

Data science in Asia (for PAKDD 2016)

James Bailey¹ · Latifur Khan² · Takashi Washio³

Published online: 23 November 2016
© Springer International Publishing Switzerland 2016

This special issue on Data Science in Asia consists of six selected papers from the 20th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2016) held in Auckland, New Zealand, in April 2016.

The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) is a leading international conference in the areas of knowledge discovery and data mining (KDD). It provides an international forum for researchers and industry practitioners to share their new ideas, original research results and practical development experiences from all KDD-related areas, including data mining, data warehousing, machine learning, artificial intelligence, databases, statistics, knowledge engineering, visualization, decision-making systems and emerging applications.

PAKDD 2016 received 307 submissions. Of these, 91 papers were selected for the conference program and proceedings, resulting in an acceptance rate below 30%. The six papers presented in this special issue have been selected from amongst the best papers accepted in PAKDD 2016, in terms of their technical contributions and paper presentation. Authors of the papers have been invited to extend and improve their conference versions significantly, based on feedback from the conference, and comments and suggestions from additional reviewers. The quality of these papers is a testament to the rapid growth of data science research in the Asia-Pacific region.

✉ James Bailey
baileyj@unimelb.edu.au

¹ Department of Computing and Information Systems, The University of Melbourne, Melbourne, VIC 3010, Australia

² The University of Texas at Dallas, 800 W Campbell Rd., Richardson, TX 75083-0688, USA

³ The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567, Japan

The paper “Intelligent medical image grouping through interactive learning” by Xuan Guo, Qi Yu, Rui Li, Cecilia Ovesdotter Alm, Cara Calvelli, Pengcheng Shi and Anne Haake proposes an approach that can be used for incorporating expert domain knowledge into a medical image grouping task. The setting is in the area of dermatological images, and they demonstrate that being able to integrate this type of knowledge can significantly improve the groupings that are discovered, compared to a non-expert-driven setting.

The paper “Classifying spatial trajectories using representation learning” by Yuki Endo, Hiroyuki Toda, Kyosuke Nishida and Jotaro Ikeda addresses the problem of feature extraction, as a pre-processing step to enable construction of a trajectory classification model. They tackle this using a deep neural network that converts the trajectory data into an image representation and show this can be an effective basis for supervised learning.

The paper “Anytime algorithm for frequent pattern outlier detection” by Arnaud Giacometti and Arnaud Soulet develops an algorithm for outlier detection based on the use of frequent patterns. An important characteristic of the approach is that the algorithm is anytime, meaning that the complete set of patterns is not needed for computing a statistic known as the frequent pattern outlier factor. The anytime property is particularly useful for deployment of outlier detection on large datasets.

The paper “A greedy algorithm to construct sparse graph by using ranked dictionary” by Shuchu Han and Hong Qin proposes an algorithm for processing a type of sparse graph known as the L_1 graph. The algorithm is based on a greedy approach in order to reduce the computation time for large graphs, whilst at the same time being able to deliver outputs that can be used for good clustering performance. The paper also makes interesting connections to the area of sparse subspace clustering.

The paper “A fast and complete algorithm for enumerating pseudo-cliques in large graphs” by Hongjie Zhai, Makoto Haraguchi, Yoshiaki Okubo and Etsuji Tomita tackles the algorithmic challenge of enumerating pseudo-cliques in graphs. It proposes a method that is both sensitive to clique size and at the same time is able to process very large networks.

The paper “Grade prediction with models specific to students and courses” by Agoritsa Polyzou and George Karypis uses techniques from matrix factorization and sparse linear models, to address the problem of estimating future grades

of students, based on their historical data. A distinctive aspect of the work is that they show how it can make effective predictions that are specific to courses, or student-course pairs.

Before closing this editorial, we want to thank all authors for submitting extended papers to this special issue and all reviewers for providing invaluable suggestions for improving the papers. We also want to express our gratitude to the JDSA Editor-in-Chief Longbing Cao for his help in making this special issue possible.

James Bailey, Latifur Khan and Takashi Washio