



Special issue on Data-driven Science

Tanu Malik¹

Accepted: 4 March 2021 / Published online: 13 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Big Data issues and challenges affect any scientific enterprise. Traditionally, most enterprises have addressed the big data issues of volume, velocity, and variety in isolation; with the advent of low-cost sensors and edge computing, the scientific enterprise must increasingly address these issues in combination. This special issue on data-driven science explores the issues of data volume, data variety, and data velocity through a single lens. It puts forward novel dynamic and analytic methods for some classical problems in scientific data management such as dependency estimation, constructing wavelet synopses, and schema evolution. The issue advances the state of statistical and scientific database management by providing data-driven guidance for knowledge discovery and analytics.

We present three papers on the topic of data-driven science. In “A Framework for Dependency Estimation in Heterogeneous Data Streams”, Fouché et al. address the problem that an incoming data stream may consist of a variety of data such as numerical, ordinal or categorical. The authors develop a dependency estimation framework for heterogeneous data streams based on Monte Carlo methods. This framework detects patterns and dependencies in heterogeneous data streams.

In “Workload-Aware Wavelet Synopses for Sliding Window Aggregates”, Mytilinis et al. address volume and velocity challenges in a data stream by summarizing it with wavelet-based methods. They show their higher-accuracy workload-aware methods to be scalable for memory-constrained edge devices.

Finally, in “CHiSEL: A User-oriented Framework for Simplifying Database Evolution”, Schuler et al. consider the problem of managing a variety of data in relational databases which necessitates a database administrator to continually undertake schema evolution. They propose a higher level algebra to assist such an administrator and show the efficacy of this algebra using benchmarks.

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

✉ Tanu Malik
tanu.malik@depaul.edu

¹ School of Computing, DePaul University, Chicago, IL 60604, USA