Dynamic Coefficient Model

Hongjian Wang

April 7, 2016

1 Background

We use various types of data to estimate the crime count in community area (CA) of Chicago. For each CA we have observations on crime count and demographics. For each pair of CAs we also have observations on the taxi flow and spatial distance. One straightforward method is to build a regression model from all the features we observed to the crime counts.

We have one interesting observations is that in the south and north part of Chicago, the significance of different features are different. Therefore, the idea is to learn a dynamic weights for different spatial region.

2 Formulation

Suppose we have n regions in total, $R = \{r_1, r_2, \cdots, r_n\}$. The following notations are used

crime count at r_i	y_i
demographics at r_i	$\mathbf{d}_i$
taxi flow between r_i and r_j	f_{ij}
taxi flow weight matrix for r_i	$\mathbf{f_i}$
spatial weight matrix for r_i	$\mathbf{g_i}$
social flow lag variable for r_i	$s_i = \mathbf{f}_i^T \mathbf{y}$
spatial flow lag variable for r_i	$p_i = \mathbf{g}_i^T \mathbf{y}$

3 Model

3.1 Baseline Model

For simplicity we use linear regression model

$$y_i = \mathbf{w}_1^T \mathbf{d}_i + w_2 s_i + w_3 p_i + w_4,$$

where $\{w\}$ are the coefficients.

To simplify notations, we use $\mathbf{x}_i$ denote all the available predictors for region r_i,

$$\mathbf{x}_i = [\mathbf{d}_i, s_i, p_i, 1].$$

Then the model becomes

$$y_i = \mathbf{w}^T \mathbf{x}_i.$$

3.2 Dynamic Model

Now we use a dynamic model, where $\mathbf{w}$ is different for various regions. This leads to

$$y_i = \mathbf{w}_i^T \mathbf{x}_i.$$

The problem with formulation is that there are too many parameters to learn. To address this issue, we use the constraint that **spatially adjacent regions share similar coefficients**.

We use S_{ij} to denote the adjacency of r_i and r_j. And the aforementioned constraint is formulated as

$$\min \sum_{i,j} S_{ij} ||\mathbf{w}_i^T - \mathbf{w}_j^T||_2^2$$

The several choice of S_{ij}

- Binary indicator. $S_{ij} = 1$ if two regions are contiguous, otherwise $S_{ij} = 0$.
- The reverse distance between r_i and r_j.

The overall objective is

$$\min_{\mathbf{W}} \sum_i ||y_i - \mathbf{w}_i^T \mathbf{x}_i||_2^2 + \eta \sum_{i,j} S_{ij} ||\mathbf{w}_i^T - \mathbf{w}_j^T||_2^2 + \theta ||\mathbf{W}||_F^2 \tag{1}$$

4 Optimization

Rewrite the Frobenius norm in the last term

$$||\mathbf{W}||_F^2 = \sum_i ||\mathbf{w}_i - \mathbf{0}||_2^2.$$

Therefore the Equation 1 is rewritten as

$$\min_{\mathbf{W}} \sum_i ||y_i - \mathbf{w}_i^T \mathbf{x}_i||_2^2 + \eta \sum_{i,j \in 0,\cdots,N} S_{ij} ||\mathbf{w}_i - \mathbf{w}_j||_2^2, \tag{2}$$

where $\mathbf{w}_0 = \mathbf{0}$ and $S_{0i} = 1$ for $\forall i$.

To solve the objective in Equation 2, we use variable splitting. Namely, when optimizing for $\mathbf{w}_i$, we assume all other $\mathbf{w}_{j,j\neq i}$ are fixed. The sub-problem is

$$\min_{\mathbf{w}_i} ||y_i - \mathbf{w}_i^T \mathbf{x}_i||_2^2 + \eta \sum_j S_{ij} ||\mathbf{w}_i - \mathbf{w}_j||_2^2. \tag{3}$$

The update on $\mathbf{w}_i$ is

$$\mathbf{w}_i = \min_{\mathbf{w}_i} ||y_i - \mathbf{w}_i^T \mathbf{x}_i||_2^2 + \eta \sum_j S_{ij} ||\mathbf{w}_i - \mathbf{w}_j^{(t)}||_2^2.$$

The closed-form solution is

$$\mathbf{w}_i = (\mathbf{x}_i^T \mathbf{x}_i + \eta \sum_j S_{ij} \mathbf{I})^{-1} (y_i \mathbf{x}_i + \eta \sum_j S_{ij} \mathbf{w}_j) \tag{4}$$

5 Inference

We use the **leave-one-out** setting to infer and evaluate the crime rate of new community area.

Suppose the we want to estimate the crime rate y_i of CA_i. During the training process, we hold everything about CA_i out (including y_i, flow coming in and leaving from CA_i). Then training the model on CA_j, $\forall j \neq i$, which gives us w_j, $\forall j \neq i$. To infer the y_i, we need estimate the model coefficient w_i first. Follow the same intuition that model on CA_i is only similar to all its neighboring models, we have

$$\min_{\mathbf{w}_i} \sum_{j, \forall j \neq i} S_{ij} ||\mathbf{w}_i - \mathbf{w}_j||_2^2 + ||\mathbf{w}_i||_2^2 \tag{5}$$

After getting $\mathbf{w}_i$, we infer y_i by

$$\hat{y}_i = \mathbf{w}_i^T * \mathbf{x}_i \tag{6}$$