On-line Spatio-Temporal Analysis of Network Data and Road Developments

CASA Conference April 13 2010

Tao Cheng (tao.cheng@ucl.ac.uk) + Team
Department of Civil, Environmental & Geomatic Engineering, UCL

Outline

• Introduction
  - Background and aim

• Methodology - integrated ST Data Mining
  - Statistical approach
  - Machining learning
  - Visualization
  - Simulation

• Programme and Progress
Background

- Large cities are increasingly crowded - population & mobility
- Traffic congestion affects both the economy and daily life.
- It is difficult and expensive to increase the capacity of the road network.

City of London

- Traffic levels in the Congestion Charging Zone are falling but congestion levels are rising.
- cost of congestion
  - £3 billion per year
- Mayor’s traffic priorities
  - reduce congestion and smooth traffic flow
- Removal of western extension of CC (27/11/2008)
- Olympic Games 2012
  - travel time to London Olympic sites
Why is this?
- Reduction in network capacity?
- Reallocations of capacity to other uses?
- Reduced resilience of the network?

**Aim** - To understand the traffic congestion in central London
- To quantitatively measure road network performance
- To understand causes of traffic congestion
  - association between traffic and interventions
    - traffic flow, speed/journey time
    - incidents, road works, signal changes and bus lane changes
- Case study - London

**Challenge (1) Network Complexity**

1) Dynamics
2) Spatial dependence
3) Spatio-temporal interactions
4) Heterogeneity
Challenge (2) - Data issues

- massive - 20GB monthly
- multi-sourced related to 5 different networks
- different scales (density & frequency)
- variable data quality
- contain conflicts, errors, mistakes and gaps

DATA COVERAGE
London Road Networks

Cordons
Central, Inner, Outer

Screenlines
Thames, Northern, five radials four peripherals

Traffic Flow Surveys

- NMC (National manual count annual data)
- ATC (Automatic Count) - 20 MB
- different time periods, intervals and accuracy
Traffic speed (and hence journey time) data

- **MCOS** (Moving Car Observer Surveys)
  - Centre, Inner, and Outer
  - least accurate of the datasets
- **ITIS** (GPS vehicle tracking system) - 2GB
  - major A roads and bus routes in town with 2000 probes
  - medium accuracy
- **ANPR** (Automatic Number Plate Reading) - 6 GB
  - main roads in the central and west extensions of CCZs
  - 5-minute intervals, 5 vehicle groups,
  - high accuracy
  - available since March 2008
- **At least 5 networks**
  - boundaries do not fully align

LTIS incident and event data - 20MB

- works, hazards, accidents, signal faults, special events, breakdowns, security, and other causes
- DfT have all these data as map or as text files

  - Minimal, Moderate, Serious or Severe subjective?
  - unrecorded?
  - not geocoded?

  - not broadcast on the traffic Link website, creating problems in analysis and reporting.

There are uncertainties and gaps in the intervention data
What’s new - (1) data-driven, top-down
Transition in data availability

- **Data scarcity:**
  - High cost
  - Low volume
  - Intensive validation

- **Data abundance:**
  - High volume
  - Multiple kinds and sources
  - Extensive application

Transition in modelling approach

- **Bottom-up:**
  - Mechanistic
  - Explicit representation of behaviour (origin, dest, model, time ...)
  - System properties by aggregation

- **Top-down:**
  - Phenomenological
  - Describe system gross of all behavioural responses
  - Direct to objectives
What’s new: (2) integrated space and time

Existing ST analysis methods
- time series analysis + spatial correlation
- spatial statistics + the time dimension
- time series analysis + artificial neural networks

ST dependence ≠ space + time
Integrated modelling of ST is needed -
- seamless & simultaneous
- ST-association/autocorrelation

What’s new: (3) hybrid/quantitative approach

- combine regression analysis with machine learning
  - improve the sensitivity and explanatory power
- study the heterogeneity and scale of road performance
  - optimal scale for monitoring
- Quantitative assessment of network performance
  - Sensible decision making & policy evaluation
Principle of ST Modelling

Space-time data = global (deterministic) space-time trends +
local (stochastic) space-time variations

\[ Z_i(t) = \mu_i(t) + e_i(t) \]

- \( z_i(t) \) - the observation of the data series at spatial location \( i \) and at time \( t \);
- \( \mu_i(t) \) - space-time patterns that explain large-scale deterministic space-time trends and can be expressed as a nonlinear function in space and time.
- \( e_i(t) \) - the residual term, a zero mean space-time correlated error that explains small-scale stochastic space-time variations.

Cheng, Wang, Li (forthcoming, Geographical Analysis)

Model 1 - STARIMA - Spatio-Temporal Auto-Regressive Integrated Moving Average

\[ z_i(t) = \sum_{k=1}^{p} \sum_{h=0}^{m_k} \phi_{kh} W^{(h)} z(t - k) - \sum_{l=1}^{q} \sum_{h=0}^{n_l} \theta_{lh} W^{(h)} e(t - l) + e(t) \]

(Pfeifer P E and Deutsch S J, 1980)
Our approach – Integrated modelling of ST

Model 1 – STARIMA

\[ z_i(t) = \sum_{k=1}^{p} \sum_{h=0}^{m} \phi_{kh} W^{(h)} z(t-k) - \sum_{l=1}^{q} \sum_{h=0}^{n} \theta_{lh} W^{(h)} \epsilon(t-l) + \epsilon(t) \]

- define weights based upon spatial distance and spatial adjacency
- consider anisotropy
- able to model spatially continued phenomena

Tao Cheng, Jiaqiu Wang, Xia Li, 2010 A hybrid approach for space-time series of environmental data, Geographical Analysis (forth coming)

Model 2 - ANN - Artificial Neural Networks

(Mandic D P and Chambers JA, 2001)

SFNN - spatial interpolation

\[ \hat{z}_i = \sum_{j=1}^{n} iw_{ij} \cdot z_j + b \]

(a) static neuron

DRNN - time series analysis

\[ \hat{z}(t) = iw \cdot z(t) + lw \cdot \hat{z}(t-1) + b \]

(b) dynamic neuron

• **ANN for space-time trend analysis**

\[ \hat{\mu}_i(t) = f(\sum_{k=1}^{n} \beta_k f(i,t) + \beta_0) \]


---

**Model 2 - STANN**

\[ \hat{z}_i(t) = \sum_{j=1}^{n} \text{lw}_j^{(i)} \cdot z_j(t-1) + \text{lw}^{(0)} \cdot \hat{z}_i(t-1) + b \]

- One step implementation of ANN+ STARIMA
- Accommodate ST associations in ANN
- Deal with nonlinearity & heterogeneity in BP learning

Model 3 – SVM - Support Vector Machines

SVC & SVR (Vapnik et al, 1996)  
Input Space  
Feature Space

Model 3 - STSVR

- Nonlinear Spatio-Temporal Regression by SVM
  \[ \phi(f_T, f_S) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K((f_T, f_S), (f_T', f_S')) \]

- Develop ST kernel function
- Overcome over-fitting in STANN
- Deal with errors
- Model nonlinearity & heterogeneity
Other methods

- Geographically Weighted Regression (GWR)
  - -> STWR?
- Permutation Scan Statistics (PSS)
  - -> STPSS? (or STC)
- Exploratory Visualization (DM) + ST+OLAP
  - -> STOEV?
- Simulation (Multi-scales)
  - -> STMSS?

Progress
GUI: A Web-base Platform for Dynamic Visualization, Simulation, Analysis (OLAP)

Tool Boxes: Integrated Spatio-Temporal Data Mining (Matlab+..?)

PatternJH/BA Clustering 2.1
Pattern Transition 2.2
Intervention Analysis 2.3
Performance Prediction 2.4
Model Updating 2.5

Database/Platform(Oracle + ArcGIS)
(ANPR, GPS, ITLS, …. based on ITN)

STANDARD Platform Structure

STARIMA for Journey Time Prediction in London

Study area

London Arterial Road Map
Pattern analysis of journey time

The distribution plot of 33 Mondays journey times link 605 during 07:00 to 19:00 (2009 Jan. – Aug.)

Space-time analysis
Accuracy

Prediction Accuracy at different prediction intervals

<table>
<thead>
<tr>
<th>Forecasting Horizon</th>
<th>5 min</th>
<th>10 min</th>
<th>15 min</th>
<th>20 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of validate prediction</td>
<td>96</td>
<td>93</td>
<td>86</td>
<td>70</td>
</tr>
<tr>
<td>Mean relative error</td>
<td>0.07%</td>
<td>0.25%</td>
<td>0.44%</td>
<td>0.81%</td>
</tr>
<tr>
<td>Standard deviation of relative error</td>
<td>0.16%</td>
<td>0.38%</td>
<td>0.77%</td>
<td>1.27%</td>
</tr>
</tbody>
</table>

Comparison of results from extended STARIMA model and standard STARIMA model (Kamarianakis and Prastacos, 2005) at 5 min interval

<table>
<thead>
<tr>
<th>Number of validate prediction</th>
<th>Mean relative error</th>
<th>Standard deviation of relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended STARIMA</td>
<td>96</td>
<td>0.07%</td>
</tr>
<tr>
<td>Standard STARIMA</td>
<td>95</td>
<td>0.11%</td>
</tr>
</tbody>
</table>
Visualization of traffic congestion in space-time

Figure 1. Delay at 9 am on 12th April 2009

Cheng, Emmonds, Tanaksaranond, Sonoiki (2010, GISUK)

Figure 2. Delay at 9:15 am on 12th April 2009
LCAP 15 January 2010 8:00-10:00 am
Isosurface

High delay value (red color)

Sideview

Topview
Cheng, Anbaroglu (2010, SDH)

James Haworth
Department of Civil, Environmental & Geomatic Engineering, UCL

Multi-scale analysis of road network performance

- Using spatio-temporal data mining techniques to look for patterns in congestion at varying spatial and temporal scales
- What patterns can be observed in inbound and outbound congestion...
  - Daily? Weekly? Seasonally?...
- Identification of recurrent and non-recurrent congestion in London
Understanding Road Congestion as an Emergent Property of Traffic Networks

**Macroscopic**

Flow and economic models based on 'known' road capacity

**Microscopic**

Individual behaviours simplistic and route choice macroscopic-driven

---

**Formal model of Emergence**

**Link level**
What causes congestion to emerge at link level?
What is the effect of road layout?

**Junction level**
Are junctions the key source of congestion?
What choices are available to drivers?

**Network level**
How does congestion spread to the whole network?

-Manley, Cheng (2010, IMCIC)
Can we predict/migrate emergence (congestion) of Road Network?

- Understand
- Detect
- Model
- Simulation

Conclusion - Network Complexity