

4th International Conference on

BIG DATA ANALYSIS AND DATA MINING

September 07-08, 2017 | Paris, France

A novel linear-time and memory saving approach for visual exploratory data analysis: Data embedding and graph visualization

Witold Dzwiniel

AGH University of Science and Technology, Poland

Data embedding (DE) and graph visualization (GV) methods are very congruent tools used in exploratory data analysis for visualization of complex data such as high-dimensional data and complex networks, respectively. However, high computational complexity and memory loads of existing DE and GV algorithms (based on t-SNE concept from one hand, and force-directed methods from the other), considerably hinders visualization of truly large and big data consisting of as many as $M \sim 10^6+$ data objects and $N \sim 10^3+$ dimensions. In this paper, we demonstrate the high computational efficiency and robustness of our approach to data embedding and interactive data visualization. We show that by employing only a small fraction of distances between data objects, one can obtain very satisfactory reconstruction of the topology of N-D data in 2D in a linear-time $O(M)$. The IVHD (Interactive Visualization of High-Dimensional Data) method quickly and properly reconstructs the N-D data topology in a fraction of computational time required for the state-of-art DE methods such as bh-SNE and all its clones. Our method can be used for both metric and non-metric (e.g. large graphs) data visualization. Moreover, we demonstrate that even poor approximations of the nearest neighbor (NN) graph, representing high-dimensional data, can yield acceptable data embedding. Furthermore, some incorrectness in the nearest neighbor list can often be useful to improve the quality of data visualization. This robustness of IVHD, together with its high memory and time efficiencies, meets perfectly the requirements of big and distributed data visualization, when finding the accurate nearest neighbor list represents a great computational challenge.

Biography

Witold Dzwiniel holds Full Professor position at AGH University of Science and Technology, Department of Computer Science in Krakow. His research activities focus on "Computer modeling and simulation methods employing discrete particles". Simultaneously, he is doing research in interactive visualization of big data and machine learning algorithms. He is the author and co-author of about 190 papers in computational science, computational intelligence and physics.

dzwiniel@agh.edu.pl

Notes: