Big Data: New Opportunities and New Challenges

Katina Michael, *University of Wollongong*

Keith W. Miller, *University of Missouri–St. Louis*

We can live with many of the uncertainties of big data for now, with the hope that its benefits will outweigh its harms, but we shouldn't blind ourselves to the possible irreversibility of changes—whether good or bad—to society.

It's no secret that both private enterprise and government seek greater insights into people's behaviors and sentiments. Organizations use various analytical techniques—from crowdsourcing to genetic algorithms to neural networks to sentiment analysis—to study both structured and unstructured forms of data that can aid product and process discovery, productivity, and policy-making. This data is collected from numerous sources including sensor networks, government data holdings, company market lead databases, and public profiles on social networking sites.

Although data mining in one form or another has occurred since people started to maintain records in the modern era, so-called big data brings together not only large amounts of data but also various data types that previously never would have been considered together. These data streams require ever-increasing processing speeds, yet must be stored economically and fed back into business-process life cycles in a timely manner.

NEW OPPORTUNITIES

Since the Internet's introduction, we've been steadily moving from text-based communications to richer data that include images, videos, and interactive maps as well as associated metadata such as geolocation information and time and date stamps. Twenty years ago, ISDN lines couldn't handle much more than basic graphics, but today's high-speed communication networks enable the transmission of storage-intensive data types.

For instance, smartphone users can take high-quality photographs and videos and upload them directly to social networking sites via Wi-Fi and 3G or 4G cellular networks. We've also been steadily increasing the amount of data captured in bidirectional interactions, both people-to-machine and machine-to-machine, by using telematics and telemetry devices in systems of systems. Of even greater importance are e-health networks that allow for data merging and sharing of high-resolution images in the form of patient x-rays, CT scans, and MRIs between stakeholders.

Advances in data storage and mining technologies make it possible to preserve increasing amounts of data generated directly or indirectly by users and analyze it to yield valuable new insights. For example, companies can study consumer purchasing trends to better target marketing. In addition, near-real-time data from mobile phones could provide detailed characteristics about shoppers that help reveal their complex decision-making processes as they walk through malls.[1]

Published by the IEEE Computer Society

Big data can expose people's hidden behavioral patterns and even shed light on their intentions.[2] More precisely, it can bridge the gap between what people want to do and what they actually do as well as how they interact with others and their environment.[3] This information is useful to government agencies as well as private companies to support decision making in areas ranging from law enforcement to social services to homeland security. It's particularly of interest to applied areas of situational awareness and the anticipatory approaches required for near-real-time discovery.

In the scientific domain, secondary uses of patient data could lead to the discovery of cures for a wide range of devastating diseases and the prevention of others.[4] By revealing the genetic origin of illnesses, such as mutations related to cancer, the Human Genome Project, completed in 2003, is one project that's a testament to the promises of big data. Consequently, researchers are now embarking on two major efforts, the Human Brain Project (EU; www.humanbrainproject.eu/vision.html) and the US BRAIN Initiative (www.whitehouse.gov/the-press-office/2013/04/02/fact-sheet-brain-initiative), in a quest to construct a supercomputer simulation of the brain's inner workings, in addition to mapping the activity of about 100 billion neurons in the hope of unlocking answers to Alzheimer's and Parkinson's. Other types of big data can be studied to help solve scientific problems in areas ranging from climatology to geophysics to nanotechnology.

NEW CHALLENGES

While big data can yield extremely useful information, it also presents new challenges with respect to how much data to store, how much this will cost, whether the data will be secure, and how long it must be maintained.

For example, both companies and law enforcement agencies increasingly rely on video data for surveillance and criminal investigation. Closed-circuit television (CCTV) is ubiquitous in many commercial buildings and public spaces. Police cars have cameras to record pursuits and traffic stops, as well as dash-cams for complaint handling. Many agencies are now experimenting with body-worn video cameras to record incidents and gather direct evidence from a crime scene for use in court, obviating the need for eyewitness versions of events.[5] Taser guns also now come equipped with tiny cameras. Because all of these devices can quickly generate a large amount of data, which can be expensive to store and time-consuming to process, operators must decide whether it is more cost-effective to let them run continuously or only capture selective images or scenes.

Big data also presents new ethical challenges. Corporations are using big data to learn more about their workforce, increase productivity, and introduce revolutionary business processes. However, these improvements come at a cost: tracking employees' every move and continuously measuring their performance against industry benchmarks introduces a level of oversight that can quash the human spirit. Such monitoring might be in the best interest of a corporation but is not always in the best interest of the people who make up that corporation.

In addition, as big multimedia datasets become commonplace, the boundaries between public and private space will blur. Emerging online apps will not only enable users to upload video via mobile social networking but will soon incorporate wearable devices in the form of a digital watch or glasses to allow for continuous audiovisual capture. People will essentially *become* a camera.[6] This publicly available data will dwarf that generated by today's CCTV cameras.

Big data will change how we live in both small and large ways.

However, unlike surveillance cameras, smartphones and wearable devices afford no privacy protection to innocent bystanders who are captured in a video at the right place at the wrong time. For example, in the wake of the recent Boston bombings, images of several people photographed at the scene were mistakenly identified as suspects on social media sites.

In fact, one of the major challenges of big data is preserving individual privacy. As we go about our everyday lives, we leave behind digital footprints that, when combined, could denote unique aspects about ourselves that would otherwise go unnoticed, akin to digital DNA.[7] Examples include our use of language and punctuation in blog and forum posts, the clothes we wear in different contexts, and the places we frequent—do we spend our Sunday mornings outdoors playing sports, indoors online, visiting friends, attending religious services, or cruising a bad part of town? Something as innocuous as when and how we use energy in our homes reveals many details about us.[8] Outside our homes, drones could well be used for ad hoc monitoring, spotting unusual changes in land use patterns and feeding data back to operation centers about emergencies.

Big data analytics will draw on aspects of our home, work, and social lives to make assumptions beyond typical "market segmentations" and delve deep into ontological questions such as, "Who are you?" This has metaphysical implications. For example, people will consciously alter their online activity, and will modify their behavior in surveilled spaces, to protect their privacy. Big data will change how we live in both small and large ways. Are we on a trajectory toward an uberveillance society?

Will pervasive and ubiquitous computing converge with underlying network infrastructure providing uber-views using advanced data analytics for convenience, care, and control purposes?[9]

Finally, many big data applications will have unintended and unpredictable results as the data scientist seeks to reveal new trends and patterns that were previously hidden. For example, genetic screening could reveal the likelihood of being predisposed to an incurable disease like Alzheimer's that leads to long-term anxiety about the future, such as being ineligible for life insurance. Likewise, technotherapeutics could assist elderly patients in one way but assert unhealthy controls on others.[10]

We can live with many of these uncertainties for now with the hope that the benefits of big data will outweigh the harms, but we shouldn't blind ourselves to the possible irreversibility of changes—whether good or bad—to society.

IN THIS ISSUE

Members of the IEEE Society for Social Implications of Technology are actively engaged in exploring big data developments and their social and ethical implications. This special issue presents some of the subjects important to SSIT.

The five articles we selected represent perspectives from diverse interests from both operational and nonoperational stakeholders in the big data value chain.

Jess Hemerly provides us with an overview of public policy considerations for a data-driven future. Hemerly, a public policy and government relations analyst at Google, emphasizes the need to tread carefully in the regulation of data flows so as not to adversely impact innovation stemming from the data sciences.

Paul Tallon addresses the need for big data governance by positing that data does have a measurable economic value and that there are technical, reputational, and economic risks to manage. Tallon also presents an important discussion on the cost of big data to organizations.

Jeremy Pitt and his coauthors write on the need to understand big data within the context of collective awareness, as a smart grid infrastructure can have a positive impact on societal transformation toward sustainability. The authors argue that computational management of common-pool resources requires a new approach—institution science.

Marcus Wigan and Roger Clarke are more circumspect about the role of big data in society, pointing to the fact that underlying problems have been in existence since the inception of automated computers. Instead, the authors point to the consequences of big data, including legality, data quality, disparate data meanings, and process quality, as just a few of the bigger issues needing attention.

Finally, we include a case study on the hopes of big data in the health informatics space in an article written by Carolyn McGregor. This article focuses on discovery and the future possibilities that monitoring real-time physiological characteristics of humans may afford to health and well-being.

We need improved powers of discernment, as well as verifiable proof, to better understand big data's opportunities and risks. It will unquestionably become an integral part of our society, used in both commercial and government applications. Our challenge will be to maximize the benefits of big data while minimizing its harms. We hope that this special issue of *Computer* inspires readers to help meet this increasingly important challenge. **C**

References

1. K. Michael and R. Clarke, "Location and Tracking of Mobile Devices: Überveillance Stalks the Streets," *Computer Law & Security Rev.*, vol. 29, 2013, pp. 216-228.
2. R. Abbas, "The Social Implications of Location-Based Services: An Observational Study of Users," *J. Location-Based Services*, vol. 5, nos. 3-4, 2011, pp. 156-181.
3. J. Pitt, "Design Contractualism for Pervasive/Affective Computing," *IEEE Technology and Society Magazine*, vol. 31, no. 4, 2012, pp. 25-28.
4. E. Strickland, "The Gene Machine and Me," *IEEE Spectrum*, Mar. 2013, pp. 26-32.
5. A. Hayes, "Cyborg Cops, Googlers and Connectivism," *IEEE Technology and Society Magazine*, vol. 32, no. 1, 2013, pp. 23-24.
6. S. Mann, "Through the Glass, Lightly," *IEEE Technology and Society Magazine*, vol. 31, no. 3, 2012, pp. 10-14.
7. K. Michael and M.G. Michael, "The Social and Behavioural Implications of Location-Based Services," *J. Location-Based Services*, vol. 5, nos. 3-4, 2011, pp. 121-137.
8. F. Sestini, "Collective Awareness Platforms: Engines for Sustainability and Ethics," *IEEE Technology and Society Magazine*, vol. 31, no. 4, 2012, pp. 54-62.
9. M.G. Michael and K. Michael, "Towards a State of Uberveillance," *IEEE Technology and Society Magazine*, vol. 29, no. 2, 2010, pp. 9-16.
10. M. Gagnon, J.D. Jacob, and A. Guta, "Treatment Adherence Redefined: A Critical Analysis of Technotherapeutics," *Nursing Inquiry*, vol. 20, no. 1, 2013, pp. 60-70.

***Katina Michael** is an associate professor in the School of Information Systems and Technology at the University of Wollongong, New South Wales, Australia. Her research focuses on emerging technologies as well as national security technologies and their corresponding social implications. Michael received a PhD in information and communication technology from the University of Wollongong. She is a senior member of IEEE. Contact her at katina@uow.edu.au.*

***Keith W. Miller** is the Orthwein Endowed Professor for Life-Long Learning in the Sciences at the University of Missouri–St. Louis. His research interests include software testing and computer ethics. Miller received a PhD in computer science from the University of Iowa. He is a member of IEEE and ACM. Contact him at kmill2@uis.edu.*