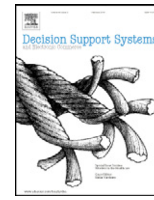
Decision Support Systems 63 (2014) 67–80

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss



Understanding the paradigm shift to computational social science in the presence of big data

Ray M. Chang [a,1], Robert J. Kauffman [b,2], YoungOk Kwon [c,*]

[a] *School of Information Systems, Singapore Management University, 80 Stamford Road, 178902, Singapore*
[b] *School of Information Systems, Singapore Management University, 80 Stamford Road, 178902, Singapore*
[c] *Division of Business Administration, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul 140-742, South Korea*

ARTICLE INFO

Available online 28 August 2013

Keywords:
Analytics
Big data
Computational social science
Data analytics
Interdisciplinary research
Managerial decision-making
Paradigm shift

ABSTRACT

The era of big data has created new opportunities for researchers to achieve high relevance and impact amid changes and transformations in how we study social science phenomena. With the emergence of new data collection technologies, advanced data mining and analytics support, there seems to be fundamental changes that are occurring with the research questions we can ask, and the research methods we can apply. The contexts include social networks and blogs, political discourse, corporate announcements, digital journalism, mobile telephony, home entertainment, online gaming, financial services, online shopping, social advertising, and social commerce. The changing costs of data collection and the new capabilities that researchers have to conduct research that leverages micro-level, meso-level and macro-level data suggest the possibility of a *scientific paradigm shift* toward *computational social science*. The new thinking related to empirical regularities analysis, experimental design, and longitudinal empirical research further suggests that these approaches can be tailored for rapid acquisition of big data sets. This will allow business analysts and researchers to achieve frequent, controlled and meaningful observations of real-world phenomena. We discuss how our philosophy of science should be changing in step with the times, and illustrate our perspective with comparisons between earlier and current research inquiry. We argue against the assertion that theory no longer matters and offer some new research directions.

"To get value from big data, 'quants' or data scientists are becoming analytic innovators who create tremendous business value within an organization, quickly exploring and uncovering game-changing insights from vast volumes of data, as opposed to merely accessing transactional data for operational reporting."
[Randy Lee, Vice President, Aster Center for Data Innovation, Teradata [81]]

"The best way to engage in … data-driven marketing is to gather more and more specific information about customer preferences, run experiments and analyses on the new data, and determine ways of appealing to [casino game] players' interests. We realized that the information in our database, couple with decision science tools that enabled us to predict individual customer's theoretical value to us, would allow us to create marketing interventions that profitably addressed players' unique preferences."
[Gary Loveman, CEO and President of Caesar's Entertainment [70]]

"Each methodology has its strengths and weaknesses. Each approach to data has its strengths and weaknesses. Each theoretical apparatus has its place in scholarship. And one of the biggest challenges in doing interdisciplinary work is being [able] to account for these differences, to know what approach works best for what question, to know what theories speak to what data and can be used in which ways."
[Danah Boyd, Senior Researcher, Microsoft; Research Assistant Professor, New York University [16]]

1. Introduction

With the rapid advances in technology, business interactions involving consumers and suppliers now generate vast amounts of information, which make it much easier to implement the kinds of data analytics that

* Corresponding author. Tel.: +82 2 2077 7907.
E-mail addresses: mrchang@smu.edu.sg (R.M. Chang), rkauffman@smu.edu.sg (R.J. Kauffman), yokwon@sm.ac.kr (Y. Kwon).
[1] Tel.: +65 6808 5227.
[2] Tel.: +65 6828 0929.

0167-9236/$ – see front matter
http://dx.doi.org/10.1016/j.dss.2013.08.008

Gary Loveman, current CEO of Caesar's Entertainment, discussed in a 2003 *Harvard Business Review* article on data mining [70]. Today, this is referred to as the *big data revolution* in the popular press, and viewed as creating challenges and opportunities for business leaders and interdisciplinary researchers. The world's volume of data doubles every eighteen months, for example, and enterprise data are predicted to increase by about 650% over the next few years [45,54]. Today, most firms have more data than they can handle, and managers recognize the potential for value, but the promise of big data still has not been realized, according to the leading academic [35,78] and business media sources [38,79].[3] The potential arises from the use of data to support the way organizations operate and serve their stakeholders. A recent article in *MIT Sloan Management Review* [62] described the use of big data by an Atlanta-based public school, for example. High school graduation rates increased due to better-informed policy decisions that were based on the application of advanced analytics capabilities to student performance data. Likewise, organizations now are embedding analytics in their operations to support data-intensive strategies.

1.1. The emergence of big data

A recent McKinsey report has referred to *big data* as "data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" [71].[4] Such data come from everywhere: pictures and videos, online purchase records, and geolocation information from mobile phones. Big data are not just about sheer volume in terabytes though. Other important aspects have been emphasized in addition to *volume*, including *variety, velocity* and *value* [76]. Big data may be unstructured too: examples are text with social sentiments, audio and video, click streams, and website log files. Such data may flow in real-time streams for analysis, which can enable a firm to maximize business value by supporting business decisions in near to real-time. This new trend in decision support is evocative of what we saw in the 1990s with the emergence of data mining, and the new emphasis on data with a large number of dimensions and much higher complexity (e.g., spatial, multimedia, XML and Internet data). Most of the data sets were "one off" opportunities, rather than data that had become available due to systemic and technological advances.

Considerable challenges are present in the quest to capture the full potential of big data. The shortage of analytics and managerial talent is a significant and pressing problem, for example. *CIO Magazine* [72] and the Corporate Executive Board [79] have reported that it is difficult for firms to find the right people. The U.S. alone is reported to face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million managers and analysts to make effective decisions [71]. (See Fig. 1.)

1.2. Toward computational social science

New perspectives in social science are now tracking the developments in big data. For example, *computational organization science* has broadened researchers' perspectives on social, organizational and policy systems, by adopting computational models that combine social science, computer science, and network science [22]. Other related developments have occurred, including the emergence of computational social science and e-social science [37,63]. *Computational social science* involves interdisciplinary fields that leverage capabilities to collect and analyze data with an unprecedented breadth, depth, and scale. Computational modeling approaches now can predict the behavior of sociotechnical systems, such as human interactions and mobility, that were previously not studied with one-time snapshots of data for very many people [83]. We see a *paradigm shift* in scientific research methods — and one that prompts new directions for research. A useful perspective in this context is attributable to Runkel and McGrath [75], who characterized research methodologies based on three goals: *generality*, *control* and *realism*. They distinguished between their *obtrusiveness* and *unobtrusiveness* for the subjects of research.

With emerging collection techniques for big data sets, there seem to be fundamental changes that are occurring related to research methods, and the ways they can be applied too [58]. In e-business, for example, the contexts include social networks, blogs, mobile telephony, and digital entertainment. The new approaches we see are based on more advantageous costs of data collection, and the new capabilities that researchers have to create research designs that were hard to implement before. The research contexts include human and managerial decision-making, consumer behavior, operational processes, and market interactions. The result is a change in our ability to leverage research methodology to achieve control and precision in measurement, while maintaining realism in application and generality in theory development.

We will discuss the causes of the paradigm shift, and explore what it means for decision support and IS research, and more broadly, for the social sciences. How can we take advantage of big data in our research? What new perspectives are needed? What will the new research practices look like? What kinds of scientific insights and business value can they deliver in comparison to past research? And what research directions are likely to be especially beneficial for the production of new knowledge?

Section 2 reviews traditional methods for research and discusses the key factors that are creating the basis for a paradigm shift. Section 3 describes the new paradigm in the era of big data, and how it relates to decision support, IS and social science research. Section 4 assesses how the research has been changing, through the use of a set of specific comparisons between research that was conducted before and after the emergence of new methods associated with big data. Section 5 offers some new research directions, and section 6 concludes.

2. How are big data supporting a research paradigm shift?

The move to computational social science in the presence of big data involves a Kuhnian *scientific paradigm shift* [60]. We will provide a background on the traditions of research inquiry, and then examine the driving forces for the paradigm shift, and why access to large stores of data is speeding the process.

2.1. Traditions of research inquiry

Churchman [27] characterized research with a set of different *inquiring systems*. They involve methods, procedures and techniques to describe and explain behavior, test hypotheses, assess causality, and

[3] A recent article in MIS Quarterly by Chen et al. [25] identified a total of 560 published articles in three areas in the IS literature, based on keywords such as business intelligence, business analytics and big data. They stated that 243 are big data-related, with 84% appearing only in 2007 or later. Also included is a table indicating that, among the major IS journals, Decision Support Systems has published the most business intelligence and business analytics articles (a total of 41), followed by Communications of the AIS (19), the Journal of Management Information Systems (12), Management Science (10), and Information Systems Research (9). This information suggests that Decision Support Systems is an important outlet for the kinds of research described in this article. The journal ranks first for publishing business intelligence and business analytics articles between 2000 and 2010. Electronic Commerce Research and Applications devoted a special issue of ten articles on "The Role of Business Analytics in E-Commerce in early 2012, and MIS Quarterly also published a "Business Intelligence" special issue in 2012, with five articles.

[4] According to Reuters, the oil company Chevron accumulates 2 terabytes of data a day from operations [48]. The Large Hadron Collider, the largest particle accelerator, generates 40 terabytes per second when it operates — about 15 petabytes annually. Online content providers generate large amounts of data too. Twitter creates more than 7 terabytes, while Facebook creates 10 terabytes, and Google produces 24 terabytes of data every day just from its search operations. Data set size for academic research is growing exponentially too. About 20% of *Science* authors now use data sets exceeding 100 gigabytes and 7% use 1 terabyte-plus data sets, for example [82].