Specialization Project

Ulrik Bernhardt Danielsen

September 13, 2022

1 Task at hand

Unsupervised clustering of behavioral patterns in rats based on 3D tracking of movements. Tracking the rats posture has already been proven successful. My task is to understand this methodology and for my thesis extend it to data including the faces of the rats mainly whiskers?.

2 First look at the methodology

Mimica et al. 2022 presents the decoding of distinct actions in figure 1. Section A shows how separate rats were fitted with the probes in two different positions. Why these positions? The leftmost figure in section B shows which movements were recorded as time series data. It records six different movements of the head, neck and back, along with the speed of the rats. Then the data is *detrended* and *decomposed* spectrally using a Morlet wavelet transform. This is maybe where I should start, no idea whats what this means. The figure shows this as two steps, while the text as one (detrended *using* Morlet). The next steps is reducing the dimensionality of the data (wavelets?). This is done by finding the principal components explaining at least 95% of the variance, before reducing non-linearly into only two dimensions using t-SNE (t-Stochastic Neighbor Embedding). Using watershed segmentation on this two-dimensional mapping the discernible actions are found—44 in total. The final part of the figure, C, shows the decoding accuracy for individual actions across animals. The decoding accuracy are shown individually for four different cortices. This I don't understand. What are cortices, and what is decoding accuracy?

3 Berman paper

The part which I will focus on is the methodology developed first in Berman et al. 2014. It develops a method for mapping distinct activities in fruit flies. As stated there: *The basis of our approach is to view behaviour as a trajectory through a high-dimensional space of postural dynamics.*

3.1 Procedure

3.1.1 Postural decomposition

Since they don't track the flies movements directly, and 40000 timeseries is a bit much, they first apply PCA to Radon transforms of the images. They find that 50 postural modes are enough to explain a sufficient amount of variance. To do this they shuffle the dataset and compare the PCA

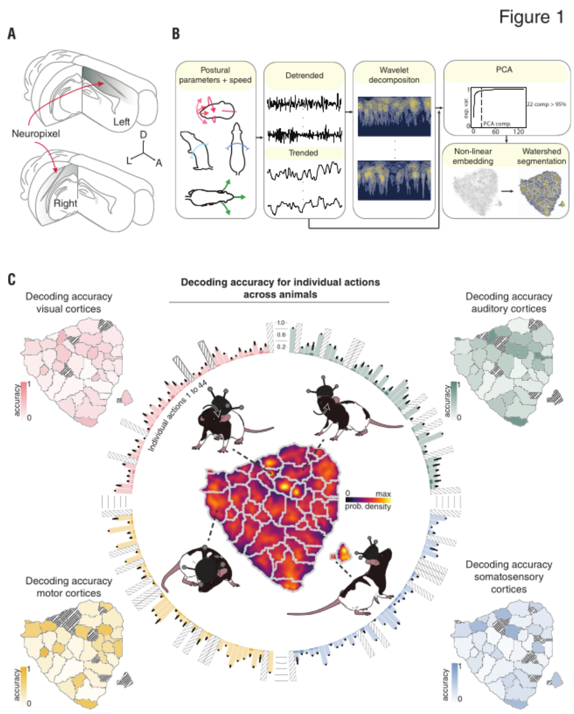

Figure 1: Figure 1 in Mimica et al. 2022.

eigenvalues of the data, to the largest one in a shuffled data set. Is this also the procedure in our case? The individual movies can thus be transformed to a 50-dimensional timeseries, which they denote

$$\mathbf{Y} \equiv \{y_1(t), y_2(t), \ldots, y_{50}(t)\}.$$

This step does not seem to be of relevance to us. I.e., no need to look up Radon transform?

3.2 Spectrogram generation

First they state that looking for repeating sub-sequences in the time series are problematic, as *certain behaviours involve multiple appendages moving at different time scales, this complicating the choice of motif length.* Thus Berman et al. 2014 chooses another path—a *spectrogram representation for the postural dynamics.* The Morlet wavelet transform is supposed to be specially suited for dynamics over multiple time scales. They back this up by citing *Daubechies I. 1992* Ten lectures on wavelets. Might be a nice resource.

3.3 Spacial embedding

The final step is to map the still very large feature vectors into a low dimensional (two-dimensional) space. For this t-SNE is chosen. Why?—because it does care much about preserving the "long" distances between the original features. Many popular dimensionality reduction methods are tweaked for the opposite purpose, to keep the overall structure in mind. This argument should definitely be looked more into.

4 Steps:

1. z-scoring (standardization although dividing by standard deviation is commented out)
2. Smoothing by 3d splines
3. Detrending around the smoothed curve (Why do we need to do this? Do we expect trends?)
4. Morlet continuous wavelet transformation
   (a) Hypothesis testing against 1st order autoregressive process
   (b) Smoothing of power across scales
   (c) Scale-averaged wavelet power
   (d) Rescaling features
   (e) Concatenating trend data and power spectrum
5. Feature vectors were downsampled in time at 1 Hz and pooled acreoss animals and conditions. Where is this done in the code? it
6. PCA, reducing to feature dimension explaining at least 95% of the variance
7. t-SNE
8. Watershed segmentation

5 Introduction

5.1 Motivation

The human brain is an incredibly complex structures that researchers have been trying to understand for a long time. One way to gain information about how the brain operates is to study its neurons. Neurons are cells which can communicate with each other through synapses. This communication are electric signals and can be recorded. Source? At Kavli Institute for Systems Neuroscience at NTNU they are interested in relating these neural spike recordings to the behavior in rats. This in turn begs the question of how rats behave. Manually labelling video recordings of rats running around seems a tedious and unfruitful endeavor. Additionally it introduces bias in our prior assumptions of how the rats behave, and which activities they engage in. Thus, a methodology for automatically detecting distinct behaviours is needed.

5.2 Previous work

Is this necessary?

6 Theory

Concise description of the mathematical concepts used in the methodology.

6.1 Time series

We define time series as a realization $y_t = \{y_{t_1}, y_{t_2}, \dots, y_{t_n}\}$ of a stochastic process $Y(w, t)$, where $w \in \Omega$, Ω being the sample space, and $t \in \mathbb{Z}$, $\mathbb{Z}$ being the chosen index set Wei 2006. It is an ordered series of random variables which can be described completely by its joint probability function

$$F_{t_1,\dots,t_n}(x_1, \dots, x_n) = \Pr\{y_{t_1} \leq x_1, \dots, y_{t_n} \leq x_n)\}.$$

The mean and variance function of a time series y are defined as

$$\mu_t = E(y_t) \tag{1}$$

and

$$\sigma_t^2 = E(y_t - \mu_t)^2.$$

Given two random variables in the series y_{t_1} and y_{t_2}, we define the covariance function and correlation function as

$$\gamma(t_1, t_2) = E[(y_{t_1} - \mu_{t_1})(y_{t_2} - \mu_{t_2})] \tag{2}$$

and

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}. \tag{3}$$

6.1.1 Stationarity

A time series y_t is nth-order stationary if for any shift h and indexes $t_1, t_2, \dots, t_n$ if

$$F_{y_{t_1},\dots,y_{t_n}}(x_1, \dots, x_n) = F_{y_{t_1+h},\dots,y_{t_n+h}}(x_1, \dots, x_n). \tag{4}$$

If (4) holds for all n, the time series is called *strictly* stationary. We also define a nth-order *weakly* stationary time series y_t if the first n joint moments are finite and time invariant. Specifically we define the second-order weakly stationary, i.e. with constant and time invariant mean function (1), and where the covariance function (2) is solely a function of the time difference, as *covariance* stationary. When the covariance function between t_1, t_2 can be written as a function of the time difference $h = |t_1 - t_2|$, i.e. $\gamma(t_1, t_2) = \gamma(h) = \gamma_h$, we call it an *autocovariance* function. The same is true for the correlation function (3), which when is a function of the time difference is called an *autocorrelation* function (ACF).

6.1.2 Detrending

Many methods for analysing and processing time series requires stationarity Shumway and Stoffer 2017. If the series is non-stationary, we can split the it into one stationary and one non-stationary part called the *trend*. Mathematically we write it as

$$y_t = \mu_t + x_t,$$

where x_t denotes the stationary part and μ_t the trend. The process of finding μ_t and then computing $x_t = y_t - \mu_t$ is called *detrending*. Detecting the trend can be done in many ways, for instance using regression techniques or smoothing. The simplest way is to assume a linear trend, $\mu_t = \beta_0 + \beta_1 t$ and estimate the parameters using least squares.

6.1.3 Fourier analysis

6.1.4 Spectral analysis

6.1.5 Wavelet transformation

Morlet wavelet—a sine wave that is "windowed" (i.e., multiplied point by point) by a Gaussian

6.2 Piecewise polynomials

Suppose we have an interval $[a, b]$ divided into M contiguous subintervals. The connecting edges of the subintervals $a = \xi_0, \xi_1, \ldots, \xi_{M-1}, \xi_M = b$ are called knots. On each of the intervals $[\xi_i, \xi_{i+1}], i = 0, \ldots, M-1$ we define a polynomial $p_i(t)$. The function

$$f(t) = \begin{cases} p_0(t), & t \in [\xi_0, \xi_1) \\ p_1(t), & t \in [\xi_1, \xi_2) \\ & \vdots \\ p_{M-1}(t), & t \in [\xi_{M-1}, \xi_M] \end{cases}$$

is called a *piecewise polynomial*.

6.2.1 Splines

In the definition of piecewise polynomials no restrictions are made on the polynomials, they are allowed to take any form. As in Quarteroni, Sacco, and Saleri 2010 we define a *spline* $s_k(t)$ of order k on the interval $[a, b]$ as a piecewise polynomial where

$$\begin{aligned} &s_k(t) \in \mathcal{P}^k, \quad t \in [\xi_i, \xi_{i+1}], \quad i = 0, 1, \ldots, M-1 \\ &s_k(t) \in \mathcal{C}^{k-1}[a, b]. \end{aligned}$$

I.e., the spline consists of piecewise polynomials of order k and has continuous derivatives up to order $k-1$. A common choice is letting $k=3$, providing continuous second derivatives over the interval. This is called *cubic* splines, and are often considered sufficiently smooth for function approximations. It is also common to add curvature constraints at the endpoints, $s_3''(a) = s_3''(b)$, arriving at the *natural* cubic splines.

6.2.2 Regression splines

Suppose now we have data points $y_{t_1}, y_{t_2}, \ldots, y_{t_n}$ on $[a = t_1, b = t_n]$. A spline of order k with chosen knots at $a = t_1 = \xi_0, \xi_1, \ldots, \xi_M = t_n = b$ can be parameterized as

$$s_k(t) = \sum_{i=1}^{M+K} \beta_i h_i(t),$$

where the functions h_i are the truncated-power basis set

$$h_j(t) = t^{j-1},\ j = 1, \ldots, k+1,$$
$$h_{k+1+l}(t) = (t - \xi_l)_+^k,\ l = 1, \ldots, M-1,$$

with $(t)_+ = \max\{t, 0\}$ Hastie, Friedman, and Tisbshirani 2017. The parameters β_i can be found using least squares.

6.3 Dimensionality reduction

6.3.1 Principal Component Analysis

6.3.2 t-Stochastic Neighbor Embedding

6.4 Watershed segmentation

7 Methodology

What is done in practice. Discussion of choices made.

7.1 Input data

The starting point of the analysis is a collection of n separate time series of equal length m right?. We denote these $\mathbf{Y} = \{y_1(t), \ldots, y_n(t)\}$, where $t \in \{t_1, t_2, \ldots, t_m\}$.

7.2 Feature extraction

7.3 Manifold embedding

References

Berman, Gordon J. et al. (2014). "Mapping the stereotyped behaviour of freely moving fruit flies". In: *Journal of The Royal Society Interface* 11.99, p. 20140672. DOI: 10.1098/rsif.2014.0672.

Hastie, Trevor, Jerome Friedman, and Robert Tisbshirani (2017). *The elements of Statistical Learning: Data Mining, Inference, and prediction*. 2nd ed. Springer.

Mimica, Bartul et al. (2022). "Behavioral decomposition reveals rich encoding structure employed across neocortex". In: *bioRxiv*. DOI: 10.1101/2022.02.08.479515. eprint: https://www.biorxiv.org/content/early/2022/02/10/2022.02.08.479515.full.pdf. URL: https://www.biorxiv.org/content/early/2022/02/10/2022.02.08.479515.

Quarteroni, Alfio, Riccardo Sacco, and Fausto Saleri (2010). *Numerical mathematics*. Springer.

Shumway, Robert H. and David S. Stoffer (2017). *Time series analysis and its applications: With R examples*. 4th ed. Springer International Publishing.

Wei, William W.S. (2006). *Time series analysis univariate and multivariate methods*. 2nd ed. Pearson.