Plan for Course STP598 Spatiotemporal Analysis

Shuyi Li

October 2020

1 Problem

Covid-19 dataset is a typical sptiotemporal data. We could only use the number of infected cases as our response y and we do statewise analysis, which means we record the the number of cases for each state and each day. If we write in formula it will be $y(x,t)$, where x is different state and t is date.

The situation is getting more and more severe, which makes it necessary to analyse via the appropriate model and see if there is any way we could do to decrease its spread rate. Specifically, if we could obtain the correlation of different states we are likely to implement corresponding action to reduce the infected cases due to states dissemination. To summarize, we'll explore the how spatial dependence changes with time goes by for the number of infected case of covid-19 dataset.

2 Model—gSTGP model

We know Gaussian process is a powerful tool and it could accurately model almost all processes. The standard Spatio Temporal Gaussian Process(STGP) with a separable joint kernel $\mathcal{C}_{\mathbf{x}} \otimes \mathcal{C}_t$ fails to characterize the temporal evolution of spatial dependence (TESD). Then researcher generalized it by introducing the time-dependence to the spatial kernel $\mathcal{C}_{\mathbf{x}}$ via a modified Mercer's representation:

$$\mathcal{C}_{\mathbf{x}|t}(\mathbf{x}, \mathbf{x}') = \sum_{\ell=1}^{\infty} \lambda_\ell^2(t)\phi_\ell(\mathbf{x})\phi_\ell(\mathbf{x}') \tag{1}$$

The generalized STGP models can be summarized in the following unified form:

$$\begin{aligned} & y(\mathbf{z}) \sim \mathcal{GP}(m, \mathcal{C}_{y|m}), \quad m(\mathbf{z}) \sim \mathcal{GP}(0, \mathcal{C}_m) \\ \text{model } O: \quad & \mathcal{C}_{y|m} = \sigma_\epsilon^2 \mathcal{I}_{\mathbf{x}} \otimes \mathcal{I}_t, \quad \mathcal{C}_m = \mathcal{C}_{\mathbf{x}} \otimes \mathcal{C}_t \\ \text{model } I: \quad & \mathcal{C}_{y|m} = \sigma_\epsilon^2 \mathcal{I}_{\mathbf{x}} \otimes \mathcal{I}_t, \quad \mathcal{C}_m = \mathcal{C}_{\mathbf{x}|t} \otimes \mathcal{C}_t \\ \text{model } II: \quad & \mathcal{C}_{y|m} = \mathcal{C}_{\mathbf{x}|t} \otimes \mathcal{I}_t, \quad \mathcal{C}_m = \mathcal{I}_{\mathbf{x}} \otimes \mathcal{C}_t \end{aligned} \tag{2}$$

Notice that the above two models specify different structures for the marginal covariance $\mathcal{C}_y = \mathcal{C}_{y|m} + \mathcal{C}_m$. $\mathcal{C}_y$ in model I is in general non-sparse but more expressive, whereas $\mathcal{C}_y$ in model II is sparse but less expensive compared with model I.
We intend to use model II on our dataset since it not only requires less computation time but also captures Temporal Evolution of Spatio Dependence(TESD) from the experiment. Furthermore, for the covid-19 data, we aim to modify model a little bit by combining log-Gaussian Cox Processes(LGCP) due to non-negative discrete response Y. Cox process is inhomogeneous Poisson process with stochastic intensity λ written in the following formula:

$$\begin{aligned} Y(\mathbf{x}_i, t_j) &\sim \text{Pois}(\lambda(\mathbf{x}_i, t_j)), \quad \lambda(\mathbf{z}) = \exp[f(\mathbf{z})] \\ f(\mathbf{z}) &\sim \text{gSTGP}(0, \mathcal{C}_{\mathbf{z}}) \end{aligned}$$

where f is a generalized STGP prior to capture the temporal change of the geographical dependence among 50 states in the covid-19 data.

3 Comparison

If things go well, we would then compare our result with CAR and SAR model.