



Introduction to the special issue on Data Science in Asia (with PAKDD'2017)

Longbing Cao¹ · Jae-Gil Lee² · Xuemin Lin¹

Published online: 9 November 2018
© Springer Nature Switzerland AG 2018

It is our great pleasure to introduce this collection of papers devoted to the special issue on Data Science in Asia. The earlier versions of these extended papers were presented at 2017 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'2017), which was held in Jeju, Korea, from 23 to 26 May 2017. In PAKDD'2017, we accepted 45 papers as long-presentation papers. After the conference, we invited the authors of the long-presentation papers to submit an extended version of their papers to this special issue with JDSA. The reviewers of JDSA conducted thorough reviews for the extended papers, and the authors revised their papers based on the reviews. Finally, we accepted seven papers for this special issue.

These seven papers are roughly classified into four categories, as below.

Deep learning in data mining

- Deep learning for detecting inappropriate content in text [1]

In this paper, the authors proposed a novel deep learning architecture called “Convolutional Bi-Directional LSTM (C-BiLSTM)” in order to detect inappropriate contents such as hurling abuses.

Scalability and parallelism in data mining

- Parallel edge-based visual assessment of cluster tendency on GPU [2]

In this paper, the authors improved the visual assessment of (cluster) tendency (VAT) algorithm such that it becomes suitable to GPU parallel processing.

- Scalable Twitter user clustering approach boosted by Personalized PageRank [3]

In this paper, the authors proposed a scalable approach for Twitter user clustering based on content and graph features with topical relevance and influence ranking by Personalized PageRank.

Parameter tuning in data mining

- Automated parameter tuning in one-class support vector machine: an application for damage detection [4]

In this paper, the authors developed an automated parameter tuning of the one-class SVM for damage detection applications.

- Stable Bayesian optimization [5]

In this paper, the authors proposed a stable Bayesian optimization technique which can be commonly used for hyper-parameter tuning in machine learning.

Similarity matching in data mining

- Drug prescription support in dental clinics through drug corpus mining [6]

In this paper, the authors suggested an approach to obtain the similarity ratio between the drug that the dentist is going to prescribe and the drug that the patient is currently taking.

- Inferring variable labels using outlines of data in Data Jackets by considering similarity and co-occurrence [7]

In this paper, the authors focused on similarity among the outlines of Data Jackets (DJs) and presented two models for inferring variable labels (VLs) based on the similarity and the co-occurrence of the VLs.

✉ Jae-Gil Lee
jaegil@kaist.ac.kr

Longbing Cao
LongBing.Cao@uts.edu.au

¹ Sydney, Australia

² Seoul, South Korea

References

1. Yenala, H., Jhanwar, A., Chinnakotla, M.K., Goyal, J.: Deep learning for detecting inappropriate content in text. *Int. J. Data. Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-017-0088-4>

2. Meng, T., Yuan, B.: Parallel edge-based visual assessment of cluster tendency on GPU. *Int. J. Data. Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-018-0100-7>
3. Naik, A., Maeda, H., Kanojia, V., Fujita, S.: Scalable Twitter user clustering approach boosted by Personalized PageRank. *Int. J. Data. Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-017-0089-3>
4. Anaissi, A., Khoa, N.L.D., Wang, Y.: Automated parameter tuning in one-class support vector machine: an application for damage detection. *Int. J. Data. Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-018-0151-9>
5. Dai Nguyen, T., Gupta, S., Rana, S., Venkatesh, S.: Stable Bayesian optimization. *Int. J. Data. Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-018-0119-9>
6. Goh, W.P., Tao, X., Zhang, J., Yong, J., Zhang, W., Xie, H.: Drug prescription support in dental clinics through drug corpus mining. *Int. J. Data. Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-018-0149-3>
7. Hayashi, T., Ohsawa, Y.: Inferring variable labels using outlines of data in Data Jackets by considering similarity and co-occurrence. *Int. J. Data. Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-018-0152-8>