Meng:

Climate modeling in spatio-temporal analysis

High-performance computing

Overview of spatial statistical analysis:

1. ESDA:
2. Spatial autocorrelation analysis

Nuisance autocorrelation: from data process

Substantive autocorrelation: real autocorrelation

LISA: identifying hot/cold spots

1. Spatial clusting
2. Extend to space-time

Data checking: detect anomalies and recording errors; Model validation and diagnostics

1. Spatial regression/econometrics
2. Spatial dependence
3. Spatial heterogeneity (Data are dependent, but the dependent relationships vary across the space)

Homogeneity

Extreme structural instability

Discrete spatial heterogeneity (spatial regimes model: build model for different region)

Continuous spatial heterogeneity: GWR

Spatial panel data models

Spatial latent variables

Bayesian spatial hierarchical models (MCMC)

Markov random field-based Bayesian classifiers (data mining)

Spatial data mining:

Supervised vs. unsupervised learning (spatial classification)

Spatial autocorrelation and heterogeneity

1. Spatial optimization
2. Location models
3. Land use planning models
4. Network design and protection models

Spatial data mining

Spatial outlier detection

Pysal

Anomalies, clustering, co-location and prediction

Cluster vs. hot spot

Hot spot analysis (LISA: HH and LL, positive correlated)

HL and LH, negative correlated.

Hot spot is more extreme than the cluster?

Hot spot with location and attribute.

Cluster contains hot spot.

Spatial data:

Exploratory spatial data analysis (Moran’s I)

Data mining: prediction

When we have a new set of spatial data, we need to check if the assumption from traditional statistics hold. If it holds, then we can apply regular regression or t-test. If not, we need to use some spatial statistics methods.

Project:

Geostatics: kriging

GWR

Cross modules: visualization, parallelization, documentation (notebook project – organization - enhancement)

Application of PySAL to your own thesis/dissertation paper

After 2 weeks, we need to have a more specific project description.

Documentation of the module.

Notebook to illustrate the code.

02012016

Point pattern analysis

First order properties: (heterogeneity)

Mean/average value

Second order properties: (interaction)

Distance and direction

Combine of first order and second order

Kernel estimate

Kernel #1 = disease case

Kernel #2 = disease control

Ratio = disease + risk factors

K function above the complete randomness, that is cluster

Space and Time in K functions

Edge effect: the same assumptions cannot be used at the boundary of a space

Correction methods:

1. Buffer zones
2. Toroidal edge correction
3. By weighting (scale up the circle to contain points outside of the study area) (doesn’t really matter if we run simulation envolopes)

CSR/K-function to access the point pattern (cluster)

Upper/lower bound of significant level

Edge effect: GWR (nearest neighbor)

Quadrats

Cluster/ Disperse

Variance: large (cluster)

Small (infirmity)

Departure from CSR (cluster/infirmity)

Simple k-function:

Function of distance

Putting circle on each of the point, count the event number in the circles.

Function d (change d)

Poisson: CSR

Counts (for positive integer + 0)

Univariate process

Taking random sample from the pool dataset, do the sampling

Re-doing the k-function with the random sample

Similar to cross-validation.

Population at-risk: correct disease with very low probabilities.

Increase the denominator: grow study area (pulling together blocks)

Point pattern: network

Detection of individual clusters (local analysis)

For local statistics, the effect of ignoring network will have more influences.

Standardize the population density (tweets are not random)

Stationarity: no randomness

Isotropic: direction/orientation doesn’t matter.

First order stationarity: intensity is constant. Variance is stable.

Second order stationarity: covariance between the counts. Distance between the points.

Isotropic:

Heat map based on kernels. Kernel represents the process that can actually happen. So sometimes it will not be isotropic.

Adaptive kernel (adaptive bandwidth)

High intensity: smaller bandwidth

Smooth: larger bandwidth