Workshop on Geospatial Data Science

@

ACM India Joint International Conference

on

Data Science & Management of Data

7th ACM IKDD CoDS and 25th COMAD, Hyderabad

[6th Jan, 2020]

Schedule

Time	Event
8:30-8:45	Introductory Remarks (Umesh Bellur, Sumit Sen)
8:45-9:45	Keynote Lecture - Siva Ravada (Oracle) Evolution of Spatial Data Management: Isolated Spatial databases to Autonomous Spatial Data Platform
9:45-10:00	KS Rajan (IIIT Hyderabad) MiSTIC, a Spatio-temporal data mining approach, based Knowledge Discovery from Observed phenomena
10:00-10:15	*Tea / Coffee Break*
10:15-11:15	Keynote Lecture - Mohamed Mokbel (QCRI/UMN) Machine Learning for Big Spatial Data and Applications
11:15-11:30	Uttam Kumar (IIIT Bangalore) Subpixel Learning Algorithms for Estimating Global Land Cover Fractions from Satellite Data Using High Performance Computing
11:30-11:45	Vishwanath Gunturi (IIT Ropar) Novel Preference Metrics in Urban Navigation
11:45-12:00	Vinod Bothale (ISRO) Spatial Data Infrastructure & Development of Enterprise Geospatial Platform
12:00-12:15	Pattabhi Rao (INCOIS) Oceanographic Data Management and Data Services
12:15-12:45	Concluding Remarks (PS Acharya, Shubha Pandey, SK Ghosh)

Keynote Lecture 1

Evolution of Spatial Data Management: Isolated Spatial databases to Autonomous Spatial Data Platform

Spatial databases came into prominence 20 years ago with the introduction of spatial data types into traditional relational databases. Over the last 20 years, many advances in spatial databases brought spatial data management into mainstream applications. With the recent developments in Big Data, NoSQL databases, Cloud Computing, Edge Computing, and AI technologies, there is disruption in the industry. The idea of consolidating all of the spatial data into one single database is being challenged. In this talk, we present the evolution of spatial data management to an autonomous spatial data platform. We present a reference architecture for this vision and describe how spatial applications can be built on top of this autonomous data platform. We also describe how Spatial Data Science applications can leverage this autonomous data platform.

Siva Ravada is the Senior Director of Development for Oracle Spatial and Graph and Mapping technologies at Oracle. Siva has been with Oracle for over 17 years leading the spatial development activities at Oracle, both for on-premise systems as well as for cloud services. Prior to joining Oracle, Siva received his PhD degree from the Univ. Of Minnesota with a thesis on parallel algorithms for spatial databases. Under his leadership, Spatial technology is incorporated into different Oracle products including the database, middleware and applications. He is also a well-known researcher in the industry with over 50 patents and publications in the area of Spatial Databases.

Keynote Lecture 2

Machine Learning for Big Spatial Data and Applications

This talk will focus on our recent efforts in adopting machine learning techniques for big spatial data and applications. This includes going for two orthogonal, but related, directions. First injecting the spatial awareness inside machine learning techniques and applications, which will result in a higher accuracy for such applications. Second, taking advantage of the recent advances in machine learning techniques to boost the usability, deployment, scalability, and accuracy of long lasting spatial and spatio-temporal data analysis techniques. For the first direction, we will present the Sya system as a full-fledged spatial machine leaning-based probabilistic knowledge base construction system. For the second direction, we will present machine-learning-based techniques for spatial autologistic regression and shortest path queries

About the Speaker

Mohamed Mokbel is Chief Scientist at Qatar Computing Research Institute and a Professor at University of Minnesota. His current research interests focus on systems and machine learning techniques for big spatial data and applications. Mohamed is an ACM Distinguished Scientist. His research work has been recognized by the VLDB 10-years Best Paper Award, four conference Best Paper Awards, and the NSF CAREER Award. Mohamed is the past elected Chair of ACM SIGPATIAL, current Editor-in-Chief for Distributed and Parallel Databases Journal, and on the editorial board of ACM Books, ACM TODS, VLDB Journal, ACM TSAS, and GoeInformatica journals. He has also served as PC Vice Chair of ACM SIGMOD and PC Co-Chair for ACM SIGSPATIAL and IEEE MDM.

Invited Talks

MiSTIC, a Spatio-temporal data mining approach, based Knowledge Discovery from Observed phenomena

KS Rajan (IIIT Hyderabad)

We present our work in developing MiSTIC - a generalised Spatio-temporal data mining approach, and how it can be used to extract spatial regions or zones which exhibit consistent temporal behaviour. In case of the Health case study, it shows how the extracted zones concur well with the topograhical and other constraints that limit the diseases sphere of influence. Further the approach can help improve our estimates of risk of these regions when there is an outbreak. In the other case study on crop production (yield) systems, MiSTIC is further extended to indicate regions and sub-regions that are consistently high and low performers across the years irrespective of the input similarity or variability in soil, climate, agricultural practices, etc. These indicate that the high and low performers might be more strongly influenced over time by other regulatory and controlling factors at a spatial/regional scale than just the input-output interactions.

Subpixel Learning Algorithms for Estimating Global Land Cover Fractions from Satellite Data Using High Performance Computing

Uttam Kumar (IIIT Bangalore)

The physical surface of the landscape is undergoing transformation either naturally or due to human interference, giving rise to Land Cover Land Use Change (LCLUC). Expanding urban regions and consequent LCLUC have emerged as one of the major anthropogenic sources of global environmental degradation, bringing numerous stresses to landscapes, vegetation, habitats,

water, etc. LCLUC at a sub-continent to global level can be monitored through high temporal and low spatial resolution data, such as those obtained from Landsat at 30 m or MODIS Terra/Aqua at 250 m spatial resolution. These satellites improve the ability to map large areas of Earth’s surface quickly due to their wider IFOV and inexpensively. However, different land cover (LC) types jointly occupy a single pixel, and the resulting spectral measurement is a composite of the individual spectra. The intrinsic scale of spatial variation in LC is usually finer than the scale of sampling imposed by the image pixels. Due to scale-resolution mismatch, the spatial resolution of the details on the ground is less than what is required, leading to sub-pixel heterogeneity, imposing limitations in modelling with these data sets. The talk will present some attempts to resolve the mixed pixel problem and also discuss some of the emerging challenges in geospatial research.

Novel Preference Metrics in Urban Navigation

Vishwanath Gunturi (IIT Ropar)

We present our work in developing MiSTIC - a generalised Spatio-temporal data mining approach, and how it can be used to extract spatial regions or zones which exhibit consistent temporal behaviour. In case of the Health case study, it shows how the extracted zones concur well with the topograhical and other constraints that limit the diseases sphere of influence. Further the approach can help improve our estimates of risk of these regions when there is an outbreak. In the other case study on crop production (yield) systems, MiSTIC is further extended to indicate regions and sub-regions that are consistently high and low performers across the years irrespective of the input similarity or variability in soil, climate, agricultural practices, etc. These indicate that the high and low performers might be more strongly influenced over time by other regulatory and controlling factors at a spatial/regional scale than just the input-output interactions.

Spatial Data Infrastructure & Development of Enterprise Geospatial Platform

Vinod Bothale (ISRO)

Spatial data infrastructure (SDI) is an infrastructure that facilitates the discovery, access, management, distribution, reuse, and preservation of digital geospatial assets. These assets include maps, data, Web services, and toolboxes. The role of SDI have become more significant in today’s big data age and also when geospatial data has become a vital component of planning , monitoring and decision making. User demands are ever increasing for diverse geospatial data and Web services to solve the problems in their application domain. Building a geospatial platform catering the large user base is a challenge in terms of many facets like generating large Scale Data Assets, creating software platforms for geospatial data handling, building the hardware platform and Optimizing the delivery mechanism . The talk focuses on these challenges and solutions for building Geospatial Platform of enterprise Scale for data visualization, delivery and governance applications at National Scale.

Oceanographic Data Management and Data Services

Pattabhi Rao, R Venkat Shesu (INCOIS)

Observations from the oceans are the backbone for providing operational oceanographic services (potential fishing zone advisory services, ocean state forecast, storm surges, cyclones, monsoon variability, tsunami etc.), research and development, calibration and validation of satellite sensors, parameterizing key processes for models and verifying model simulations. The Ocean Observing System (OOS) comprises of in-situ and remote sensing platforms measuring a suite of marine meteorological and oceanographic parameters on a broad spectrum of spatial and temporal time scales.

Preservation of the long-term oceanographic data and its availability from a single source would facilitate multi-disciplinary approach in understanding the oceans in a better way and bringing out new insights. Further,

technological advances in the last two decades in oceanographic sensors; communication and computing systems facilitated the ocean scientists to acquire data in real-time from a variety of ocean observing platforms. Management of heterogeneous and voluminous oceanographic data is imperative to ensure high-quality data for research and for data-driven decision making in operational oceanographic services.

Ocean data management is becoming an integral part of ocean observation programmes with emerging technologies such as sensor observing service, web-based data services to serve the data with analytical and visualization features on the fly. We briefly present the ocean observing systems in the Indian Ocean, operational oceanographic services provided India and development of ocean data and information system, an end-to-end data management system to provide data services to the stakeholders and the recent developments.

Organizers

Umesh Bellur,
Head - Department of Computer Science and Engineering, IIT Bombay
Email: umesh.bellur@iitb.ac.in

Soumya K Ghosh,
Professor, Department of Computer Science and Engineering, IIT Kharagpur
Email: skg@cse.iitkgp.ac.in

PS Acharya,
CEO NSDI -India and Scientist ‘G’, DST, Ministry of Science & Technology, New Delhi
Email: psa@nic.in
Sumit Sen,
GISE Lab, Department of Computer Science and Engineering, IIT Bombay
Email: sumitssen@gmail.com