
Cyrus Shahabi
University of Southern California
Los Angeles, CA 90089-0781
shahabi@usc.edu
http://infolab.usc.edu
OUTLINE

• Introduction to IMSC
• Intelligent Transportation
• Project: TransDec
• Applications: Time-Dependent kNN
• Demonstration
• Future Plans
IMSC Background

- One of the 54 existing National Science Foundation (NSF) Engineering Research Centers, focusing on “Multimedia”
- ERC emphasis: Multidisciplinary, Industry Presence
- New management, team, and vision in July 2010
- Currently collaborating with 12+ faculty from:
  - Viterbi School of Engineering (CS, EE, ISE, Civil-E, Petroleum-E)
  - Keck School of Medicine
  - Annenberg School for Communication & Journalism
  - Price School of Public Policy
  - Dornsife College of Letters, Arts and Sciences
Remote Sensing: Satellite & Multi-Spectral Imagery

Aerial Sensing: LiDAR, aerial imagery, UAV video

Ground Sensing: Traffic loop detectors, CCTV, pollution stations

People Sensing: smart phones, GPS, navigation

Human Body Sensing

Exploiting the common coordinates of time and space to fuse these geo-data into actionable knowledge in order to deal with various natural, man-made and socioeconomic crises more effectively.
IMSC Projects

Vision: Geo-Immersion

Use Cases

1. Intelligent Transportation
2. Intelligent Surveillance
3. Intelligent Campus

Fundamental Research
System Integration Experiments (SIE)
Real-World Applications
OUTLINE

• Introduction to IMSC
• Intelligent Transportation
• Project: TransDec
• Applications: Time-Dependent kNN
• Demonstration
• Future Plans
Intelligent Transportation

PROBLEM

• Traffic congestion is a $87.2 billion annual drain on the U.S. economy:
  • 4.2 billion lost hours (one work week for every traveler)
  • 2.8 billion gallons of wasted fuel (three weeks worth of gas for every traveler)

1 Texas Transportation Institute Urban Mobility Report, 2007 data

Location data could save consumers worldwide more than $600 billion annually by 2020.
The biggest single consumer benefit will be 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.
Intelligent Transportation

PROBLEM

• Traffic congestion is a $87.2 billion annual drain on the U.S. economy\(^1\):
  • 4.2 billion lost hours (one work week for every traveler)\(^1\)
  • 2.8 billion gallons of wasted fuel (three weeks worth of gas for every traveler)\(^1\)

\(^1\) Texas Transportation Institute Urban Mobility Report, 2007 data

GOAL

• To improve the performance of the surface transportation network through:
  • Capturing real-time data from infrastructure and vehicles
  • Developing data-driven solutions to improve mobility by leveraging optimization opportunities (e.g., path planning for commuter groups)
Traffic Data Lifecycle

• **Loop Detectors**
  - Most commonly used traffic sensors
  - The data is collected in Detector Cabinet and relayed to the service provider
  - Provide two data fields: volume (count) and occupancy (% time a vehicle is over the sensor)
Traffic Data Lifecycle: Loop Detectors

Loop inductance decreases when a car is on top of it.
Traffic Data Lifecycle: Loop Detectors

- Single loops can measure:
  - Occupancy ($O$): % of time loop is occupied (had a car on it) per interval
  - Volume ($N$): vehicles per interval
  - Speed = $(N*L)/O$ where $L$ is a constant proportional to the average length of a car
Traffic Data Lifecycle: Data Aggregator

RIITS (Regional Integration of Intelligent Transportation Systems)

- A data network affiliated with Los Angeles County Metropolitan Transportation Authority (Metro)
- Collects and serves data from Caltrans, City of Los Angeles Department of Transportation (LADOT), California Highway Patrol (CHP), Long Beach Transit (LBT), Foothill Transit (FHT) and Metro

http://www.riits.net/
# Traffic Data Lifecycle

## A Data Management Problem

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Hourly (in KB)</th>
<th>Daily (in KB)</th>
<th>Annual (in KB)</th>
<th>3 Years (in KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bus_mta_inv2.xml</td>
<td>0.96</td>
<td>23.00</td>
<td>8,395.00</td>
<td>25,185.00</td>
</tr>
<tr>
<td>bus_mta_rt2.xml</td>
<td>532.50</td>
<td>3,195.00</td>
<td>76,680.00</td>
<td>279,882,000.00</td>
</tr>
<tr>
<td>cctv_inv.xml</td>
<td>0.04</td>
<td>57.00</td>
<td>20,805.00</td>
<td>62,415.00</td>
</tr>
<tr>
<td>cms_inv.xml</td>
<td>2.17</td>
<td>52.00</td>
<td>18,980.00</td>
<td>56,940.00</td>
</tr>
<tr>
<td>cms_rt.xml</td>
<td>38.40</td>
<td>55,296.00</td>
<td>20,183,040.00</td>
<td>60,549,120.00</td>
</tr>
<tr>
<td>event_d7.xml</td>
<td>8.80</td>
<td>12,672.00</td>
<td>4,625,280.00</td>
<td>13,875,840.00</td>
</tr>
<tr>
<td>rail_mta_inv.xml</td>
<td>0.04</td>
<td>1.00</td>
<td>365.00</td>
<td>1,095.00</td>
</tr>
<tr>
<td>rail_rt.xml</td>
<td>8.00</td>
<td>11,520.00</td>
<td>4,204,800.00</td>
<td>12,614,400.00</td>
</tr>
<tr>
<td>rms_inv.xml</td>
<td>36.04</td>
<td>865.00</td>
<td>315,725.00</td>
<td>947,175.00</td>
</tr>
<tr>
<td>rms_rt.xml</td>
<td>1.45</td>
<td>87.29</td>
<td>2,095.00</td>
<td>5,159,139.840.00</td>
</tr>
<tr>
<td>signal_inv.xml</td>
<td>988.80</td>
<td>59,328.00</td>
<td>1,423,872.00</td>
<td>3,541,926,400.00</td>
</tr>
<tr>
<td>signal_rt.xml</td>
<td>3,514.67</td>
<td>50,61,120.00</td>
<td>1,847,308,800.00</td>
<td>5,419,264,000.00</td>
</tr>
<tr>
<td>tt_d7_inv.xml</td>
<td>0.52</td>
<td>31.08</td>
<td>746.00</td>
<td>2,722,900.00</td>
</tr>
<tr>
<td>tt_d7_rt.xml</td>
<td>152.00</td>
<td>912.00</td>
<td>218,880.00</td>
<td>798,91,200.00</td>
</tr>
<tr>
<td>vds_art_d7_inv.xml</td>
<td>0.08</td>
<td>4.79</td>
<td>115.00</td>
<td>41,979.00</td>
</tr>
<tr>
<td>vds_art_d7_rt.xml</td>
<td>45.00</td>
<td>2,700.00</td>
<td>64,800.00</td>
<td>23,652,000.00</td>
</tr>
<tr>
<td>vds_art_ladot_inv.xml</td>
<td>1.76</td>
<td>105.75</td>
<td>2,539.00</td>
<td>926,370.00</td>
</tr>
<tr>
<td>vds_art_ladot_rt.xml</td>
<td>969.00</td>
<td>58,140.00</td>
<td>1,395,360.00</td>
<td>5,093,064,000.00</td>
</tr>
<tr>
<td>vds_fr_d7_inv.xml</td>
<td>0.66</td>
<td>39.88</td>
<td>957.00</td>
<td>349,305.00</td>
</tr>
<tr>
<td>vds_fr_d7_rt.xml</td>
<td>722.00</td>
<td>43,320.00</td>
<td>1,039,680.00</td>
<td>379,483,200.00</td>
</tr>
<tr>
<td>Total KB from XML data</td>
<td>13980</td>
<td>6,985.28</td>
<td>10,057,449.00</td>
<td>36,709,688.855.00</td>
</tr>
</tbody>
</table>

**Heterogeneous (gps, video, loop sensor, events)**

**Continuous**

**Large**
OUTLINE

• Introduction to IMSC
• Intelligent Transportation
• Project: TransDec
• Applications: Time-Dependent kNN
• Demonstration
• Future Plans
IMSC TransDec

2000: Outside-IMSC fundamental research

USC Infolab: Acquisition, storage, query and analysis of data streams

Aug 2010: Los Angeles Metropolitan Transportation Authority (LAMTA)

ProBio: An End-to-End Bioinformatics Platform for the Analysis of DNA Microarray Experiments

Dec 2010: Microsoft

U2lization of Stream-Insight and Azure cloud

August 2011: NSF

Real-World Traffic Data Management for Time-Dependent Spatial Queries

January 2012: USC Stevens

Proof-of-concept prototype & market analysis

Spinoff – ClearPath (Amplify Incubator?)
TransDec System

Input Traffic Data

Data Processing

Storage

Query Retrieval & Visualization

Highway
Arterial
Bus & Rail
Ramp meter
Event
CMS

MICROSOFT STREAMINSIGHT

BIG DATA!

MOUNTAIN

SQL Azure

ORACLE DATABASE 11G
OUTLINE

• Introduction to IMSC
• Intelligent Transportation
• Project: TransDec

• Applications: Time-Dependent kNN
• Demonstration
• Future Plans
Outline

• KNN Search Past and Present
• Motivation
• Related Work
• Time-Dependent Spatial Network
• TD-kNN Search Approaches
  – Time-Expanded Networks
  – Incremental Network Expansion
  – Indexing Time-Dependent Networks
• Conclusion
Evolution of kNN Search

Euclidean Space

Time-Dependent Spatial Network (2003-2010)
Outline

• KNN Search Past and Present
• Motivation
• Related Work
• Time-Dependent Spatial Network
• TD-kNN Search Approaches
  – Time-Expanded Networks
  – Incremental Network Expansion
  – Indexing Time-Dependent Networks
• Conclusion
Motivation

• Existing k-NN research (2003-2010)
  – Find k-NN based on the constant edge weights for each edge, (i.e., usually the maximum allowed speed-> minimum travel-time)

• In Real-world
  – The weight of an edge is a function of time, i.e., time-dependent.
  – Arrival-time to an edge determines the travel-time on that edge.

Pictures courtesy: http://www.wfrc.org/cms

Monday travel-time on a segment of I-10 in LA
(generated based on two years of historical traffic sensor data)
Time-dependent kNN (TD-kNN)

- Given a query point $q$ and a set of objects $P$, find the $k$ objects in $P$ that are closest to $q$ in time-dependent cost (e.g., travel-time)

- The shortest path from source to candidate hospitals may change depending on the departure-time

- The result of kNN depend on when the query is issued
Outline

• KNN Search Past and Present
• Motivation
• Related Work
• Time-Dependent Spatial Network
• TD-kNN Search Approaches
  – Time-Expanded Networks
  – Incremental Network Expansion
  – Indexing Time-Dependent Networks
• Conclusion
Related Work

Spatial Networks

Shortest Path

- Dijkstra
- A*

kNN Search

- Query processing in SNDB: Papadias et al., [VLDB03]
- Voronoi-based kNN in SNDB: Shahabi et al., [VLDB04]
- CNN queries in RN: Cho et al., [VLDB05]
- CNN monitoring in RN: Mouratidis et al., [VLDB06]
- S-Grid: Huang et al., [SSTD07]
- Scalable network distance browsing: Samet et al., [SIGMOD08]

Most kNN approaches are based on precomputation of static SP algorithms
Related Work

Time-Dependent Spatial Networks

Time-Dependent Shortest Path

Fixed Edge Cost (FC)
- Cook and Halsey, [JMA’66] – Dynamic Programming
- Kohler et al., [ESA’02] – Time-Expanded Graphs

Variable Edge Cost (VC)
- Dreyfus, [JOR’69] – Dijkstra Variant in FIFO
- Harpen, [MMO’69] - Non-FIFO, NP Hard
- Orda and Rom, [JACM’90] – Bellman Ford Algorithm
- George and Shekhar, [SSTD’07] - Time-Aggregated Graphs
- Bolin et al, [EDBT’08] - Dijkstra Variant in FIFO and non-FIFO

Recent Interest
- Foschini et al, [SODA’11]
- Malviya et al, [ICDE’11]
- Nannicini and Delling, [INFORMS’10]
- Demiryurek et al, [SSTD’11]

Time-Dependent kNN Search

Not Studied

Both of FC and VC algorithms can be extended to address TD-kNN. However, they are either approximate or do not scale.

Precomputation in time-dependent networks is a big challenge.
Outline

• KNN Search Past and Present
• Motivation
• Related Work
• Time-Dependent Spatial Network
• TD-kNN Search Approaches
  – Time-Expanded Networks
  – Incremental Network Expansion
  – Indexing Time-Dependent Networks
• Conclusion
Time-dependent Spatial Network

- $w_{12}(t), w_{34}(t)$
- $w_{23}(t)$
- $w_{24}(t)$
- $w_{13}(t)$

- $G(V,E,T)$: For every edge $e(v_i; v_j)$, there is a cost function $w_{ij}(t)$ which specifies the cost of traveling from $v_i$ to $v_j$ at time $t$.
- First-In-First-Out (FIFO) Network: Moving objects exit from an edge in the same order as they entered into that edge.
Time-dependent Spatial Network (Cont)

- **Path (s-d) Cost Functions**
  - Composition of edge arrival functions.
  - Arrival-time (and cost) to destination is a function of departure-time from the source.

\[
p_1 = (v_1, v_2, v_4) \quad f_{p1} = f_{24}(f_{12}(t)) \\
p_2 = (v_1, v_2, v_3, v_4) \quad f_{p2} = f_{34}(f_{23}(f_{12}(t))) \\
p_3 = (v_1, v_3, v_4) \quad f_{p3} = f_{34}(f_{13}(t))
\]
Time-dependent Spatial Network (Cont)

• Path (s-d) Cost Functions
  – Shortest path is **not unique** and changes based on the departure time
  – The lower-envelope of $f_p(t)$
    • Each piece gives the shortest path for the corresponding time-interval
    • Exponential pieces [Dean04] $\rightarrow$ exponential number of paths

**Pre-computation**
One needs to consider all possible paths (since SP is not unique) between all possible source and destination nodes
Time-dependent Spatial Network (Cont)

- Path (s-d) Cost Functions  (Static Networks)
  - Shortest path is unique
  - The lower-envelope of $f_p(t)$
    - There is only one piece i.e., the shortest path
Outline

• KNN Search Past and Present
• Motivation
• Related Work
• Time-Dependent Spatial Network
• TD-kNN Search Approaches
  – Time-Expanded Networks
  – Incremental Network Expansion
  – Indexing Time-Dependent Networks
• Conclusion
TD-kNN Approach-1
Time Expanded Networks (TEN)

- Given $G(V,E,T)$, discretize the time domain $T$ into $n$ points of time, and construct $n$ $G(V,E)$ graphs. [Koehler02]
TD-kNN Approach-1
Time Expanded Networks (TEN)

• The weight of an edge in TEN is the time difference between the time events associated with its endpoints

• A time-dependent edge cost is represented as a static flow in the corresponding TEN
TD-kNN Approach-1
Time Expanded Networks (TEN)

• Pros
  – Enables time-dependent kNN problem to be solved by applying techniques developed for static networks hence recomputation is possible

• Cons
  – **High storage:** The network size is increased proportional to the number of snapshots \( n \)
  – **Approximate Results:** The state of the network between two snapshots is not captured
    • The query time can be between two snapshots \( t_1 \) and \( t_2 \) (e.g., \( t=12 \)). However, only the edge weights at \( t_1 \) or \( t_2 \) can be used, hence causing errors. The error is accumulated along the path and is unbounded.
TD-kNN Approach-2

Incremental Network Expansion (INE)

• Modified Dijkstra algorithm for time-dependent spatial networks \textbf{[Dreyfus 69]}
  - Analogous to shortest path distances, arrival-time to nodes is used as the labels that form the basis of the greedy algorithm

• Expand the network based on the arrival-time to each node around \( q \) until \( k \) objects are found
TD-kNN Approach-2
Incremental Network Expansion (INE)

• Pros
  – INE provides exact results as compared to TEN

• Cons
  – Slow response time: the overhead of network expansion is very high particularly in large networks with a sparse set of data objects, hence not applicable to online apps
TD-kNN Approach-3
Indexing Time-Dependent Spatial Network [DEXA2010]

- Edge cost function

Lower-bound travel-time (LTT) of an edge is traversing that edge with Maximum possible speed

Grow SP trees from each site simultaneously using UTT for one site and LTT for the other sites

Repeat the process for all sites and find Tight Cells (TC)
TD-kNN Approach-3
Indexing Time-Dependent Spatial Network

• Loose Cells (LC)

• Loose Cells cover the entire network.
• Any query point $q$ outside of the Loose Cell of $p$ is guaranteed not to have $p$ as its NN
• If $q$ is not inside any TC, it must be inside one (or more) LC(s) and the generator of those LCs are the only NN candidates.

Index TCs and LCs with a spatial index (e.g., R-tree, Quad-tree) to expedite the process of finding the tight/loose cell(s) that contain $q$. 

Grow SP trees from each site simultaneously using LTT for one site and UTT for the other sites
TD-kNN Approach-3
Indexing Time-Dependent Spatial Network

• Pros
  – Provides exact results
  – Localize the NNs and minimize the need for time-dependent SP calculation
  – Scalable and efficient for large set of query and data objects, and large networks

• Cons
Experimental Results [DEXA’10]

Naïve Approach = INE with Dreyfus’ Dijkstra

k vs Response Time

k vs Network Node Access
Outline

• KNN Search Past and Present
• Motivation
• Related Work
• Time-Dependent Spatial Network
• TD-kNN Search Approaches
  – Time-Expanded Networks
  – Incremental Network Expansion
  – Indexing Time-Dependent Networks
• Conclusion
Conclusion

• In real-world edge weights are time-dependent.
  – New techniques needed to extend spatial query processing (such as kNN queries) in RN to a new family of time-dependent query processing solutions.

• Existing approaches (TEN and INE) can be extended to address time-dependent kNN; not efficient and exact.

• Indexing time-dependent network (on TCs and LCs) is an efficient and scalable approach.
OUTLINE

• Introduction to IMSC
• Intelligent Transportation
• Project: TransDec
• Applications: Time-Dependent kNN
• Demonstration
• Future Plans
OUTLINE

• Introduction to IMSC
• Intelligent Transportation
• Project: TransDec
• Applications: Time-Dependent kNN
• Demonstration
• Future Plans
Future Plans

- Data Reduction & Modeling
  - Spatio-temporal compression: DCT, SVD, Sampling, Wavelets, ...
  - Use models to generate traffic data for un-equipped roads
- Time-dependent shortest path
  - Evaluate the proposed indexing techniques
  - Other spatial queries: RkNN, spatial-skyline, ...
- The CT Project: Closed-loop, holistic, real-time transportation optimization
  - Complementing analysis with decision-making and actuation to close the optimization loop
  - Developing scalable optimization solutions that consider the integrated transportation system as a whole
  - Introducing efficient data analysis techniques to enable on-the-fly decision making
References


• Ugur Demiryurek, Bei Pan, Farnoush Banaei-Kashani, and Cyrus Shahabi. Towards Modeling the Traffic Data on Road Networks, ACMGIS 2009 Workshop on Computational Transportation Science, Seattle, Washington, November 2009


• F. Dehne, M.T. Omran, and J. Sack. Shortest paths in time-dependent FIFO networks using edge load forecasts. ACMGIS-IWCTS, 2009


• N. Malviya, S. Madden, and A. Bhattacharya A Continuous Query System for Dynamic Route Planning. In ICDE, 2011


• D. Delling and G. Nannicini. Core routing on dynamic time-dependent road networks. INFORMS 2010
Thanks!