

1 データと手法

1.1 データ

1.1.1 P-POTEKA データセット

1.1.2 内挿処理

P-POTEKA のデータは 35 個の観測点におけるそれぞれの位置での気象データである。本研究では機械学習モデルが学習可能なデータ形式にするために、生の P-POTEKA に対してガウス過程回帰と呼ばれる手法を用いてグリッドデータに変換した。

まずガウス過程とは関数 $f(x)$ を確率変数と見立てた確率分布のことである。そしてガウス過程回帰とは、データから関数 $f(x)$ の確率分布をガウス過程の形で求める方法のことである。関数 $f(x)$ の任意個数の入力点 x_1, \dots, x_N

に対する出力 $f(x_1), \dots, f(x_N)$ が N 次元の多変量ガウス分布でモデル化される。したがってその予測値 $f(x_*)$ もガウス分布に従い、期待値と分散で表現される。ガウス過程回帰では予測値の不確実性を分散で評価できるという特徴を持つ。さらに線形回帰などとは異なり、非線形な関数に対してもフィッティングできる柔軟性を持つ。

多変量ガウス分布とは D 次元のベクトル $x = (x_1, \dots, x_D)$ の変数がそれぞれ独立にガウス分布に従う場合における同時分布のことである。ベクトル x が平均 μ 、共分散行列 Σ のガウス分布 $\mathcal{N}(x|\mu, \Sigma)$ に従っているとき、確立密度関数は以下のように表される。

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \exp \left\{ -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right\} \quad (1)$$

$\mu = (\mu_1, \dots, \mu_N)$ は x の期待値を表す平均ベクトル、 Σ は $D \times D$ の共分散行列でその要素 (i, j) が x_i と x_j の共分散を表している。

一般的な線形回帰モデルでは基底関数 $\phi_0(x), \phi_1(x), \dots, \phi_H(x)$ によって x の写像を並べた特徴ベクトル $\phi(x) = (\phi_0(x), \phi_1(x), \dots, \phi_H(x))^T$ とパラメータ w を用いて以下のように表される。

$$\hat{y} = w^T \phi(x) \quad (2)$$

\hat{y} は観測データ y に近づけるあてはめ値である。複数のデータ点 x_1, \dots, x_N の場合に拡張すると以下のような行列形式で表すことができる。

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \dots \\ \hat{y}_N \end{bmatrix} = \begin{pmatrix} \phi_0(x_1) & \phi_1(x_1) & \dots & \phi_H(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \dots & \phi_H(x_2) \\ \dots & \dots & \ddots & \vdots \\ \phi_0(x_N) & \phi_1(x_N) & \dots & \phi_H(x_N) \end{pmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_H \end{bmatrix} \quad (3)$$

この方法では入力 x の次元数が小さい場合でしか使えないという問題がある。なぜなら学習すべきパラメータ w の次元が入力次元の増大に対して指数的に増加してしまうためである。この問題は一般的に次元の呪いと呼ばれている。ガウス過程回帰ではパラメータ w に対して期待値をとり、積分消去することでこの問題を回避している。 $\hat{y} = (\hat{y}_1, \dots, \hat{y}_N)$ 、 $w = (w_0, \dots, w_H)$ 、および $\Phi_{nh} = \phi_h(x_n)$ を要素とする計画行列 Φ を用いると、先ほどの行列形式は以下のように書ける。

$$\hat{\mathbf{y}} = \Phi \mathbf{w} \quad (4)$$

ガウス過程では重み \mathbf{w} が平均 $\mathbf{0}$ 、分散 $\lambda^2 \mathbf{I}$ のガウス分布

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \lambda^2 \mathbf{I}) \quad (5)$$

から生成されたものと仮定する。この場合式 (4) は $\hat{\mathbf{y}}$ は、ガウス分布に従うベクトル \mathbf{w} を定数行列 Φ で線形変換したものとなっている。したがって $\hat{\mathbf{y}}$ もガウス分布に従う。 $\hat{\mathbf{y}}$ の期待値、共分散行列は以下のように求められる。

$$\mathbb{E}[\hat{\mathbf{y}}] = \mathbb{E}[\Phi \mathbf{w}] = \Phi \mathbb{E}[\mathbf{w}] = \mathbf{0} \quad (6)$$

$$\begin{aligned} \Sigma &= \mathbb{E}[\hat{\mathbf{y}} \hat{\mathbf{y}}^T] - \mathbb{E}[\hat{\mathbf{y}}] \mathbb{E}[\hat{\mathbf{y}}]^T \\ &= \mathbb{E}[(\Phi \mathbf{w})(\Phi \mathbf{w})^T] \\ &= \Phi \mathbb{E}[\mathbf{w} \mathbf{w}^T] \Phi^T \\ &= \lambda^2 \Phi \Phi^T \end{aligned} \quad (7)$$

結果として $\hat{\mathbf{y}}$ の分布は、多変量ガウス分布

$$\hat{\mathbf{y}} \sim \mathcal{N}(\mathbf{0}, \lambda^2 \Phi \Phi^T) \quad (8)$$

となることがわかる。この式からわかるように重み \mathbf{w} は期待値がとられることによって消去されている。したがって入力 \mathbf{x} や $\phi(\mathbf{x})$ の次元がどれだけ高くとも、対応する高次元の重み \mathbf{w} を求める必要はなく、 \mathbf{y} の分布はデータ数 N に依存する共分散行列 $\lambda^2 \Phi \Phi^T$ で決まる。

式 (8) の共分散行列を

$$\mathbf{K} = \lambda^2 \Phi \Phi^T \quad (9)$$

とみると、この (n, n') 要素は

$$K_{nn'} = \lambda^2 \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_{n'}) \quad (10)$$

で与えられる。すなわち \mathbf{x}_n と $\mathbf{x}_{n'}$ の特徴ベクトル $\phi(\mathbf{x}_n)$ と $\phi(\mathbf{x}_{n'})$ の内積の定数倍が共分散行列 \mathbf{K} の (n, n') 要素 $K_{nn'}$ になっている。多変量ガウス分布において2つの変数間の共分散が大きいということは似た値を取りやすいことを意味する。すなわち、 \mathbf{x}_n と $\mathbf{x}_{n'}$ が似ているなら、対応する \hat{y}_n と $\hat{y}_{n'}$ も似た値を持つことになる。さらにガウス過程回帰では共分散行列 \mathbf{K} の要素 $\phi(\mathbf{x}_n)^T \phi(\mathbf{x}_{n'})$ を明示的に求めることはせず、カーネル関数 $k(\mathbf{x}_n, \mathbf{x}_{n'})$ で以下のように置き換える。

$$k(\mathbf{x}_n, \mathbf{x}_{n'}) = \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_{n'}) \quad (11)$$

こうすることで特徴ベクトル $\phi(\mathbf{x})$ を明示的に表現することなしに、カーネル関数のみに基づく簡単な計算で計画行列を求めることができる。カーネル関数を適切に設計することで、 \mathbf{K} の逆行列が計算できない場合を避けることができる。代表的なカーネル関数として線形カーネル (式 12) や指数カーネル (式 13, θ はパラメータ) がある。カーネル関数はデータ間の距離に応じた出力の類似度を表現していることから、出力の特徴に応じて設計することも可能である。

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}' \quad (12)$$

$$k(\mathbf{x}, \mathbf{x}') = \exp(-|\mathbf{x} - \mathbf{x}'|/\theta) \quad (13)$$

ここで以下の N 個のデータ点がすでに観測されているとする。

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\} \quad (14)$$

今までの議論により、ガウス過程回帰において \mathbf{y} はガウス分布に従い、データ \mathcal{D} とカーネル関数から求められる共分散行列 \mathbf{K} を用いて以下のように表される。

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}) \quad (15)$$

この分布を用いて新しい観測点 $\mathbf{X}^* = (\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_M^*)$ での予測値 $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_M^*)$ を求めたい場合を考える。 \mathbf{y}^* も多変量ガウス分布に従うことからすでに得られている分布 \mathbf{y} との同時分布もガウス分布に従うことが分かる。この同時分布を以下に示す。

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \\ y_1^* \\ \vdots \\ y_M^* \end{bmatrix} = \mathcal{N} \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_N) & k(\mathbf{x}_1, \mathbf{x}_1^*) & \dots & k(\mathbf{x}_1, \mathbf{x}_M^*) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \dots & k(\mathbf{x}_N, \mathbf{x}_N) & k(\mathbf{x}_N, \mathbf{x}_1^*) & \dots & k(\mathbf{x}_N, \mathbf{x}_M^*) \\ k(\mathbf{x}_1^*, \mathbf{x}_1) & \dots & k(\mathbf{x}_1^*, \mathbf{x}_N) & k(\mathbf{x}_1^*, \mathbf{x}_1^*) & \dots & k(\mathbf{x}_1^*, \mathbf{x}_M^*) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_M^*, \mathbf{x}_1) & \dots & k(\mathbf{x}_M^*, \mathbf{x}_N) & k(\mathbf{x}_M^*, \mathbf{x}_1^*) & \dots & k(\mathbf{x}_M^*, \mathbf{x}_M^*) \end{pmatrix} \right) \quad (16)$$

新しい観測点 \mathbf{X}^* における出力 \mathbf{y}^* の周辺分布 $p(\mathbf{y}^* | \mathbf{X}^*, \mathcal{D})$ を求めることで未観測点の出力を推定することができる。同時分布がガウス分布に従っていることから周辺分布もガウス分布となることがわかる。予測値を求めていることから、この周辺分布は予測分布とも呼ばれる。以下のように計算される。

$$p(\mathbf{y}^* | \mathbf{X}^*, \mathcal{D}) = \mathcal{N}(\mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{y}, \mathbf{k}_{**} \mathbf{K}^{-1} \mathbf{k}_*) \quad (17)$$

ただし $\mathbf{k}_*(n, m) = k(\mathbf{x}_n, \mathbf{x}_m^*)$ 、 $\mathbf{k}_{**}(m, m') = k(\mathbf{x}_m, \mathbf{x}_{m'}^*)$ である。

本研究においては、 \mathbf{x} は緯度・経度の2次元で y は地点 \mathbf{x} における気象パラメータである。また、気象パラメータの特徴ごとにカーネル関数を設定した。具体的には、雨や風はデータの距離に近いほど似たデータになるが距離が大きくなると途端に類似度は下がると仮定し指数カーネルを用いた。気温や湿度など雨や風程には地域間で差が生じないようなデータには線形カーネルを用いた。

1.2 Self-Attention ConvLSTM

1.3 学習と評価

1.3.1 学習

1.3.2 評価

参考文献

- [1] Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. NIPS 2015, 802–810.
- [2] Lin Z.; Li M.; Zheng Z.; Cheng Y.; and Yuan C. 2020. Self-Attention ConvLSTM for Spatiotemporal Prediction. Association for the Advancement of Artificial Intelligence
- [3] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need., NIPS, 2017.
- [4] Su A.; Li H.; Cui L.; and Chen Y. 2020. A Convection Nowcasting Method Based on Machine Learning. Hindawi.
- [5] Hubel D.H. and Wiesel T.N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology.
- [6] Hochreiter S. and Schmidhuber J. 1997. Long Short-Term Memory. Neural Computation.
- [7] Ko C.M.; Jeong Y.Y.; Lee Y.M.; and Kim B.S. 2020. The Development of a Quantitative Precipitation Forecast Correction Technique Based on Machine Learning for Hydrological Applications. MDPI.
- [8] 気象庁. 2018. 気候変動監視レポート 2018 世界と日本の気候変動および温室効果ガスとオゾン層等の状況. 気象庁.
- [9] 気象庁. 2020. 気象庁業務評価レポート（令和 2(2020) 年度版）. 気象庁.
- [10] 佐藤正樹;. 2020. 近年における降雨状況の実態：極端豪雨は増えているか 水環境学会誌 第 43 巻（A）第 5 号 pp.142 147. 公益社団法人 日本水環境学会.
- [11] Fumikai F.; Nobuo Y.; and Kenji K. 2006. Long-Term Changes of Heavy Precipitation and Dry Weather in Japan(1901-2004). Meteorological Society of Japan.