Research Institute
Cyber Defence
Bundeswehr University Munich

Report – Workshop 1 "Geoinformation Systems"

During the CODE JT 2019, a Workshop entitled "Geoinformation Systems" was organized by Martin Werner (DLR / Universität der Bundeswehr, Big Geospatial Data). This workshop brought together researchers and practitioners from the fields of spatial data acquisition, spatial data management, and spatial data science in order to discuss the current status of the field from different perspectives. This included a technical point of view related to managing big and complex spatial data including trajectories of moving objects, a perspective of open source for spatial applications, perspectives of spatiotemporal statistics, remote sensing and Earth observation, and data science and data fusion.
With this multidisciplinary perspective, the speakers covered a wide portion of the field of working with spatiotemporal data with significant differences in data, terminology, demand, and abilities from significantly different perspectives. Discussions with the audience (around 40 people) showed that geoinformation systems are interesting for a wide range of application (e.g., space, medicine, social networks) fostering emerging research. Thus, abstracts of the given talks are presented in the following concluding with the key message of the individual research topics to point out in which direction the development and research will go.

Michael Schmitt (Technical University of Munich) has given a nice introduction to Remote Sensing as a prime area of big data processing. Not only that the current satellite archives are filling with petabytes of information, it is as well the variety of sensors, the rising interest in global and time-continuous applications, the quick development of novel algorithms and the advent of artificial intelligence for remote sensing that makes it an exciting field. He showed examples related to colorization of otherwise difficult to understand imagery from synthetic aperture radar observations (SAR) lowering barriers for the exploitation of SAR sensors in the wild.

Anita Graser (Austrian Institute of Technology) continued highlighting that big geospatial data is more than remote sensing. Movement data analysis, she says, is computationally challenging due to complexity of the data and algorithms. Dataset sizes are increasing steadily, data sources are diversifying, and analysis approaches are attempting to incorporate additional contextual information. While GIS tools for big spatio-temporal data do already exist (including open source tools such as GeoMesa and GeoWave which build on the popular Hadoop stack), they apply an event-based perspective rather than a continuous process perspective, which would be more appropriate for movement data. Her expectation for the future of GIS is that these systems need to become better at exploration and modeling of time, movement, and processes using large datasets

Philipp Otto (Universität Hannover) provides a perspective of spatio-temporal statistics in which high-dimensional, high-volume datasets still are largely unexplored. With growing availability of high-resolution spatial data like high-definition images, 3d point clouds of LIDAR scanners, or communication and sensor networks, he says, it might become challenging to timely detect changes and simultaneously account for spatial interactions. To detect local changes in the mean of isotropic spatiotemporal processes with a locally constraint dependence structure, we propose a monitoring procedure, which can completely be run on parallel processors. This allows for a fast detection of local changes, i.e., only a few spatial locations are affected by the change. Due to parallel computation, high-frequency data could also be monitored. We, therefore, additionally focus on the processing time required to compute the control statistics. Finally, the performance of the charts is analyzed by a series of Monte-Carlo simulation studies. In data science, he claims, it is important that the computer scientists and statisticians work closely together. Nowadays, the number of joint projects is growing, although we still observe that these two scientific fields and communities stay separated to a certain extent.

With respect to the future of Geoinformation system, he says that while machine learning methods and artificial intelligence show a very good performance for classification, these models are non-linear and highly complex, such that it is hard to gain new knowledge about the driving factors and dynamics of process. Thus, the extraction of knowledge from, e.g., neural networks should be one important point for future research. This could either be done by simplification of the models, innovative ways of visualizations, or new ways of communicating results.

Matthias Renz (Universität Kiel) contributed a computer science view introducing the notion of Heterogeneous Information Networks as a vehicle to create more flexible ways of schema inference and data conflation. In traditional databases, he says, diverse datasets need to be mapped to a joint schema and finding this schema is far from trivial when the involved data sources get more diverse. Heterogeneous information networks represent data in a fundamentally different way in which a schema does not represent the form of a data item, but rather its likely relations to other objects. That is, data is represented as a knowledge graph and queries can be contextualized from this graph. This strategy helps to work especially with data that is only weakly linked or where the links are not obvious taking into account many other data items.

Based on the presented talks and discussions a panel discussion took part with the speakers asking, “What is going to happen in the domain of spatial data collection, analysis, and application. “ Figure 1 shows the panel.

Figure 1: Panel @ Workshop „Geoinformation Systems" (from left: Philipp Otto, Anita Graser, Matthias Renz, Michael Schmitt, and Martin Werner)

Summing up the talks and discussions in the workshop, research has to face manifold challenges ranging from technical challenges (e.g., big data/exascale computing, complex data and relations, visualization and interaction) to societal challenges (e.g., democratizing technology, controlling technology, and applying spatial technology for social goods). In order to develop a spatial data ecosystem we need to simplify technology, educate people, learn about positive and negative impacts avoiding drawbacks, and support positive effects. Overall, the ecosystem has four dimensions to look on that can be support by different technologies shown in Figure 2.

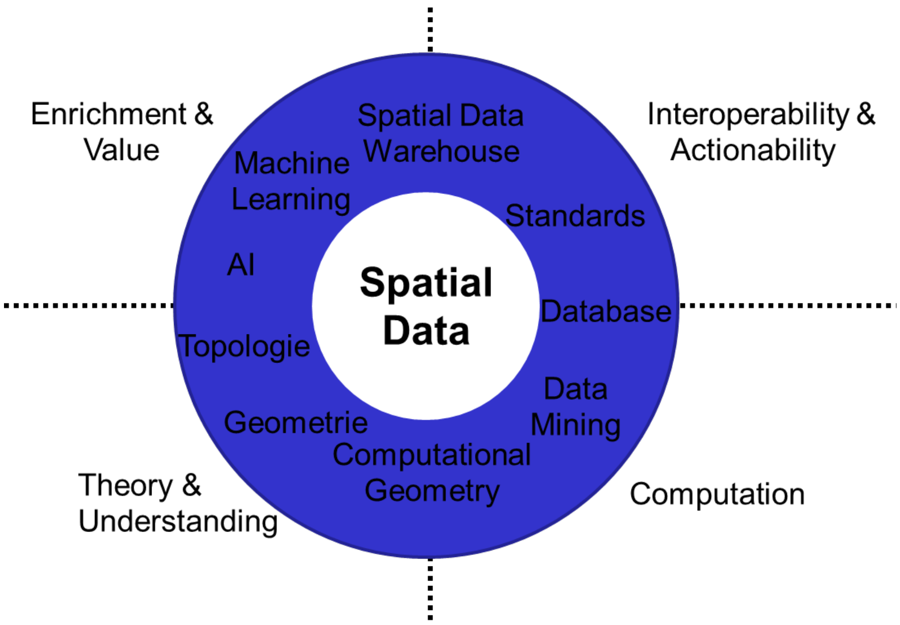

Figure 2: Spatial Data Ecosystem

Minutes of panel:

Will AI help us or even introduce risks?

- Powerful – useful as we have too much data in place to be analyzed
- Combine with machine learning and neuronal networks
- Q: But does it make sense to get the human out of the analysis? Humans have knowledge that is hard to be trained to a machine
- Resulting maps can only answer close questions by purpose, deeper questions can only be answered with a combination of knowledge and deep analysis by machine of complete data source
- Coloring can improve visualization of data for humans, but still data is multi-dimensioned and this has to be kept in mind always, because the view influences the interpretation
- A map is a tool, but we humans need to learn how to use the information for individual purposes and interactions in the community

Simplifying models or reducing data?

- For sure, models need to be understandable, but depending on the area, they will stay complex
- Point-clouds are nice, but does everyone understand it, so based on data different models will be in place and used by personal favorites