From GPS, Google Maps and Uber to Spatial Computing

December 2019

Shashi Shekhar
McKnight Distinguished University Professor
Computer Sc. and Eng., University of Minnesota
www.cs.umn.edu/~shekhar

Thanks: Sponsors: NSF, USDOD, NASA, USDOT, USDOE, Ford URP, IBM, Microsoft, …
UMN: CTS, IonE, DTC, Spatial Computing Group students
Spatial Computing: Recent Examples
## The Changing World of Spatial Computing

<table>
<thead>
<tr>
<th></th>
<th>Last Century</th>
<th>Last Decade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Map User</strong></td>
<td>Well-trained few</td>
<td>Billions</td>
</tr>
<tr>
<td><strong>Mappers</strong></td>
<td>Well-trained few</td>
<td>Billions</td>
</tr>
<tr>
<td><strong>Software, Hardware</strong></td>
<td>Few layers, e.g., Applications: Arc/GIS, Databases: SQL3/OGIS</td>
<td>Almost all layers</td>
</tr>
<tr>
<td><strong>User Expectations &amp; Risks</strong></td>
<td>Modest</td>
<td>Many use-case &amp; Geo-privacy concerns</td>
</tr>
</tbody>
</table>
Q: Which agencies sowed seeds for Google Maps?


![Table 1. Members of the Federal Geographic Data Committee (FGDC)]

<table>
<thead>
<tr>
<th>Dept. of Agriculture</th>
<th>Environmental Protection Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept. of Commerce</td>
<td>Federal Emergency Management Agency</td>
</tr>
<tr>
<td>Dept. of Defense</td>
<td>General Services Administration</td>
</tr>
<tr>
<td>Dept. of Energy</td>
<td>Library of Congress</td>
</tr>
<tr>
<td>Dept. of Health and Human Services</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>Dept. of Housing and Urban Development</td>
<td>National Archives and Records Administration</td>
</tr>
<tr>
<td>Dept. of the Interior (Chair)</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>Dept. of Justice</td>
<td>Tennessee Valley Authority</td>
</tr>
<tr>
<td>Dept. of State</td>
<td>Office of Management and Budget (Co-Chair)</td>
</tr>
</tbody>
</table>
Deconstructing Precision Agriculture

#AgInnovates2015

Wednesday, March 4, 2015
Reception | 5:00 to 7:00 pm
House Agriculture Committee Room,
1300 Longworth House Office Building,
Washington, DC

Think Moon landing. Think Internet. Think iPhone and Google. Think bigger.

Come hear U.S. farmers, leading agriculture technology companies, and scientists tell how they work together to fuel U.S. innovation and the economy to solve this global challenge.

The event will exhibit three essential technologies of precision agriculture that originated from a broad spectrum of federally funded science: Guidance Systems and GPS, Data & Mapping with GIS, and Sensors & Robotics.

Moderator
Raj Khosla, Professor of Precision Agriculture at Colorado State Univ.

Farmers
David Hula, of Renwood Farms in Jamestown, Virginia
Rod Weimer, of Fagerberg Produce in Eaton, Colorado
Del Unger, of Del Unger Farms near Carlisle, Indiana

Speakers
Mark Harrington, Vice President of Trimble
Carl J. Williams, Chief of the Quantum Measurement Division at NIST
Bill Raun, Professor at Oklahoma State Univ.
Marvin Stone, Emeritus Professor at Oklahoma State Univ.
J. Alex Thomasson, Professor at Texas A&M Univ.
Dave Gebhardt, Director of Data and Technology at Land O'Lakes/WinField
Shashi Shekhar, Professor at the Univ. of Minnesota

Hosted by the Congressional Soils Caucus
In partnership with
Agricultural Retailers Association
American Society of Plant Biologists
American Physical Society
American Society of Agronomy
Association of Equipment Manufacturers
Coalition for the Advancement of Precision Agriculture
Computing Research Association
CropLife America
Crop Science Society of America
PrecisionAg Institute
Soil Science Society of America
Task Force on American Innovation
Texas A&M AgriLife
Trimble
WinField

RSVP
http://bit.ly/1CoOYoa

This is about feeding the world.
McKinsey Global Institute

The study estimates that the use of personal location data could save consumers worldwide more than $600 billion annually by 2020. Computers determine users' whereabouts by tracking their mobile devices, like cellphones.

The biggest single consumer benefit, the study says, is going to come from time and fuel savings from location-based services tapping into real-time traffic and weather data that help drivers avoid congestion and suggest alternative routes. The location tracking, McKinsey says, will work either from drivers' mobile phones or GPS systems in cars.

The New York Times
Published: May 13, 2011
From GPS and Virtual Globes to Spatial Computing-2020

About the workshop
This workshop outlines an effort to develop and promote a unified agenda for Spatial Computing research and development across US agencies, industries, and universities. See the original workshop proposal here.

Spatial Computing
Spatial Computing is a set of ideas and technologies that will transform our lives by understanding the physical world, knowing and communicating our relation to places in that world, and navigating through those places.

The transformational potential of Spatial Computing is already evident. From Virtual Globes such as Google Maps and Microsoft Bing Maps to consumer GPS devices, our society has benefitted immensely from spatial technology. We’ve reached the point where a hiker in Yellowstone, a schoolgirl in DC, a biker in Minneapolis, and a taxi driver in Manhattan know precisely where they are, nearby points of interest, and how to reach their destinations. Large
Workshop Highlights

Agenda

- Identify fundamental research questions for individual computing disciplines
- Identify cross-cutting research questions requiring novel, multi-disciplinary solutions

Organizing Committee

- Peggy Agouris, George Mason University
- Walid Aref, Purdue University
- Michael F. Goodchild, University of California - Santa Barbara
- Erik Hoel, Environmental Systems Research Institute (ESRI)
- John Jensen, University of South Carolina
- Craig A. Knoblock, University of Southern California
- Richard Langley, University of New Brunswick
- Ed Mikhail, Purdue University
- Shashi Shekhar, University of Minnesota
- Ouri Wolfson, University of Illinois
- May Yuan, University of Oklahoma
2012 CCC Workshop: Spatial Computing Visioning

[PDF] Spatial Thinking: A missing building block in STEM education Spatial ... scienceoflearning.jhu.edu/assets/documents/spatial_thinking_FINAL.pdf

by K Gagnier - Related articles

One critical building block of success in STEM fields, however, is often overlooked: the ability to think spatially. Spatial thinking refers to a set of mental skills that ...

• Ten Opportunities
  1. Spatial Abilities Predict STEM Success
  2. Emerging Spatial Big Data
  3. Augmented Reality Systems
  4. Time-Travel in Virtual Globes
  5. Spatial Predictive Analytics
  6. Persistent Environment Hazard Monitoring
  7. Geo-collaborative Systems, Fleets, and Crowds
  8. Localizing Cyber Entities
  9. GPS Deprived Environment
  10. Beyond Geo
Outline

- Introduction
- Broad Interest Examples
  - GPS
    - Outdoors => Indoors
  - Spatial Database Management Systems
  - Location Based Services
  - Spatial Data Science
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions
Global Positioning Systems (GPS)

- Positioning ships
  - Latitude \( f(\text{compass, star positions}) \)
  - Longitude Prize (1714) => marine chronometer
  - accuracy in nautical miles

- Global Navigation Satellite Systems
  - Use: Positioning, Clock synchronization
  - Infrastructure: satellites, ground stations, receivers, …

Trilateration


http://answers.oreilly.com/topic/2815/how-devices-gather-location-information

Positioning Precision

- **Emergency Location**
- **Precision Agriculture**
- **Earthquake Displacements**

**Positioning Precision**

- **Decadal Survey Missions**
- **Sea Level**
- **Geodynamics**
- **Seismic Hazard**
- **Volcanic Hazards**
- **Hydrology**
- **Satellite Orbit Determination**
- **Leveling**
- **Lidar/SAR**
- **Airborne Surveys**
- **Weather Forecasting***
- **Precision Timing**
- **Tsunami Warning**
- **Glacial Flow**
- **Autonomous Navigation**
- **Spacecraft Navigation**
- **Aircraft Navigation**
- **Aircraft Landing**

**Applied Geodesy**

- **Ocean Navigation**

**Time Scale**

- Seconds
- Minutes
- Hours
- Days
- Months
- Years
- Decades
Spatial Computing is a Critical Infrastructure Today!

- 2 billion GPS receivers in use, will hit 7 billion by 2022.
- Besides location, it reference time for critical infrastructure
  - Telecommunications industry, Banks, Airlines...
- GPS is the single point of failure for the entire modern economy.
- 50,000 incidents of deliberate (GPS) jamming last two years
  - Against Ubers, Waymo’s self-driving cars, delivery drones from Amazon
- Ground based alternatives appearing in S. Korea, USA, …

Trends: Localization Without GPS

• GPS works outdoors, but,
  – It can be jammed or spoofed
  – We are indoors 80% of time!
  – Ex. malls, hospitals, airports, …

• Indoor infrastructure
  – Location: Wi-Fi, Blue Tooth, …
  – How to represent indoor for navigation?
Even before cable news outlets began reporting the tornadoes that ripped through Texas on Tuesday, a map of the state began blinking red on a screen in the Red Cross' new social media monitoring center, alerting weather watchers that something was happening in the hard-hit area. (AP, April 16th, 2012.)
Outline

• Introduction
• Broad Interest Examples
  – GPS
  – Spatial Database Management Systems
    • Point Location => Spatial
    • Scalability => Privacy
  – Location Based Services
  – Spatial Data Science
  – Virtual Globes & Remote Sensing
  – Geographic Information Systems
• Conclusions
Q? What should Google return for the following questions?
- Distance between Gujarat and Rajasthan
- Distance between Gujarat and India
Is GIS just Location?

- Spatial Relationships
  - Ex. Topological, Metric, …
  - OGC Simple Features Standards
- Help feature selection for machine learning & modeling
  - Ex. Distance to key geographic features
  - Ex. Neighbor relationships
Spatial Big Data Curation

- Meta-data, Schema, DBMS (SQL, Hadoop)
- Challenge: One size does not fit all!

- Ex. Spatial Querying
  - Geo-tag, Checkin, Geo-fence

- Spatial Querying Software
  - OGC Spatial Data Type & Operations
  - Data-structures: B-tree => R-tree
  - Algorithms: Sorting => Geometric
  - Partitioning: random => proximity aware
Geo-Security & Geo-Privacy

- Operational Security Advice by US Army: **Avoid Geo-tags!**
  - Q. Why?

**Geo-tags can show enemies your location**

*ArmyTimes*  
Monday Dec 20, 2010

The Army is warning troops to be careful when using Facebook and other popular social networking sites because their geo-tagging features may show where U.S. forces are located in war zones.

---

**Insurgents Used Cell Phone Geotags to Destroy AH-64s in Iraq ...**

[https://www.defensetech.org](https://www.defensetech.org) › Aircraft  
Mar 15, 2012 - From an Army press release warning of the dangers of *geotags*: ... location of the helicopters inside the compound and conduct a mortar attack, ...
“I ran a little experiment. On a sunny Saturday, I spotted a woman in Golden Gate Park taking a photo with a 3G iPhone.

Because iPhones embed geo-data into photos that users upload to Flickr or Picasa, iPhone shots can be automatically placed on a map.

At home I searched the Flickr map, and score—a shot from today. I clicked through to the user’s photostream and determined it was the woman I had seen earlier.

After adjusting the settings so that only her shots appeared on the map, I saw a cluster of images in one location.

Clicking on them revealed photos of an apartment interior—a bedroom, a kitchen, a filthy living room. Now I know where she lives.”
Challenge: Geo-privacy, ...

- Emerging personal geo-data
  - Trajectories of smart phones, Google map search, ...
- Privacy: Who gets my data? Who do they give it to? What promises do I get?
- Groups: Civil Society, Economic Entities, Public Safety, Policy Makers

Table 4.2: Geo-privacy Policy Conversation Starters

1. Emergencies are different (E-911)
2. Differential geo-privacy can improve safety (E-911 → PLAN, CMAS)
3. Send apps to data, not vice-versa (e.g., eco-routing)
4. Transparent transactions for location traces for increased consumer confidence
5. Responsible entities for location traces (Credit-bureau/census, HIPPA++ for responsible parties)
Outline

• Introduction
• Broad Interest Examples
  – GPS
  – Spatial Database Management Systems
  – Location Based Services
    • Queries => Persistent Monitoring
  – Spatial Statistics
  – Virtual Globes & Remote Sensing
  – Geographic Information Systems
• Conclusions
Location Based Services

- **Location**: Where am I? (street address, <latitude, longitude>)
- **Directory**:
  - What is around me?
  - Where is the nearest clinic (or ambulance)?
- **Routes**: What is the shortest path to reach there?
Models: Spatial Graphs & Flow Networks

- Ex.: Roadmaps, Electric grid, Supply chains, …
- Graphs: Nodes, Edges, Routes, …
- Flow networks: Capacity constrain
- Operations:
  - Geo-code, Map-matching, …
  - Connectivity, shortest path, nearest neighbor
  - Logistics: Site selection, Allocation, Max-flow, …

Graph Data for UMN Campus
Courtesy: Bing

<table>
<thead>
<tr>
<th>NID</th>
<th>EID</th>
<th>From</th>
<th>To</th>
<th>Speed</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>E1</td>
<td>N1</td>
<td>N2</td>
<td>35mph</td>
<td>0.075mi</td>
</tr>
<tr>
<td>N2</td>
<td>E2</td>
<td>N1</td>
<td>N4</td>
<td>30mph</td>
<td>0.075mi</td>
</tr>
<tr>
<td>N3</td>
<td>E3</td>
<td>N2</td>
<td>N3</td>
<td>35mph</td>
<td>0.078mi</td>
</tr>
<tr>
<td>N4</td>
<td>E4</td>
<td>N2</td>
<td>N5</td>
<td>30mph</td>
<td>0.078mi</td>
</tr>
<tr>
<td>N5</td>
<td>E5</td>
<td>N3</td>
<td>N6</td>
<td>30mph</td>
<td>0.077mi</td>
</tr>
<tr>
<td>N6</td>
<td>E6</td>
<td>N4</td>
<td>N1</td>
<td>30mph</td>
<td>0.075mi</td>
</tr>
<tr>
<td>N7</td>
<td>E7</td>
<td>N4</td>
<td>N7</td>
<td>30mph</td>
<td>0.078mi</td>
</tr>
<tr>
<td>N8</td>
<td>E8</td>
<td>N5</td>
<td>N2</td>
<td>30mph</td>
<td>0.078mi</td>
</tr>
<tr>
<td>N9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dynamic Nature of Transportation Network
Next Generation Navigation Services

- Eco-Routing
- Best start time
- Road-capacity aware, e.g., evacuation route planning

Why UPS trucks (almost) never turn left - CNN.com
www.cnn.com/2017/02/16/world/ups-trucks-no-left-turns/

Feb 23, 2017 - Left-hand turns are dangerous and wasteful, data shows. By avoiding them, UPS saves 10 million gallons of fuel each year. ... pedestrians than right ones, according to data collected by New
Next Generation Navigation Services

- Eco-Routing
- Best start time
- Road-capacity aware

<table>
<thead>
<tr>
<th>Static</th>
<th>Time-Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which is the shortest travel time path from downtown Minneapolis to airport?</td>
<td>Which is the shortest travel time path from downtown Minneapolis to airport at different times of a work day?</td>
</tr>
<tr>
<td>What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis?</td>
<td>What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis at different times in a work day?</td>
</tr>
</tbody>
</table>
Routing Challenges: Lagrangian Frame of Reference

Q? What is the cost of Path $<A,C,D>$ with start-time $t=1$? Is it 3 or 4?

Snapshots of a Graph

<table>
<thead>
<tr>
<th>Path</th>
<th>T = 0</th>
<th>T = 1</th>
<th>T = 2</th>
<th>T = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt;A,C,D&gt;$</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>$&lt;A,B,D&gt;$</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

**Spatio-temporal Graphs: Computational Challenges**

**Ranking changes over time**

- Violates stationary assumption in Dynamic Programming

<table>
<thead>
<tr>
<th>Time</th>
<th>Preferred Routes</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30am</td>
<td>Via Hlawatha</td>
</tr>
<tr>
<td>8:30am</td>
<td>Via Hiawatha</td>
</tr>
<tr>
<td>9:30am</td>
<td>via 35W</td>
</tr>
<tr>
<td>10:30am</td>
<td>via 35W</td>
</tr>
</tbody>
</table>

**Waits, Non FIFO Behavior**

- Violates assumption of Dijkstra/A*

<table>
<thead>
<tr>
<th>Time</th>
<th>Route</th>
<th>Flight Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30am</td>
<td>via Detroit</td>
<td>6 hrs 31 mins</td>
</tr>
<tr>
<td>9:10am</td>
<td>direct flight</td>
<td>2 hrs 51 mins</td>
</tr>
<tr>
<td>11:00am</td>
<td>via Memphis</td>
<td>4 hrs 38 mins</td>
</tr>
<tr>
<td>11:30am</td>
<td>via Atlanta</td>
<td>6 hrs 28 mins</td>
</tr>
<tr>
<td>2:30pm</td>
<td>direct flight</td>
<td>2 hrs 51 mins</td>
</tr>
</tbody>
</table>

*Flights between Minneapolis and Austin (TX)

**Details:** A Critical-Time-Point Approach to All-Start-Time Lagrangian Shortest Paths, IEEE Transactions on Knowledge and Data Engineering, 27(10):2591-2603, Apr. 2015 (A summary in Proc. Intl. Symp. on Spatial and Temporal Databases, Springer LNCS 6849, 2011), (w/ V. Gunturi et al.)
Trends: Persistent Geo-Hazard Monitoring

- Environmental influences on our health & safety
  - air we breathe, water we drink, food we eat

- Large Area Surveillance
  - Passive > Active > Persistent
  - How to economically cover all locations all the time?
  - Crowd-sourcing, e.g., smartphones, tweets,
  - Wide Area Motion Imagery, UAVs, …
Outline

• Introduction
• Broad Interest Examples
  – GPS
  – Spatial Database Management Systems
  – Location Based Services
  – Spatial Data Science
    • Limitations of Traditional Data Science
    • Novel approaches
  – Virtual Globes & Remote Sensing
  – Geographic Information Systems
• Conclusions
Spatial Data Science: A Historical Example

1854: What causes Cholera?

Collect & Curate Data  
Discover Patterns, Generate Hypothesis  
Test Hypothesis (Experiments)  
Develop Theory  

Impact:
sewage system, drinking water supply...

Q? What are Choleras of today?  
Q? How may Spatial Data Sc. Help?
Limitations of Traditional Data Science

• A. High cost of missed or spurious patterns
  – Pr.[Self-driving car sensors fail to detect a red traffic light] > 0
  – Loss of life, stigmatization, economic loss

• B. Gerrymandering risks
  – Spatial partitioning choice may alter results

• C. Spatial data violates assumptions of traditional data science
  – Data samples: independent and identically distributed (i.i.d)
  – Nearby spatial data samples are not independent
  – No two places on Earth are exactly alike!
A. Reducing Spurious Patterns

- **SatScan (National Cancer Institute)**
  - Compare with complete spatial random
  - Monte Carlo simulation

- **Spatial Statistics**
  - Quantify uncertainty, confidence, …
  - Model Auto-correlation, heterogeneity, …
Hotel That Enlivened the Bronx Is Now a ‘Hot Spot’ for Legionnaires’

Contaminated Cooling Towers

Five buildings have been identified as the potential source of the Legionnaires’ disease outbreak in the South Bronx.

- Possible sources of Legionnaires’ outbreak
- Additional sites found with legionella bacteria
- Locations of people with Legionnaires’

The Opera House Hotel is at the center of the outbreak. Edwin J. Torres for The New York Times
Legionnaires’ Disease Outbreak in New York

(a) Legionnaire’s in New York (2015)
(b) Output of SaTScan

Significant Hotspot (Arbitrary Shape)

Problem definition
- **Inputs**: A set of points; DBSCAN parameters; Test statistic; Significance level
- **Output**: Significant clusters
- **Objective**: Computational efficiency

**Contributions**
- Significance modeling in DBSCAN
- A fast dual-convergence algorithm

**Trends**
- DBSCAN cannot avoid chance patterns
- SaTScan cannot detect arbitrary shapes

**Details**: Significant DBSCAN towards Statistically Robust Clustering, Y. Xie and S. Shekhar. Proc. 16th International Symposium on Spatial and Temporal Databases (SSTD ’19), 2019, ACM (Best Paper Award)
B. Neighbor Graph Reduces Gerrymandering Risks

(d) Neighbor graph

(a) a map

(b) Partition A

(c) Partition C

<table>
<thead>
<tr>
<th>Participation Index</th>
<th>Ripley’s Cross-K</th>
<th>Pattern</th>
<th>Pearson Correlation</th>
<th>Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.33</td>
<td><img src="image" alt="Pattern" /></td>
<td>(-) 0.9</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td><img src="image" alt="Pattern" /></td>
<td>1</td>
<td>(-) 0.9</td>
</tr>
</tbody>
</table>
Co-locations/Co-occurrence

• Given: A collection of different types of spatial events
• Find: Co-located subsets of event types

Details: Discovering colocation patterns from spatial data sets: a general approach, (w/ H. Yan et al.), IEEE Transactions on Knowledge and Data Engineering, 16(12), Dec. 2004.
Colocation Mining

Participation ratio $\text{pr}(f_i, c)$ of feature $f_i$ in colocation $c = \{f_1, f_2, \ldots, f_k\}$:
fraction of instances of $f_i$ with feature $\{f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_k\}$ nearby

**Participation index** $\text{PI}(c) = \min\{\text{pr}(f_i, c)\}$

**Properties:**

1. **Computational**: Non-monotonically decreasing like support measure
   Allows scaling up to big data via pruning
2. **Statistical**: Upper bound on Cross-K function

**Comparison with Ripley’s K-function (Spatial Statistics)**

<table>
<thead>
<tr>
<th>K-function $(B, A)$</th>
<th>2/6 = 0.33</th>
<th>3/6 = 0.5</th>
<th>6/6 = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI $(B, A)$</td>
<td>2/3 = 0.66</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Cascading spatio-temporal pattern (CSTP)

- **Input**: Urban Activity Reports
- **Output**: CSTP
  - Partially ordered subsets of ST event types.
  - Located together in space.
  - Occur in stages over time.
- **Applications**: Public Health, Public Safety, ...

**Details**: Cascading Spatio-Temporal Pattern Discovery, (w/ P. Mohan et al.), IEEE Transactions on Knowledge and Data Engineering, 24(11), Nov. 2012.
MDCOP Motivating Example: Input

- Manpack stinger (2 Objects)
- M1A1_tank (3 Objects)
- M2_IFV (3 Objects)
- Field_Marker (6 Objects)
- T80_tank (2 Objects)
- BRDM_AT5 (enemy) (1 Object)
- BMP1 (1 Object)
MDCOP Motivating Example : Output

- Manpack stinger
  (2 Objects)
- M1A1_tank
  (3 Objects)
- M2_IFV
  (3 Objects)
- Field_Marker
  (6 Objects)
- T80_tank
  (2 Objects)
- BRDM_AT5
  (enemy) (1 Object)
- BMP1
  (1 Object)

C1. Modeling Auto-correlation in Prediction Models

- **Traditional Models**
  - Linear Regression (e.g., Logit), Bayes Classifier, Neural Networks, Decision Trees
- **Semi-Spatial**: auto-correlation in regularizer
- **Spatial Models**
  - $W = \text{neighbor matrix (row-normalized)}$
  - Spatial autoregressive model (SAR)
  - Markov random field (MRF) based Bayesian Classifier

$$\varepsilon = \|y - X\beta\|^2 + \|y - y_{\text{neighbor}}\|^2$$

<table>
<thead>
<tr>
<th>Traditional</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y = X\beta + \varepsilon$</td>
<td>$y = \rho Wy + X\beta + \varepsilon$</td>
</tr>
<tr>
<td>$\Pr(C_i \mid X) = \frac{\Pr(X \mid C_i) \Pr(C_i)}{\Pr(X)}$</td>
<td>$\Pr(c_i \mid X, C_N) = \frac{\Pr(C_i) \Pr(X, C_N \mid c_i)}{\Pr(X, C_N)}$</td>
</tr>
</tbody>
</table>
Ex.: Spatial Auto-Regression Parameter Estimation

\( \rho \) : the spatial auto-regression (auto-correlation) parameter

\( W \) : \( n \)-by-\( n \) neighborhood matrix over spatial framework

<table>
<thead>
<tr>
<th>Name</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Linear Regression</td>
<td>( y = x\beta + \varepsilon )</td>
</tr>
<tr>
<td>Spatial Auto-Regression</td>
<td>( y = \rho W y + x\beta + \varepsilon )</td>
</tr>
</tbody>
</table>

• Maximum Likelihood Estimation

\[
\ln(L) = \ln|I - \rho W| - \frac{n \ln(2\pi)}{2} - \frac{n \ln(\sigma^2)}{2} - SSE
\]

• Computing determinant of large matrix is a hard (open) problem!
  – size(\( W \)) is quadratic in number of locations/pixels.
  – Typical raster image has Millions of pixels
  – \( W \) is sparse but not banded.

C2. Modeling Spatial Heterogeneity: GWR

- Geographically Weighted Regression (GWR)
  - Goal: Model spatially varying relationships
  - Example: \( y = X\beta + \varepsilon \)
    Where \( \beta \) and \( \varepsilon \) are location dependent

Source: resources.arcgis.com
Quiz

Which are addressed in Convolutional Neural Networks (CNN)?
- Statistical significance to reduce chance Patterns
- Spatial Auto-correlation
- Spatial Heterogeneity
Trends: Civil Society Concerns

- Was Tesla “self-driving” claim fatal? : https://www.youtube.com/watch?v=o02H2xGlecc
- Are automated face recognition software fair?
- NSF DCL 19-016: Fairness, Ethics, Accountability, and Transparency: Enabling Breakthrough Research to Expand Inclusivity in CISE Research
- Books:
  - Weapons of Math Destruction, Cathy O’Neil, 2016 (2019 Euler Book Prize, TED talk)
  - Automating Inequality, V. Eubanks, 2018.
Outline

• Introduction
• Broad Interest Examples
  – GPS
  – Spatial Database Management Systems
  – Location Based Services
  – Spatial Data Science
  – Virtual Globes & Remote Sensing
    • Quilt => Time-travel & Depth
  – Geographic Information Systems
• Conclusions
A downpour in Las Vegas at the CES technology show a year and a half ago may prove to have been a watershed moment in the race to develop autonomous cars. While other companies promoting their experimental self-driving vehicles had to keep them parked in the rain, one company, AdaSky, demonstrated how its sensors could see people hundreds of feet ahead even in a downpour, and even when they were wearing white and standing against a white background.

“Thermal imaging is the best sensor at detecting people, day or night,” Chris Posch of FLIR Systems said.
ElectroMagnetic (EM) radiation

- Emitted by all objects above absolute zero (0° Kelvin (K), −273°C)

Source: directthermography.co.uk

Source: imagine.gsfc.nasa.gov
Spectral Indices

- **Index** = a summary across multiple bands to predict a feature, e.g., vegetation, water, ...

- **Example using near-infra-red (NIR) and red (RED)**
  - Ratio vegetation index (RVI) or Simple ratio index (SRI) = \( \frac{\text{NIR}}{\text{RED}} \)
  - Normalized Difference Vegetation Index (NDVI) = \( \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \)
  - Physical interpretation: energy absorption, photosynthetic capacity

- **Ex.: NDVI for Healthy vegetation (NIR = 50%, RED 8%)**
  - Stressed/sparse vegetation (NIR = 40%, RED = 30%)
  - Q? How may a farmer use NDVI to monitor crops?

We walked to this location to ground-truth the aerial images and found much sparser rows in the red areas shown in the image at left.
Virtual Globes & Volunteered Geo-Information

- **Virtual Globes**: Geo distribution, patterns
  - 1995: UMN Map Server
  - 1998: Al Gore’s Digital Earth Speech
  - 1999: Microsoft Terra-server
  - 2004: Keyhole (Google Earth) : Fly-through

- **Volunteered Geo-Information**
  - Allow citizens to make maps & report
  - 2009 Haiti Post-Earthquake Maps
  - Road maps, Traffic maps, …
Remote Sensing – Agriculture Monitoring

Countries:
- AMIS Countries
- Non-AMIS Countries

Crops:
- Maize
- Wheat
- Soybean
- Rice

Conditions:
- Exceptional
- Favourable
- Watch
- Poor
- Out-of-Season
- No Data

AMIS
Agricultural Market Information System

GEOGLAM
Global Agricultural Monitoring
Opportunities: Time-Travel and Depth in Virtual Globes

- Virtual globes are (quilt) snapshots

- **How to add time?**
  - Ex. NASA NEX, Google Earth Engine,
  - Ex. Google Timelapse: 260,000 CPU core-hours for global 29-frame video

- **Spatio-temporal Resolution**
  - Planet Labs. : daily 1m scan (visual bands)
  - USDA VegScape / CropScape

- **Small Satellites**
  - CubeSat (10cm x 10cm x 11.35cm)
Large Constellations of Small Satellites

- Hi-frequency (e.g., daily or hourly) time-series of imagery of entire earth
  - Monitor illegal fishing, forest fires, crops (2017 DARPA Geospatial Cloud Analytics)
- Large Constellations
  - 2017: Planet Labs: 100 satellites: daily scan of Earth at 1m resolution in visible band

Cheap (or free) satellite data on cloud computers

- 2008: USGS gave away 35-year LandSat satellite imagery archive
  - Analog of public availability of GPS signal in late 1980s
- 2017: Many cloud-based Virtual collaboration environment
  - Explosion in machine learning on satellite imagery to map crops, water, buildings, roads, …

<table>
<thead>
<tr>
<th>Elevation, Landsat, LOCA, MODIS, NAIP</th>
<th>Google Earth Engines</th>
<th>NEX</th>
<th>AWS Earth</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOOA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVHRR, FIA, GIMMM, GlobCover, NARR, TRIMM, Sentinel-1</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>IARPA, GDELT, MOGREPS, OpenStreetMap, Sentinel-2, SpaceNet (building/road labels for ML)</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>CHIRPS, GeoScience Australia, GSMap, NASS, Oxford Map, PSDI, WHRC, WorldClim, WorldPop, WWF,</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCCA, FLUXNET</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From Earth Observation to Geo-Dashboards

Aral Sea Shrinkage (1978-2014) Due to Cotton Farms

Alerts

Geo Dashboard

State

Trends

Global Temperature

Global Population

Sea-Surface Temperature Anomaly

NOAA/NESDIS SST Anomaly (degrees C), 6/25/2015

World Population Growth (in billions of people)
Outline

• Introduction
• Broad Interest Examples
  – GPS
  – Location Based Services
  – Spatial Data Science
  – Spatial Database Management Systems
  – Virtual Globes & Remote Sensing
  – Geographic Information Systems
    • Geo => Beyond Geo
• Conclusions
**Geographic Information Systems & Geodesy**

- **GIS**: An umbrella system to
  - capture, store, manipulate, analyze, manage, and present diverse geo-data.
  - SDBMS, LBS, Spatial Statistics, …
  - Cartography, Map Projections, Terrain, etc.

- **Map Projections**
  - Which countries in North Korea missile range?
  - Spherical coordinates vs. its planar projections
**Trend:** 3 Dimensions, e.g., High Definition (HD) Roadmaps

- Trucks and bridges (https://www.youtube.com/watch?v=USu8vT_tfdw)
- Self-Driving Cars, HD-Roadmaps (Cm accuracy)
  - Base Map: road centerlines
  - (3D) Geometry: overpass, walls, curbs, slope, …
  - Semantic: known traffic lights, stop sign, …
  - Map priors: known object types (e.g., people, cars, …
  - Real-time: Camera, LiDAR, …

Trends: From Maps to Models

- Ex. Climate Projections
- Uncertainty Quantification
  - Scenarios, Confidence Bands

Higher Emissions Scenario - Projected Temperature Change (°F)
From 1961-1979 Baseline

Mid-Century (2040-2059 average)  End-of-Century (2080-2099 average)
Facilitate Collaboration & Interaction

• Example: Collaborative Geodesign

• Goal: Improving water quality under limited budget

• Features
  – **Collaboration** to resolve conflicts
  – **Interactive** land allocation
  – Real-time visualization and feedback
  – Iterate till convergence
Opportunities: Beyond Geographic Space

• Spaces other than Earth
  – Challenge: reference frame?

• Ex. Human body
  – What is Reference frame?
    • Adjust to changes in body
    • For MRIs, X-rays, etc.
  – What map projections?
  – Define path costs and routes to reach a brain tumor?

<table>
<thead>
<tr>
<th></th>
<th>Outer Space</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moon, Mars, Venus, Sun, Exoplanets, Stars, Galaxies</td>
</tr>
<tr>
<td>Geographic</td>
<td>Terrain, Transportation, Ocean, Mining</td>
</tr>
<tr>
<td>Indoors</td>
<td>Inside Buildings, Malls, Airports, Stadiums, Hospitals</td>
</tr>
<tr>
<td>Human Body</td>
<td>Arteries/Veins, Brain, Neuromapping, Genome Mapping</td>
</tr>
<tr>
<td>Micro / Nano</td>
<td>Silicon Wafers, Materials Science</td>
</tr>
</tbody>
</table>

Outline

• Introduction
• GPS
• Spatial Database Management Systems
• Location Based Services
• Spatial Data Science
• Virtual Globes & Remote Sensing
• Geographic Information Systems
• Conclusions
Summary

• Spatial Data are ubiquitous & important

• Spatial Computing has transformed our society
  – It is only a beginning!
  – It promises an astonishing array of opportunities in coming

• Current Data Science Tools are inadequate
  – Gerrymandering, Spatial Auto-correlation, …

• Ask: Data Science Degrees should include
  – Spatial Data Science Methods…

A UCGIS Call to Action:
Bringing the Geospatial Perspective to Data Science Degrees and Curricula

University Consortium for Geographic Information Science

Summer 2018
References : Surveys, Overviews

- **Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data**, IEEE Transactions on Knowledge and Data Mining, 29(10):2318-2331, June 2017. (DOI: 10.1109/TKDE.2017.2720168).
- **Parallel Processing over Spatial-Temporal Datasets from Geo, Bio, Climate and Social Science Communities: A Research Roadmap**, IEEE BigData Congress 2017: 232-250.
## References: Details

### Colocations
- Discovering colocation patterns from spatial data sets: a general approach, *IEEE Trans. on Know. and Data Eng.*, 16(12), 2004 (w/ Y. Huang et al.).
- A join-less approach for mining spatial colocation patterns, IEEE Trans. on Know. and Data Eng., 18 (10), 2006. (w/ J. Yoo).

### Spatial Outliers
- Detecting graph-based spatial outliers: algorithms and applications (a summary of results), *Proc.: ACM Intl. Conf. on Knowledge Discovery & Data Mining*, 2001 (with Q. Lu et al.)

### Hot Spots

### Location Prediction

### Change Detection
Facial Recognition Is Accurate, if You’re a White Guy

Color Matters in Computer Vision

Facial recognition algorithms made by Microsoft, IBM and Face++ were more likely to misidentify the gender of black women than white men.

Gender was misidentified in up to 1 percent of lighter-skinned males in a set of 385 photos.

Gender was misidentified in up to 7 percent of lighter-skinned females in a set of photos.

Gender was misidentified in up to 12 percent of darker-skinned males in a set of 318 photos.

Gender was misidentified in 35 percent of darker-skinned females in a set of 271 photos.

Feb. 9, 2018

The New York Times
Trust and Ethics: FATE debate

- **Government View:** Security, Balance prosperity and civil society
- **Business View:** Innovation critical for prosperity but carries risks
- **Civil Society View:** Risks should be disclosed
  - Fairness (or equity): Reduce bias across gender, race, age, ...
  - Accountability: Determine and assign responsibility for a machine judgement
  - Transparency (or explainability): Be open and clear about (prediction) process
  - Ethics:
    - Privacy-preserving, Use case specific dilemmas
    - Trustworthy: Safe (Do no harm), Secure (Guard against malicious behavior)


(ii) *Data for Good: FATES, Elaborated*, J. Wing, Jan. 23, 2018. (iii) [FAT ML](https://www.fatml.org/) and [FATES](https://www.fatml.org/) Workshop

https://www.fatml.org/
Gerrymandering Risk in Traditional Data Science

- Traditional methods not robust in face of
  - Spatial continuity
    - Gerrymandering risk: Spatial partitioning affects Results (Modifiable Areal Unit Problem)
  - Auto-correlation, Heterogeneity, Edge-effect, …
  - Noise challenge data mining methods

<table>
<thead>
<tr>
<th>Partition A: Pearson’s Correlation</th>
<th>Pairs</th>
<th>Partition B: Pearson’s Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Red Circle" /> <img src="image2.png" alt="Blue Circle" /></td>
<td>- 0.90</td>
</tr>
<tr>
<td>- 0.90</td>
<td><img src="image3.png" alt="Yellow Circle" /> <img src="image2.png" alt="Blue Circle" /></td>
<td>1</td>
</tr>
</tbody>
</table>
Limitations of Traditional Data Mining: Association Rules

(a) Map of 3 item-types

(b) Spatial Partition P1

(c) Spatial Partition P2

(d) Spatial Partition P3

<table>
<thead>
<tr>
<th>Partitioning</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transactions</td>
<td>T11, T12, T13, T14</td>
<td>T21, T22, T23, T24</td>
<td>T31, T32, T33, T44</td>
</tr>
<tr>
<td>Associations with support &gt;= 0.5</td>
<td>( ▲ ● )</td>
<td>( □ ▲ )</td>
<td>( □ ▲ ● )</td>
</tr>
</tbody>
</table>