Last updated: 2018-03-01

Code version: b6ae3c8

Introduction

The goal of this project was to extract the common behavioural patterns of bees in an experiment which shared similarities to the morris water maze experiment. Cluster analysis has previously been successfully applied to the morris water maze experiment to provide insight into how rodents behaviour at a fine grain resolution. The goal of this project was to investigate whether similar techniques were generalizable to different experiments and species. In particular we wanted to analyse the behaviour of bees in an adapted experiment of the detour paradigm.

In this experiment bees were placed in a circular arena in which they could only move in the x and y directions, i.e. not vertically. Lights were then switched off and on at different positions along the edge of the arena. The bee's position was tracked for the duration of the experiment, producing a dataset of x, y coordinates at uniform time intervals. The experiments were conducted for 3 different treatment groups (1) no treatment (2) treated with saline (3) treated with saline.

Feature Extraction

One of the central challenges of this project was to extract features from the recorded x,y coordinates of the bees so that the data could be represented in a format that would enable typical behaviours and strategies to be detected using cluster analysis. Features were selected that were likely to capture the geometric and positional characteristics of a segment.

Features

Median distance from the centre

This is the median Euclidean distance of all points of a segment scaled by the arena radius

Interquartile range of the distance from the centre

This is the difference between upper and lower quartile distances from the centre scaled by the arena radius

Focus

The focus is specified as:

$$f = 1 - \frac{4A}{\Pi d^2}$$

Where A is minimum enclosing ellipsoid around the segment and d is the segment length.

Eccentricity

The eccentricity is specified as:

E=1-b$^{2/a}$2

$$E = 1 - \frac{b^2}{a^2}$$

Where a and b is semi-major axis and the semi-minor axis of the enclosing ellipse.

Inner radius variation

The inner radius variation is specified as:

$$v = \frac{(c_{UQ} - c_{LQ})}{c_{MED}}$$

Where c_{UQ}, c_{LQ} and c_{MED} are the upper quartile, lower quartile and median for the distance from the centre of the minimum enclosing ellipse.

Central displacement

The central displacement is specified as the distance between the centre of the arena and the centre of the minimum enclosing ellipse.

Path efficiency

The path efficiency is specified as:

$$P = \frac{\sqrt{(x_n - x_1)^2 + (y_n - y_1)^2}}{\sum_{i=1}^{n} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}$$

Where x_i and y_i are ith x and y coordinates of the segment respectively and n is the number of points in the segment. intuitively, it can be thought of as how directly the segment moves from one location to another. (https://www.sciencedirect.com/science/article/pii/S0889159115300398)

Sum of absolute angles

The sum of absolute angles is specified as:

$$S = \sum \arccos\left(\frac{\langle \bar{u}, \bar{v} \rangle}{||\bar{u}||.||\bar{v}||}\right)$$

Where $\bar{u}$ and $\bar{v}$ are sequential vectors. (https://www.sciencedirect.com/science/article/pii/S0889159115300398)

Intermediate calculations

Calculating the position of the centre of the arena

The centre of the arena was taken to be the midpoint between the max and min x and y coordinates in the entire dataset of points.

Calculating the diameter of the arena

The diameter of the arena was taken to be either the distance between the max and min Y coordinate or the max and min X coordinate, whichever was the greatest.

Determining the minimum enclosing ellipse

The minimum enclosing ellipse was computed using the Khachiyan Algorithm. Code from https://ecogenomics.github.io/GroopM/dev_docs/groopm.ellipsoid.EllipsoidTool-class.html was used with some modifications to work for 2d data.

Voting

In reality a bees path can be broken down into shorter joined up segments where each is internally consistent in behaviour. Hereafter we will refer to these segments as the true segments to distinguish them from the arbitrary segments specified for the cluster analysis.

Since the purpose of the this analytical study was to determine the behaviour of bees and how it changes within a single path, it is impossible to know in advance where these true segments start and end. Therefore some arbitrary segment length and start and end position had to be selected to perform the analysis. As a consequence, it is highly unlikely that their beginning and ending will coincide with those of the true segments. In order to more accurately identify the true segment's endpoints, segments were overlapped and voting performed. More specifically, for any given point, the number of segments assigned to each cluster were counted, and the point assigned to the cluster with the highest count. In some cases there was no outright winner, but instead there were two or more clusters equally tied. In this case the point was considered unidentified.

Session information

```
sessionInfo()
```

```
R version 3.4.3 (2017-11-30)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 7 x64 (build 7601) Service Pack 1

Matrix products: default

locale:
[1] LC_COLLATE=English_United Kingdom.1252
[2] LC_CTYPE=English_United Kingdom.1252
[3] LC_MONETARY=English_United Kingdom.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United Kingdom.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
 [1] compiler_3.4.3  backports_1.1.1 magrittr_1.5    rprojroot_1.3-2
 [5] tools_3.4.3     htmltools_0.3.6 yaml_2.1.16     Rcpp_0.12.13
 [9] stringi_1.1.6   rmarkdown_1.8   knitr_1.17      git2r_0.21.0
[13] stringr_1.3.0   digest_0.6.12   evaluate_0.10.1
```