Editorial: Machine intelligence for data-driven decisions: data and techniques

Daniel Sun*

Software and Computational Systems,
DATA61, CSIRO Canberra Research Laboratory,
7 London Circuit, Canberra ACT 2601, Australia
Email: daniel.sun@data61.csiro.au
*Corresponding author

Jianquan Liu

System Platform Research Laboratories,
NEC Corporation,
1753, Shimonumabe, Nakahara-ku,
Email: j-liu@ct.jp.nec.com

Jagan Sankaranarayanan

Google Inc.,
1200 Crittenden Lane, Mountain View,
CA, Mountain View, 94043, USA
Email: sjagan@gmail.com

Tatsuhiro Tsuchiya

Graduate School of Information
Science and Technology,
Osaka University,
1-5 Yamadaoka, Suita, 565-0871, Japan
Email: t-tutiya@ist.osaka-u.ac.jp

Biographical notes: Daniel Sun is a Research Scientist in Data61, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia. He is also a Conjoint Lecturer in the School of Computer Science and Engineering, the University of New South Wales, Australia, and Visiting Researcher at the Shanghai Jiao Tong University. He received his PhD in Information Science from the Japan Advanced Institute of Science and Technology (JAIST) in 2008. From 2008 to 2012, he was an Assistant Research Manager in the NEC Central Laboratories in Japan. From 2013 to 2016, he was a researcher in National ICT Australia (NICTA). His research interests inclued data mining, big data, dependable computing, real-time systems and distributed systems. He is a member of IEEE, the ACM, and SIAM.

Jianquan Liu received his BE degree from the Shantou University, China, ME and PhD degrees from the University of Tsukuba, Japan, in 2005, 2009, and 2012, respectively. He was a Development Engineer in Tencent Inc. from 2005 to 2006, and was a Visiting Research Assistant at the Chinese University of Hong Kong in 2010. He joined NEC Corporation in 2012, and is currently a Senior Researcher and an Assistant Manager at the System Platform Research Laboratories. He is also an Adjunct Assistant Professor at Hosei University, Japan. His research interests include multimedia databases, data mining, information retrieval, cloud computing, and social network analysis. Currently, he is/was serving as the PC Co-chair of IEEE conferences (ICSC2018, ISM2017, ICSC2017, ICRC2017, and BigMM2016), and the Workshop Co-chair of ICSC2016. He is a member of ACM and the Database Society of Japan (DBSJ).

Jagan Sankaranarayanan is with the Data Infrastructure team at Google where he works on problems related to large-scale analytics. In his previous role, he was the Head of Data Management at the NEC Labs America. He received his Doctorate in Computer Science from the University of Maryland in 2008. He has published close to 50 papers in the areas of databases, GIS and computer graphics, which have been cited more than 2000 times. He is the recipient of best paper awards at ACM SIGMOD 2008, ACM SIGSPATIAL GIS 2008, Computers &

Graphics Journal 2007, a best paper nomination at the ICDE 2009 Conference and an 'Excellent Invention Award of 2014' by NEC. His dissertation was nominated for the ACM Doctoral Dissertation and Jim Gray Dissertation Awards by the Computer Science Department of the University of Maryland. He is also in the editorial board of *Distributed and Parallel Databases Journal*.

Tatsuhiro Tsuchiya received his ME and PhD degrees from the Osaka University, Japan in 1995 and 1998, respectively. Currently, he is a Professor of the Graduate School of Information Science and Technology at Osaka University. His research interests are in the areas of dependable systems, model checking, and software testing. He served as a Program Co-chair 16th Workshop on Dependable Parallel, Distributed and Network-Centric Systems (DPDNS 2011) and 21st IEEE Pacific Rim International Symposium on Dependable Computing (PRDC'15).

1 Background

Data is the carrier of information of everything in the universe, and the interface between the physical world and the virtual world inside networks and computers. Although people may not be aware of data when people are seeing, hearing, and even thinking, when we use machines to create intelligence, data is only object we are manipulating. Today, it has been well-known that vision, image, voice, and neuro and brain activities can all be represented by data in different forms, as the advance of artificial intelligence is continuously generating a wave and another of life changing. This will go on to the future inevitably, promisingly, and excitingly.

2 Our view

The purpose of analysing data is to make decisions, which should be better than made by human. Between data and decisions, the techniques born from broad mathematics become the bridges, such as data mining, machine learning, statistics, semantics and logic. However, there is no magic, since simply feeding data to a particular technique does not usually make better decisions than human does. We need to carefully study data, select correct techniques, and tune techniques to produce usefulness, and more importantly understand the unique structures and features of concrete problems in practice. That is, human intelligence creates machine intelligence. Hence, it is essential to learn from case by case for improving our art of generating machines intelligence for decision-making.

3 This issue

In this special issue, we carefully selected and accepted nine papers from 31 submissions. In the accepted papers, typical data such as text, time series, images, and social media are studied, and the analysis techniques include data mining, machine learning, and time series analysis, and semantics. We selected these papers in order to clearly show and support our view on data-driven intelligence. For example, in 'Sightseeing value estimation by analysing geosocial images', composed by Yizhu Shen, Min Ge, Chenyi Zhuang and Qiang Ma, the success of recommending points of interests (POI) makes use of geo-tagged images and collar harmony in the images to produce better sight-seeing values; in 'Document stream classification based on transfer learning using latent topics', from Masato Shirai, Jianquan Liu, Takao Miura and Yi-cheng Chen, the proposed transfer learning framework successfully constructs appropriate intermediate domain to connect source and target domains by latent topic and updating dependence on the change of features of document stream, thus it can classify documents in high accuracy.