

ガウス過程に基づく時空間データ解析

田中 佑典¹・上田 修功²

(受付 2024 年 6 月 28 日；改訂 11 月 22 日；採択 11 月 26 日)

要 旨

ガウス過程はデータから関数を推定するための確率モデルの一つであり、古くから時空間データのモデリングに広く使われてきた。本稿では、時空間データ解析に関連する、二つの重要な問題を取り上げ、ガウス過程を用いたモデリングについて述べる。一つ目の問題は、集約データのモデリングである。都市などで取得される時空間データ(貧困度など)は、座標や時刻など「点」に紐づくデータのことを指すが、プライバシー保護や行政上の理由で、行政区画など「領域」に紐づくデータとして扱われることが少なくない。本稿では、ガウス過程の領域に対する積分を考えることによって、集約データを自然に扱う方法について紹介する。二つ目の問題は、常微分方程式で表される力学系のモデリングである。力学系のダイナミクスは状態の時間微分によって定義され、相空間上における「ベクトル場」として表現される。素朴には多出力ガウス過程を用いることによってベクトル場をモデリングできるが、物理法則を満たすような学習結果が必ずしも得られるとは限らない。本稿では、解析力学の一形式である「ハミルトン力学」を融合したガウス過程により、エネルギーの保存・散逸則を満たすベクトル場をモデリングする方法について紹介する。

キーワード：ガウス過程、時空間データ、集約データ、力学系、ハミルトン力学。

1. はじめに

機械学習の研究はますます発展を続けており、近年では様々な分野に応用されている。機械学習とは、一言でいうと、解きたい問題に応じた入出力関係を満たす関数を、大量のデータから自動的に獲得する方法である。本稿では、関数に対する確率分布であるガウス過程(Gaussian process: GP) (Rasmussen and Williams, 2006) を用いた、時空間データのモデリングについて述べる。ガウス過程は、関数の事前分布として用いられ、ベイズ推論の枠組みに基づいて学習および予測を行う。ガウス過程は、関数形を陽に仮定しないノンパラメトリックな手法であり、共分散関数を指定することにより柔軟な関数を表現できる。また、データが与えられたときの事後分布を用いて、予測分散を評価できることもガウス過程の利点の一つである。

本稿では、時空間データ解析のためのガウス過程について、比較的新しい二つの重要な研究トピックを紹介する。一つ目は、集約データのためのガウス過程である。標準的な時空間回帰では、点(座標や時刻)と属性の組でサンプルが表され、その背後にある関数を推定する。一方、ここで扱う問題設定では、サンプルが領域(行政区画や時間帯)と属性の組で表されることを仮定する。例えば、都市において収集された貧困度や犯罪数のようなデータは、プライバ

¹ NTT コミュニケーション科学基礎研究所：〒619-0237 京都府相楽郡精華町光台 2-4

² 理化学研究所 革新知能統合研究センター：〒103-0027 東京都中央区日本橋 1-4-1

シー保護や行政上の理由から、行政区画や管轄区域において平均化などの統計処理がなされ、公開されることが多い。従来のガウス過程では、このような集約データを自然に扱うことが困難であった。Smith et al. (2018)において、単一の集約データを、ガウス過程の領域における積分としてモデリングする方法が提案された。これにより、領域の形や大きさを反映した共分散関数を表現でき、集約データから背後にある連続関数の推定を可能とした。その後、複数の集約データを同時に解析することを目的として、重回帰分析に基づく拡張 (Law et al., 2018; Tanaka et al., 2019a) や、多出力ガウス過程 (Multi-output GP) に基づく拡張 (Tanaka et al., 2019b; Yousefi et al., 2019; Hamelijncx et al., 2019) がなされ、様々な解像度を持つ集約データ間の相関を考慮してデータの予測が可能な手法へと発展している。本稿では、Tanaka et al. (2019b) の内容を中心に解説を行う。

二つ目のトピックは、力学系のためのガウス過程である。多くの力学系は常微分方程式で表され、状態の時間微分をモデリングすることで系のダイナミクスを記述することができる。この時間微分を相空間 (状態空間と同義) 上に表したものを「ベクトル場」と呼び、機械学習モデルを用いてベクトル場をパラメタライズすることで、データからダイナミクスの学習を可能とする (Chen et al., 2018)。素朴には多出力ガウス過程を用いることによりベクトル場のモデリングが可能であるが、物理法則を満たすような学習結果が必ずしも得られるとは限らない。そこで、物理ダイナミクスをより適切にモデリングするために、物理学に由来する「数理構造」を組み込むことが重要である。ダイナミクスを表すベクトル場は、物理法則の下である種の普遍性を持つ。例えば、質量保存則に基づき、非圧縮流体の速度場は発散しない (Divergence-free) ことが知られている。また、電磁気学においては、エネルギーの保存則に基づき、磁場が時間に対して不変なとき電場は回転をもたない (Curl-free) ことが知られている。このような数理構造を、ベクトル解析の知識を利用して共分散関数に組み込む方法が提案されている (Narcowich and Ward, 1994; Macêdo and Castro, 2010)。さらに、ハミルトン力学の理論 (Goldstein, 1980; 畑, 2014) に基づき、エネルギーの保存・散逸則を満たすベクトル場に対するガウス過程が提案されている (Rath et al., 2021; Tanaka et al., 2022; Tanaka, 2024; Boffi et al., 2022)。これらのモデリングにおいては、Greydanus et al. (2019) におけるアイデアを援用し、ベクトル場を直接モデリングする代わりに、系の全体エネルギーである「ハミルトニアン (Hamiltonian)」をガウス過程によってモデリングする。ベクトル場は、ハミルトンの運動方程式に基づき、ハミルトニアンの勾配を用いて導出される。ハミルトン力学に基づくモデリングは、エネルギーの保存・散逸則を導入可能であることに加え、様々な拡張の可能性を持つという点において重要である。例えば、Beckers et al. (2022) では、制御入力を持つポート・ハミルトン系 (Port-Hamiltonian system) への拡張がなされている。本稿では、Tanaka et al. (2022) や Tanaka (2024) の内容を中心に解説を行う。

本稿の構成は以下のとおりである。2 章では、ガウス過程の基本事項について整理し、本稿を理解する上で重要な多出力ガウス過程について述べる。3 章では、集約データのためのガウス過程について、4 章では、力学系のためのガウス過程について、問題設定、手法、および、実験結果を述べる。最後に 5 章で結論を述べる。

2. ガウス過程

2.1 ガウス過程回帰

本節では、1 次元の回帰問題を例にして、ガウス過程 (Rasmussen and Williams, 2006) の基本事項をまとめる。 $x, y \in \mathbb{R}$ とし、回帰モデル

$$(2.1) \quad y = f(x) + \epsilon$$

を考えよう．ここで， ϵ は平均 0 で分散 σ^2 のガウスノイズとする． N 個のサンプル $\{(x_n, y_n) \mid n = 1, \dots, N\}$ が与えられたとき，関数 $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ を推定したいとする．

ガウス過程は「関数を生成する確率分布」であり，推定したい関数 $f(x)$ の事前分布として用いられる． $f(x)$ がガウス過程に従うことを

$$(2.2) \quad f(x) \sim \mathcal{GP}(0, k(x, x'))$$

と表す．ここで，簡単のため $\mathbb{E}[f(x)] = 0$ を仮定した ($\mathbb{E}[\cdot]$ は期待値を表す)．また， $k(x, x') : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ はガウス過程のパラメータであり，共分散関数 (Covariance function) と呼ばれ，

$$(2.3) \quad k(x, x') = \mathbb{E}[f(x)f(x')]$$

を意味する．共分散関数には，正定値カーネル (Positive definite kernel) (福水, 2010) が用いられる．正定値カーネルは以下のように定義される．

定義 (正定値カーネル)．以下の二つの条件を満たすとき $k(x, x')$ は正定値カーネルである．

- 対称性： $k(x, x') = k(x', x)$ が成り立つ．
- 正定値性： 任意の正の整数 N ，任意の点 $x_1, \dots, x_N \in \mathbb{R}$ および任意の実数 $c_1, \dots, c_N \in \mathbb{R}$ に対して以下が成り立つ．

$$(2.4) \quad \sum_{i=1}^N \sum_{j=1}^N c_i c_j k(x_i, x_j) \geq 0.$$

この条件は，後述のグラム行列 (式 (2.7)) が半正定値であることを意味する．

最もよく使われるものの一つは二乗指数カーネル (Squared exponential kernel) であり，

$$(2.5) \quad k(x, x') = \alpha^2 \exp\left(-\frac{1}{2\beta^2}(x - x')^2\right)$$

と表される．ここで， $\alpha^2 \in \mathbb{R}_{>0}$ は分散パラメータであり， $\beta^2 \in \mathbb{R}_{>0}$ はスケールパラメータである．正定値カーネルを用いてガウス過程は以下のように定義される．

定義 (ガウス過程)． N 個の点 $\mathbf{X} = (x_1, \dots, x_N)^\top$ における関数の値を $\mathbf{f} = (f(x_1), \dots, f(x_N))^\top$ とする．任意の自然数 N に対して， \mathbf{f} が N 次元ガウス分布

$$(2.6) \quad \mathbf{f} \sim \mathcal{N}(\mathbf{0}, k(\mathbf{X}, \mathbf{X}))$$

に従うとき， \mathbf{f} はガウス過程である．ここで， $k(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{N \times N}$ は正定値カーネルで計算される対称行列 (グラム行列 (Gram matrix) と呼ばれる) であり，

$$(2.7) \quad k(\mathbf{X}, \mathbf{X}) = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{pmatrix}$$

と表される．

式 (2.7) からわかるように，カーネル $k(x, x')$ は任意の二点間における関数値の類似度を表していると解釈できる．したがって，カーネルを適切に設計することにより，関数の滑らかさについての事前知識を導入することができる．

次に，観測 $\mathbf{y} = (y_1, \dots, y_N)^\top$ が与えられたときの事後分布を計算することにより，予測

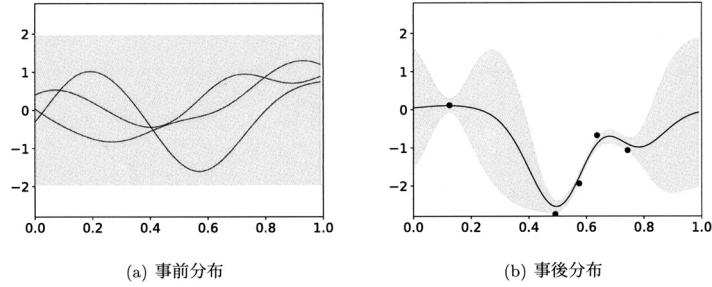


図 1. ガウス過程の事前分布と事後分布. 網掛け部分は 95% 信頼区間を表す. (a) カーネルパラメータを $\alpha^2 = 1$ および $\beta^2 = 0.2$ に設定し, 事前分布から 3 つの関数を生成した. (b) 事後分布は黒点で示す 5 つのサンプル ($\sigma^2 = 0.1$) から学習した. 黒線は予測平均 $m^*(x)$ を表す.

を行う方法について述べる. 図 1 に, ガウス過程の事前分布と事後分布を可視化する. 上記の回帰問題では, 尤度はノイズ分散を σ^2 とするガウス分布であり, 尤度と事前分布が共役 (Conjugate) となり, 事後分布が解析的に計算できる. 事後分布もガウス過程となり,

$$(2.8) \quad f^*(x) | \mathbf{y} \sim \mathcal{GP}(m^*(x), k^*(x, x'))$$

と表される. 事後分布の平均関数 $m^*(x) : \mathbb{R} \rightarrow \mathbb{R}$ と共分散関数 $k^*(x, x') : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ は,

$$(2.9) \quad m^*(x) = k(x, \mathbf{X})^\top \mathbf{C}^{-1} \mathbf{y}$$

$$(2.10) \quad k^*(x, x') = k(x, x') - k(x, \mathbf{X})^\top \mathbf{C}^{-1} k(x, \mathbf{X})$$

と表される. ここで, $k(x, \mathbf{X}) = (k(x, x_1), \dots, k(x, x_N))^\top$ および $\mathbf{C} = k(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}$ である. \mathbf{I} は単位行列を表す. ガウス過程回帰では, 式 (2.9) を用いて任意の点における関数値の予測を行い, 式 (2.10) を用いて予測に対する分散を評価する.

最後に, ノイズ分散 σ^2 およびカーネルパラメータ α^2, β^2 の決め方について述べる. 最もよく使われるのは, 周辺尤度 $p(\mathbf{y})$ が最大となるようにパラメータを最適化する方法である. この方法は第二種最尤推定 (Type II maximum likelihood) と呼ばれる. 上記の回帰問題では, 事後分布と同様に周辺尤度も解析的に計算でき, $p(\mathbf{y}) = \mathcal{N}(\mathbf{0}, \mathbf{C})$ と表される.

2.2 ベクトル値関数への拡張

前章の議論を, D 次元ドメインにおいて M 個の出力を持つベクトル値関数 (Vector-valued function) $\mathbf{f}(\mathbf{x}) : \mathbb{R}^D \rightarrow \mathbb{R}^M$ へと拡張する. $\mathbf{x} \in \mathbb{R}^D$ および $\mathbf{y} \in \mathbb{R}^M$ とし, 回帰モデル

$$(2.11) \quad \mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\epsilon}$$

を考える. ここで, $\boldsymbol{\epsilon}$ は平均 0 で共分散行列 $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$ のガウスノイズとする. 観測データ $\{(\mathbf{x}_{m,n}, y_{m,n}) \mid m = 1, \dots, M; n = 1, \dots, N_m\}$ が与えられたとき, 関数 $\mathbf{f}(\mathbf{x})$ を推定したいとする. ここで, N_m は m 番目の出力に対するサンプル数を表す.

ベクトル値関数に対するガウス過程は, 多出力ガウス過程 (Multi-output GP) と呼ばれ,

$$(2.12) \quad \mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}(\mathbf{x}, \mathbf{x}'))$$

と表される. ここで, $\mathbf{K}(\mathbf{x}, \mathbf{x}') : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}^{M \times M}$ は行列値カーネル (Matrix-valued kernel) (Álvarez et al., 2012) と呼ばれる. 行列値カーネルも対称性および正定値性を満たすように構成され, 以下のように定義される.

定義(行列値カーネルに対する対称性と正定値性). 以下の二つの条件を満たすとき $K(\mathbf{x}, \mathbf{x}')$ は正定値カーネルである.

- 対称性: $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}', \mathbf{x})^\top$ が成り立つ.
- 正定値性: 任意の正の整数 N , 任意の点 $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^D$ および任意の実数ベクトル $\mathbf{c}_1, \dots, \mathbf{c}_N \in \mathbb{R}^M$ に対して以下が成り立つ.

$$(2.13) \quad \sum_{i=1}^N \sum_{j=1}^N \mathbf{c}_i^\top K(\mathbf{x}_i, \mathbf{x}_j) \mathbf{c}_j \geq 0.$$

この条件は、後述のグラム行列(式(2.15))が半正定値であることを意味する.

$K(\mathbf{x}, \mathbf{x}')$ の具体的な構成法については後続の段落にて述べる. 行列値カーネルを用いて多出力ガウス過程は以下のように定義される.

定義(多出力ガウス過程の定義). M 個の出力に対する全サンプル数を $N = \sum_{m=1}^M N_m$ とする. $\mathbf{X}_m = (\mathbf{x}_{m,1}, \dots, \mathbf{x}_{m,N_m})$ とし, 全ての入力をまとめて $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_M)^\top \in \mathbb{R}^{N \times D}$ とする. \mathbf{X}_m における関数値を $\mathbf{f}_m = (f_m(\mathbf{x}_{m,1}), \dots, f_m(\mathbf{x}_{m,N_m}))$ とし, M 個の出力からの関数値をまとめて $\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_M)^\top \in \mathbb{R}^N$ とする. 任意の自然数 N に対して, \mathbf{f} が N 次元ガウス分布

$$(2.14) \quad \mathbf{f} \sim \mathcal{N}(\mathbf{0}, K(\mathbf{X}, \mathbf{X}))$$

に従うとき, \mathbf{f} は多出力ガウス過程である. ここで, $K(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{N \times N}$ は行列値カーネルで計算されるグラム行列であり,

$$(2.15) \quad K(\mathbf{X}, \mathbf{X}) = \begin{pmatrix} k_{1,1}(\mathbf{X}_1, \mathbf{X}_1) & \cdots & k_{1,M}(\mathbf{X}_1, \mathbf{X}_M) \\ \vdots & \ddots & \vdots \\ k_{M,1}(\mathbf{X}_M, \mathbf{X}_1) & \cdots & k_{M,M}(\mathbf{X}_M, \mathbf{X}_M) \end{pmatrix}$$

と表される. ここで, $K(\mathbf{X}, \mathbf{X})$ は $M \times M$ のブロック行列であり, (m, m') 番目のブロック $k_{m,m'}(\mathbf{X}_m, \mathbf{X}_{m'}) \in \mathbb{R}^{N_m \times N_{m'}}$ は, 行列値カーネル $K(\mathbf{x}, \mathbf{x}')$ の (m, m') 番目の要素 $k_{m,m'}(\mathbf{x}, \mathbf{x}')$ を \mathbf{X}_m と $\mathbf{X}_{m'}$ について計算したグラム行列である.

以下では, 行列値カーネル $K(\mathbf{x}, \mathbf{x}')$ を構成する方法について述べる. 行列値カーネルを構成する最も簡単な方法は, 出力間の依存関係を表す変数を直接導入することにより,

$$(2.16) \quad K(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \mathbf{Q}$$

とすることである. ここで, $k(\mathbf{x}, \mathbf{x}') : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ は任意の正定値カーネルであり, $\mathbf{Q} \in \mathbb{R}^{M \times M}$ は出力の依存関係を表す半正定値行列である. 行列 \mathbf{Q} を半正定値とすることで, 適当な行列 \mathbf{A} を用いて $\mathbf{Q} = \mathbf{A} \mathbf{A}^\top$ のように分解することが可能である. これにより, 式(2.16)が式(2.13)の正定値性を満たすことが示される. $k(\mathbf{x}, \mathbf{x}')$ には, 入力次元毎にスケールパラメータを推定可能な ARD (Automatic relevance determination) カーネル (Rasmussen and Williams, 2006) などが使われる. ARD カーネルは,

$$(2.17) \quad k(\mathbf{x}, \mathbf{x}') = \alpha^2 \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}') \right)$$

と表され, $\mathbf{B} = \text{diag}(\beta_1^2, \dots, \beta_D^2)$ によって各入力次元に対するスケールが調節される.

この他にも, 行列値カーネルを構成する発展的手法は複数提案されているが, その中でも最もよく使われる手法の一つが Linear model of coregionalization (LMC) (Teh et al., 2005; Álvarez

et al., 2012)である。LMC では、複数の潜在的なガウス過程を用意し、それらの線形結合により多出力ガウス過程を表現する。\$L\$ 個の潜在ガウス過程を

$$(2.18) \quad g_\ell(\mathbf{x}) \sim \mathcal{GP}(0, k_\ell(\mathbf{x}, \mathbf{x}')), \quad \ell = 1, \dots, L$$

とする。ここで、\$k_\ell(\mathbf{x}, \mathbf{x}') : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}\$ は \$\ell\$ 番目の潜在ガウス過程の共分散関数である。\$\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_L(\mathbf{x}))^\top\$ として、重み行列 \$\mathbf{W} \in \mathbb{R}^{M \times L}\$ による線形変換を考え、

$$(2.19) \quad \mathbf{f}(\mathbf{x}) := \mathbf{W}\mathbf{g}(\mathbf{x})$$

によって \$M\$ 個の出力を持つ関数 \$\mathbf{f}(\mathbf{x})\$ を定義する。ガウス過程の線形変換は、またガウス過程になることが知られており、式(2.19)のように定義された関数 \$\mathbf{f}(\mathbf{x})\$ は多出力ガウス過程に従う。このとき、行列値カーネルは、

$$(2.20) \quad \mathbf{K}(\mathbf{x}, \mathbf{x}') = \mathbf{W}\mathbf{K}^{\text{diag.}}(\mathbf{x}, \mathbf{x}')\mathbf{W}^\top$$

と表される。ここで、\$\mathbf{K}^{\text{diag.}}(\mathbf{x}, \mathbf{x}') = \text{diag}(k_1(\mathbf{x}, \mathbf{x}'), \dots, k_L(\mathbf{x}, \mathbf{x}'))\$ である。\$\mathbf{K}^{\text{diag.}}(\mathbf{x}, \mathbf{x}')\$ は明らかに正定値カーネルであり、式(2.20)を式(2.13)の不等式の左辺に代入すれば、上記のように定義された行列式カーネルが正定性を満たすことが直ちにわかる。また、\$\mathbf{K}(\mathbf{x}, \mathbf{x}')\$ の \$(m, m')\$ 番目の要素は、

$$(2.21) \quad k_{m,m'}(\mathbf{x}, \mathbf{x}') = \sum_{\ell=1}^L w_{m,\ell} w_{m',\ell} k_\ell(\mathbf{x}, \mathbf{x}')$$

と表される。式(2.20)は、式(2.16)を一般化し、複数のカーネルの重み付き和によって行列値カーネルを構成することに対応する。これにより、関数の共分散構造に対する表現力を向上させることができ、\$L < M\$ を仮定することにより過学習を軽減することも可能である。

前章と同様、尤度と事前分布が共役であるため、事後分布が解析的に計算でき、

$$(2.22) \quad \mathbf{f}^*(\mathbf{x}) \mid \mathbf{y} \sim \mathcal{GP}(\mathbf{m}^*(\mathbf{x}), \mathbf{K}^*(\mathbf{x}, \mathbf{x}'))$$

と表される。ここで、\$m\$ 番目の出力に対する観測を \$\mathbf{y}_m = (y_{m,1}, \dots, y_{m,N_m})\$ としたとき、全ての観測をまとめて \$\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_M)^\top\$ とした。また、事後分布の平均関数 \$\mathbf{m}^*(\mathbf{x}) : \mathbb{R}^D \to \mathbb{R}^M\$、および、共分散関数 \$\mathbf{K}^*(\mathbf{x}, \mathbf{x}') : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^{M \times M}\$ は、

$$(2.23) \quad \mathbf{m}^*(\mathbf{x}) = \mathbf{K}(\mathbf{X}, \mathbf{x})^\top \mathbf{C}^{-1} \mathbf{y},$$

$$(2.24) \quad \mathbf{K}^*(\mathbf{x}, \mathbf{x}') = \mathbf{K}(\mathbf{x}, \mathbf{x}') - \mathbf{K}(\mathbf{X}, \mathbf{x})^\top \mathbf{C}^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x}')$$

となる。ここで、\$\mathbf{K}(\mathbf{X}, \mathbf{x}) \in \mathbb{R}^{N \times M}\$ は、\$\mathbf{X}\$ と任意の点 \$\mathbf{x}\$ に対して行列値カーネル \$\mathbf{K}(\mathbf{x}, \mathbf{x}')\$ の値を並べたものである。ただし、\$N = \sum_{m=1}^M N_m\$ である。また、\$\mathbf{C} \in \mathbb{R}^{N \times N}\$ は共分散行列であり、\$M \times M\$ のブロック行列で表され、\$(m, m')\$ 番目のブロックを \$\mathbf{C}_{m,m'}\$ とすると、

$$(2.25) \quad \mathbf{C}_{m,m'} = k_{m,m'}(\mathbf{X}_m, \mathbf{X}_{m'}) + \delta_{m,m'} \sigma_m^2 \mathbf{I}$$

と表される。ここで、\$k_{m,m'}(\mathbf{X}_m, \mathbf{X}_{m'}) \in \mathbb{R}^{N_m \times N_{m'}}\$ は、式(2.21)の行列値カーネルの \$(m, m')\$ 番目の要素を、\$\mathbf{X}_m\$ と \$\mathbf{X}_{m'}\$ に対して並べたグラム行列である。式(2.23)が \$M\$ 個の出力に対する全観測 \$\mathbf{y}\$ の線形変換によって表されることからわかるように、出力間の依存関係を考慮しつつ予測が行われる。

最後に、多出力ガウス過程回帰におけるパラメータ推定について述べる。上述の LMC に基づくモデル化において、推定すべきパラメータは、ノイズ分散 \$\Sigma\$、\$L\$ 個の潜在ガウス過程のカーネルパラメータ \$\{\alpha_\ell, \mathbf{B}_\ell\}_{\ell=1}^L\$、および、重み行列 \$\mathbf{W}\$ である。これらのパラメータをまとめ

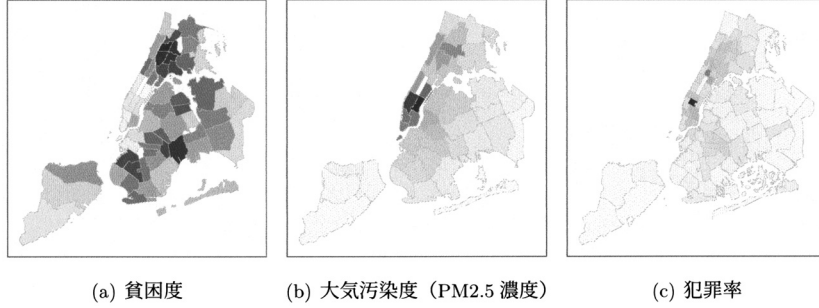


図 2. NYC Open Data が公開する集約データ (<https://opendata.cityofnewyork.us>). 属性によって都市の分割の仕方が異なることがわかる。

て θ と書くことにすると, θ は第二種最尤推定に基づき, 以下の最適化問題を解くことにより求めることができる.

$$(2.26) \quad \underset{\theta}{\text{maximize}} \log p(\mathbf{y}).$$

ここで, $p(\mathbf{y}) = \mathcal{N}(\mathbf{0}, \mathbf{C})$ であり, 解析的に計算が可能である.

3. 集約データのためのガウス過程

3.1 問題設定

前章で述べた標準的な回帰問題では, 点(座標や時刻)と属性の組でサンプルが表され, その背後にある関数を推定する. 一方, ここで扱う問題設定では, サンプルが領域(行政区画や時間帯)と属性の組で表されることを仮定する. 例えば, 都市において収集された貧困度や犯罪数のようなデータは, プライバシー保護や行政上の理由から, 行政区画や管轄区域において平均化などの統計処理がなされ, 公開されることが多い. 図 2 にニューヨーク市が公開する集約データの例を示す. 属性によって都市の分割の仕方が異なり, 様々な解像度を持つことがわかる.

以下に, 問題設定を数学的に定義する. 本稿では, 二次元空間における集約データを想定するが, 時空間における集約データなども同様に扱うことが可能である. 入力ドメインを $\mathcal{X} \in \mathbb{R}^2$ とし, $\mathbf{x} \in \mathcal{X}$ を入力変数とする. 属性の数を M とし, m 番目の属性に対するサンプルを $(\mathcal{R}_{m,n}, y_{m,n})$ と表す. ここで, $\mathcal{R}_{m,n} \subset \mathcal{X}$ は, m 番目の属性における n 番目のサンプルが紐づく領域を表し, 全ての領域は共通部分を持たないものとする. また, $y_{m,n} \in \mathbb{R}$ は属性の値である. m 番目の属性におけるサンプル数を N_m としたとき, 観測データは $\{(\mathcal{R}_{m,n}, y_{m,n} \mid m = 1, \dots, M; n = 1, \dots, N_m)\}$ である. 以下では, このような複数の集約データのモデリングについて考え, 背後にある M 個の属性に対する関数 $f(\mathbf{x}) : \mathbb{R}^D \rightarrow \mathbb{R}^M$ の推定問題を扱う.

3.2 手法

集約データのモデリングを難しくしているのは, 図 2 のように属性によって入力ドメインの分割の仕方が異なり, また領域の形は様々であることが挙げられる. そのため, 異なる領域分割に紐づいた複数のデータ間の依存関係を適切に捉える方法は自明ではない.

本稿では, Tanaka et al. (2019b)において提案された Spatially aggregated Gaussian process (SAGP) について紹介する. Yousefi et al. (2019)や Hamelijnck et al. (2019)においても, 同時

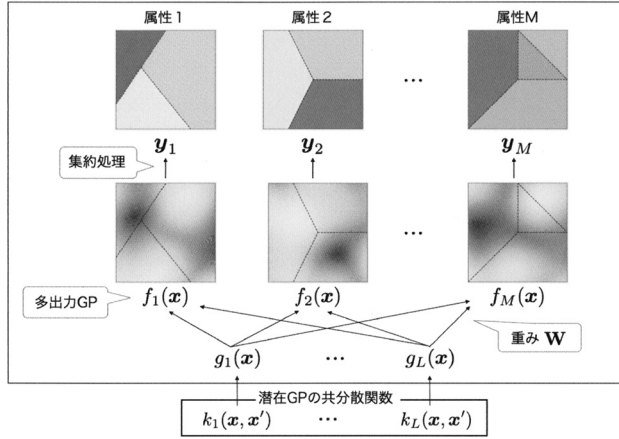


図 3. SAGP における集約データの生成モデル.

期に独立に類似手法が提案された．確率モデリングのアプローチに従い，観測データの生成プロセスを設計する．図 3 に，SAGP における集約データの生成モデルを示す．SAGP では，集約データの背後に滑らかな関数が存在することを仮定し，式 (2.19) における LMC に基づく多出力ガウス過程 $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_M(\mathbf{x}))^\top$ によって表す．次に，Smith et al. (2018), Burgess and Webster (1980) と同様の方法により，各集約データに対する集約処理をガウス過程の積分によってモデリングし，集約データの観測モデルを

$$(3.1) \quad \mathbf{y} \mid \mathbf{f}(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{y} \mid \int_{\mathcal{X}} \mathbf{A}(\mathbf{x}) \mathbf{f}(\mathbf{x}) d\mathbf{x}, \boldsymbol{\Sigma}\right)$$

とする．ここで，ガウス過程 $\mathbf{f}(\mathbf{x})$ は積分可能であると仮定する（ガウス過程の可積分性についての議論は Tanaka et al. 2019b の付録を参照）．また， $\mathbf{y}_m = (y_{m,1}, \dots, y_{m,N_m})$ として， $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_M)^\top \in \mathbb{R}^N$ とした． $\mathbf{A}(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}^{N \times M}$ は，

$$(3.2) \quad \mathbf{A}(\mathbf{x}) = \begin{pmatrix} \mathbf{a}_1(\mathbf{x}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_2(\mathbf{x}) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{a}_M(\mathbf{x}) \end{pmatrix}$$

と表され， $\mathbf{a}_m(\mathbf{x}) = (a_{m,1}(\mathbf{x}), \dots, a_{m,N_m}(\mathbf{x}))^\top$ である．ここで，各要素 $a_{m,n}(\mathbf{x})$ は，領域 $\mathcal{R}_{m,n}$ におけるデータ集約のための非負の重み関数であり，LMC によって表現された関数 $\mathbf{f}(\mathbf{x})$ の集約の範囲と集約の仕方を指定する．例えば，

$$(3.3) \quad a_{m,n}(\mathbf{x}) = \frac{\mathbb{1}(\mathbf{x} \in \mathcal{R}_{m,n})}{\int_{\mathcal{X}} \mathbb{1}(\mathbf{x}' \in \mathcal{R}_{m,n}) d\mathbf{x}'}$$

とすると， $y_{m,n}$ は領域 $\mathcal{R}_{m,n}$ における $f_m(\mathbf{x})$ の平均値を表す．ここで， $\mathbb{1}(\cdot)$ は指示関数であり， Z が真のとき $\mathbb{1}(Z) = 1$ となり，そうでないとき $\mathbb{1}(Z) = 0$ となる．重み関数を変更することで，単純な合計や人口に対する重み付け和などを表現することも可能である．また，式 (3.1) の $\boldsymbol{\Sigma}$ はノイズ分散であり， $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2 \mathbf{I}, \dots, \sigma_M^2 \mathbf{I})$ である．

事前分布を LMC とし，尤度を式 (3.1) としたときの事後分布について述べる．Tanaka et al. (2019b) において，事後分布がガウス過程として解析的に計算できることが示された．導出に

については (Tanaka et al., 2019b) の付録に譲り、ここでは結果だけを紹介する。事後分布は、式 (2.22) と同様に表され、平均関数 $\mathbf{m}^*(\mathbf{x})$ と共分散関数 $\mathbf{K}^*(\mathbf{x}, \mathbf{x}')$ は、

$$(3.4) \quad \mathbf{m}^*(\mathbf{x}) = \mathbf{H}(\mathbf{x})^\top \mathbf{C}^{-1} \mathbf{y},$$

$$(3.5) \quad \mathbf{K}^*(\mathbf{x}, \mathbf{x}') = \mathbf{K}(\mathbf{x}, \mathbf{x}') - \mathbf{H}(\mathbf{x})^\top \mathbf{C}^{-1} \mathbf{H}(\mathbf{x}')$$

と表される。ここで、 $\mathbf{H}(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}^{N \times M}$ は、

$$(3.6) \quad \mathbf{H}(\mathbf{x}) = \int_{\mathcal{X}} \mathbf{A}(\mathbf{x}') \mathbf{K}(\mathbf{x}', \mathbf{x}) d\mathbf{x}'$$

であり、任意の点 \mathbf{x} と各領域との間の共分散を表す。また、 $\mathbf{C} \in \mathbb{R}^{N \times N}$ は共分散行列であり、

$$(3.7) \quad \mathbf{C} = \iint_{\mathcal{X} \times \mathcal{X}} \mathbf{A}(\mathbf{x}) \mathbf{K}(\mathbf{x}, \mathbf{x}') \mathbf{A}(\mathbf{x}')^\top d\mathbf{x} d\mathbf{x}' + \Sigma$$

と表される。 \mathbf{C} は $M \times M$ のブロック行列であり、 (m, m') 番目のブロックを $\mathbf{C}_{m, m'}$ とすると、

$$(3.8) \quad \mathbf{C}_{m, m'} = \iint_{\mathcal{X} \times \mathcal{X}} k_{m, m'}(\mathbf{x}, \mathbf{x}') \mathbf{a}_m(\mathbf{x}) \mathbf{a}_{m'}(\mathbf{x}')^\top d\mathbf{x} d\mathbf{x}' + \delta_{m, m'} \sigma_m^2 \mathbf{I}$$

である。ここで、式 (3.8) における $k_{m, m'}(\mathbf{x}, \mathbf{x}')$ は、式 (2.21) で表される LMC の共分散関数であり、「二点間 $(\mathbf{x}, \mathbf{x}')$ の共分散」を規定している。また、 $k_{m, m'}(\mathbf{x}, \mathbf{x}')$ においては、重み \mathbf{W} を通じて出力間 $(f_m(\mathbf{x}), f_{m'}(\mathbf{x}))$ の依存関係が導入されている。一方、式 (3.8) の $\mathbf{C}_{m, m'}$ は、領域の各ペアに対して $k_{m, m'}(\mathbf{x}, \mathbf{x}')$ の二重積分を計算しており、これによって「二つの領域間 $(\mathcal{R}, \mathcal{R}')$ の共分散」を規定している。この二重積分は、直感的には、二つの領域に含まれる無限個の点に対する共分散を計算し、集約することを意味する。ここで、式 (3.8) を実装する際には、カーネルの積分を数値的に近似計算する必要があることに注意する。

以上の定式化により、領域の形や大きさを考慮した共分散を評価可能であり、それを用いてデータの予測を実現する。集約データのための共分散行列 \mathbf{C} を用いた周辺尤度に基づき、パラメータ最適化も可能である。

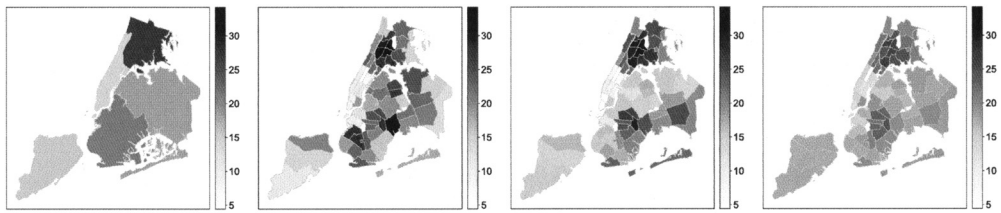
3.3 実験

本節では、2.2 節で述べた LMC と SAGP との比較実験について述べる。LMC は点に紐づくデータを前提としているため、集約データの値を領域の重心に紐付けることで適用した。潜在ガウス過程の共分散関数には ARD カーネルを用い、潜在ガウス過程の数は leave-one-out 交差検定により、 $\{1, \dots, M-1\}$ から決定した。モデルパラメータは、周辺尤度を最大にするように L-BFGS 法 (Liu and Nocedal, 1989) を用いて最適化した。また、SAGP におけるカーネルの領域積分は、都市全体を $300\text{m} \times 300\text{m}$ のグリッドで覆い、合計 9,352 個のグリッド点を用いて近似した。

実験で用いたデータは、ニューヨーク市が公開している 10 種類の集約データで、NYC Open Data (<https://opendata.cityofnewyork.us>) からダウンロードが可能である。表 1 に、実験で用いた集約データをまとめる。ここで、全てのデータの値を平均 0、分散 1 に正規化して使用した。本実験では、様々な解像度を持つ複数の集約データを用いて、集約データを高解像度化するタスクを解いた。高解像度な集約データの予測精度を評価するために、まずターゲットとする集約データを 1 つ決め、そのデータを低解像度化したもの (図 4(a) 参照) とその他のデータを用いてモデルパラメータを学習し、学習済みモデルを用いて元の高解像度なデータを予測した。評価指標は、Mean absolute percentage error (MAPE) であり、 m をターゲットの集約データの属性としたとき、

表 1. 集約データ. データは年に一度収集されており, 複数年のデータがある場合は平均値を計算して実験に用いた.

データ	分割の仕方	領域数	取得年
PM2.5	UHF42	42	2009 – 2010
Poverty rate	Community district	59	2009 – 2013
Unemployment rate	Community district	59	2009 – 2013
Mean commute	Community district	59	2009 – 2013
Population	Community district	59	2009 – 2013
Recycle diversion rate	Community district	59	2009 – 2013
Crime	Police precinct	77	2010 – 2016
Fire incident	Zip code	186	2010 – 2016
311 call	Zip code	186	2010 – 2016
Public telephone	Zip code	186	2016



(a) 低解像度データ (b) 高解像度データ (c) SAGP の予測結果 (d) LMC の予測結果

図 4. Poverty rate データを高解像度化したときの可視化結果.

表 2. 高解像度データの予測に対する MAPE と標準誤差. 括弧内の数字は交差検定によって選ばれた潜在ガウス過程の数 L を示す. 太文字は t 検定 (p 値は 0.05) において SAGP と LMC との間に有意差があることを示す.

	LMC	SAGP
PM2.5	0.036 ± 0.005 (6)	0.030 ± 0.005 (5)
Poverty rate	0.207 ± 0.025 (4)	0.177 ± 0.019 (3)
Unemployment rate	0.195 ± 0.024 (3)	0.165 ± 0.020 (3)
Mean commute	0.057 ± 0.007 (4)	0.050 ± 0.007 (6)
Population	0.337 ± 0.039 (3)	0.295 ± 0.033 (3)
Recycle diversion rate	0.222 ± 0.032 (4)	0.211 ± 0.029 (4)
Crime	0.401 ± 0.053 (2)	0.379 ± 0.055 (3)
Fire incident	0.500 ± 0.052 (4)	0.396 ± 0.038 (3)
311 call	0.061 ± 0.004 (6)	0.052 ± 0.003 (3)
Public telephone	0.086 ± 0.008 (4)	0.080 ± 0.007 (6)

$$(3.9) \quad \frac{1}{N_m} \sum_{n=1}^{N_m} \left| \frac{y_{m,n}^{\text{true}} - y_{m,n}^*}{y_{m,n}^{\text{true}}} \right|$$

と表される. ここで, $y_{m,n}^{\text{true}}$ は真の値を表し, $y_{m,n}^*$ は予測値を表す. SAGP の予測値は, 式 (3.4) の平均関数 $m^*(x, x')$ を高解像度な各領域において積分することで求めた.

表 2 に SAGP と LMC に対する MAPE と標準誤差を示す. すべてのデータセットにおいて, SAGP は LMC と同等かそれ以上の精度を達成した. 特に, 表 2 における太文字は, t 検定の結果, SAGP と LMC との差が統計的に有意であったことを表す. また, 図 4 に Poverty rate

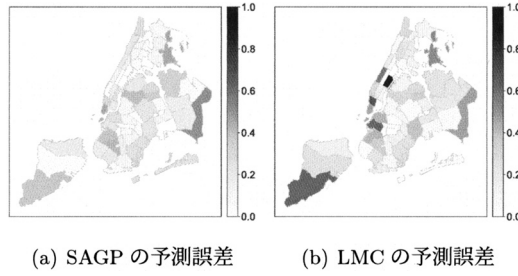


図 5. Poverty rate データに対する予測誤差の可視化結果.

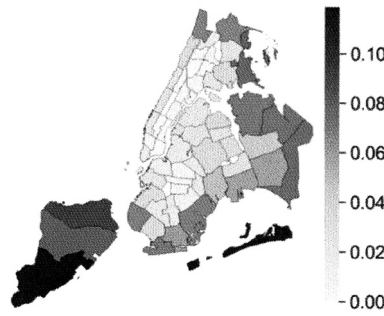


図 6. SAGP の予測分散.

データを高解像度化した際の可視化結果を示し、図 5 に SAGP と LMC に対する予測誤差の可視化結果を示す。ここで、予測誤差は Absolute percentage error であり、 $|(y_{m,n}^{\text{true}} - y_{m,n}^*)/y_{m,n}^{\text{true}}|$ を各領域について計算した。これらの結果より、SAGP が様々な種類の集約データを活用して、高解像度データをより正確に予測できることがわかる。

SAGP を用いると、領域に紐づくデータを予測した際の分散を評価することも可能である。図 6 に SAGP の予測分散を示す。予測分散は、式(3.5)を領域について積分することで計算できる。図 6 より、都市の端に位置する地域の分散が、都市の内部に位置する地域に比べて大きな値をとる傾向があることがわかる。一般に外挿は内挿よりも難しいので、合理的な結果であると言える。

4. 力学系のためのガウス過程

4.1 問題設定

本節では、力学系の時間発展を表すダイナミクスをガウス過程によりモデリングすることを考える。まず対象とする力学系について述べる。本稿では、解析力学の一形式である「ハミルトン力学」によって表現される系(ハミルトン系と呼ぶ)を対象とする。ハミルトン力学とは、ニュートン力学をエネルギーをベースに再定式化したものであり、ニュートン力学に基づく系を表現可能である。ハミルトン力学では、系の自由度を $M/2$ としたとき、一般化座標 $\mathbf{u}^q \in \mathbb{R}^{M/2}$ と一般化運動量 $\mathbf{u}^p \in \mathbb{R}^{M/2}$ を用いて、状態 $\mathbf{u} = (\mathbf{u}^q, \mathbf{u}^p) \in \mathbb{R}^M$ のダイナミクスを考える。 \mathbf{u} によって生成される空間を「相空間」(状態空間と同義)と呼ぶ。ハミルトン力学に基づく定式化では、系全体のエネルギー関数に対応する「ハミルトニアン」 $H(\mathbf{u}) : \mathbb{R}^M \rightarrow \mathbb{R}$ を設計する。ハミルトニアンが与えられれば、ハミルトンの運動方程式にしたがって、ダイナミクスは、

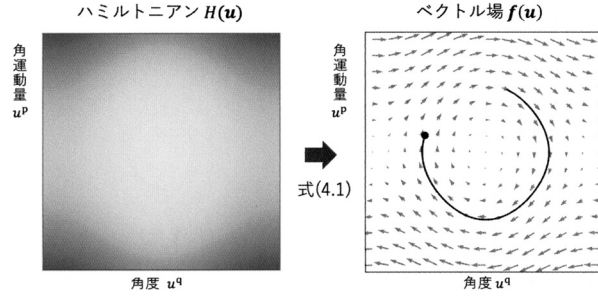


図 7. 単振り子のハミルトニアンとベクトル場. 左図の色の濃淡はエネルギーの大きさを表す. 式(4.1)によって定義されるベクトル場(右図)はシンプレクティック構造を持ち, 任意の初期条件からの軌跡(黒線)はエネルギーの値を保存しつつ時間発展する.

$$(4.1) \quad \dot{\mathbf{u}} = (\mathbf{S} - \mathbf{R})\nabla H(\mathbf{u}) =: \mathbf{f}(\mathbf{u})$$

と表される. ここで, $\dot{\mathbf{u}}$ は状態の時間微分 $d\mathbf{u}/dt$ を表し, $\nabla H(\mathbf{u}) : \mathbb{R}^M \rightarrow \mathbb{R}^M$ はハミルトニアン H の状態 \mathbf{u} についての勾配を表す. また, $\mathbf{S} \in \mathbb{R}^{M \times M}$ は歪対称行列

$$(4.2) \quad \mathbf{S} = \begin{pmatrix} \mathbf{O} & \mathbf{I} \\ -\mathbf{I} & \mathbf{O} \end{pmatrix}$$

であり, $\mathbf{R} \in \mathbb{R}^{M \times M}$ はエネルギーの散逸を表す半正定値行列である. 式(4.1)の右辺を「ベクトル場」と呼び, $\mathbf{f}(\mathbf{u}) : \mathbb{R}^M \rightarrow \mathbb{R}^M$ と定義する. $\mathbf{R} = \mathbf{O}$ のとき, ハミルトンの運動方程式によって定義されたベクトル場 $\mathbf{f}(\mathbf{u})$ は, シンプレクティック構造(Symplectic structure)という幾何学的構造を持ち, 式(4.1)に従うダイナミクスはハミルトニアン $H(\mathbf{u})$ を保存(すなわちエネルギー保存)することが知られている. また, $\mathbf{R} = \text{diag}(0, \dots, 0, r_1, \dots, r_{M/2})$ とすることで, 摩擦係数を $r_d \geq 0$ とする加法的な散逸系を表すこともできる.

本稿で扱う問題を定義する. ϵ を M 次元のガウスノイズとして, 観測 $\hat{\mathbf{y}}$ が, $\hat{\mathbf{y}} = \dot{\mathbf{u}} + \epsilon$ によって得られるとする. N 個の状態に対する微分観測(derivative observation) $\{(\mathbf{u}_n, \hat{\mathbf{y}}_n) \mid n = 1, \dots, N\}$ が与えられたとしたとき, ダイナミクスを表すベクトル場 $\mathbf{f}(\mathbf{u})$ を推定する問題を扱う. 推定したベクトル場と, 時刻 $t = 0$ における任意の初期条件 \mathbf{u}_0 が与えられたとき,

$$(4.3) \quad \mathbf{u}(t) = \mathbf{u}_0 + \int_0^t \mathbf{f}(\mathbf{u}) dt$$

を数値的に求めることで, 力学系のシミュレーションが可能である.

以下にハミルトン系の例として単振り子のハミルトニアンとベクトル場について述べる.

例(単振り子). 図 7 に, 単振り子を例としたときのハミルトニアンとベクトル場を可視化する. 単振り子の自由度は 1 であり, u^q は振り子の角度を, u^p は振り子の角運動量を表す. 単振り子のハミルトニアンは,

$$(4.4) \quad H(\mathbf{u}) = 2mg\ell(1 - \cos u^q) + \frac{\ell^2(u^p)^2}{2m}$$

と表される. ここで, g は重力定数, m は質量, ℓ は振り子の長さを表す. 図 7 の可視化では, $g = 3$ および $m = \ell = 1$ とし, $\mathbf{R} = \mathbf{O}$ とした.

4.2 手法

ベクトル場 $\mathbf{f}(\mathbf{u})$ は, 素朴には 2.2 節で述べた多出力ガウス過程によりモデリングすること

が可能である。しかし、単純なガウス過程では、物理法則を満たすような学習結果が必ずしも得られるとは限らない。そこで、本節では、ハミルトン力学の理論を組み込むことで、エネルギーの保存・散逸則を満たすように制約を加える方法について述べる。

Rath et al. (2021)や Tanaka et al. (2022)では、ベクトル場を直接モデリングする代わりに、ハミルトニアン $H(\mathbf{u})$ を単一出力を持つガウス過程を用いて、

$$(4.5) \quad H(\mathbf{u}) \sim \mathcal{GP}(0, k(\mathbf{u}, \mathbf{u}'))$$

とモデル化する。式(4.1)における微分作用素を $\mathcal{L} := (\mathbf{S} - \mathbf{R})\nabla$ とすると、ベクトル場は $\mathbf{f}(\mathbf{u}) = \mathcal{L}H(\mathbf{u})$ によって与えられる。微分作用素 \mathcal{L} は線形であるため、 $\mathbf{f}(\mathbf{u})$ は多出力ガウス過程で表される (Solak et al., 2003)。このとき、行列値カーネルは、

$$(4.6) \quad \mathbf{K}(\mathbf{u}, \mathbf{u}') = \mathcal{L}k(\mathbf{u}, \mathbf{u}')\mathcal{L}^\top$$

$$(4.7) \quad = (\mathbf{S} - \mathbf{R})[\nabla^2 k(\mathbf{u}, \mathbf{u}')](\mathbf{S} - \mathbf{R})^\top$$

と表される。ここで、 ∇^2 は Hessian operator を表す。式(4.7)は、シンプレクティックカーネル (Symplectic kernel) と呼ばれ (Boffi et al., 2022)、シンプレクティックカーネルに基づくガウス過程回帰のことをシンプレクティックガウス過程 (symplectic Gaussian process: SGP) (Rath et al., 2021) と呼ぶ。式(4.6)の右辺は、式(2.20)の LMC に基づく行列値カーネルと同様の形をしていることがわかる。LMC における線形変換 \mathbf{W} が線形作用素 \mathcal{L} に置き換わっているが、線形作用素は線形変換を無限次元に一般化したものとみなせることに注意すれば、式(4.7)のシンプレクティックカーネルは正定値カーネルであることがわかるであろう。以上より、ハミルトン力学の理論を活用し、エネルギーの保存・散逸則に従うダイナミクスを表現するのに適した多出力ガウス過程が導出された。カーネルパラメータの学習や予測は、2.2 節で述べた方法で実行可能である。

本稿では割愛するが、Tanaka et al. (2022)や Boffi et al. (2022)では、シンプレクティックカーネルをフーリエ基底によって近似するためのシンプレクティック乱択化フーリエ特徴 (symplectic random Fourier feature: S-RFF) が提案されており、サンプル数が多いときに学習の効率化を図ることができる。さらに、Tanaka et al. (2022)では、数値ソルバを活用した変分学習アルゴリズムを開発することにより、状態の軌跡だけが与えられる状況でもハミルトン系の学習を可能とした。詳細は (田中, 2023) にも解説されている。

4.3 実験

本節では、SGP の評価を行うため、ハミルトニアンニューラルネットワーク (Hamiltonian neural network: HNN) (Greydanus et al., 2019) との比較を行う。HNN は、ハミルトニアンをニューラルネットワークにより近似し、自動微分を活用してベクトル場を導出する方法である。HNN は 3 層の多層パーセプトロンを用いて実装し、ユニット数は 200、活性化関数は \tanh を用いた。SGP および HNN におけるパラメータ推定には、Adam (Kingma and Ba, 2015) を用い、学習率は $1e-3$ とした。

実験では、4.1 節で述べた単振り子、および、調和振動子を対象にした。調和振動子のハミルトニアンは、

$$(4.8) \quad H(\mathbf{u}) = \frac{1}{2}k(u^q)^2 + \frac{(u^p)^2}{2m}$$

と表される。ここで、バネ定数 $k = 1$ 、質量 $m = 1$ に設定した。また、どちらの系も摩擦はないものとし、 $\mathbf{R} = \mathbf{O}$ とした。学習のために $N = \{40, 60, 100\}$ 個のサンプル $\{(\mathbf{u}_n, \dot{\mathbf{y}}_n)\}$ を作成

表 3. ベクトル場に対する MSE と標準偏差. 太文字は t 検定 (p 値は 0.05) において SGP と HNN の間に有意差があることを示す.

(a) 調和振動子			
サンプル数	40	60	100
HNN	0.600±0.098	0.476±0.048	0.368±0.034
SGP	0.041±0.053	0.040±0.021	0.011±0.001

(b) 単振り子			
サンプル数	40	60	100
HNN	2.253±0.107	1.996±0.099	1.863±0.089
SGP	0.323±0.161	0.237±0.056	0.200±0.030

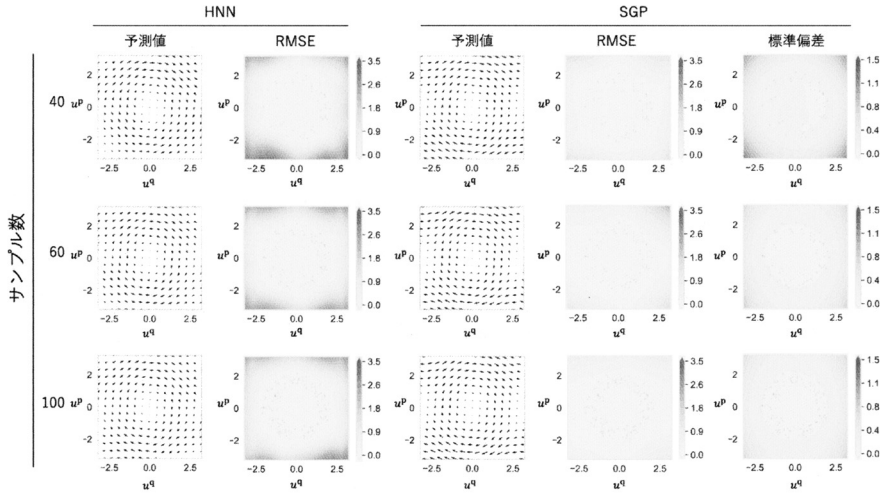


図 8. 予測されたベクトル場の可視化. 灰色の点は学習点を表す.

した. ここで, 状態 \mathbf{u}_n は, エネルギーの値が $[1.3, 2.0]$ の範囲に含まれるように一様分布から生成した. また, 観測 \mathbf{y}_n は, ハミルトンの運動方程式により $\dot{\mathbf{u}}_n = \mathbf{S} \nabla H(\mathbf{u}_n)$ を計算し, ノイズ分散を $\sigma^2 = 0.1$ とするガウスノイズを加えることで生成した.

テスト時には, $[-3.2, 3.2]^2$ で定義された相空間を 15×15 のグリッドに区切り, テスト点を $\{\mathbf{x}_i^* \mid i = 1, \dots, I\}$ とした. ここで, I はテストデータのサンプル数である. 評価指標は mean squared error (MSE) であり,

$$(4.9) \quad \frac{1}{I} \sum_{i=1}^I \|\dot{\mathbf{u}}_i^{\text{pred}} - \dot{\mathbf{u}}_i^{\text{true}}\|^2$$

と表される. ここで, $\dot{\mathbf{u}}_i^{\text{true}}$ は各テスト点における真のダイナミクスを表し, $\dot{\mathbf{u}}_i^{\text{pred}}$ はその予測値を表す.

表 3 に, HNN と SGP によって予測されたベクトル場に対する MSE と標準偏差を示す. この結果から, SGP がすべての場合において HNN よりも高精度にベクトル場の予測ができていることがわかる. SGP はサンプル数が少ない場合に特に有効であることがわかる. これはガウス過程が不確実性を考慮できるモデルであり, ノイズや少量データに有効であるためである. 図 8 に, 単振り子のデータに対して HNN と SGP が予測したベクトル場, RMSE (Root mean squared error), および, SGP の予測に対する標準偏差を示す. 真のベクトル場 (図 7 の右図)

とその予測値(図 8)を比較すると、特にサンプル数が少ない場合に SGP が HNN よりも高精度であることがわかる。SGP の予測に対する標準偏差はサンプル数をよく反映しており、RMSE と比較することで、標準偏差が予測値の信頼度を正確に表現していることがわかる。さらに、図 8 の RMSE からわかるように、学習点がない領域においても、SGP は HNN よりも誤差が小さくロバストであることがわかる。

5. 結論

本稿では、ガウス過程の基礎事項、特に多出力ガウス過程について解説し、それに関連する二つの研究トピックについて紹介した。一つ目は、集約データに対するガウス過程であり、点ではなく領域に紐づくデータのモデリングである。ガウス過程の領域積分により集約データを表現することで、事後分布が解析的に求まることについて述べた。これにより、領域の形や大きさを考慮した共分散関数を定義でき、それに基づいた予測が可能となる。二つ目は、力学系のためのガウス過程であり、ダイナミクスを表現するベクトル場のモデリングである。ハミルトン力学の理論に基づき、エネルギー関数をガウス過程でモデリングし、ハミルトンの運動方程式にしたがって導出されたガウス過程は、エネルギーの保存・散逸則を満たすことについて述べた。二つのトピックにおいて共通かつ重要であることは、各トピックにおけるモデリングを通じて、多出力ガウス過程の行列値カーネルにデータ集約過程や物理法則といった事前知識が埋め込まれたことである。これにより、データと知識とを同時に活用し、高精度な予測を可能とした。近年は深層学習のブームが続いている状況であるが、データに含まれる不確実性を考慮でき、ノイズや少量データからの学習に強みがあるガウス過程は、今後も注目すべき機械学習手法の一つであると言える。

参 考 文 献

- Álvarez, M. A., Rosasco, L. and Lawrence, N. D. (2012). Kernels for vector-valued functions: A review, *Foundations and Trends® in Machine Learning*, **4**(3), 195–266.
- Beckers, T., Seidman, J., Perdikaris, P. and Pappas, G. J. (2022). Gaussian process port-Hamiltonian systems: Bayesian learning with physics prior, *2022 IEEE 61st Conference on Decision and Control*, 1447–1453, <https://doi.org/10.1109/CDC51059.2022.9992733>.
- Boffi, N. M., Tu, S. and Slotine, J.-J. E. (2022). Nonparametric adaptive control and prediction: Theory and randomized algorithms, *Journal of Machine Learning Research*, **23**(281), 1–46.
- Burgess, T. M. and Webster, R. (1980). Optimal interpolation and isarithmic mapping of soil properties I: The semi-variogram and punctual kriging, *Journal of Soil Science*, **31**(2), 315–331.
- Chen, R. T. Q., Rubanova, Y., Bettencourt, J. and Duvenaud, D. K. (2018). Neural ordinary differential equations, *Advances in Neural Information Processing Systems*, **31**, https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf (最終アクセス日 2024 年 11 月 22 日).
- 福水健次 (2010). 『カーネル法入門—正定値カーネルによるデータ解析—』, 朝倉書店, 東京.
- Goldstein, H. (1980). *Classical Mechanics*, Addison-Wesley, Reading, Massachusetts.
- Greydanus, S., Dzamba, M. and Yosinski, J. (2019). Hamiltonian neural networks, *Advances in Neural Information Processing Systems*, **32**, https://proceedings.neurips.cc/paper_files/paper/2019/file/26cd8ecadce0d4efd6cc8a8725cbd1f8-Paper.pdf (最終アクセス日 2024 年 11 月 22 日).
- Hamelijnck, O., Damoulas, T., Wang, K. and Girolami, M. (2019). Multi-resolution multi-task Gaussian processes, *Advances in Neural Information Processing Systems*, **32**, https://proceedings.neurips.cc/paper_files/paper/2019/file/0118a063b4aae95277f0bc1752c75abf-Paper.pdf (最終アクセス日 2024 年 11 月 22 日).

- 畑浩之 (2014). 『解析力学, 基幹講座物理学』, 東京図書, 東京.
- Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization, *International Conference on Learning Representations*, 1–15.
- Law, H. C. L., Sejdinovic, D., Cameron, E., Lucas, T. C. D., Flaxman, S., Battle, K. and Fukumizu, K. (2018). Variational learning on aggregate outputs with Gaussian processes, *Advances in Neural Information Processing Systems*, **31**, https://proceedings.neurips.cc/paper_files/paper/2018/file/24b43fb034a10d78bec71274033b4096-Paper.pdf (最終アクセス日 2024 年 11 月 22 日).
- Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization, *Mathematical Programming*, **45**(1–3), 503–528.
- Macêdo, I. and Castro, R. (2010). Learning divergence-free and curl-free vector fields with matrix-valued kernels, Technical Report, Instituto Nacional de Matematica Pura e Aplicada, Brazil.
- Narcowich, F. J. and Ward, J. D. (1994). Generalized hermite interpolation via matrix-valued conditionally positive definite functions, *Mathematics of Computation*, **63**(208), 661–687.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Massachusetts.
- Rath, K., Albert, C. G., Bischl, B. and von Toussaint, U. (2021). Symplectic Gaussian process regression of maps in Hamiltonian systems, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **31**(5), <https://doi.org/10.1063/5.0048129>.
- Smith, M. T., Álvarez, M. A. and Lawrence, N. D. (2018). Gaussian process regression for binned data, arXiv, <https://arxiv.org/abs/1809.02010> (最終アクセス日 2024 年 11 月 22 日).
- Solak, E., Murray-smith, R., Leithhead, W., Leith, D. and Rasmussen, C. (2003). Derivative observations in Gaussian process models of dynamic systems, *Advances in Neural Information Processing Systems*, **15**, https://proceedings.neurips.cc/paper_files/paper/2002/file/5b8e4fd39d9786228649a8a8bec4e008-Paper.pdf (最終アクセス日 2024 年 11 月 22 日).
- 田中佑典 (2023). ガウス過程と物理現象のモデル化, 人工知能, **38**(3), 318–325.
- Tanaka, Y. (2024). Learning Hamiltonian dynamics under uncertainty via symplectic Gaussian processes, *International Conference on Scientific Computing and Machine Learning*, <https://scml.jp/2024/paper/15/CameraReady/scml2024.pdf> (最終アクセス日 2024 年 11 月 22 日).
- Tanaka, Y., Iwata, T., Tanaka, T., Kurashima, T., Okawa, M. and Toda, H. (2019a). Refining coarse-grained spatial data using auxiliary spatial data sets with various granularities, *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**(01), 5091–5100.
- Tanaka, Y., Tanaka, T., Iwata, T., Kurashima, T., Okawa, M., Akagi, Y. and Toda, H. (2019b). Spatially aggregated Gaussian processes with multivariate areal outputs, *Advances in Neural Information Processing Systems*, **32**, https://proceedings.neurips.cc/paper_files/paper/2019/file/a941493eeea57ede8214fd77d41806bc-Paper.pdf (最終アクセス日 2024 年 11 月 22 日).
- Tanaka, Y., Iwata, T. and Ueda, N. (2022). Symplectic spectrum Gaussian processes: Learning Hamiltonians from noisy and sparse data, *Advances in Neural Information Processing Systems*, **35**, https://proceedings.neurips.cc/paper_files/paper/2022/file/82f05a105c928c10706213952bf0c8b7-Paper-Conference.pdf (最終アクセス日 2024 年 11 月 22 日).
- Teh, Y. W., Seeger, M. and Jordan, M. I. (2005). Semiparametric latent factor models, *International Conference on Artificial Intelligence and Statistics*, 333–340.
- Yousefi, F., Smith, M. T. and Álvarez, M. A. (2019). Multi-task learning for aggregated data using Gaussian processes, *Advances in Neural Information Processing Systems*, **32**, https://proceedings.neurips.cc/paper_files/paper/2019/file/64517d8435994992e682b3e4aa0a0661-Paper.pdf (最終アクセス日 2024 年 11 月 22 日).

Spatio-temporal Data Analysis Using Gaussian Processes

Yusuke Tanaka¹ and Naonori Ueda²

¹NTT Communication Science Laboratories

²RIKEN Center for Advanced Intelligence Project

A Gaussian process (GP) is a probabilistic model for estimating a function from data and has long been widely used in spatio-temporal data modeling. This paper addresses two significant problems related to spatio-temporal data analysis and GP models. The first problem is the modeling of aggregate data. Due to privacy concerns and administrative reasons, spatio-temporal data (e.g., poverty rates) obtained in cities are not associated with points but with regions. This paper introduces a way to handle aggregate data by an integral of the GPs over regions. The second problem is the modeling of dynamical systems defined by ordinary differential equations. The system's dynamics are defined by the time derivative of the state and are represented as a vector field in phase space. Although vector fields can be modeled naively using multi-output GPs, the vanilla GPs do not always satisfy physical laws. In this paper, we introduce a method for modeling vector fields that adhere to energy conservation or dissipation laws by combining the theory of Hamiltonian mechanics with GPs.