EDITORIAL

Introduction to the special issue on data science in Asia (with PAKDD'2020)

Hady W. Lauw[1] · Raymond Chi-Wing Wong[2] · Alexandros Ntoulas[3]

It is our great pleasure to introduce the special issue on Data Science in Asia. This special issue consists of extended papers from the 2020 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2020), which was held from 11 to 14 May 2020. 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). PAKDD 2020 received 628 submissions. Of these, 135 papers were accepted for the conference program and proceedings, resulting in an acceptance rate of 21.5%. The authors of the five papers in this issue have been invited to extend and improve their conference versions significantly, based on feedback from the conference reviewers. They further went through the rigorous and thorough reviewing and revision process of JDSA.

The paper "C-MemMAP: Clustering-driven Compact, Adaptable, and Generalizable Meta-LSTM Models for Memory Access Prediction" by Pengmiao Zhang, Ajitesh Srivastava, Ta-Yang Wang, Cesar A. F. De Rose, Rajgopal Kannan, and Viktor K. Prasanna proposes clustering-driven compact LSTM models that can predict the next memory access with high accuracy. This has the potential to improve the performance of data prefetching for memory-bound applications.

The paper "Probabilistic Exact Adaptive Random Forest for Recurrent Concepts in Data Streams" by Ocean Wu, Yun Sing Koh, Gillian Dobbie, and Thomas Lacombe adapts random forests to the dynamic nature of data streams, by introducing a novel framework called PEARL, which uses both an exact technique and a probabilistic graphical model with Lossy Counting.

The paper "Graph Sparsification with Graph Convolutional Networks" by Jiayu Li, Tianyun Zhang, Hao Tian, Shengmin Jin, Makan Fardad, and Reza Zafarani proposes a neural network graph sparsifier that sparsifies a graph by pruning some edges. This has a positive effect on node classification on large graphs.

The paper "Exploiting Feature Fusion and Long-term Context Dependencies for Simultaneous ECG Heartbeat Segmentation and Classification" by Xi Qiu, Shen Liang, Lingxiao Meng, Yanchun Zhang, and Fan Liu proposes a novel deep learning model that simultaneously conducts heartbeat segmentation and classification. This is in the context of classifying ECG heartbeats for the purpose of arrhythmia detection.

The paper "Learning Low-dimensional Manifolds under the L0-norm Constraint for Unsupervised Outlier Detection" by Yoshinao Ishii, Satoshi Koide, and Keiichiro Hayakawa proposes a novel reconstruction-based method to learn low-dimensional manifolds that capture the nonlinear features of the data, and uses a novel optimization algorithm that can decompose the data under the L0-norm constraints on the error matrix. This shows greater effectiveness at unsupervised outlier detection without the need for clean data.

Our thanks go to the authors for submitting their excellent extended papers to this special issue, to the reviewers for providing useful suggestions and advices for improving the papers, as well as to the JDSA Editor-in-Chief Longbing Cao for making this special issue possible.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

✉ Hady W. Lauw
hadywlauw@smu.edu.sg

Raymond Chi-Wing Wong
raywong@ust.hk

Alexandros Ntoulas
antoulas@di.uoa.gr

[1] School of Computing and Information Systems, Singapore Management University, Singapore, Singapore

[2] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China

[3] Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece