Visual clustering and ensemble clustering methods

R.W. Oldford

November, 2011

The purpose of cluster analysis is to conjecture plausible differences in kind amongst a given collection of instances. This is also what our human visual system excels at: quick and considered detection of the visually like and unlike through a wide variety of display cues e.g. location and relative proximity, movement, shape, colour, texture, and matching against predetermined patterns. Unfortunately, it is difficult to view more than a few dimensions at a time and the clusters we see depend on which dimensions we view and also on how we view them.

This talk introduces a new means of combining the results of any number of clustering methods into a single ensemble method. The method is called TREC (for Tree Reduced Ensemble Clustering) and allows combination of clusterings from any method (partition or hierarchy) whether the result of an algorithm or interaction. The theory on which this ensemble method is based is introduced and applied to several examples.

In particular, I will show how high dimensional data spaces may be explored via low-dimensional trajectories where some clusters might be identified. These low-dimensional clusterings are then combined into a single clustering by application of TREC.

This talk is based on joint work with Adrian Waddell and Wu Zhou.