Chapter 16
Philosophising Data:
A Critical Reflection On The 'Hidden' Issues

Jackie Campbell
Leeds Beckett University, UK

Victor Chang
Leeds Beckett University, UK

Amin Hosseinian-Far
Leeds Beckett University, UK

ABSTRACT

This chapter aims to critically reflect on the processes, agendas and use of Big Data by presenting existing issues and problems in place and consolidating our points of views presented from different angles. This chapter also describes current practices of handling Big Data, including considerations of smaller scale data analysis and the use of data visualisation to improve business decisions and prediction of market trends. The chapter concludes that alongside any data collection, analysis and visualisation, the 'researcher' should be fully aware of the limitations of the data, by considering the data from different perspectives, angles and lenses. Not only will this add the validation and validity of the data, but it will also provide a 'thinking tool' by which to explore the data. Arguably providing the 'human skill' required in a process apparently destined to be automated by machines and algorithms.

1. INTRODUCTION AND BACKGROUND

"Data is the new oil" (Thorpe, 2012) and Big Data can be used to know us better than we know ourselves (Mayer-Thurman and Cukier, 2013). Based on our Google searches, flu outbreaks can be predicted (known as nowcasting) (Google, 2014; Dugas; 2012); based on our social lifestyle data, the authorities will know when we will commit a crime before it has even been committed (Berk, 2009). These kinds of claims and threats are the current strap lines promoting the use and potential for Big Data and Business Intelligence (BI). Alongside this new research efforts in economics, Kahneman questions our basic skills

DOI: 10.4018/978-1-4666-9840-6.ch016

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

as decision makers, claiming that human decisions are naturally biased and usually flawed (Kahneman, 2013). Kahneman argues (amongst other things) that if recent financial decisions had been based on data, the current financial crisis in the UK, US and now apparently China would never have happened, which have been supported by analyses undertaken by Chang (2014) in his studies.

The power of Big Data combined with the ability to ignore our own intuition could provide an entirely new paradigm to the processes by which we do business, research and make decisions. However, data still comes with its flaws: bias data, data with quality issues, interpreted data, assumed data, mismatched data, subjective data, dated data, ethically questionable data, data from 'chosen' samples, data designed to prove a point, lies and statistics. Data need to be reprocessed, reorganised and restructured before any forms of analysis can be conducted. Contemplation of systems as black boxes and since the number of inflows and outflows through the black boxes is massive, a data flaw would be very challenging to spot. Hence, it would make it more difficult to bring about a new skill and area of concern; one which academics and professionals alike should develop and consider. That would be the ability to know the data and results for what they are, to understand the version of the 'truth' that this data represents, and to ensure the piece of information is not lost in the subsequent use and promotion of the findings.

This chapter aims to critically reflect on the processes and use of Big Data by returning to the issues and considerations of smaller scale data analysis and research. The issues are unpacked with respect to the data findings and the search for the 'truth' or a perspective on the truth. The paper concludes that alongside any data collection and analysis the 'researcher' needs to be fully aware of the limitations of the data, by considering the data from different perspectives, angles, and lenses. Not only will this add the validation and validity of the data, but it will also provide a 'thinking tool' by which to explore the data. Arguably providing the 'human skill' required in a process apparently destined to be automated by machines and algorithms.

2. BIG DATA: THE NEW OIL OR FOOLS' GOLD?

There are some amazing data mining discoveries based on the leverage of Big Data; UPS used predictive analysis by monitoring and replacing specific parts they had saved on repair costs (FieldLogix, 2014). Deadly manhole explosions were predicted in New York (Ehrenberg, 2010). Walmart discovered that just before a hurricane, people in America bought an unusually large number of pop tarts (Hays, 2004). Data usually serves a purpose to prove or disprove hypotheses and theories, and excellent example of which is an investigation into the relationship between mobile phone usage and brain tumours (Frei et al, 2011). The Danish cell phone operators and health care service worked together to provide the data for a study into the relationship between brain tumours and cancer. The results showed no direct correlation. The benefits of adopting Big Data are as follows. First, the process of selecting data or population sampling is not required, as researchers can just take it 'all' and being persuaded to ignore data validity issues based on the 'general trend' provided by such a vast dataset. Second, Big Data processing can highlight the part of the datasets which reflects the core part of the problem. For example, medical data analysis can find the direct correlation between the lifestyle and genetic history with breast cancer.

It is interesting that 70% of companies say they are keeping data, but they do not know what for (Avanade, 2012). In terms of the amount of data being produced, current estimates are that 90% of the world's data as of 2013 was created in 2012 and 13 alone, and there will be 44 times as much by the year 2020 (ScienceDaily, 2013). It is being called 'the new oil', yet clearly companies are aware they

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/philosophising-data/150171

Related Content

Application of Big Data in Healthcare: Opportunities, Challenges and Techniques
Md Rakibul Hoqueand Yukun Bao (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1189-1208).*
www.irma-international.org/chapter/application-of-big-data-in-healthcare/150211

Optimization of a Hybrid Methodology (CRISP-DM)
José Navaand Paula Hernández (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1998-2020).*
www.irma-international.org/chapter/optimization-hybrid-methodology-crisp/73532

Multi-Criteria Decision Making in Manufacturing Systems: Identification of Critical Factors for Establishing a Smart Factory Using ISM and MICMAC Approach
Hande Erdoan Aktanand Ömür Tosun (2019). *Optimizing Big Data Management and Industrial Systems With Intelligent Techniques (pp. 80-107).*
www.irma-international.org/chapter/multi-criteria-decision-making-in-manufacturing-systems/218741

Loan Default Prediction Based on Convolutional Neural Network and LightGBM
Qiliang Zhu, Wenhao Ding, Mingsen Xiang, Mengzhen Huand Ning Zhang (2023). *International Journal of Data Warehousing and Mining (pp. 1-16).*
www.irma-international.org/article/loan-default-prediction-based-on-convolutional-neural-network-and-lightgbm/315823

A Novel Hybrid Algorithm Based on K-Means and Evolutionary Computations for Real Time Clustering
Taha Mansouri, Ahad Zare Ravasanand Mohammad Reza Gholamian (2014). *International Journal of Data Warehousing and Mining (pp. 1-14).*
www.irma-international.org/article/a-novel-hybrid-algorithm-based-on-k-means-and-evolutionary-computations-for-real-time-clustering/116890