Towards a New Science of Big Data Analytics, based on the Geometry and the Topology of Complex, Hierarchic Systems

Fionn Murtagh*

*Royal Holloway, University of London
fionn@cs.rhul.ac.uk

1 Summary

My work is concerned with pattern recognition, knowledge discovery, computer learning and statistics. I address how geometry and topology can uncover and empower the semantics of data. In addition to the semantics of data that can be explored using Correspondence Analysis and related multivariate data analyses, hierarchy is a fundamental concept in this work. I address not only low dimensional projection for display purposes, but carry out search and pattern recognition, whenever useful, in very high dimensional spaces. High dimensional spaces present very different characteristics from low dimensions, I have shown that in a particular sense very high dimensional space becomes, as dimensionality increases, hierarchical. I have also shown how in hierarchy, and hence in an ultrametric topological mapping of information space, we track change or anomaly or rupture.

In this presentation, the first theme discussed is that of linear time hierarchical clustering with application to sky survey data in astronomy, and to chemo-informatics. The second theme discussed is computational text analysis. It is interesting to note that J.P. Benzécri's original motivation was in language and linguistics. In my text analysis work, I have taken the dictum of McKee (Story : Substance, Structure, Style and the Principles of Screenwriting, Methuen, 1999) that "text is the sensory surface of a work of art" and show just how this insight can be rendered in computational terms. This leads to demarcating, tracking, statistical modelling, visualizing, and pattern recognition of narrative. In an application to collaborative writing, I developed an interactive framework for critiquing, and assessing fit and appropriateness of content, on the basis of semantics, leading to books that were published as e-books, having been written by school children in a few days of collaborative class work. In many aspects of this work, hierarchy expresses both continuity and change in the textual narrative or in the narrative of chronological events.

2 Biography

Fionn Murtagh is Professor of Computer Science in the University of London, at Royal Holloway (http ://www.cs.rhul.ac.uk/home/fionn). For the past 5 years he was directing Science Foundation Ireland's funding programmes across a wide area including semantic and sensor web, renewable energy, nanotechnology, and telecommunications. In the past he has held Full

Professor of Computer Science positions at Queen's University Belfast, and at the University of Ulster. For 12 years he served in the Space Science Department of the European Space Agency, based at the European Southern Observatory, in Munich. For many years he was an Adjunct Professor at Strasbourg Astronomical Observatory, Université de Strasbourg. Extensive visiting positions were held at the Joint Research Centre, Ispra, Italy, and in the Department of Statistics, University of Washington.

Fionn Murtagh's degrees are in Mathematics and Engineering Science (BA, BAI), an MSc in Computer Science, all from Trinity College Dublin, a Doctorat de 3ème Cycle in Mathematical Statistics from Université P.M. Curie, Paris 6, and an HDR in Pattern Recognition in Astronomy from (now) Université de Strasbourg. Fionn has been President of the Classification Society (formerly the Classification Society of North America), and President of the British Classification Society. He is an elected Member of the Royal Irish Academy, a Fellow of the International Association for Pattern Recognition, and a Fellow of the British Computer Society.

Fionn Murtagh is Editor-in-Chief of the Computer Journal, the flagship journal of the British Computer Society (published by Oxford University Press). He is a member of the editorial boards of various journals including Pattern Recognition, Journal of Classification, and Neurocomputing. He has published 6 research monographs, 125 journal papers, 135 other papers in book compilations and conference proceedings, and he has edited many books and journal special issues. His Erdös number is 2.