Machine Learning Approaches in Big Data Visualization

Nikos Bikakis, ATHENA Research Center, Greece

Panos K. Chrysanthis, University of Pittsburgh, USA

George Papastefanatos, ATHENA Research Center, Greece

Tobias Schreck Graz University of Technology, Austria

Welcome to the 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 to invite 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, eleven submissions on different applications of visual data analysis techniques were received. From these, four 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 Ko-Chih Wang and Cheng Sun 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*, Haoyang Zhang and co-authors 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*, Xiaoyang Han and co-authors 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 is developed along requirements obtained from domain users and demonstrated by a use case discussion.

In *Supporting Visual Exploration of Iterative Job Scheduling*, Gennady Andrienko and co-authors 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.

In *Visualizing Internal Sustainability Efforts in Big Companies*, Chiara Ceccarini and co-authors 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 on 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 accepted paper authors for their timely revisions. We sincerely thank Pak Chung Wong and Torsten Moeller 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.

Guest Editors

Nikos Bikakis

Dr. Nikos Bikakis is a research associate at ATHENA Research Center, Greece. He obtained a PhD in Computer Science from the Technical University of Athens. His research interests lie within the areas of Human-Data Interaction, Data Structures and Indexing, Scalable Visual Analytics, Hard Computational Problems, and Personalization Systems.

Panos K. Chrysanthis

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. He is also an adjunct Professor at Carnegie-Mellon University and at the University of Cyprus. He earned his PhD in Computer and Information Sciences in 1991 from the University of Massachusetts at Amherst. His research interests lie within the areas of data management (Big Data, Databases, Data Visualization & Analytics, and Data Streams, Mobile & Sensor networks/IoT).

George Papastefanatos

Dr. George Papastefanatos is a principal researcher at ATHENA Research Center, 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. George obtained his Diploma and his PhD in Computer Science from the Department of Electrical and Computer Engineering of the National Technical University of Athens.

Tobias Schreck

Tobias Schreck is a Professor and head of the Institute of Computer Graphics and Knowledge Visualization at Graz University of Technology, Austria. Before that, he was an Assistant Professor with University of Konstanz and a Postdoc Fellow with Technische Universitaet Darmstadt, both in Germany. He obtained a PhD in Computer Science in 2006 from the University of Konstanz. Tobias Schreck works in the areas of Visual Analytics, Information Visualization, and Applied 3D Object Retrieval.