

ガウス過程の学習の高速化手法の検討

Acceleration of Gaussian Process Regression

寺岡悠一
Yuichi Teraoka

神野健哉
Kenya Jinno

東京都市大学大学院 総合理工学研究所 情報専攻
Division of Informatics, Graduate School of Integrative Science and Engineering, Tokyo City University

1 まえがき

ガウス過程 (Gaussian Process) とは、入力変数 x から変数 y を推定するモデルである。そしてその特徴は非線形性にあり、一般的な線形回帰では上手くフィッティングできない場合に有効である。

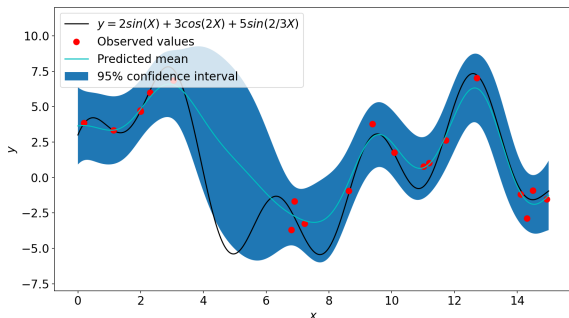


図 1: Gaussian Process の例

ガウス過程回帰の主目的は、予測分布に基づく出力 y の期待値を求めることである。以下ガウス過程による予測分布を示す。

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

(1) 式からデータ数 N が大きい場合のカーネル行列やその逆行列の求めるためのコストがかかることがわかる。この手法は計算コストが $O(N^3)$ となってしまう、現実問題へ適用するのに実用的でないという問題がある。本研究では、多次元データに KD 木を使用してガウス過程の学習の高速化を行う。

2 データ分割手法

ガウス過程回帰モデルのトレーニングでは、線形行列 $\mathbf{A}\mathbf{x} = \mathbf{y}$ を解くが、共役勾配法を適用することで近似解を効率的に得られる。共役勾配法では、各反復で \mathbf{A} の行列ベクトル積が必要となる。そこで行列 \mathbf{A} を、 $\mathbf{A} = \mathbf{K} + \sigma^2 \mathbf{I}$ と分解して、カーネル行列 \mathbf{K} の効率的な操作を可能にするため KD 木を使用する。

KD 木の各ノードに含まれる bounding box に関する情報から、クエリ \mathbf{x}^* との距離に基づき、 \mathbf{K} 行列の一部の計算を制限する。クエリに対して距離が設定した閾値 ((2) 式) 以下になるようなノードを選択し、このプロセスを再帰的に適用する。

$$\text{if } \frac{\text{average_distance}}{\text{distance}(\text{query_point})} < \text{threshold, then process} \quad (2)$$

3 実験

これらの方法を用いて効率的に学習できたか確認するため実験を行った。実験は使用するデータは、イタリアの都市の大気質データセット [3], 都市サイクルの燃料消費量に関するオート MTG データセット [4], ポルトガル北東部の森林火災の焼失地域のデータセット [5] である。推定する際には、クロスバリデーションより設定した新しいクエリに対して、学習済みのガウス過程モデルを使用し、本手法に基づき行列 \mathbf{K} の計算を行い、計算コストを確認した。精度は予測値と観測値との二乗誤差を表す。

表 1: 実験結果

データセット・本手法	計算時間 [s]	精度 (MSE)
大気質データ・なし	1232.95 秒	12.976
大気質データ・あり	10.20 秒	750.308
森林火災データ・なし	4019.78 秒	720.443
森林火災データ・あり	19.09 秒	3215.321
MTG データ・なし	3979.88 秒	820.413
MTG データ・あり	19.33 秒	3027.970

表 1 から結果から、KD 木を導入することで、カーネル行列の計算が効率的になったことを確認できる。

4 結論

提案した高速化手法はガウス過程のスケラビリティ向上に寄与することを確認した。

参考文献

- [1] 持橋大地, 大羽成征. ガウス過程と機械学習. 機械学習プロフェッショナルシリーズ. 講談社, 2019.
- [2] Yirong Shen, Andrew Y. Ng, and Matthias Seeger. Fast gaussian process regression using kd-trees. In Advances in Neural Information Processing Systems 18. MIT Press, 2006
- [3] Vito, Saverio. (2016). Air Quality. UCI Machine Learning Repository. <https://doi.org/10.24432/C59K5F>.
- [4] Quinlan, R.. (1993). Auto MPG. UCI Machine Learning Repository. <https://doi.org/10.24432/C5859H>.
- [5] Cortez, Paulo and Morais, Anbal. (2008). Forest Fires. UCI Machine Learning Repository. <https://doi.org/10.24432/C5D88D>.