

Prof. Fionn Murtagh: clinical and all medical practice, and the health context, the major benefits and some challenges of Big Data science

Received: 16 July 2018; Accepted: 27 July 2018; Published: 03 August 2018.

doi: 10.21037/jmai.2018.07.03

View this article at: <http://dx.doi.org/10.21037/jmai.2018.07.03>

Editor's note

Fionn Murtagh (*Figure 1*) is Professor of Data Science and was Professor of Computer Science, including Department Head, in many universities. Fionn was Editor-in-Chief of the *Computer Journal* (British Computer Society) for more than 10 years, and is an Editorial Board member of many journals. With over 300 refereed articles and 30 books authored or edited, his fellowships and scholarly academies include: Fellow of: British Computer Society (FBCS), Institute of Mathematics and Its Applications (FIMA), International Association for Pattern Recognition (FIAPR), Royal Statistical Society (FRSS), Royal Society of Arts (FRSA), Senior Fellow, Higher Education Academy (SFHEA). Elected Member: Royal Irish Academy (MRIA), Academia Europaea (MAE). Senior Member IEEE. Board member, EuADS, European Association of Data Science.

It is a great honor for us to have an interview with Prof. Murtagh.

Interview

JMAI: *As a data scientist, how did you get into the field of computer science?*

Prof. Murtagh: Well, following my primary, Bachelors, degree studies, in Engineering Science and Mathematics, I had a job as Programmer-Statistician in an educational research centre, in a teacher training higher education institute in Dublin, Ireland. There were quite a few links of colleagues, between it and Boston College in the US. While my primary interest and focus was on methodology, for data analytics, certainly with implementation, and computer science as such, all was closely associated. While my first year's job, full time, as a Programmer-Statistician, my second year there was part time, because I was doing an MSc degree in Computer Science, that was entirely information retrieval, implementation.



Figure 1 Prof. Fionn Murtagh.

JMAI: *What role does computer science play in interdisciplinary research, such as in medical research?*

Prof. Murtagh: May I take this as, in effect, computational science and data science. As a summary expression, Data Science is, firstly the integration of data sources and analytical and related data processing methodologies, and, secondly and quite fundamentally, arising from the convergence of disciplines. Convergence of disciplines can be so very beneficial in practice. That is, beneficial in regard to addressing and solving problems, and also in regard to the cooperation yielded by cross-disciplinarity.

JMAI: *What do you think are the advantages and limitations of AI applied to medical diagnosis and treatment?*

Prof. Murtagh: In general, it may so very well be the case that machine learning will perform excellently, when the situation and context is very well specified. Therefore it

can be, or it is, justified to experimentally assess that in practice.

Given the data that is at issue, it can be considered that data analysis ultimately serves the goal of: data synthesis, with the aim of: actionable decision-making, based on structured and unstructured data. A derived point from that may be, having the best and more relevant and most fundamental data sources.

Jean-Paul Benzécri whose Correspondence Analysis and hierarchical clustering, and so many related domains of semantic analysis and lots, lots more, for inductive reasoning in data analytics, that included a book in 1992, with the French title translated as *In Medicine, Pharmacology, Clinical Physiology—Data Analysis Practice*. In my 2005 book on Correspondence Analysis and Data Coding with Java and R, an extensive Preface by Jean-Paul Benzécri starts with the very great importance for all of analytics and decision-making that is enabled by computers.

Included in that Preface is how extremely important, context is. Here is a little extract from that Preface: “As far as the philosophy of numbers is concerned, the distinction between qualitative and quantitative seems to us to be still not always understood. In brief, it should not be held that: (I) continuous numerical value implies quantitative value; and (II) value with a finite set of modalities implies qualitative value, because at the level of the statistical item (e.g., patient’s dossier) a numerical value (age, or even artery pressure or glycemy) is not generally to be taken with its given precision, but according to its significance. In particular, to compare an observation to another, one must consider not two sets of primary data between which global similarities are not in evidence, but the synthesis of these sets, ending up with a few gradations, or discontinuities, hence ultimately with diagnostics.”

From the context of the analyses to be undertaken, at issue is the data sourcing in many practical aspects. Considering the various analyses carried out, and the inductive reasoning that is pursued, then here is an important perspective on all that I am expressing here. It is a characterization of data analysis in the sense of not just automating the carrying out of the analysis. In debate some years ago, this statement was formulated: “correlation is not causation”. This was to counterpose the claim that determining similarity and best matching, including correlation, could replace analytical reasoning. What this implies is that the pure data properties must be related to the context, with the context encompassing decision-

support and diagnosis, etc., and also all that is to do with, and is related to, the data sources.

JMAI: What is the future of computer science?

Prof. Murtagh: There is a good deal of current descriptions of automated medical support, perhaps using smart phone apps, where as a patient, and as someone suffering with health and physical problems, one will get both advice and decisions made, relating to medical treatment. Such apps will be very much used. There is also some described progress towards having robotic implementation of medical treatment. Contemporary central themes in computer science include Internet of Things, smart cities and smart homes. The latter can have a lot of relevance for medical issues, as well as health and lifestyle issues.

A beneficial and important aspect of Big Data is how ancillary and contextual data sources are to be both associated and integrated. The following article, focused on bias through uses of social media, if that is the data source, and, let it be assumed, elderly suffering individuals are not likely to communicate a lot on social media. This article includes how Big Data can calibrate sampled data: N Keiding and TA Louis, “Perils and potentials of self-selected entry to epidemiological studies and surveys”, *Journal of the Royal Statistical Society, Series A*, 179, Part 2, pp. 319–376, 2016. I am a contributor to the discussion in this article, to stress the importance, for computational science here, and for all analytics, of eminent social scientist, Pierre Bourdieu, whose field and homology concepts relate to calibration and contextualization, that relate to decision-support and policy-making, and, in the medical domain, diagnosis.

JMAI: Can you talk about the ethical and regulatory challenges in AI?

Prof. Murtagh: To begin with here, there may be the stressing of data confidentiality and data rights. I would hope that calibrating or contextualizing data and information, can lead to the benefits of, and importance of, open data sources.

Apart from bias in the data sources, it is known how aggregated data can be used, if required, for individual-related analysis. It such resolution and scale effects in the data that are sourced, computationally and methodologically there can be significant benefits, but some ethical effects also arise in resolution and scale effects. Such is the case

when the individual with his/her characteristics have simply become linked to the mean or the average. A citation in my 2017 book, *Data Science Foundations: Geometry and Topology of Complex Hierarchic Systems and Big Data Analytics*, is this: “Rehabilitation of individuals. The context model is always formulated at the individual level, being opposed therefore to modelling at an aggregate level for which the individuals are only an ‘error term’ of the model.”

Overall at issue here is how ethical as well as methodological issues can arise in scale effects, representation and expression, and particular context effects. Sometimes challenges for methodology developments are good, as long as both innovation is a key objective, and also outcomes, for deployment and practical application and use.

JMAI: *What is the best path for physicians without data science background to conduct computer-related knowledge learning?*

Prof. Murtagh: I feel that there can and should be some degree of collaboration or awareness, and having information about innovative and important developments. For computer-related knowledge, there can well be presentations that will follow from what is recognized as important innovation, and the latter must be published.

For submitted journal articles that are dealing with health and medical topics, my view is that there must be some, at least, commentary on how the work can or will result in practical deployment, or use, or application. So based on recognized achievements and innovation, there ought to be dissemination and access to all such important outcomes.

JMAI: *What are the skills needed to become a data scientist in your opinion?*

Prof. Murtagh: In our teaching courses, and in our engagements, in whatever form they are, important aspects are to communicate, and to observe and understand, good practice and best practice in the application domains, and clear and convincing motivation and justification for general orientation pursued. There is the viewpoint that Data Science integrates, at least in evolution or sequencing, the following: data and information, knowledge and wisdom.

Miscellaneous:

- (I) Themes here have included: integrating data sources; bridging and inherently associating the

data sources and the analytics carried out; being directed towards innovation in methodology and also having, at least potentially, comprehensive evaluation, of what one is undertaking, to be carried out; taking care of bias in data sources, and taking care of purely and only aggregating the data sources, following the benefits of resolution scale effects for computational reasons and even for conceptual purposes. Also important are the outcomes, as well as the processes followed, and the general motivation and justification, in general.

- (II) Extremely important are: the analysis context to be set up; and the underlying and underpinning causation to be discovered, giving rise to the data properties and characteristics that are observed.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Medical Artificial Intelligence* for the series “Meet the Professor”. The article did not undergo external peer review.

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jmai.2018.07.03>). The series “Meet the Professor” was commissioned by the editorial office without any funding or sponsorship. NL reports that she is a full-time employee of AME Publishing Company (publisher of the journal). The author has no other conflicts of interest to declare.

Ethical Statement: The author is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with

the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license).

See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

(Science Editor: Nikki Ling, JMAI, jmai@amegroups.com)

doi: 10.21037/jmai.2018.07.03

Cite this article as: Ling N. Prof. Fionn Murtagh: clinical and all medical practice, and the health context, the major benefits and some challenges of Big Data science. *J Med Artif Intell* 2018;1:6.