Spatial Interpolation using Multiple Regression

ICDM'2012 - Brussels

Orlando Ohashi and **Luís Torgo**

LIAAD - INESC TEC / DCC - Faculty of Sciences
University of Porto - Portugal

13-Dec-2012

Spatial Interpolation/Imputation

Problem Definition/Motivation

- Filling in unknown values in geo-referenced data sets
- Data collection is not fully controllable and it is prone to failures
- Data incompleteness may be caused by poor data collection, measurement errors, costs management and many other factors

Illustrative Application Areas

Wind speed forecasting, oil resources analysis, water quality assessment, satellite images, pictures and/or paintings repair, surveillance, security, etc.

State of the art : Spatial Interpolators

1^{st} Law of Geography

Everything is related to everything else, but near things are more related than distant things.

State of the art : Spatial Interpolators

1^{st} Law of Geography

Everything is related to everything else, but near things are more related than distant things.

Inverse Distance Weighing - IDW

Approximates values with the weighted average of the known neighbourhood values - weights inversely proportional to the distance from the target location.

Kriging

Kriging uses the same basic principle as IDW - weights are calculated using the covariation between known data at various spatial locations.

Our Proposal

Key Idea

Allow the use of data from faraway regions provided these neighbourhoods have similar spatial dynamics to the target location

Our Proposal

Key Idea

Allow the use of data from faraway regions provided these neighbourhoods have similar spatial dynamics to the target location

How to achieve this?

- Map the spatial interpolation problem into a regression task
- Propose a series of spatial indicators to better describe the spatial dynamics of a region

Describe the Spatial Dynamics for a Location

- Inspired by financial technical indicators
  - Summary statistics describing certain properties of the time series
    - e.g.: tendency, acceleration, momentum, volatility, etc.

Proposed Spatial Indicators

For a given variable of interest Z, a location o and its spatial neighbourhood $\mathcal{N}_o^\beta$:

- Centrality (averages and weighted averages)

$$\overline{Z}(\mathcal{N}_o^\beta) \qquad \widetilde{Z}(\mathcal{N}_o^\beta)$$

- Variability/Spread

$$\sigma_Z(\mathcal{N}_o^\beta)$$

- **Spatial Tendency**

$$\overline{Z}_o^{\beta_1,\beta_2} = \frac{\overline{Z}(\mathcal{N}_o^{\beta_1})}{\overline{Z}(\mathcal{N}_o^{\beta_2})} \qquad \widetilde{Z}_o^{\beta_1,\beta_2} = \frac{\widetilde{Z}(\mathcal{N}_o^{\beta_1})}{\widetilde{Z}(\mathcal{N}_o^{\beta_2})}$$

Spatial Interpolation as a Multiple Regression Task

- Target: the variable of interest value at location o
- Predictors: a description of the spatial dynamics of the variable in the neighbourhood of o

Spatial Interpolation as a Multiple Regression Task

- Target: the variable of interest value at location o
- Predictors: a description of the spatial dynamics of the variable in the neighbourhood of o
- An illustrative formalization of the prediction task:

$$
\begin{aligned}
Z_o = f(&\overline{Z}(\mathcal{N}_o^{k_1}), \overline{Z}(\mathcal{N}_o^{k_2}), \overline{Z}(\mathcal{N}_o^{k_3}), \overline{Z}_o^{k_1,k_2}, \overline{Z}_o^{k_2,k_3}, \\
&\widetilde{Z}(\mathcal{N}_o^{k_1}), \widetilde{Z}(\mathcal{N}_o^{k_2}), \widetilde{Z}(\mathcal{N}_o^{k_3}), \widetilde{Z}_o^{k_1,k_2}, \widetilde{Z}_o^{k_2,k_3}, \\
&\sigma_Z(\mathcal{N}_o^{k_1}), \sigma_Z(\mathcal{N}_o^{k_2}), \sigma_Z(\mathcal{N}_o^{k_3}))
\end{aligned}
$$

Spatial Interpolation as a Multiple Regression Task

- Target: the variable of interest value at location o
- Predictors: a description of the spatial dynamics of the variable in the neighbourhood of o
- An illustrative formalization of the prediction task:

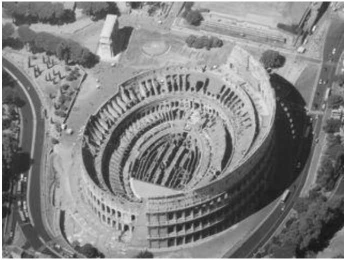
$$Z_o = f(\overline{Z}(\mathcal{N}_o^{k_1}), \overline{Z}(\mathcal{N}_o^{k_2}), \overline{Z}(\mathcal{N}_o^{k_3}), \overline{Z}_o^{k_1,k_2}, \overline{Z}_o^{k_2,k_3},$$
$$\widetilde{Z}(\mathcal{N}_o^{k_1}), \widetilde{Z}(\mathcal{N}_o^{k_2}), \widetilde{Z}(\mathcal{N}_o^{k_3}), \widetilde{Z}_o^{k_1,k_2}, \widetilde{Z}_o^{k_2,k_3},$$
$$\sigma_Z(\mathcal{N}_o^{k_1}), \sigma_Z(\mathcal{N}_o^{k_2}), \sigma_Z(\mathcal{N}_o^{k_3}))$$

- Procedure: (i) build a data set with the available data; (ii) use a regression method to approximate the unknown function $f()$; (iii) use the obtained model to carry out spatial interpolation

An Illustrative Application - Image Repair

Data and experimental methodology

- **Task:** forecast missing pixels on images - degree of gray [0,255];
- **The used source images**:



- For each picture we randomly removed: 10%, $20\%, \cdots, 90\%$ of the pixels; process repeated 10 times;
- Total of 180 data sets - 10 times for each 9 settings for the 2 pictures;
- Used error metric:

$$Mean\ Absolute\ Error = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

Competing Methods

Distance Interpolator (DI) - mean value of a circular neighbourhood region, with radius of: 10, 20 and 30

Inverse Distance Weighted (IDW) - weighted mean value of a circular neighbourhood region, with radius of: 10, 20 and 30

Ordinary Kriging (OK) - function `learn.autokrige`, R package `automap`, maximum neighborhood size 90.

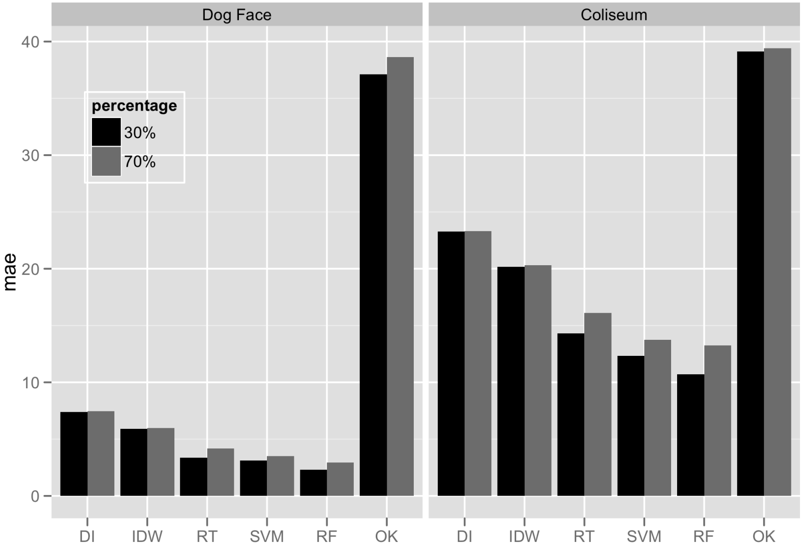
Regression Trees (RT) - function `rpartXse`, R package `DMwR`
se: 0, 0.5, 1 and 1.5

Support Vector Machines (SVM) - function `svm` R package `e1071`
cost: 1, 10 and 100, *gamma*: 0.1 and 0.5

Random Forest (RF) - R package `randomForest`
ntree: 500, 1000 and 1500

Results

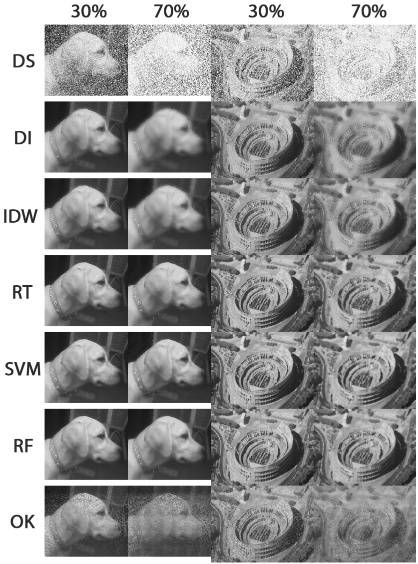
Datasets without 30% and 70% of the pixels



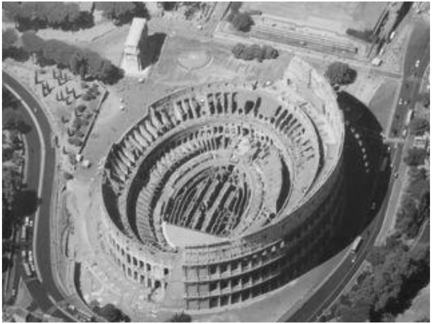
Visualization of the Results

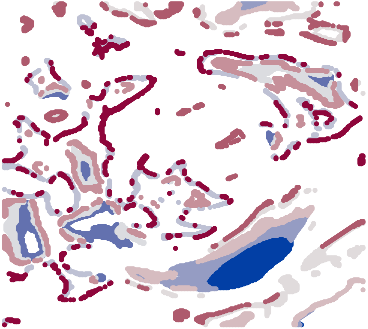
Best Models:

- DI_{10};
- IDW_{10};
- RT (se=0);
- SVM (cost=100, gamma=0.1);
- RF (ntree=1000);
- OK.



- The leafs of a regression tree on the coliseum data





Conclusions

- General methodology for spatial interpolation
- Two key ideas: i) map problem into regression task; ii) use predictors describing spatial dynamics
- Proposed some new spatial indicators for describing spatial tendency
- Approach *allows* models to use data from faraway regions going against the intuition behind the 1^{st} law of Geography
- Very interesting results on image repair tasks

Spatial Interpolation using Multiple Regression

ICDM'2012 - Brussels

Orlando Ohashi and **Luís Torgo**

LIAAD - INESC TEC / DCC - Faculty of Sciences
University of Porto - Portugal

13-Dec-2012

Full results and code at `http://www.dcc.fc.up.pt/~ltorgo/ICDM12`