Rokos Summer Project Report - Topological Data Analysis

Charlie Pugh

Outline
I’ve enjoyed doing an online research project, during 6 weeks of this summer, in Topological Data Analysis (TDA), supervised by Siddharth Unnithan Kumar, a PhD student in the field. Although we carried this project out remotely, it was an engaging topic and I greatly enjoyed the zoom discussions and online research. I felt it was the best way to be stuck at home. I’m extremely grateful to Siddharth who was amazing to work with and helped me to appreciate both some new Mathematics and the natural world, to WildCRU and Dr Andrew Loveridge for allowing me access to such fascinating data and also to the Rokos Foundation who generously made this possible.

The Topic
TDA is in the intersection between pure maths and data science, using ideas from algebraic topology to study large data sets. One of the main techniques in TDA is to generate a simplicial complex from a point cloud of data, of which you can compute the persistent homology. This means we take data, which comes as discrete points, and then convert it into a ‘continuous’ shape (the simplicial complex), and then study its topological features (the persistent homology). These topological features are to do with the ‘shape’ of the data, disregarding size and particular geometry. For instance in topology we would consider a coffee cup and a doughnut to be equivalent, in the sense one could be morphed continuously into the other. One reason we might be interested in doing this is that it can be a good way to study high dimensional data. This is data we cannot possibly visualise, and using TDA we can study it without taking some kind of projection into lower dimensional space, which can lose valuable information.

The Application
My supervisor Siddharth, is interested in applying TDA to ecology, the study of animals, plants, their environments and interactions. Ecology is changing as our we develop new ways of collecting ecological information such as though camera traps and even satellite scans of land. This means that large data sets are more common in the field, which require more sophisticated techniques of analysis, but can yield more informative results when used properly.

Fig. 1: The Sunda Clouded Leopard (Neofelis Diardi) native to Borneo featured in the Borneo study.

Fig. 2: A plot of sightings of four species of wild cat in the island of borneo, from the Borneo study.
Two data sets which I was given access to were from a study of different species in Sabah, a region of Borneo\(^1\), and a study tracking the movement of 52 lions over a period of around 20 years in total\(^2\).

The Borneo study was a look at a number of species in Sabah a region in the North of Borneo. It recorded sightings of species in different positions and data about these positions including elevation, a human impact index and other important factors. The lion study gave insight into the varying behaviour of lions based on gender. For instance the category of young male lions can be observed to travel far more and over a larger area than other demographic groups.

One way I tried using TDA was with various lions from this study of their movements. I took the point cloud of where each lion had been in x, y co-ordinates to create a 2d map. I then computed the persistent homologies of these maps with a view to comparing the topological features of the various lions’ paths.

---


---

Fig. 3: The path of an adult male lion with the axes representing the lion’s universal transverse mercator (UTM) coordinates.

Fig. 4: The ‘Betti curve’ of the persistent homology of this path (for dimension 1). This is a measure of how many ‘loops’ there are in the data (on the y-axis) as we vary a parameter (x-axis) which varies the topological surface we create. We can see a ‘bump’ at the right side of the curve which represents a loop which lasts for a reasonably long time. This sort of loop is interesting and represents the big loop in the top right of the lions path. We might interpret this as an exploratory journey to a new area for the lion.

---

Fig.1: Wright S (2009) Clouded Leopard (*Neofelis spec.*), Santiago. Wikipedia