LITHUANIAN COMPUTER SOCIETY

VILNIUS UNIVERSITY
INSTITUTE OF DATA SCIENCE AND DIGITAL TECHNOLOGIES

LITHUANIAN ACADEMY OF SCIENCES

12th Conference on

DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

Druskininkai, Lithuania, Hotel "Europa Royale"
http://www.mii.lt/DAMSS

December 2–4, 2021

VILNIUS UNIVERSITY PRESS

Vilnius, 2021

Co-Chairmen:

Dr. Saulius Maskeliūnas (Lithuanian Computer Society)
Prof. Gintautas Dzemyda (Vilnius University, Lithuanian Academy of Sciences)

Programme Committee:

Prof. Juris Borzov (Latvia)
Prof. Robertas Damaševičius (Lithuania)
Prof. Janis Grundspenkis (Latvia)
Prof. Janusz Kacprzyk (Poland)
Prof. Ignacy Kaliszewski (Poland)
Prof. Yuriy Kharin (Belarus)
Prof. Tomas Krilavičius (Lithuania)
Prof. Julius Žilinskas (Lithuania)

Organizing Committee:

Dr. Jolita Bernatavičienė
Dr. Olga Kurasova
Dr. Viktor Medvedev
Dr. Martynas Sabaliauskas
Laima Paliulionienė

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Dr. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. +370 5 2109 315

https://doi.org/10.15388/DAMSS.12.2021

ISBN 978-609-07-0673-2 (print)
ISBN 978-609-07-0674-9 (digital PDF)

On the Parallelization of the Geometric Multidimensional Scaling

Gintautas Dzemyda, Viktor Medvedev,
Martynas Sabaliauskas

Institute of Data Science and Digital Technologies
Vilnius University
viktor.medvedev@mif.vu.lt

A well-known procedure for mapping data from a high-dimensional space to a lower-dimensional space is Multidimensional Scaling (MDS). This algorithm keeps the distances in the low-dimensional space as close as possible to the distances in the original multidimensional space. Although MDS demonstrates great versatility, it is computationally demanding. Conventional approaches to the MDS method are limited when analyzing very large datasets, as they require very long computation times and large amounts of memory. The MDS method is used in many data analysis applications. However, despite a number of studies in which MDS methods have been applied to data mining, the number of data points to be analyzed has been limited by the high computational complexity of MDS. A new Geometric MDS method with lower computational complexity has been developed (G. Dzemyda and M. Sabaliauskas. Geometric multidimensional scaling: A new approach for data dimensionality reduction. *Appl. Math. Comput*. 409, 125561 (2021). https://doi.org/10.1016/j.amc.2020.125561; G. Dzemyda and M. Sabaliauskas. 2021. New capabilities of the geometric multidimensional scaling. In *Trends and Applications in Information Systems and Technologies*. WorldCIST 2021, A. Rocha et al. (Ed.). *Advances in Intelligent Systems and Computing*, Vol. 1366. Springer, 264–273. https://doi.org/10.1007/978-3-030-72651-5_26), making MDS more applicable to large-scale datasets. The results allow us to extend the application of this method to new and efficient ways of visualizing large-scale multidimensional data and reducing its dimensionality. A way to minimize MDS stress has been developed using the ideas of Geometric MDS, where all points in a low-dimensional space change their coordinates simultaneously and independently during a single iteration of stress minimization. Its efficiency depends on the choice of parallelization strategy associated with different ways of grouping the analyzed data, simultaneous computation of multiple coordinates in low-dimensional space, and the number of multi-core processors used for the computation. The proposed Geometric MDS allows the implementation of parallel computing for the dimensionality reduction process of large-scale data using multithreaded multi-core processors or parallel coprocessors such as GPUs. We examine how the computational speed of multidimensional data visualization varies depending on the strategy chosen and the number of processors.

This research has received funding from the Research Council of Lithuania (LMTLT), agreement No S-MIP-20-19.