A fused nearest neighbor Gaussian process model for large datasets

Yewen Chen

October 13, 2020

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 THE NEAREST NEIGHBOR GAUSSIAN PROCESS

Similar to the pre-defined knots for Gaussian predictive processes, the starting point of nearest neighbor Gaussian process (NNGP) approach is to choose a fixed collection (e.g., S) of distinct locations in D, where S need not coincide with or be apart of the observed locations, so its size k need not equal the size of the dataset n (or even larger than the size n). The set S is called as reference set by Datta et al. [2016a].

A directed acyclic graph is defined on the reference set S, and then the joint distribution of spatial process $w(s)$ from the reference set is represented by the product of conditional densities which is motivated by Vecchia's approximation [Vecchia, 1988] ideas, where a careful choice of suitable conditional sets is required, and this conditional set is constrained to be the m-nearest neighbors by the NNGP approach, thereby facilitating computation in estimation and prediction problem (for more details on those problems, see Appendix of Datta et al. [2016a] and [Finley et al., 2017] and Finley et al. [2019]); the distributed Implementation.

Based on these results above, a NNGP can be well-defined from a parent Gaussian process $GP(0, C(\cdot;\theta))$.

1) Total operations is $O((n+k)m^3)$ where $m(\approx 20)$ is the size of conditioning set or neighbor set, and several processes can run in parallel (e.g, computations of weights in Kriging equation (**??**)).

2) The other major advantage is that the precision matrix of the NNGP is sparse with at most $km(m+1)/2$ nonzero entries.

3) NNGP can also be extended to large spatio-temporal data (Datta et al. [2016b]) and non-stationary process(Konomi et al. [2019]), [Risser and Turek, 2020].

2 TWO EXPLORATORY LIMITATIONS OF THE NNGP APPROACH

2.1 Exploratory analysis

1) Problem 1: Since the joint distribution of spatial process $w(s)$ from the reference set is represented by using the product of conditional densities and those conditional sets are constrained to be the m-nearest neighbors, so m-nearest neighbors may fail to capture all the information about the covariance parameters when there is a true **large scale dependence** in the dataset.

2) Problem 2: From an application point of view, the reference set S cannot be infinite because we need to solve neighbor set for every location, so the performance of the NNGP approximation depends on the size of the spatial dependence range relative to the spacing of the reference set, that maybe lead to the quality of the NNGP approximation gets worse when **the spatial dependence range gets shorter**.

Combing a NNGP $w(s)$ used to capture the local, small scale dependence structure with its spatial smooth version $f(s)$ used to capture the large scale dependence structure.

2.2 Model specification

$$y(s) = \mathbf{x}(s)\boldsymbol{\beta} + w(s) + \epsilon(s). \tag{1}$$

Furthermore, by decomposing $\boldsymbol{w}$ into $\boldsymbol{v}$ and $\mathbf{f}$, then

$$y(s) = \mathbf{x}(s)\boldsymbol{\beta} + v(s) + f(s) + \epsilon(s). \tag{2}$$

2.2.1 *NNGP Component in $\boldsymbol{v}$ on observed level*

Denote the reference set as $\mathbf{S}^\star = \{s_i^\star : i = 1, \cdots, n\}$ which coincides with the observed locations, and denote $v(s_j)$ as v_j. And then assume that $\boldsymbol{v}^\star = \left(v_1^\star, \cdots, v_n^\star\right)' \sim \text{NNGP}\left(0, \tilde{\mathbf{C}}^\star(\cdot;\theta^\star)\right)$ derived from a parent Gaussian process $\boldsymbol{w} \sim \text{GP}\left(0, \mathbf{C}^\star(\cdot;\theta^\star)\right)$.

2.2.2 *NNGP Component in $\mathbf{f}$ on grid level*

Denote the other reference set as $\mathbf{S}^{\star\star} = \{s_j^{\star\star} : j = 1, \cdots, k\}$ was determined through the grid of the domain, and let $h_k, k = 1, \cdots, K$ be a sequence of fixed basis functions and $f(s) = \sum_{k=1}^K h_k(s)\lambda_k$ where $\lambda_k^{\star\star} \equiv \lambda(s_k^{\star\star})$. Furthermore, the basis coefficients $\boldsymbol{\lambda}^{\star\star} = \left(\lambda_1^{\star\star}, \cdots, \lambda_K^{\star\star}\right)' \sim \text{NNGP}\left(0, \tilde{\mathbf{C}}^{\star\star}(\cdot;\theta^{\star\star})\right)$ derived from a parent Gaussian process $\boldsymbol{\lambda} \sim \text{GP}\left(0, \mathbf{C}^{\star\star}(\cdot;\theta^{\star\star})\right)$.

Let $\boldsymbol{\epsilon} = (\epsilon(s_1), \cdots, \epsilon(s_n))' \sim N(0, \tau^2\mathbf{I})$, according to the normal properties, we have

$$\mathbf{y} \sim \text{GP}\left(0, \boldsymbol{\Sigma}(\cdot;\theta)\right) \tag{3}$$

where $\boldsymbol{\Sigma}(\cdot;\theta,\tau^2) = \tilde{\mathbf{C}}^\star(\cdot;\theta^\star) + \mathbf{H}\tilde{\mathbf{C}}^{\star\star}(\cdot;\theta^{\star\star})\mathbf{H}' + \tau^2\mathbf{I}$.

2.2.3 *Inference*

With customary prior specifications, we obtain the joint distribution:

$$\begin{aligned}&\prod_{i=1}^n N\left(\mathbf{y}\left(s_i\right) \mid \mathbf{x}\left(s_i\right)\boldsymbol{\beta} + v\left(s_i\right) + \mathbf{h}'(s_i)\boldsymbol{\lambda}^{\star\star}\right) \times N\left(v_{(s_i)} \mid \mathbf{c}_{s_i}^{\star\prime}\mathbf{v}^\star, \xi_i\right) \times \\ &N\left(\mathbf{v}^\star \mid 0, \tilde{\mathbf{C}}_1\right) \times N\left(\boldsymbol{\lambda}^{\star\star} \mid 0, \tilde{\mathbf{C}}_2\right) \times N\left(\boldsymbol{\beta} \mid \mu_\beta, \mathbf{V}_\beta\right) \times \prod_{j=1}^l IG\left(\tau_j^2 \mid a_{\tau_j}, b_{\tau_j}\right) \times p(\theta)\end{aligned} \tag{4}$$

where $\mathbf{c}_{s_i}^{\star\prime} = \mathbf{c}_{s_i,\mathcal{N}(i)}^\star \mathbf{C}_{\mathcal{N}(i)}^{\star-1}$ and $\xi_i = c^\star(s_i, s_i) - \mathbf{c}_i^{\star\prime}\mathbf{c}_{\mathcal{N}(i),s_i}^\star$.

2.2.4 *Computational complexity*

2.2.5 *Robustness under stationary covariance families*

1.Matérn:

$$C(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\phi}\right) \mathcal{K}_\nu \left(\sqrt{2\nu}\frac{d}{\phi}\right)$$

2.Spherical:

$$C(d) = \sigma^2 \left\{1 - 1.5\left(\frac{d}{\phi}\right) + 0.5\left(\frac{d}{\phi}\right)^3\right\} \mathbf{I}(d < \phi)$$

2.2.6 *Non-stationary performance*

A deterministic function g is used to generate $\boldsymbol{w}$ by defining $w(s) = -10g(x_1)g(x_2)$, where

$$g(s) = \exp\left\{-(s/750-1)^2\right\} + \exp\left\{-0.8\,(s/750+1)^2\right\} - 0.05\sin\left(8(s/750+0.1)\right), \quad (5)$$

adapted from [Guhaniyogi et al., 2017].

2.3 Non-stationary performance

The spatial effects are each assigned a different prior distribution $F(s)$, the prior for $F(s)$ is the potentially infinite mixture, i.e.

$$f(s) \sim F(s) \stackrel{d}{=} \sum_{l=1}^{nGrid} p_l(s)\delta(c_l)$$

where $p_1(s) = U_1(s), p_l(s) = U_l(s)\prod_{j=1}^{l-1}\left(1-U_j(s)\right)$ for $i > 1$ and $U_l(s) = \mathcal{K}_l(s)u_l$, and u_l is given the priors $u_l \sim Beta(a,b)$, each independent across l; and $\boldsymbol{c} \sim N(0, \tilde{\boldsymbol{C}}^{**})$.

However, the distributions vary spatially according to the kernel functions $\mathcal{K}_l(s)$, which are centered at some lattice knots and restricted to the interval $[0,1]$.

REFERENCES

A. Datta, S. Banerjee, A. O. Finley, and A. E. Gelfand. Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111 (514):800–812, 2016a. URL https://chenyw68.github.io/Literature/[2016]Hierarchicalnearest-neighborGaussianprocessmodelsforlargegeostatisticaldatasets.pdf.

A. Datta, S. Banerjee, A. O. Finley, N. A. Hamm, and M. Schaap. Nonseparable dynamic nearest neighbor gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *The annals of applied statistics*, 10(3):1286, 2016b. URL https://chenyw68.github.io/Literature/[2016]NonseparabledynamicNNGPmodelsforlargespatio-temporaldata.pdf.

A. O. Finley, A. Datta, B. C. Cook, D. C. Morton, H. E. Andersen, and S. Banerjee. Applying nearest neighbor gaussian processes to massive spatial data sets forest canopy height prediction across tanana valley alaska.". *arXiv preprint arXiv:1702.00434*, 2017. URL https://chenyw68.github.io/Literature/[2017]ApplyingNearestNeighborGaussianProcessestoMassiveSpatialDataSets.pdf.

A. O. Finley, A. Datta, B. D. Cook, D. C. Morton, H. E. Andersen, and S. Banerjee. Efficient algorithms for bayesian nearest neighbor gaussian processes. *Journal of Computational and Graphical Statistics*, 28(2):401–414, 2019. URL https://chenyw68.github.io/Literature/[2019]Efficientalgorithmsforbayesiannearestneighborgaussianprocesses.pdf.

R. Guhaniyogi, C. Li, T. D. Savitsky, and S. Srivastava. A divide-and-conquer bayesian approach to large-scale kriging. *arXiv preprint arXiv:1712.09767*, 2017.

B. Konomi, A. Hanandeh, P. Ma, and E. Kang. Computationally efficient nonstationary nearest-neighbor gaussian process models using data-driven techniques. *Environmetrics*, 30(8):e2571, 2019. URL https://chenyw68.github.io/Literature/[2020]ComputationallyefficientnonstationaryNNGPmodelsusingdata-driventechniques.pdf.

M. D. Risser and D. Turek. Bayesian inference for high-dimensional nonstationary gaussian processes. *Journal of Statistical Computation and Simulation*, pages 1–27, 2020. URL https://chenyw68.github.io/Literature/[2020-formal]Bayesianinferenceforhighdimensionalnonstationary Gaussianprocesses.pdf.

A. V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312, 1988.