Big data and analytics in healthcare: Introduction to the special section

Atreyi Kankanhalli[1] · Jungpil Hahn[1] · Sharon Tan[1] · Gordon Gao[2]

Aging populations and lifestyle changes pose increasing pressures on healthcare systems around the world. These trends accompanied by the digitization of health and patient data through advances in information technology, including medical sensors, have led to the generation of large volumes of primary and secondary data in the healthcare domain. The demand for big data is also spurred by a shift to evidence-based medicine as opposed to subjective clinical decisions. While the trove of data offers significant opportunities for improving healthcare delivery, management, and policy making, new information systems and approaches are needed to make effective use of the big data. Indeed, big data has been referred to as data that is too large and complex to be analyzed and managed by traditional computing tools.

The complexity of analyzing big data arises from its three dimensions, i.e., variety, velocity, and volume (Gartner 2011). 'Variety' implies that big data is made up of many different kinds of data, both structured and unstructured, e.g., physician's notes. Healthcare data appears in a variety of formats and representations from diverse sources that include: (1) clinical data in Electronic Health Records, medical images, machine created/sensor data, and genomics, (2) pharmaceutical and R&D data e.g., from clinical trials and journal articles, (3) activity claims and cost data from healthcare providers and insurance companies, and (4) patient behaviour and sentiment data from wearable devices and social media posts, including Twitter feeds, blogs, status updates on Facebook, healthcare communities, and web pages (Raghupathi and Raghupathi 2014). 'Velocity' refers to big data being transmitted and available in real-time, e.g., vital signs, and often arriving in bursts rather than at a constant rate. 'Volume' implies that big data, as per the term itself, is extremely large in size. For instance, a 3D CAT scan typically takes up 1 GB while a single human genome accounts for about 3 GB of data.

Healthcare analytics refers to the systematic use of health data and related business insights developed through applying analytical, e.g. statistical, contextual, quantitative, predictive, cognitive, and other models, to drive fact-based decision making for planning, management, measurement, and learning in healthcare (Cortada et al. 2012). Big data analytics has the ability to go beyond improving profits and cutting down on waste, to be able to predict epidemics, cure diseases, improve the quality of life and reduce preventable deaths (Marr 2015). Among these applications, predictive analytics is believed to be the next revolution both in statistics and medicine around the world (Winters-Miner 2014). Predictive analytics involves using empirical methods (statistical and other) to generate data predictions as well as methods for assessing predictive power (Shmueli and Koppius 2011). It uses a variety of statistical techniques such as modeling, machine learning, and data mining that analyze current and historical data to make predictions about the future. For instance, predictive analytics could be used to identify high-risk patients and provide them treatment to reduce unnecessary hospitalizations or readmissions.

As noted above, big healthcare data analytics presents great potential for transforming healthcare, yet there are manifold challenges ahead. These challenges include not only technological hurdles but also organizational, social, economic, and policy barriers that accompany the application of analytics to big healthcare data. On the technological front, challenges

✉ Atreyi Kankanhalli
   atreyi@comp.nus.edu.sg

1   Department of Information Systems, National University of
    Singapore, Singapore, Singapore

2   Decision, Operations and Information Technologies Department,
    University of Maryland, College Park, MD, USA

include integrating and/or analyzing various forms of healthcare data to address impending problems. In terms of organizational barriers, prior studies have reported how organizations (Yang et al. 2015) and healthcare professionals (Yang et al. 2012) may resist the introduction of technologies that facilitate data capture for analytics but change their work processes. Additionally, social issues, such as privacy concerns, surround the use of new technologies such as wearables that enable personal data analytics. Economics scholars, on the other hand, are concerned about how these technologies and analytics outcomes may impact healthcare costs for various stakeholders. Finally, technologies that allow healthcare data capture and analytics entail policy implications such as changes in privacy and data protection laws. These myriad challenges around the use of technology for big healthcare data analytics present a fertile ground for IS researchers of technical, behavioral/organizational, and economics streams.

Motivated thus, the Centre for Health Informatics at the National University of Singapore organized the 1st and 2nd "International Conference on Big Data and Analytics in Healthcare" (BDAH) in 2013[1] and 2014[2] to provide a forum for researchers, practitioners, and policy makers to share cutting-edge research and practice on this important phenomenon. In the 2nd BDAH conference, both academic/research and practice focus paper submissions were invited. This special section includes original research contributions both from participants of the 2nd BDAH as well as from the wider research community. This section comprises two articles that address specific technical and behavioral challenges, respectively, of healthcare data analytics.

The first paper by Timsina et al. (2016) provides an advanced analytics solution for automation of medical systematic reviews. While systematic reviews are an essential element of modern evidence-based medical practice, the creation and update of these reviews is resource intensive. The authors leveraged advanced analytics techniques for automatically classifying articles for inclusion and exclusion for systematic reviews. Specifically, they used soft-margin polynomial Support Vector Machine (SVM) as a classifier, exploited Unified Medical Language Systems (UMLS) for medical terms extraction, and examined various techniques to resolve the class imbalance issue. Through an empirical study, they demonstrated that the proposed soft-margin polynomial SVM achieves better classification performance than the existing algorithms used in current research, and the performance of the classifier was further improved by using UMLS to identify medical terms in articles and by applying re-sampling methods to resolve the class imbalance issue.

The second paper by Yang and Lee (2016) investigates the antecedents of healthcare information protection intention (HIPI) of healthcare information systems (HIS) users, which is a prerequisite for healthcare information analytics. They proposed and tested a model to explain HIPI, which incorporates constructs from general deterrence theory and protection motivation theory. Their results show that: (1) a clear awareness of the consequences of security threats increases HIS users' understanding on the severity of healthcare information leakage, and thus may decrease abuse of HIS by users, (2) user satisfaction with the security system enhances their self-efficacy that they can handle the medical information leakage issue by themselves, and (3) although HIS users are starting to realize the consequences of healthcare information leakage, they think that they are unlikely to encounter such situations. The findings imply that ongoing security education is needed to increase HIPI of HIS users, and it is important to motivate users to protect healthcare information through their satisfaction with the security system.

References

Cortada, J. W., Gordon, D., & Lenihan, B. (2012). The value of analytics in healthcare: from insights to outcomes. IBM Global Business Services, Life Sciences and Healthcare, Executive Report.

Gartner. (2011). Gartner says solving 'Big Data' challenge involves more than just managing volumes of data. STAMFORD, Con. http://www.gartner.com/newsroom/id/1731916. Accessed 1 Dec 2015.

Marr, B. (2015). How big data is changing healthcare. http://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/print/. Accessed 1 Dec 2015.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems, 2*, 3.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly, 35*(3), 553–572.

Timsina, P., Liu, J., & El-Gayar, O. (2016). Advanced analytics for the automation of medical systematic reviews. *Information Systems Frontiers, 18*, 2. doi:10.1007/s10796-015-9589-7.

Winters-Miner, L. A. (2014). Seven ways predictive analytics can improve healthcare. Elsevier Connect. https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare. Accessed 1 Dec 2015.

Yang, C., & Lee, H. (2016). A study on the antecedents of healthcare information protection intention. *Information Systems Frontiers, 18*, 2. doi:10.1007/s10796-015-9594-x.

_____

Yang, Z., Ng, B. Y., & Kankanhalli, A. (2012). Workarounds in the use of IS in healthcare: a case study of an electronic medication administration system. *International Journal of Human Computer Studies, 70*(1), 43–65.

Yang, Z., Kankanhalli, A., Ng, B. Y., & Lim, J. T.-Y. (2015). Examining the pre-adoption stages of healthcare IT: a case study of vital signs monitoring systems. *Information Management, 52*(4), 454–467.

Atreyi Kankanhalli is Associate Professor of Information Systems at the National University of Singapore (NUS). She serves as Assistant Dean (Research) and Coordinator of the Service Systems Innovation Research Laboratory at her School. She obtained her B. Tech. from IIT Delhi and Ph.D. from NUS. Her research interests are in knowledge management, online communities, and IT-enabled innovation in service sectors such as Government and healthcare. Her work has appeared in premium journals such as MISQ, ISR, JMIS, JASIST, CACM, DSS, IJHCS, and EJIS among others. She has served or is serving as Associate Editor at ISR, MISQ, and Information and Management and Senior Editor at JAIS. She has received the IBM Faculty Award and ACM-SIGMIS 2003 Best Doctoral Dissertation award among other recognitions.

Jungpil Hahn is Associate Professor and Head of the Dept. of Information Systems at the National University of Singapore (NUS). He received his Bachelors and MBA from Yonsei University and his PhD from the University of Minnesota. His current research interests include open innovation, software project management and organizational learning. His research has appeared in such journals as Mgmt. Science, Org. Science, ISR, JMIS, DSS, ACM Trans. on CHI, and, Electronic Markets. Prior to moving to NUS, he was on the faculty at Purdue University and visited at CMU in between. He has been recognized as an Outstanding Reviewer by IEEE Internet Computing. He also received the Excellence in Teaching Award from the PhD Program at the Carlson School at Minnesota. He was recognized as a Krannert Distinguished Teacher at Purdue and was awarded as ISR Best Reviewer in 2009.

Sharon Tan is Assistant Professor of Information Systems at the National University of Singapore (NUS). She received her Ph.D. and M.Sc. from Carnegie Mellon University, and her B.Sc. Honours from NUS. Her research interests are in the areas of Health Informatics and Analytics and Knowledge Management. Her work has been published in leading academic journals such as MISQ and ISR, and the proceedings of the International Conference of Information Systems (ICIS), the Academy of Management (AoM) Conference, and the Annual Symposiums of the American Medical Informatics Association among others. Her papers have won several awards including the Emerald Management Review Citation of Excellence award, the Best Paper Award of the 2009 Pre-ICIS workshop on HCI, and the Best Paper (Runners' Up) Award of the 2002 ICIS. She is serving on the editorial boards of Journal of AIS and Journal of Database Management.

Guodong (Gordon) Gao is Associate Professor in the Decision, Operations and Information Technologies Department at the Robert H. Smith School of Business at the University of Maryland. His research interests include IT's impact on Health care and innovation, and the transparency in service quality. His paper has been published at leading academic journals like Management Science, ISR, MIS Quarterly, Manufacturing & Service Operations Management, and JAMA Internal Medicine. He has presented extensively at various conferences like ICIS, WISE, INFORMS, CIST, AcademyHealth, AshEcon, and iHEA. His research has won several grants and best paper awards, including the National Science Foundation's CAREER Award. He received his B.S. in Electrical Engineering and B.A. in Economics from Tsinghua University, his M.B.A. from the Tsinghua and MIT Sloan Joint Program, and PhD in IS and Economics from the Wharton School of University of Pennsylvania.