

Special Issue on Machine Learning Approaches in Big Data Visualization

Nikos Bikakis , Athena Research Center, 15125, Marousi, Greece

Panos K. Chrysanthis , University of Pittsburgh, Pittsburgh, PA, 15260-9161, USA

George Papastefanatos , Athena Research Center, 15125, Marousi, Greece

Tobias Schreck , Graz University of Technology, 8042, Graz, Austria

Welcome to the *IEEE Computer Graphics and Applications* (IEEE CG&A) Special Issue on Machine Learning Approaches in Big Data Visualization. Data visualization is now one of the cornerstones of data science, turning the abundance of big data being produced through modern systems into actionable knowledge. Data visualization in the big data era raises the need to co-design and more closely align the underlying data management systems with the user-oriented techniques that state-of-the-art visualization systems now offer. In addition, the tight integration of suitable machine learning approaches with data visualization and their control by users-in-the-loop promises to enhance scalability, effectiveness, and adaptivity of the interactive visual data analysis process.

This special issue attracted and publishes research work on multidisciplinary research areas from the human-computer interaction, computer graphics, and data management communities. In addition to the normal submissions, this special issue considered some of the best papers from the 4th International Workshop on Big Data Visual Exploration and Analytics, held in conjunction with the 24th International Conference on Extending Database Technology & 24th International Conference on Database Theory (EDBT/ICDT 2021).

In response to the call for papers, 11 submissions on different applications of visual data analysis techniques were received. From these, five works were accepted after a two-stage review process supported by a reviewer board of internationally renowned experts in the field.

The paper "DLA-VPS: Deep-Learning-Assisted Visual Parameter Space Analysis of Cosmological Simulations," by Sun and Wang, targets large-scale cosmological simulations using deep learning techniques. It presents an interactive visualization tool, which enables the

exploration of the parameters space of large-scale cosmological data. The system generates simulations based on GAN-based surrogate models aiming at reducing the time of reconstructing simulation outputs, and allows users to interactively discover and select from alternative simulations.

In "Giga Graph Cities: Their Buckets, Buildings, Waves, and Fragments," Abello *et al.* introduce a novel visual representation for very large graph data, relying on a combination of graph analysis, data reduction, and scalable visual representations following a city metaphor. The aggregation is based on the graph degree distribution, and the technique provides fine-grained representations of buildings by partitions of the degree distribution. The work stands as an innovative visual graph aggregation with detailed interactive visual representations.

In "Narrative *In Situ* Visual Analysis for Large-Scale Ocean Eddy Evolution," Han *et al.* present a visual analytics approach to explore large simulation data on eddies, important flow phenomena in oceans. Their study is of great importance for the understanding of earth environment developments. The approach adequately combines several complementary views, including maps, vector fields, descriptors, and timeline, and supports detail on demand analysis. The approach was developed along requirements obtained from domain users and demonstrated by a use case discussion.

In "Supporting Visual Exploration of Iterative Job Scheduling," Andrienko *et al.* contribute an approach for visual analysis of job schedules, as required in many important application domains such as industrial production or transportation. The authors structure the scheduling design space along jobs, machines, and steps. Based on a task analysis, they introduce a set of effective scheduling views, helpful for exploring and understanding scheduling alternatives (e.g., in terms of schedule evolution, machine utilization, hotspots, and delays). The authors obtain user feedback from scheduling algorithm developers and domain experts, indicating the approach is very useful to their tasks.

DISCLAIMER

Peer review and decision-making for "Supporting Visual Exploration of Iterative Job Scheduling," by Gennady Andrienko, Natalia Andrienko, Jose Manuel Cordero Garcia, Dirk Hecker, and George A. Vouros were handled by an independent, nonconflicted *IEEE Computer Graphics and Applications* editorial board member; the guest editors had no role in determining the outcome of this article.

In "Visualizing Internal Sustainability Efforts in Big Companies," Ceccarini *et al.* compute and visualize corporate sustainability measures, using open corporation review texts and an appropriate machine learning model. The results are visualized and matched against user preferences of these measures. This is a nice example of providing a widely usable visual interface to understand publicly available data, fostering comparison and transparency, in this application on corporate structures. As a crowd-based user study shows, stakeholders can learn the meaning of sustainability indices, compare them with their preferences, and explore their preferences and corporation assessments.

We thank all authors who have submitted their work to this special issue, and the authors of the accepted papers for their timely revisions. We sincerely thank Pak Chung Wong and Torsten Möller of IEEE CG&A for hosting this special issue and supporting us during all stages of the call and review process. We are very grateful to all expert reviewers for their time and valuable constructive comments, which helped the authors improve their work and enabled us to select the contributions for this special issue.

NIKOS BIKAKIS is a research associate at ATHENA Research Center, 15125, Marousi, Greece. His research interests include

human-data interaction, data structures and indexing, scalable visual analytics, hard computational problems, and personalization systems. Bikakis received his Ph.D. degree in computer science from the Technical University of Athens. Contact him at bikakis@athenarc.gr.

PANOS K. CHRYSANTHIS is a professor of computer science and a founder and director of the Advanced Data Management Technologies Laboratory at the University of Pittsburgh, Pittsburgh, PA, 15260-9161, USA. He is also an adjunct professor at Carnegie Mellon University, Pittsburgh, and the University of Cyprus, Nicosia, Cyprus. His research interests include data management (big data, databases, data visualization and analytics, and data streams, mobile, and sensor networks/IoT). Chrysanthis received his Ph.D. degree in computer and information sciences from the University of Massachusetts Amherst. He is a senior member and distinguished contributor of the IEEE. Contact him at panos@cs.pitt.edu.

GEORGE PAPASTEFANATOS is a principal researcher at Athena Research Center, 15125, Athens, Greece, working in the field of big data management and analytics, in problems related to scalable visual analytics and interactive exploration, data integration, web data management, and data evolution. Papastefanatos received his Ph.D. degree in computer science from the Department of Electrical and Computer Engineering of the National Technical University of Athens. He is a senior member of IEEE. Contact him at gpapas@athenarc.gr.

TOBIAS SCHRECK is a professor and head of the Institute of Computer Graphics and Knowledge Visualization at Graz University of Technology, 8042, Graz, Austria. His research interests include visual analytics, information visualization, and applied 3-D object retrieval. Schreck received his Ph.D. degree in computer science from the University of Konstanz. He is a member of IEEE. He is the corresponding author of this article. Contact him at tobias.schreck@cgvtugraz.at.