Motivation for Spatial Computing

- Societal:
  - Google Earth, Google Maps, Navigation, location-based service
  - Global Challenges facing humanity – many are geo-spatial!
  - Future of Computer Science (CS) is to address societal challenges!

by the Millennium Project of WFUNA
www.millennium-project.org
Spatial thinking in Business Analytics

• Experience of an e-commerce pioneer
  – Mid 1990s: Geography is dead in post-internet era.
  – Late 1990s: Logistics and delivery are our biggest challenge!

• Spatial questions are central to many businesses!
  – Where are our customers? Suppliers? Stake-holders? …
  – Where should we do business?
    • Which countries? (Globalization), Which cities? …
  – Where should we locate? – store, warehouse, factories, offices, …
  – Petroleum, Mining – where should one drill / dig ?
  – Where is geographic event, e.g. storm, earth-quake, flood ?
  – What is impact of this event on our organization?
  – …

• Should business analytics address spatial questions?
  – Where is a pattern prevalent ?
  – Context: Patterns involving geographic events, e.g. storm, climate…
  – Ex. Association (diaper, beer)
Spatial Thinking in Consumer Applications

• Trends:
Consumers account for two-third of US economy
Cell phone outnumber personal computers
Spatial apps dominate the Google android app. contest winner
Research Challenges in Spatial Computing

• Is spatial computing just an application of well-known CSE techniques?

• Are there CSE research challenges and opportunities?

• Dynamic Programming is a popular algorithm design paradigm
  • Shortest Path Algorithm
  • DBMS Query Optimization
  • Sequence alignment,
  • Viterbi algorithm, …

• However, DP assumes stationary ranking of candidate solutions
  • Is DP appropriate for longitudinal problems?
Outline

• Motivation

• Case 1: Infrastructure:
  • Database Management Systems (DBMS)
  • Routing
  • Evacuation route planning

• Case 2: Intelligence: Statistics, Data Mining
Relation DBMS to Spatial DBMS

• 1980s: Relational DBMS
  • Relational Algebra
  • Query Processing, e.g. sort-merge equi-join algorithm, …
  • B+ Tree index
• Spatial customer (e.g. NASA, USPS) got interested
• But faced challenges
  • Semantic Gap
    • Spatial concepts: distance, direction, overlap, inside, shortest paths, …
    • SQL representation was quite verbose
    • Relational algebra can not represent Transitive closure
  • Performance challenge due to linearity assumption
    • Is B+ tree appropriate for geographic data?
    • Is sorting natural in geographic space?
• New ideas emerged in 1990s
  • Spatial data types and operations (e.g. OGIS Simple Features)
  • R-tree, space partitioning, …
Spatial Databases: Representative Projects

Parallelize Range Queries

Evacuation Route Planning

- only in old plan
- Only in new plan
- In both plans

Shortest Paths

Storing graphs in disk blocks
Revisit Shortest Path Problem

New Routing Questions
- Best start time to minimize time spend on network
- Account for delays at signals, rush hour, etc.

Time-Variant Flow Network Questions

<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>

U.P.S. Embraces High-Tech Delivery Methods (July 12, 2007)
By “The research at U.P.S. is paying off. …….— saving roughly three million gallons of fuel in good part by mapping routes that minimize left turns.”
Routing in ST Networks

Predictable Future

Stationary

Non-stationary

Unpredictable Future

Dijkstra’s, A*…. 
Broader Implication of Stationary Assumption

• Dynamic Programming is a popular algorithm design paradigm
  • Shortest Path Algorithm
  • DBMS Query Optimization
  • Sequence alignment,
  • Viterbi algorithm, …

• However, DP assumes *stationary ranking* of candidate solutions
  • Is DP appropriate for longitudinal spatial problems?
Evacuation Route Planning - Motivation

- No coordination among local plans means
  - Traffic congestions on all highways
  - e.g. 60 mile congestion in Texas (2005)

- Great confusions and chaos

"We packed up Morgan City residents to evacuate in the a.m. on the day that Andrew hit coastal Louisiana, but in early afternoon the majority came back home. The traffic was so bad that they couldn't get through Lafayette."
Mayor Tim Mott, Morgan City, Louisiana
(http://i49south.com/hurricane.htm)
A Real Scenario

Nuclear Power Plants in Minnesota

Twin Cities

Prairie Island

Monticello
Monticello Emergency Planning Zone

Emergency Planning Zone (EPZ) is a 10-mile radius around the plant divided into sub areas.

Monticello EPZ
Subarea  Population
2          4,675
5N         3,994
5E         9,645
5S         6,749
5W         2,236
10N        391
10E        1,785
10SE       1,390
10S        4,616
10SW       3,408
10W        2,354
10NW       707
Total      41,950

Estimate EPZ evacuation time:
Summer/Winter (good weather): 3 hours, 30 minutes
Winter (adverse weather): 5 hours, 40 minutes

Data source: Minnesota DPS & DHS
Web site: http://www.dps.state.mn.us
http://www.dhs.state.mn.us
A Real World Testcase

Congestion is likely in old plan near evacuation destination due to capacity constraints. Our plan has richer routes near destination to reduce congestion and total evacuation time.

Experiment Result

Total evacuation time:
- Existing Plan: 268 min.
- New Plan: 162 min.
Problem Statement

Given
- A transportation network, a directed graph $G = (N, E)$ with
  - Capacity constraint for each edge and node
  - Travel time for each edge
- Number of evacuees and their initial locations
- Evacuation destinations

Output
- Evacuation plan consisting of a set of origin-destination routes
  - and a scheduling of evacuees on each route.

Objective
- Minimize evacuation egress time
  - time from start of evacuation to last evacuee reaching a destination

Constraints
- Route scheduling should observe capacity constraints of network
- Reasonable computation time despite limited computer memory
- Capacity constraints and travel times are non-negative integers
- Evacuees start from and end up at nodes
Summary of Related Works & Limitations

A. Capacity-ignorant Approach
- Simple shortest path computation, e.g. A*, Dijkstra’s, etc.
- e.g. EXIT89 (National Fire Protection Association)
**Limitation:** Poor solution quality as evacuee population grows

B. Operations Research: Time-Expanded Graph + Linear Programming
- Optimal solution, e.g. EVACNET (U. FL), Hoppe and Tardos (Cornell U).
**Limitation:**
- High computational complexity => Does not scale to large problems
- Users need to guess an upper bound on evacuation time
  Inaccurate guess => either no solution or increased computation cost!

<table>
<thead>
<tr>
<th>Number of Nodes</th>
<th>50</th>
<th>500</th>
<th>5,000</th>
<th>50,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVACNET Running Time</td>
<td>0.1 min</td>
<td>2.5 min</td>
<td>108 min</td>
<td>&gt; 5 days</td>
</tr>
</tbody>
</table>

C. Transportation Science: Dynamic Traffic Assignment
- Game Theory: Wardrop Equilibrium, e.g. DYNASMART (FHWA), DYNAMIT(MIT)
**Limitation:** Extremely high compute time
- Is Evacuation an equilibrium phenomena?
Representations of (Spatio-)temporal Networks

(1) **Snapshot Model** [Guting 04]

Node: \( N_i \)

Edge: Travel time

(2) **Time Expanded Graph (TEG)** [Ford 65]

Holdover Edge

Transfer Edges

(3) **Time Aggregated Graph (TAG)** [Our Approach]

olumbia aggregating over edges and nodes.
Performance Evaluation

Setup: fixed number of evacuees = 5000, fixed number of source nodes = 10 nodes, number of nodes from 50 to 50,000.

- CCRP produces high quality solution, solution quality increases as network size grows.
- Run-time of CCRP is scalable to network size.

---

**Figure 1** Quality of solution

**Figure 2** Run-time

<table>
<thead>
<tr>
<th>Number of Nodes</th>
<th>CCRP (unit)</th>
<th>NETFLO (unit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td>500</td>
<td>200</td>
<td>250</td>
</tr>
<tr>
<td>5000</td>
<td>300</td>
<td>350</td>
</tr>
<tr>
<td>50,000</td>
<td>350</td>
<td>400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Nodes</th>
<th>CCRP (second)</th>
<th>NETFLO (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>500</td>
<td>1.5</td>
<td>25.5</td>
</tr>
<tr>
<td>5000</td>
<td>23.1</td>
<td>962.2</td>
</tr>
<tr>
<td>50,000</td>
<td>316.4</td>
<td></td>
</tr>
</tbody>
</table>
Routing in ST Networks: Scalable Methods

Predictable Future

Stationary

Special case (FIFO)

Non-stationary

General Case

TEG: LP, Label-correcting

TAG: Transform to Stationary TAG

Dijkstra’s, A*….

travel times → arrival times at end node → Min. arrival time series

Non-stationary TAG

Stationary TAG
Outline

• Motivation
• Case 1: Infrastructure:
• Case 2: Intelligence
  • Data Mining
  • Statistics
Case 2: Data Mining (DM) to Spatial DM

• 1990s: Data Mining
  • Scale up to traditional models to large relational databases
    • Linear regression, Decision Trees, …
  • New pattern families
    • Association rules
    • Which items are bought together? E.g. (Diaper, beer)

• Spatial customers
  • Walmart
    • Which items are bought just before/after events, e.g. hurricanes?
    • Where is (diaper-beer) pattern prevalent?
  • Global climate change

• But faced challenges
  • Independence Assumption
  • Transactions, i.e. disjoint partitioning of data
Spatial Data Mining: Representative Projects

### Location prediction: nesting sites
- Nest locations
- Distance to open water
- Vegetation durability
- Water depth

### Spatial outliers: sensor (#9) on I-35

### Co-location Patterns

### Tele connections
Association Patterns

- Association rule e.g. (Diaper in T \implies Beer in T)

<table>
<thead>
<tr>
<th>Transaction</th>
<th>Items Bought</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{socks, milk, beef, egg, …}</td>
</tr>
<tr>
<td>2</td>
<td>{pillow, toothbrush, ice-cream, muffin, …}</td>
</tr>
<tr>
<td>3</td>
<td>{pampers, pacifier, formula, blanket, …}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>{battery, juice, beef, egg, chicken, …}</td>
</tr>
</tbody>
</table>

- Support: probability (Diaper and Beer in T) = 2/5
- Confidence: probability (Beer in T \mid Diaper in T) = 2/2

- Algorithm Apriori [Agarwal, Srikant, VLDB94]
  - Support based pruning using monotonicity

- Note: **Transaction is a core concept!**
Spatio-temporal Co-occurrence of Crime-Types
Pattern Family 4: Co-locations/Co-

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

Answers: 🌳🔥👥🏡
Co-occurrence in space and time!

- 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)
Co-occurring object-types

- 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)
Co-location: A Neighborhood based Approach

<table>
<thead>
<tr>
<th></th>
<th>Association rules</th>
<th>Colocation rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>underlying space</td>
<td>discrete sets</td>
<td>continuous space</td>
</tr>
<tr>
<td>item-types</td>
<td>item-types</td>
<td>events /Boolean spatial features</td>
</tr>
<tr>
<td>collections</td>
<td>Transactions</td>
<td>neighborhoods</td>
</tr>
<tr>
<td>prevalence measure</td>
<td>support</td>
<td>participation index</td>
</tr>
<tr>
<td>conditional probability measure</td>
<td>Pr.[ A in T</td>
<td>B in T ]</td>
</tr>
</tbody>
</table>

Challenges:

1. Computational Scalability
   Needs a large number of spatial join, 1 per candidate colocation
2. Spatio-temporal Semantics
   Spatio-temporal co-occurrences
   Emerging colocations
   ...

...
Spatial Prediction

Nest locations

Distance to open water

Vegetation durability

Water depth
Autocorrelation

• First Law of Geography
  – “All things are related, but nearby things are more related than distant things. [Tobler, 1970]”

• Autocorrelation
  – Traditional i.i.d. assumption is not valid
  – Measures: K-function, Moran’s I, Variogram, …
Implication of Auto-correlation

<table>
<thead>
<tr>
<th>Name</th>
<th>Model</th>
<th>Classification</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Linear Regression</td>
<td>[ y = x\beta + \epsilon ]</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Spatial Auto-Regression</td>
<td>[ y = \rho Wy + x\beta + \epsilon ]</td>
<td></td>
<td>High</td>
</tr>
</tbody>
</table>

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

\[ W : n \times n \] neighborhood matrix over spatial framework

**Computational Challenge:**
Computing determinant of a very large matrix in the Maximum Likelihood Function:

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

• Spatial Computing is critical to many societal grand challenges
  • Sustainable development, Environment
  • Energy, Water, Public Safety …

• It Challenges many CSE assumptions
  • Linearity assumption in relational DBMS
    • B+ tree, Sort-merge equi-join, …
  • Stationary assumption behind Dynamic Programming
    • Shortest Path problem
    • DBMS query optimization (Selinger style)
  • Independence assumption in Statistics, Machine Learning, …
    • Decision trees, Linear Regression, …

• Many disciplines are addressing spatial challenges
  • Spatial Statistics, Spatial Economics, Environmental Epidemiology …

• Time is ripe for broader participation from Business Analytics!
Spatial Thinking Across Disciplines!
Spatial Computing Questions

• How do we conceptualize spatio-temporal (ST) worlds?

• How do we measure ST concepts, recognize them in (remotely) sensed information or in the field, and identify their accuracy and quality?

• How do we represent ST concepts with incomplete/ uncertain information, with alternative data models, and possibly with multiple representations for the same data, in digital environments?

• How do we store, access, and transform ST concepts, facilitating data sharing, data transfer, and data archiving, while ensuring minimum information loss?

• How do we explain ST phenomena through the application of appropriate methods of forward or inverse models of physical and human processes?

• How do we visualize ST concepts on a variety of media such as maps on electronic displays or animated displays?  

• How do we use ST concepts to think about spatio-temporal phenomena, and to seek explanations for spatio-temporal patterns and phenomena?

• What ST issues or business organizations? How do we use ST concepts to think about business issues?

(Source: A daptation from NCGIA proposal to NSF by Goodchild et al.)