

Linear Projectionによる多重解像度近似を用いた 大規模空間データに対する予測について

関東学院大学 経済学部 平野 敏弘

人工衛星を通じて広範囲から得られた空間データの規模は非常に大きいですが、このような空間データは雲の影響といった何らかの原因による欠損を含んでいることがある。例として大気中の水蒸気の総量が挙げられるが、この場合、豪雨の発生を検知するといった目的のために未観測地点における予測が必要とされている。空間データに対する1つの予測方法として、観測された空間データで条件付けした予測分布を求めることが考えられる。具体的には

$$Z(\mathbf{s}) = Y_0(\mathbf{s}) + \epsilon(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d,$$

というモデルを考える。ここで、 $Y_0(\mathbf{s})$ は共分散関数 C_0 を持つ平均 0 の正規確率場、 $\epsilon(\mathbf{s})$ は観測誤差を表しており、 $\epsilon(\mathbf{s}) \sim \text{i.i.d. } \mathcal{N}(0, \tau^2)$ とする。このとき、予測分布として

$$Y_0(S^P) | Z(S_0) \sim N \left(C_0(S^P, S_0) \{C_0(S_0, S_0) + \tau^2 \mathbf{I}_n\}^{-1} Z(S_0), \right. \\ \left. C_0(S^P, S^P) - C_0(S^P, S_0) \{C_0(S_0, S_0) + \tau^2 \mathbf{I}_n\}^{-1} C_0(S_0, S^P)^\top \right),$$

を考える。ただし、 $S_0 = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ は観測地点、 $S^P = \{\mathbf{s}'_1, \dots, \mathbf{s}'_m\}$ は予測地点、 $Y_0(S^P) = (Y_0(\mathbf{s}'_1), \dots, Y_0(\mathbf{s}'_m))^\top$, $Z(S_0) = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^\top$, $(C_0(S_0, S_0))_{ij} = C_0(\mathbf{s}_i, \mathbf{s}_j)$ ($i, j = 1, \dots, n$), $(C_0(S^P, S^P))_{ij} = C_0(\mathbf{s}'_i, \mathbf{s}'_j)$ ($i, j = 1, \dots, m$), $(C_0(S^P, S_0))_{ij} = C_0(\mathbf{s}'_i, \mathbf{s}_j)$ ($i = 1, \dots, m, j = 1, \dots, n$) である。しかし、予測分布の平均ベクトルと共分散行列に含まれる $C_0(S_0, S_0) + \tau^2 \mathbf{I}_n$ の逆行列に関する計算量は、サンプルサイズが n の場合 $O(n^3)$ となるため、 n が大きいときの計算負荷は非常に大きくなる。

この問題に対処するため、本発表では、Linear Projectionによる多重解像度近似 (Multi-Resolution Approximation via Linear Projection, M -RA-lp) を用いた予測分布の高速計算方法を提案する。Banerjee et al. (2013) によって提案された Linear Projection は、正規確率場をサンプルサイズより低い次元の空間へ射影し、真の共分散関数を非定常共分散関数で近似することで、共分散行列の逆行列に関する計算を高速化する。 M -RA-lp は、空間相関を Linear Projection で近似した後、サンプリング領域を分割して各領域上で近似の差分部分をさらに Linear Projection で近似して元の近似に加えるという操作を繰り返すことで、既存手法より真の共分散関数 C_0 に対する近似精度が高い $C_{M\text{-RA-lp}}$ を出力する。 M -RA-lp は、Hirano (2017) における Modified Linear Projection の解像度を増大させる形で拡張していると共に、Katzfuss (2017) における多重解像度近似を部分的に一般化している。本発表では、 C_0 を $C_{M\text{-RA-lp}}$ で置換した予測分布における平均ベクトルと共分散行列を高速計算するアルゴリズムを提案し、シミュレーションを通じて評価を行う。

参考文献

- Banerjee, A., D. B. Dunson, and S. T. Tokdar (2013). Efficient Gaussian process regression for large datasets. *Biometrika* 100, 75–89.
- Hirano, T. (2017). Modified linear projection for large spatial datasets. *Communications in Statistics - Simulation and Computation* 46, 870–889.
- Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association* 112, 201–214.