



Introduction to the special issue on Big Data, IoT Streams and Heterogeneous Source Mining

Jesse Read¹ · Albert Bifet² · Wei Fan³ · Qiang Yang⁴ · Philip Yu⁵

Published online: 2 October 2019
© Springer Nature Switzerland AG 2019

Recent years have witnessed a dramatic increase in our ability to collect data from sensors and devices, across different formats, from connected applications and enormous dynamic networks including social networks, as well as many other sources. This data flood has outpaced our capability to process, analyze, store and understand these datasets using traditional methods. In all these areas, we are facing significant challenges in leveraging the vast amount of data, and dealing with its speed of arrival and its heterogeneous and evolving nature. This includes challenges in system capabilities, storage and processing, algorithmic design and business models. Approaches for dealing with data in a streaming fashion are thus becoming increasingly relevant in many tasks, including mining and analysis, data representation and visualization, incremental learning and anomaly detection.

We are pleased to introduce this collection of papers in the special issue on Big Data, IoT Streams and Heterogeneous Source Mining. Earlier versions of these extended papers were presented at BigMine 17—a KDD 2017 workshop, which was held in Halifax, Canada, on 14 August 2017. After the workshop, 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 a thorough review process for the extended papers. Finally, we accepted five papers for this special issue. These papers involve a number of applications from evolving news streams to event detection in time-series data, covering a range of techniques from clustering to deep learning. These five papers are roughly classified into two categories, as follows.

Evolving Data Streams

- BFSPMiner: An Effective and Efficient Batch-Free Algorithm for Mining Sequential Patterns over Data Streams [1] presents a method for efficiently mining patterns in streams without being constrained to the traditional batch-based processing.
- Analyzing Evolving Stories in News Articles [2] detects the origin of an event in news streams and can segment the timeline into disjoint groups of coherent news articles.

Process Models and Anomaly Detection

- AdaHash: Hashing-Based Scalable, Adaptive Hierarchical Clustering of Streaming Data on Map-Reduce Frameworks [3] proposes a hierarchical clustering method on Map-Reduce which responds rapidly to new data.
- Online Conformance Checking: Relating Event Streams to Process Models using Prefix-Alignments [4] provides a novel approach to incrementally compute prefix-alignments, thus paving the way for real-time online conformance checking.
- dLSTM: a new approach for anomaly detection using deep learning with delayed prediction [5] develops a method based on several LSTMs that can detect anomalies in time-series data and dynamically select the most appropriate model for prediction.

We believe that these papers offer an excellent snapshot of the variety of methods and applications currently studied

✉ Jesse Read
jesse.read@polytechnique.edu

Albert Bifet
abifet@waikato.ac.nz

Wei Fan
davidwfan@tencent.com

Qiang Yang
qyang@cse.ust.hk

Philip Yu
psyu@cs.uic.edu

¹ École Polytechnique, Paris, France
² University of Waikato, Hamilton, New Zealand
³ Tencent Medical AI Lab, Palo Alto, USA
⁴ HKUST Clearwater Bay, Kowloon, Hong Kong
⁵ University of Illinois at Chicago, Chicago, USA

in the data stream literature, and are relevant to researchers as well as practitioners in industry who are looking to meet the challenges posed in this modern area of ubiquitous data streams.

References

1. Hassani, M., Töws, D., Cuzzocrea, A., et al.: BFSPMiner: an effective and efficient batch-free algorithm for mining sequential patterns over data streams. *Int. J. Data Sci. Anal.* (2017). <https://doi.org/10.1007/s41060-017-0084-8>
2. Camacho Barranco, R., Boedihardjo, A.P., Hossain, M.S.: Analyzing evolving stories in news articles. *Int. J. Data Sci. Anal.* (2017). <https://doi.org/10.1007/s41060-017-0091-9>
3. Teffer, D., Srinivasan, R., Ghosh, J.: Adahash: hashing-based scalable, adaptive hierarchical clustering of streaming data on mapreduce frameworks. *Int. J. Data Sci. Anal.* (2018). <https://doi.org/10.1007/s41060-018-0145-7>
4. van Zelst, S.J., Bolt, A., Hassani, M., et al.: Online conformance checking: relating event streams to process models using prefix-alignments. *Int. J. Data Sci. Anal.* (2017). <https://doi.org/10.1007/s41060-017-0078-6>
5. Maya, S., Ueno, K., Nishikawa, T.: dLSTM: a new approach for anomaly detection using deep learning with delayed prediction. *Int. J. Data Sci. Anal.* **8**, 137 (2019). <https://doi.org/10.1007/s41060-019-00186-0>

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