Inferring Time Varying Processes from Cross-Sectional Data

Birkbeck UNIVERSITY OF LONDON

Department of Computer Science and Information Systems

Research Student
Tom Nealon

Supervisors
David Weston
Allan Tucker

Research Area

Time varying processes are often difficult to model accurately because the data sets that describe such processes are often cross sectional without temporal ordering. For example, biomedical data is often very expensive to gather and priority is given to testing to determine a diagnosis rather than model the progression of a condition. This research aims to determine how best to build models of these processes from such data that can then be used for process prediction. In the case of medical conditions such models could be used for earlier diagnosis than is currently possible.

Research Methodology

Initially simple methods of data point ordering implemented such as the pruned minimum spanning tree (MST) method described in "Reconstructing the temporal ordering of biological samples using microarray data" (Magwene, Lizardi, & Kim, 2003). This helped prepare for the implementation of more involved methods of analysis that also used paths calculated using a data set defined by distances and positions in high multidimensional Euclidean space.

Research Approach

The main method to be investigated in this research will be the Pseudotemporal Bootstrap (PTB). This method is described in "The Pseudotemporal Bootstrap for Predicting Glaucoma From Cross-Sectional Visual Field Data" (Tucker & Garway-Heath, 2010). Where trajectories are inferred by using a network created from a (MST) built using the distances between data points. These, so called Pseudotemporal Trajectories are subsequently used to generate a predictive Hidden Markov Model (HMM). The PTB method has been developed to analyse data sets that contain two or more classes. Trajectories have a start point in one class and an end point in a different class.

1. Input: Cross Sectional Data and Class Labels
2. Standardise the data to be zero-centred with a standard deviation of 1.0
3. Construct a weighted graph. First calculate the distance between all points in n-dimensional space then construct a minimal spanning tree across all the points.
4. Randomly select a start position in class 1 and an end point in a class other than 1. Then select the shortest path between those points on the weighted network. Store this trajectory.
5. Repeat step 4 an appropriate number of times.
6. Use these stored trajectories to train a HMM predictive model.
7. Output: TBS Temporal Model

A data set of genomic information taken from biopsies of diagnosed prostate cancer patients and a control group was used to produce TBS trajectories. A visualization of a trajectory is shown in *Figure 1*.

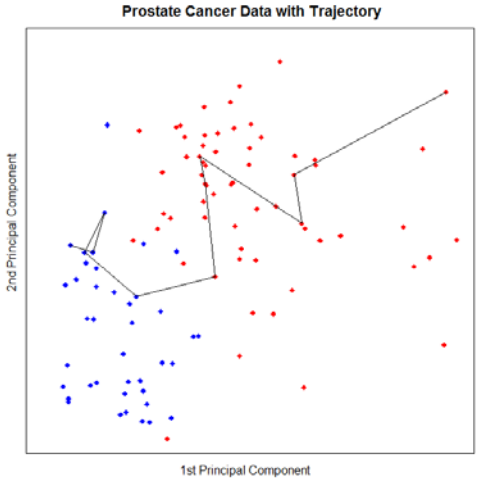

Figure 1. A trajectory from an undiagnosed to a diagnosed sample.

References

Magwene, P. M., Lizardi, P., & Kim, J. (2003). Reconstructing the temporal ordering of biological samples using microarray data. *Bioinformatics*, *19*(7), 842–850.

Tucker, A., & Garway-Heath, D. (2010). The Pseudotemporal Bootstrap for Predicting Glaucoma From Cross-Sectional Visual Field Data. *IEEE Transactions on Information Technology in Biomedicine*, *14*(1), 79–85.