Nearest Neighbour Method for Head Cell Decoding

Rasmus Munter

May 11, 2018

Head direction(HD) cells in mice are a neuron population that have a clear relationship with the compass head direction of mice. This paper decodes HD cell response using the nearest neighbor method. The model performance was generally accurate but not precise with a 20.29 RMSE and an correct angle bin prediction accuracy of 21%. A possible route for improvement is to use an average of the nearest neighbors to increase precision. Further analysis showed the choice of smoothing window had a relatively large effect on performance. When using mutual information to assess smoothing window size the HD cells showed signs of temporal coding. However it is suggested that this was simply an artifact caused by mouse movement across the large smoothing window used and that it is primarily rate coding HD cells rely on.

Introduction

Neural decoding is the process of trying to calculate the stimulus that caused a certain measured response. Performing decoding can give insight into how the brain encodes information. The complexity of this issues varies widely with both the method used to decode and the response that is being decoded. This paper addresses a simple variation of this problem by attempting to decode head direction(HD) cells in mice. HD cells are neurons where the firing rate is directly related to the compass heading of the mouse, with the firing rate forming a clear peak at a certain angle, while remaining low otherwise[1]. Since this is a simple encoding of head direction the reverse process, decoding, should be simple as well. This paper applies and analyses one of the simpler decoding methods, the nearest neighbor algorithm[2], on a population of neurons in areas of the mouse brain known to contain HD cells. It then proceeds to use this model to discuss the type of encoding head cells might primarily use.

Methods

The cell population recordings were taken from an experiment by Adrien Peyrache[3] and consists of measurements from the thalamus and post subiculum. Due to time and computing power constraints the data was constrained to recordings from mouse 28 in session 14038 consisting of 80 cells. The continuous head angle data, which can be viewed as the stimulus, was split into 40 non-overlapping 9 degree bins. The session was then was split into two unshuffled halves, with the first half as the training set and the other half as the test set.

Reference vectors for each of the 40 angle bins was calculated. This was done by calculating the average firing rate in each bin and dividing by the occupancy using only data from the training set. The population vector was then calculated on the test set by calculating the average firing rate based on a sliding 100 ms smoothing window. Finally population vectors for timestamps that had no angle data were removed before performing predictions.

The predictions were calculated using a variant of the nearest neighbor algorithm. The pairwise Pearson correlation was calculated between each population vector and the 40 reference vectors. The prediction was then set to be the angle corresponding to the the reference vector with the highest correlation with the population vector.

Part of the analysis of the results uses mutual information between single HD neurons response and the stimulus. The HD cells were found by plotting the tuning curves for all 80 cells that had a minimum firing rate under 10. Then tuning curves with distinct single peaks with a low firing rate otherwise were classified as head direction cells.

Mutual information is a measure of the reduction in uncertainty about a variable given knowledge about another variable[2]. Denoting the neuron response as R and the stimulus as S the equation for mutual information is

$$I(S; R) = \sum_{s,r} P(r)P(s|r)log_2\frac{P(s|r)}{P(s)}.$$

This however requires knowledge about the distributions of the response, stimulus and stimulus conditioned on the response. The simplest approximation is to use the normalized frequency of the responses and stimulus from the experimental data. This is known as the plug-in method[4].

Results

Model perfomance

Running the described method on the test set resulted in an root mean square error(RMSE) of 20.29 and a correct prediction accuracy at 21%.

In figure 1a the distribution shows that this is in general due to a lot of predictions being a few bins off the target value rather than a few extreme errors. 51.6% of the predictions were correct or one bin off the target. However, the confusion matrix in 1b does show that certain angles caused the model to frequently predict a far off angle.
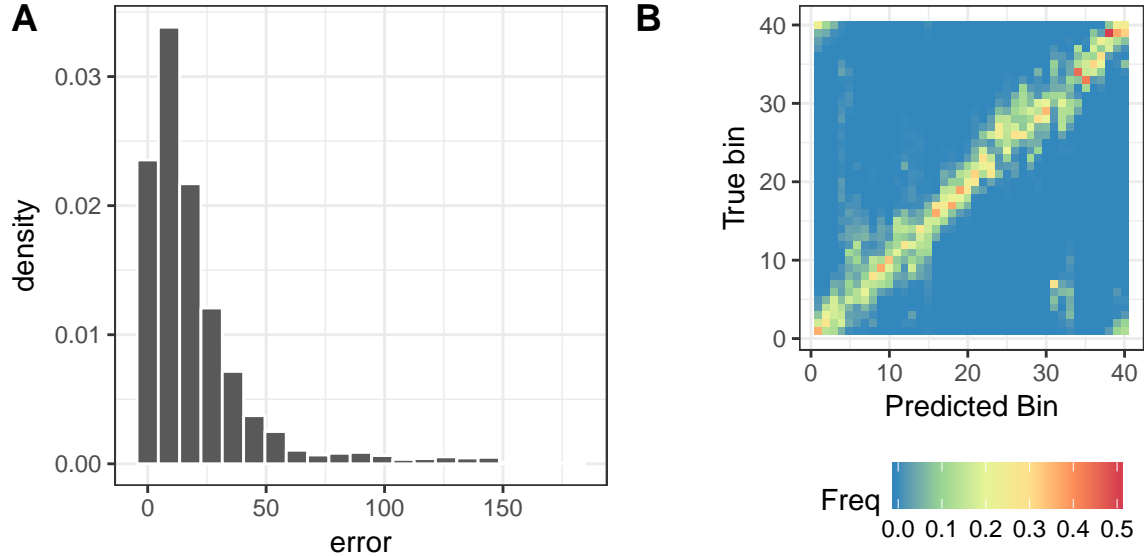

Figure 1: **Prediction error overview:** A: The distribution of prediction error binned into the 40 angle bins. B: Confusion matrix
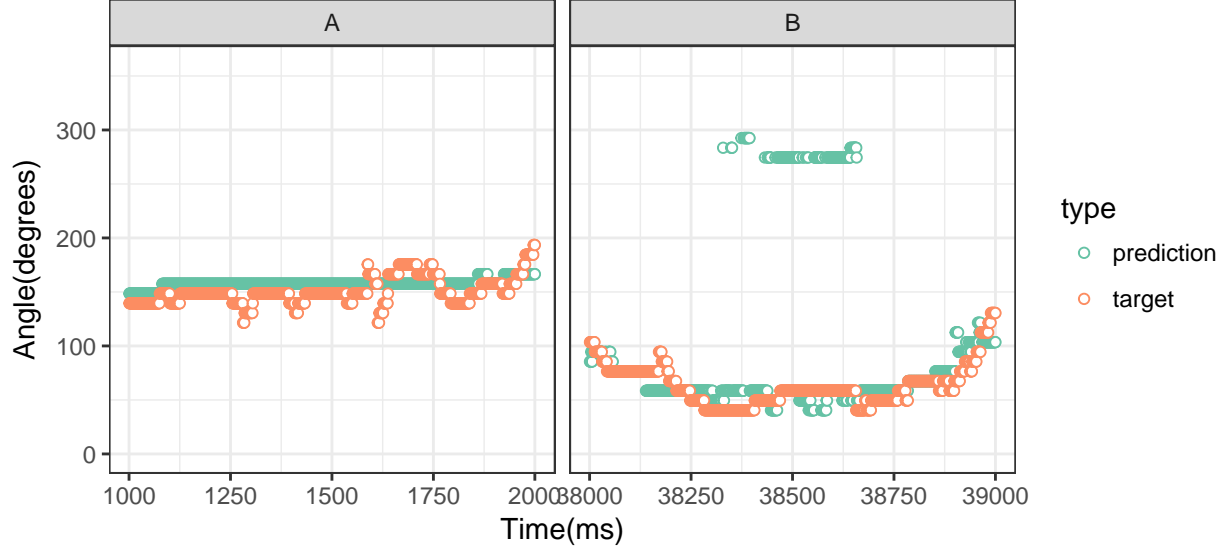
2

Figure 2: **Visualization of model performance on test set:** Two 1000 ms long time periods with the true head direction labeled as target and the corresponding model prediction

Figure 2 shows samples of the decoders output versus the true head angle. Here the same issues appear, with most predictions being one to two bins off the correct angle. It's worth noting that in Figure 2A the prediction value varies up to 7 bins despite the target value staying constant. Figure 2B also shows results from a target angle from where the confusion matrix showed large errors in prediction. Here large jumps between the correct angle and a far off angle are observed. Once again this changes in the prediction occur despite no change in the target angle.

Smoothing window size

The choice of smoothing window effects the model as it removes the temporal information of when the timespikes occurred. The reason smoothing windows are used are to capture firing rate encoding and to decrease random noise by averaging over a longer time period. However, choosing larger smoothing window sizes leads to a loss of the temporal relation between the spikes that were recorded[2]. Figure 3 shows the change in the RMSE and the mutual information for different choices of bin size. In figure 3A, as the smoothing window increases, the RMSE decreases up until a bin size of ~300 ms, before it starts increasing again. Figure 3B shows the mutual information generally decreases with bin size. However it is also important to keep in mind the scale of the information rate. The decrease in information rate is relatively small, especially after passing 300 ms bins where some HD cells mutual information flatten out. This could explain why the RMSE decreases despite the mutual information rate decreasing as well.
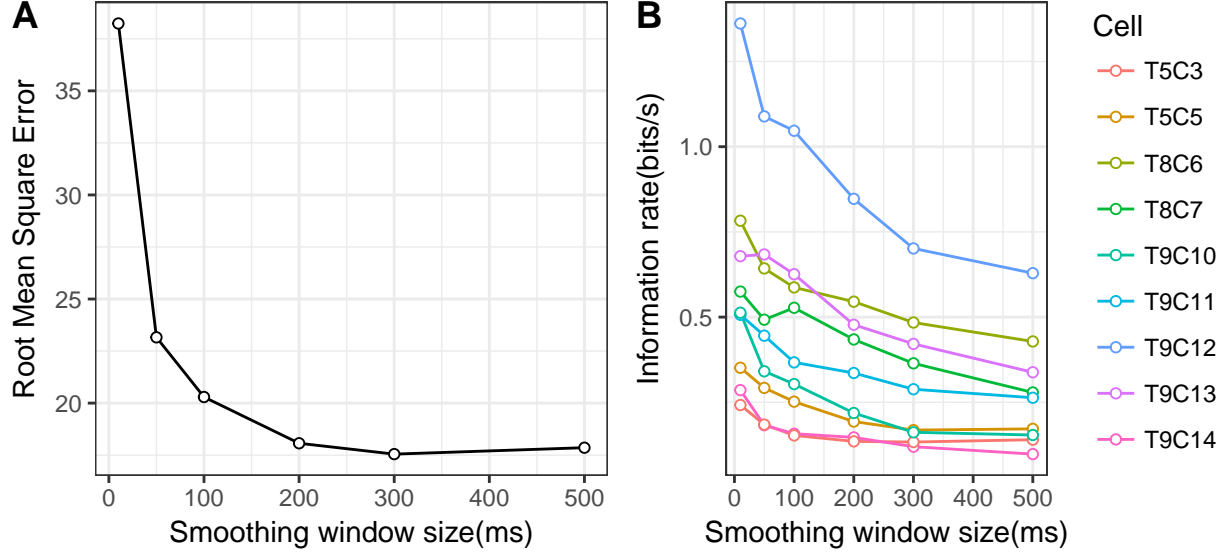
Figure 3: **Effect of smoothing window size:** A: Root mean square error and B: Informaton rate for HD cells plotted against the smoothing window used. Cellnames are denoted by the tetrode number and then the corresponding cell number

Discussion

The performance of the nearest neighbor method on HD cell decoding was in general accurate but not precise. In addition it was shown that the models predictions were unstable and at certain angles the method could predict values almost 180 degrees apart despite the head angle staying the same. This is to be expected from the nearest neighbor algorithm as it is a method with high variance meaning small changes in the model input can lead to large changes in the output[5]. One simple modification that could lower the variance of the model would be use what is known as the K nearest neighbor algorithm. This algorithm bases its prediction on the average of a set of nearest neighbors instead of only the closest one.

It is important to note that using K nearest neighbors would most likely not fix the large mispredictions at certain angles. It is harder to pinpoint exactly why the model fails at certain angles as there are multiple reasons this could be happening. One educated guess is that too many of the neurons in the population are not head direction cells but that coincidentally correlate well at certain angles. Picking out the head cells and creating the model purely on these might yield better results.

The choice of smoothing window was shown to have an effect on the models performance. The 300 ms smoothing window had the lowest RMSE on the test set. However, the information rate steadily decreased for most neurons upon increasing the smoothing window. This suggests that there is some temporal encoding in the head direction cell[2]. Intuitively this makes little sense as there is a clear relation between firing rate and head angle[1] and so what additional benefit temporal coding would have is unclear. It also seems unlikely since the RMSE is much lower for large window sizes. I searched to find relevant literature on the topic but found no results where the information rate had been used to investigate the head cell encoding.

The most likely cause of this issue is the movement speed of mice. With a smoothing window of 100 ms one is calculating the average firing rate for 200 ms. In efforts to find angular head velocity cells, mice have been recorded changing head direction with a velocity over 270 degrees per second[6]. This means a mouse could possibly change its head direction by 6 angle bins over the 200 ms time period. This means as the smoothing window is increased the relation between the firing rate and angle is weakened which would explain a lower calculated information rate. Attempting to ensure that the mouse remains in the same angle bin during one smoothing window would then require a 15 ms time period assuming the mouse started with it's

head direction in the middle of the bin. This corresponds to a 7.5 ms smoothing window which matches the relatively large increase in mutual information from a smoothing window of 5 ms to 50 ms.

Looking past this issue, another possible problem with the mutual information rate is the sampling bias. Sampling bias is a known issue when using the plug in principle to estimate the mutual information as there is an underlying assumption that the data samples recorded represent the true distribution. There exist multiple bias correction methods that could be used to check the effect this has on the information rate relation to smoothing window size[4].

Finally it is important to note that despite the decrease in information rate with smoothing window size the best model by the RMSE metric used a large smoothing window of 300 ms. This indicates that the noise reduction of averaging the firing rate over a larger time period played a larger role than the issues that come with the larger smoothing window. Based on the results from this analysis it seems likely that the head direction cells depend more on rate encoding than temporal encoding.

References

1. Taube JS, Muller RU, Ranck JB Jr. (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J Neurosci. 10(2):420-35.

2. Quiroga RQ, Panzeri S (2013) Principles of Neural Coding. In: Decoding and Information Theory in Neuroscience. Quiroga RQ, Panzeri S. CRC Press, Boca Raton, USA.

3. Peyrache A, Lacroix MM, Petersen PC et al. (2015) Internally organized mechanisms of the head direction sense. Nat Neurosci 18(4):560-575.

4. Panzeri S, Senatore R, Montemurro MA et al. (2007) Correcting for the sampling bias problem in spike train information measures. J Neurophysiol. 98(3):1064-72

5. Hastie T, Tibshirani R, Friedman J. (2009) The Elements of Statistical Learning. In: Overview of Supervised Learning. Hastie T, Tibshirani R, Friedman J. Springer, New York, USA.

6. Stackman RW, Taube JS. (1998) Firing Properties of Rat Lateral Mammillary Single Units: Head Direction, Head Pitch, and Angular Head Velocity. J Neurosci. 15;23(4):1555-1556