Geoinformatica (2016) 20:797–799
DOI 10.1007/s10707-016-0269-7

Guest editorial: big spatial data

Raju Vatsavai[1] **· Varun Chandola**[2]

Published online: 10 September 2016

We are living in the era of 'Big Data.' Big data is currently the hottest topic for data researchers and scientists with huge interests from the industry and federal agencies alike, as evident in the recent White House initiative on "Big data research and development". Big data, often characterized in terms of volume, velocity, variety, and veracity, is impacting the traditional data storage and processing frameworks. As the data sets are becoming large and complex, big data is posing challenges to traditional data storage and processing workflows including but not limited to data capture, transfer, storage, curation, search, query, analysis, and visualization. Spatial and spatiotemporal data, whether captured through remote sensors (e.g., remote sensing imagery, Atmospheric Radiation Measurement (ARM) data) or large scale simulations (e.g., climate data) has always been 'Big.' However, recent advances in instrumentation and computation, and advent of social media is making the spatiotemporal data even bigger, putting several constraints on data analytics capabilities. For example, Google generates more than 25 PB of data per day, a significant portion of which is spatiotemporal (images and videos) data. The rate at which spatiotemporal data is being generated clearly exceeds our ability to organize and analyze them to extract patters critical for understanding dynamically changing world. Spatial computation needs to be transformed to meet the challenges posed by the big spatial and spatiotemporal data.

The purpose of this special issue is to showcase some of the recent developments and novel applications of the big spatial data field. The open call for big spatial data has attracted eleven papers covering broad range of spatial big data technologies and applications. After two rounds of peer-reviews by a team of international experts, four papers were selected to be included in this special issue.

✉ Raju Vatsavai
rrvatsav@ncsu.edu

Varun Chandola
chandola@buffalo.edu

[1] North Carolina State University, Raleigh, NC, USA

[2] State University of New York at Buffalo, Buffalo, NY, USA

Spatiotemporal co-occurring patterns represent subsets of event types that occur together in both space and time. These patterns reflect the spatiotemporal overlap relationships among two or more spatiotemporal instances both in spatial and temporal dimensions. The discovery of these patterns have broad ranging applications in various scientific domains, including but not limited to climate science, earth sciences, astronomy, biology, and geosciences. The computational complexity of the spatiotemporal co-occurrence pattern mining algorithms coupled with large and complex datasets originating from the above-described scientific domains makes it harder to follow the traditional route of centralized data processing. The "Mining Spatiotemporal Co-occurrence Patterns in Non-Relational Databases" paper addresses this limitation by developing a cloud-based distributed spatiotemporal data mining system for discovering spatiotemporal co-occurring patterns (DOI 10.1007/s10707-016-0255-0). Instead of traditional relational databases, this paper utilizes a column-oriented database for managing spatiotemporal trajectories. Authors developed in-memory join-index structure and join algorithm for efficient processing of join operations on spatiotemporal trajectories. Experimental evaluation shows the efficacy of the author's spatiotemporal co-occurrence pattern-mining framework.

Volunteered geographic information (VGI) refers to assertive methods of collecting geographic data as opposed the traditional authoritative methods employed by private companies and government agencies. With the advent and proliferation of social media technologies, recent years have witnessed the growth of VGI websites, including OpenStreetMap, WikiMapia, and Google Map Maker. In addition, user generated geo-tagged messages (e.g., Twitter), photos (e.g., Flickr), and videos (e.g., YouTube) are leading to big spatial data repositories. It is challenging to index and search unstructured video content in large repositories. However, one can exploit the geotagging information for indexing and efficient retrieval of videos using spatial database principles. The authors of "Efficient Indexing and Retrieval of Large-scale Geo-tagged Video Databases" paper presented a framework consisting of new R-tree index structures for effectively harnessing each video frame's (Field-of-View) camera locations, orientations, and view-distance attributes (DOI 10.1007/s10707-016-0250-5). The authors of this paper also presented various search strategies and efficient range and directional queries over new index structures.

Remote sensing data represents a prime example of big spatial data. Hyper-spectral sensors mounted on unmanned aerial systems (UAS) are capable of producing on demand data critical for monitoring agriculture and environment to critical infrastructures like oil and gas distribution networks. However, processing hyper-spectral data poses several computational challenges. The paper "The index array approach and the dual tiled similarity algorithm for UAS hyper-spatial image processing" presents an index array approach for lens distortion correction and geo-referencing on graphics processing units (GPU) (DOI 10.1007/s10707-016-0253-2). The index array approach presented in this paper supports parallel image I/O. Experimental results showed 10× speedups on GPUs relative to CPU processing.

Finally, the "The Big Data of Violent Events: Algorithms for Association Analysis Using Spatio-Temporal Storytelling" paper presents three methods of association analysis for event mining in large geotemporally-encoded text streams (DOI 10.1007/s10707-016-0247-0). The distance-based Bayesian inference method exploits proximity to relate similar events. The spatial association index and spatio-logical inference methods measure the influence of violent events in different locations. These methods are capable of dealing with high-volume and variability of these datasets. Extensive experiments on social unrest in Mexico and wars in the Middle East showed that these methods could achieve precision and recall as high as 80 % in retrieval tasks that use both keywords and geospatial information as search criteria.

Together these four papers showcase various aspects of the emerging big spatial data. We hope that this special issue will be appealing to both experts and application developers of big spatial and spatiotemporal data.

Raju Vatsavai is a Chancellor's Faculty Excellence Program Cluster Associate Professor in Geospatial Analytics in the Department of Computer Science, North Carolina State University (NCSU). He works at the intersection of spatial and temporal bigdata management, analytics, and high performance computing with applications in the national security, geospatial intelligence, natural resources, climate change, location-based services, and human terrain mapping. Before joining NCSU, Raju was the Lead Data Scientist for CSE Division at ORNL. As the Associate Director of the Center for Geospatial Analytics (CGA), Raju plays a leadership role in the center's strategic vision for spatial computing research. He holds MS and PhD in Computer Science from the University of Minnesota.

Varun Chandola is a tenure-track Assistant Professor at University at Buffalo (SUNY) in the Computer Science Department and the Center for Computational and Data-Enabled Science and Engineering (CDSE). His research covers the application of data mining and machine learning to problems involving big and complex data, focusing on anomaly detection finding surprising patterns, connections, associations, and trends in data. Before joining UB, he was a scientist in the Computational Sciences and Engineering Division at the OakRidge National Laboratory. He holds a PhD in Computer Science from the University of Minnesota.