Research and Reviews: Journal of Global Research in Computer Science

Extended Abstract
Vol. 9, Iss. 1
2018

Data Mining 2016: Geometric data analysis: Analytics of processes and behaviors- Fionn Murtagh - University of Derby

Fionn Murtagh
University of Derby, UK

Geometric data analysis allows for â€œletting the data speakâ€□ and integrates qualitative and quantitative analytics. Scope and potential are major in many fields. Case studies here are large scale social media analytics, associated with a neighborhood of social practice and a neighborhood of health and well-being. The interesting survey of Keiding and Louis, â€œPerils and potentials of selfselected entry to epidemiological studies and surveysâ€□ points to very interesting issues in big data analytics. My contribution is in the discussion part of this paper. Through the geometry and topology of knowledge and knowledge , with inclusion of context, of chronology and of frame-models, we are addressing such problems with sampling and representativity. The case studies to be discussed in this presentation are related to mental health and to social entertainment events and contexts in the latter case with many millions of Twitter tweets, using many languages. Particular consideration is given to use and implementation of our analytical perspectives. This includes determining the knowledge content of our data clouds, and of mapping onto Euclidean-distance endowed semantic factor spaces, also because the ultrametric or hierarchical topology, that is characteristic of all forms of complex systems.

Geometric Data Analysis (GDA) is that the name suggested by P. Suppes (Stanford University) to designate the approach to Multivariate Statistics initiated by Benzécri as Correspondence Analysis, an approach that has become more and more used and appreciated over the years. This book presents the complete formalization of GDA in terms of algebra - the foremost original and far-reaching consequential feature of the approach - and shows also how to integrate the quality statistical tools like Analysis of Variance, including Bayesian methods. Chapter 9, Research Case Studies, is almost a book in itself; it presents the methodology in action on three extensive applications, one for medicine, one from politics , and one from education (data borrowed from the Stanford computer-based program for presented Youth). Thus the readership of the book concerns both mathematicians curious about the applications of mathematics, and researchers willing to master an exceptionally powerful approach of statistical data analysis.

Data analysis is that the process of cleaning, transforming, modelling or comparing data, so as to infer useful information and gain insights into complex phenomena. From a geometrical perspective, when an instance (a natural phenomenon , a private , etc.) is given as a fixed-sized collection of real-valued observations, it's naturally identified with a geometrical point having these observations as coordinates. Any collection of such instances is then seen as a point cloud sampled in some metric or normed space.

Big data have 4V characteristics of volume, variety, velocity, and veracity, which authentically involves big data analytics. However, what are the dominant characteristics of massive data analysis? Here, the analytics is said to the whole methodology instead of the individual specific analysis. In this paper, six techniques concerning big data analytics are proposed, which include: (1) Ensemble analysis associated with an outsized volume of knowledge , (2) Association analysis associated with unknown data sampling, (3) High-dimensional analysis associated with a spread of knowledge , (4) Deep analysis associated with the veracity of knowledge , (5) Precision analysis associated with the veracity of knowledge , and (6) Divide-and-conquer analysis associated with the speed of knowledge . The essential of massive data analytics is that the structural analysis of massive data in an optimal criterion of physics, computation, and human cognition. Fundamentally, two theoretical challenges, ie the violation of independent and identical distribution, and therefore the extension of general set-theory, are posed. In particular, we've illustrated three sorts of association in geographical big data, ie geometrical associations in space and time, spatiotemporal correlations in statistics, and space-time relations in semantics. Furthermore, we've illustrated three sorts of spatiotemporal data analysis, ie measurement (observation) adjustment of geometrical quantities, human spatial behavior analysis with trajectories, data assimilation of physical models and various observations, from which spatiotemporal big data analysis could also be largely derived.

Biography

Fionn Murtagh is Professor of Data Science and previously he was into Big Data in Education, Astrophysics and Cosmology. He was the Director of National Research Funding across many domains including Computing & Engineering, Energy, Nanotechnology and Photonics. He has been the Professor of Computer Science, including Head of Department, and Head of School at many universities. He was the Editor-in-Chief of the Computer Journal for more than 10 years, and is a Member of the Editorial Boards of many other journals.

Email: F.Murtagh@gold.ac.uk