 3 major areas: Northeast, Southeast, and West (Fig 1). Data from each individual was treated as a single separate data stream. Missing data (including overnights) was ignored.

Bayesian Structural Inference

Using the Bayesian Epsilon Machine, I set a prior of all unifiler HMM structures of up to 3 states with a 3 letter alphabet with $\beta=4$. I then calculated a separate posterior for the movement of each individual baboon. Two thousand samples of machines were taken from each posterior the entropy rate was then calculated for each sample machine.

Results

Despite being estimated independently, the structure of the machines with the highest posterior probability are very similar (Fig 2). This is almost certainly because the underlying data are very similar. Except for individual #16 and individual #25 for whom we have very little data, each machine has three states, two of which do not have transitions between them. This mimics the spatial Structure in Figure 1 where for most

individuals the Western cluster is contiguous with both the Northeastern and the Southeast clusters, but the Northeastern and Southeastern Clusters do not border Each other. This offers us a first look into how epsilon machine structure mimics spatial structure.

The entropy rates of the machines in the posterior distributions are much lower than the entropy rates of the machines in the prior distribution (Fig 3). This suggests a high level of predictability in these movement patterns at least at this coarse level. However, this predictability could simply be a byproduct of spatial structure itself rather than something specific to animal movement.

Conclusion

Bayesian Structural Inference can be used to select hidden markov models that reflect the spatial structure in an animal movement data set. It is still unclear whether this representation will be useful to the animal movement community.

Future Directions

This research could greatly benefit from some thoughtful Null Models. For example, simulating random random walks and then pushing the simulated data through this pipeline. This would allow us to identify the features common to epsilon machines build from spatially structured data and distinguish those from features of a particular data set or species.

Future work could also look at the mutual information in movement patterns between individuals within a group, and help uncover social ties.

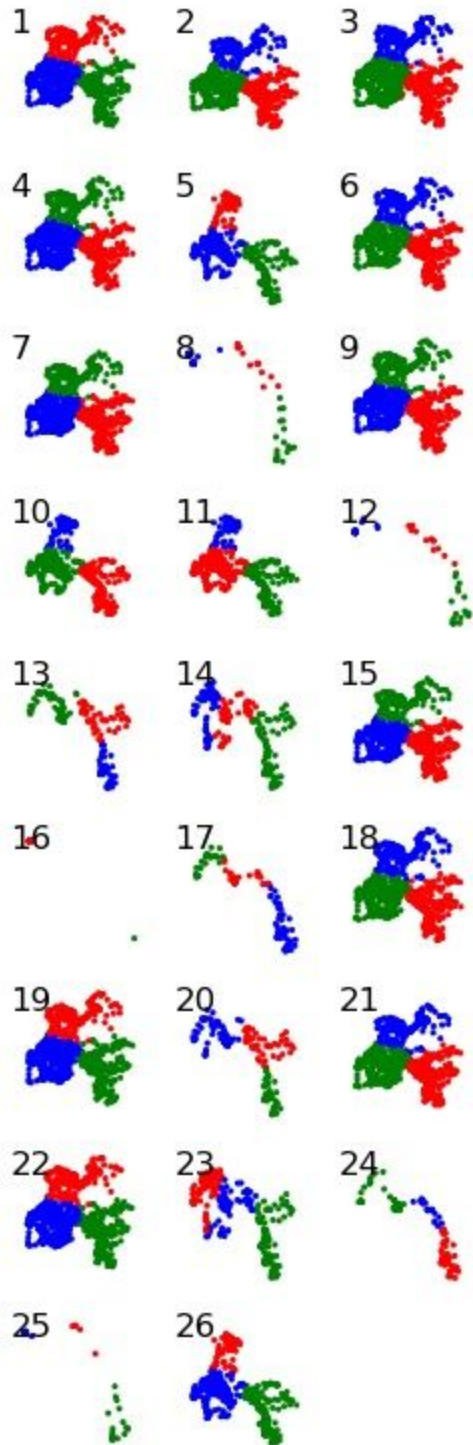


Figure 1

Plot of downsampled data colored by cluster using K means, $k=3$. Each plot represents one individual. Clustering was done separately on each individual's data so there is no correspondence in color scheme among different individuals.

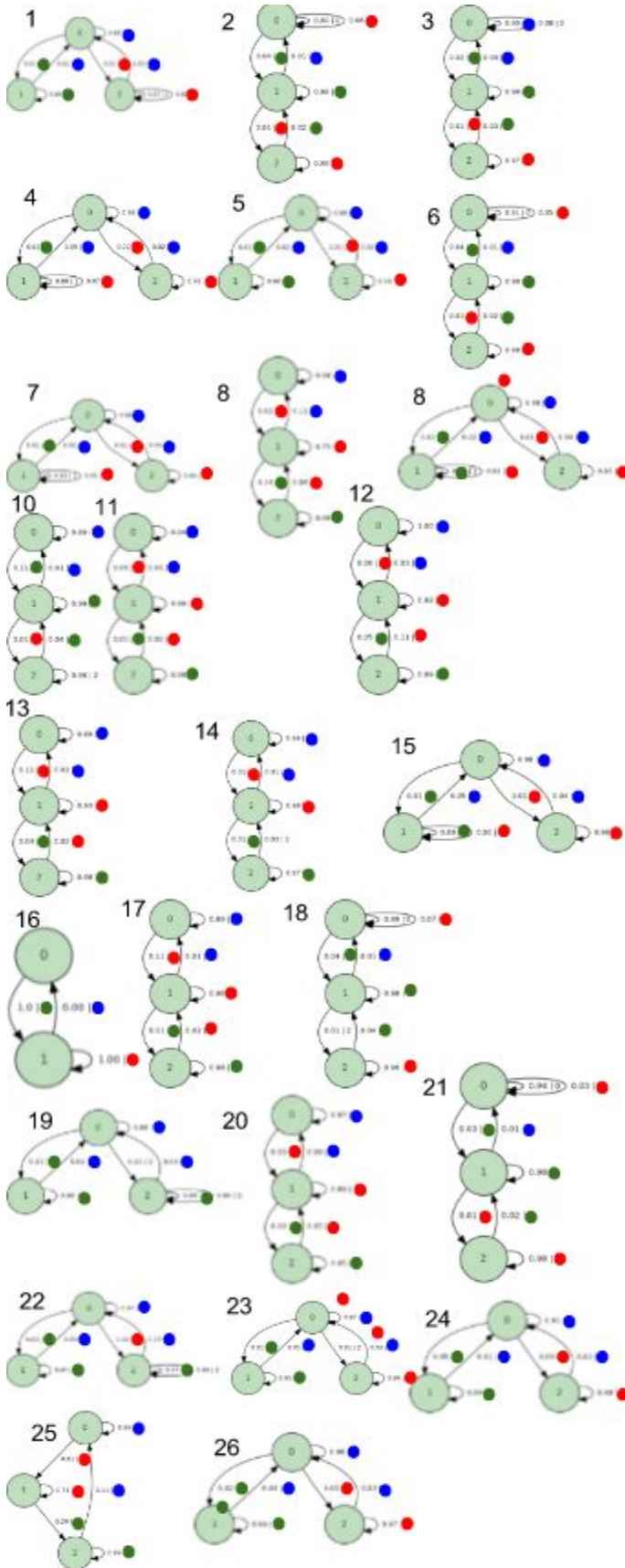


Figure 2

Machine structure with highest probability for each individual. Each machine is for a different individual. Each transition outputs a symbol that corresponds to the individual being located in one of the location clusters above. Within each individual, colors are consistent with Figure 1.

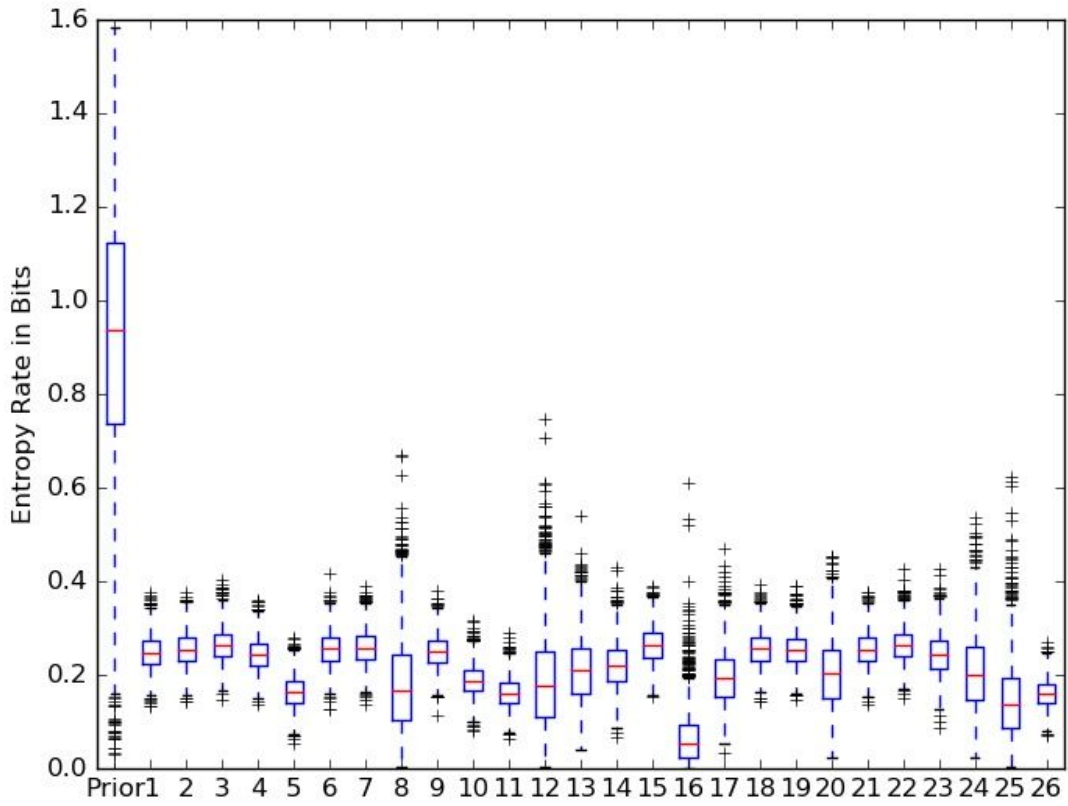


Figure 3

Entropy rate (h_{μ}) of the prior (for all individuals) compared to the entropy rate of a sample of machines from the posterior for each individual.