Note on parallel universes

David J. Hand, Niall M. Adams

Department of Mathematics
Imperial College London
{d.j.hand|n.adams}@imperial.ac.uk

Abstract. The parallel universes idea is an attempt to integrate several aspects of learning which share some common aspects. This is an interesting idea: if successful, insights could cross-fertilise, leading to advances in each area.

The 'multi-view' perspective seems to us to have particular potential.

We have investigated several aspects of this, including the following:

1 Adverse drug reactions

One of the difficulties in detecting adverse drug reactions is that there is often no well-defined baseline against which to run a statistical test to see if the number of events is unexpectedly large. Because of this, almost by default, it is common to adopt an independence model, which assumes that the drugs and reactions are independent. This very simply model is all very well, but fails to take account of information containing in the parallel universe describing taxonomies, similarities between, and relationships between drugs, and the similar parallel universe for side effects. We explored the use of a similarity matrix for drugs, based on their chemical structures, provided by our collaborators GlaxoSmithKline. This has the immediate advantage that it can be more sensitive in detecting drug reactions. In particular, there are situations where the simple independence model will fail to recognise that certain small counts are really special cases of a drug type. Using the similarity matrix allows us to group these instances together, yielding a larger overall count, and hence statistical significance. (David Hand and ZhiCheng Zhang)

2 Patient safety

A particularly important problem of multiple universes is the combination of numerical and text data. In collaboration with the National Patient Safety Agency, we are developing methods for detecting causes of accidents to patients. The data for the incidents includes the parallel universes of numerical and small text descriptions. We are currently developing methods for converting the text descriptions into vector space representations using multimensional scaling methods, so that the two universes can be combined. Other areas in which numerical and text data must be combined are clinical trials and medicine more generally. (David Hand and James Bentham)

3 Behaviour of credit card users

As part of a larger project on fraud detection in retail banking, we have developed methods for combining (i) numerical data describing individual transactions with (ii) nominal data with many categories (e.g. individual ATM machines), using information about the overall behaviour patterns of the population over the nominal variables. (David Hand, Niall Adams, and Piotr Juszczak)

Once numerical representations have been found, a class of tools which may often be effective is Procrustes Analysis. These were originally developed for image matching, but can be applied to distributions of data points. (J.C. Gower, G.B. Dijksterhuis, Procrustes Problems, Oxford University Press (2004).)