Spatio-Temporal Analysis for Wireless Network Data

G. Subramaniam, C. Goodall, D. Lin & R. Varadhan

AT&T Labs - Research
Johns Hopkins University
Introduction

Data

- AT&T collects mobility traffic data from thousands of cell sites.
- This is very high dimensional data (by markets, handset tech, upload/download, Data/Mins. etc.
- The data (kb) used in the analysis is average monthly volumes observed over 2 years by location (lat & long)

Goal is to explore spatial characteristics for the mobility data:

- Spatial attributes of various recorded mobility data (kilobytes – MB and GB, minutes of use etc.)
- Spatial temporal properties of site time series.
- Prediction
Distribution of KB - Data Usage & Minutes - Voice Usage

Points (size) Proportional to Quartiles
We are averaging the time series of data volumes and MOU across a time interval (29 months) to get purely spatial data.

Data consist of finite collection of spatial locations \( \{s_1, \ldots, s_n\} \) where \( s_i \) are coordinates in the space \( \mathbb{R}^2 \) and \( \{z(s_1), z(s_2), \ldots, z(s_n)\} \) are observed values at these locations.

Variogram used instead of covariance function for spatial data.

Parameters of variogram: range, sill and nugget effect.

Experimental Variogram is:

\[
2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum N(h)[z(s_i) - z(s_j)]^2
\]

Here \( N(h) \) is the set of all distinct pairs of points \( z(s_i), z(s_j) \) such that \( |s_i - s_j| = \pm h \)

Variogram cloud: the set of all pair-wise distance vectors combined with the squared differences of the observations \( z(s_1), z(s_n) \) a variogram cloud is then \( (s_j - s_i, (z(s_j) - z(s_i))^2) \) (\( i \neq j, i, j = 1, \ldots, n \)). So this is all pairs, without binning.
R Demo: Variograms Cloud

Spatial Analysis of Mobility Data
Regional Variograms
Variograms without Spatial Outliers

G. Subramaniam, C. Goodall, D. Lin & R. Varadhan (AT&T Labs - Research Johns Hopkins University)

Spatial Analysis of Mobility Data
When you have space-time data, the analysis can be done in steps:

1. Separate spatial analyses for each time point
2. Separate temporal analyses for each spatial location
3. Spatio-temporal analysis: Let $z(s_i, t)$ denote the measurement of data/voice usage at site $s_i$ and at time $t$ where $i = 1, \ldots n$ and $t = 1, \ldots T$.
4. Assuming the process is stationarity in space and time, we have
   \[ \gamma(h, t) = \frac{1}{2} \text{Var}(z(s, t) - z(s + h, t + k)) \]
5. The empirical spatio-temporal semivariogram estimator is given by
   \[ \hat{\gamma}(h, t) = \frac{1}{2|N(h, k)|} \sum_{N(h, k)} \{z(s_i, t_i) - z(s_j, t_j)\}^2 \]
6. Models for spatio temporal analysis
For selected sites (marked red) on the cloud that are close to each other have high semi variance in Winter (2009) months.

For other seasons in 2009 into 2010 the semi variance declined (correlation increased) for the same sites.

Provides an example of spatial & temporal anomaly.
Upgraded tower & neighboring towers

Time Series Plot & Semivariance over Time

KBs (scaled)

Months

0.0e+00 5.0e+15 1.0e+16 1.5e+16

Semivariance

Months

G. Subramaniam, C. Goodall, D. Lin & R. Varadhan (AT&T Labs - Research Johns Hopkins University)

Spatial Analysis of Mobility Data
Space time Variogram

Gamma

SpaceLag

TimeLag

G. Subramaniam, C. Goodall, D. Lin & R. Varadarajan

Spatial Analysis of Mobility Data
Next Steps

- Automate the detection of spatial outliers.
- Introduce formal models for spatio-temporal analysis in R
- Package in R for interactively visualizing spatio-temporal effects, outliers etc.
Prediction on Grid

interpolation Points & Sample grid

Kb inverse distance weighted interpolations
Variogram Maps