The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

多環境における低精度センサーを用いた空間分布のベイズ推定

Bayesian Estimation for Spatial distribution using Low precision Sensors in Multi-Environment

太田 真人 *1 Masato Ota　花房 諒 *1 Ryo Hanafusa　岡留 剛 *1 Takeshi Okadome

*1関西学院大学大学院 理工学研究科
Graduate School of Science and Technology, Kwansei Gakuin University

The method proposed in this paper enables us to estimate the spatial distribution of physical quantities in multiple geographical regions, where many low-precision sensors are densely placed and a small number of (or no) high-precision sensors are positioned. For a region that has high-precision sensors, the method determines the biases of the low-precision sensors placed in the regions accurately using the values of the high-precision sensors and it corrects the values of the low-precision sensors precisely. Furthermore, the method divides the regions into the clusters using the sensor data similarity as a similarity measure. Then, for a region that has no high-precision sensor, it estimates the spatial distribution of physical quantities in the region by a novel multi-task learning that transfers the regional shared-information in the cluster which the region belongs to. Some experiments show that the method estimates the spatial distribution of physical quantities accurately.

1. はじめに

生態の環境モニタリングや農業や漁業・都市における環境センシングにおいて，コストがかかる高精度センサーは多く配置することができず配置する場所が限られる．一方，低精度センサーは誤差は大きいが，安価で多地点に配置でき，それを用いて大量のセンサデータを入手することができる．例えば，AMeDAS を高精度センサーとみなし，Raspberry Pi を使ったセンサノードや衛星から撮ったサーモ画像を低精度センサーとみなして，ある領域の各点での温度を予測するのに用いることができる．

環境から得られるセンサデータはその総数も種類も増えてきており，多数のセンサデータを用いた空間分布の推定は多く研究されている [Xu 17]．さらに，センシングする環境の対象も増え，本研究で対象とする多環境は，複数の都市や，各地域の工場内や港，各地方の森林が挙げられる．しかし，高精度センサーがある環境は少なく，低精度センサーしかおけない環境は多くある．その場合，低精度センサーだけで高精度に環境をセンシングする必要がある．本研究では，各環境における物理量の空間分布を推定するタスクにおいてマルチタスク学習を行い，低精度センサーだけの環境においても精度高く推定することを目的とする．マルチタスク学習は，関連する複数のタスクを解く際に，タスク間で情報を共有することで推定精度を上げる学習手法である．

地理空間データにおけるマルチタスク学習として，タスク間を関係付けさせるタスク関係行列に地形などの景観類似性や，環境間の場所が近い空間的近接性をパラメータの正則化に用いた研究がされている [Atluri 18]．しかし，位置が互いに近く，景観は似ていてるが，環境周りの外的要因や遠方の影響を受けている場合があり，位置の近傍関係は正しいタスク間の関係に必ずしも反映していない．例えば，地形は似ていてるが，一方の環境では工場地区が近くにあり他方に比べて CO_2 量や雲量が多い場合，または一方だけが風下の影響を受けている場合が挙げられる．本研究では，多数のセンサデータや地理情報を含む空間的補助データを用いて，各環境の観測データから多環境の類似性を推定しクラスタリングするマルチタスク学習を行うことにより，各環境の物理量の空間分布を精度高く推定する．

連絡先: 関西学院大学，兵庫県三田市学園 2 丁目 1 番地，masato190@kwansei.ac.jp

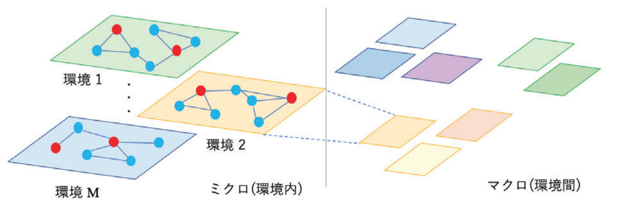

図 1: 本提案モデルの概念図．左：環境内の少数の高精度センサーと多数の低精度センサーの観測位置における空間分布の真値を重み付きエッジで結び，無向グラフを構成する．右：環境間は観測データと補助データの類似性からクラスタリングし，マルチタスク学習により，物理量の空間分布を推定する．

本研究は，多環境のクラスタリングに，[Iwata 13] らが提案した iWMM（infinite warped mixture model）を用いる．iWMM は，多次元の観測データに対して GPLVM（GP Latent Variable Model）による非線形な次元削減を行い，低次元空間内で iGMM（infinite Gaussian mixture model）による観測データの特徴をクラスタ化しつつ，クラス数を自動決定する手法である．iWMM は汎用性が高くクラスタリングや密度推定・多様体学習・視覚化に適応可能である．本提案手法における iWMM の位置付けを示すために，各環境内のモデルを以下で説明する．

本研究では観測データの信頼性を確率変数で表現し，教師データは高精度センサーの観測値であり，低精度センサーの観測データはドリフトを含み誤差が大きく擬似教師データとして考える．このような観測データに対して，一般的な回帰手法のニューラルネットワークやガウス過程（GP：Gaussian Process）では推定誤差が大きくなる．提案手法は，環境内において高精度センサー値を頼りに低精度センサー値を補正するため，画像分野の [Geman 84] の画素の近傍は相関しているという考えに基づく．したがって，観測地点の空間分布はマルコフ確率場（MRF：Markov Random Field）の構造を持つ多

次元ガウス分布に従う．また，各観測位置が等距離ではないため，ノード間のエッジに重みを与える．重みの与え方として，空間統計学の一般的なアプローチに，距離の逆数の重みづけや k 近傍法を用いた空間重み行列 [Stakhovych 09] がある．しかし提案手法では空間重み行列を用いず，補助データ間の類似度で与える．ただし，類似度は一般にわからず，類似度のパラメータを事後確率最大化からベイズ推定する．iWMM はその類似度を定めるパラメータを生成するために用いる．iWMM の性質上，同じクラスでは似たパラメータベクトルを持つように正則化が働くため，類似した環境と共同したマルチタスク学習になる．ここで提案モデルの概念図を図 1 に示す．最後に，本研究の貢献を以下に述べる．

- 環境に配置されたドリフトを含む多数の低精度センサー値と，環境内の少数の高精度センサー値に基づいて物理量の空間分布を高精度に推定.
- 多環境においてタスク間の関係性に空間近接性を用いず，観測データの類似性に基づき環境をクラスタリングし，マルチタスク学習を行うことにより，低精度センサーのみの環境においても空間分布の推定精度の向上.

2.　物理量の空間分布の推定手法

ある環境のセンサーの観測位置を含む補助データは，$\mathbf{X}^m = (\mathbf{x}_1, \ldots, \mathbf{x}_{S_m})^{\mathrm{T}}$ とする．$\mathbf{x}_i \in \mathbb{R}^{D_h}$ (D_h：補助データの種類) とし, $S_m = L_m + H_m$, H_m は高精度センサー, L_m は低精度センサーの総数である．m はある環境を表し，全環境数は M ある．低精度センサーの観測値は $y_i^{(n)} = z_i + \mu_i + \epsilon_n$, 高精度センサーのそれは $y_i^{(n)} = z_i + \eta_n$ で表される．ここで μ_i は i 個目の低精度センサーの系統誤差, ϵ_n は低精度センサーのノイズで期待値 0 で精度が α のガウス確率変数であり, η_n は高精度センサーのノイズで期待値 0 で精度が β のガウス確率変数である．ただし $\alpha < \beta$ とする．センサーを置いた位置での空間分布の真値を z_i とし，その集合を $\mathbf{z}^m = (z_1, \ldots, z_{S_m})^{\mathrm{T}}$ とする．環境 m の低精度と高精度センサーの観測変数集合を合わせて $\mathbf{Y}^m \in \mathbb{R}^{N_m \times S_m}$ とする．ここで，N_m は計測回数である．全環境での観測値集合は $\mathbf{Y} = \{\mathbf{Y}^1, ..., \mathbf{Y}^M\}$ である．

エネルギー関数を用いて環境内のミクロなモデルの外観を説明し，次にエネルギー関数を確率分布として表す．最後に，環境間を考慮するマクロな確率モデルに拡張していく．

2.1　エネルギー関数

まず環境内のエネルギー関数は次のように定義される．

$$\begin{aligned} E_m &= \sum_{j \in \mathcal{N}(i)} k(\mathbf{x}_i, \mathbf{x}_j)(z_i - z_j)^2 \\ &+ \frac{\alpha}{2} \sum_{i=1}^{L_m} (y_i - z_i - \mu_i)^2 + \frac{\beta}{2} \sum_{i=L_m+1}^{S_m} (y_i - z_i)^2, \quad (1) \end{aligned}$$

ただし,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{1}{2}\mathbf{w}_m^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)^2).$$

ここで，$\mathbf{w}_m \in \mathbb{R}^{Dh}$ は類似度を構成するパラメータである．エネルギー関数の 1 項目は平滑化項，2 項目は低精度センサー値に関する誤差を最小化する項，3 項目は高精度センサーのそれに対応する項である．このエネルギー関数の最小化における特徴は 2 つある．近傍内の補助データの類似性を用いた重み $k(\mathbf{x}_i, \mathbf{x}_j)$ により，近傍内の空間分布の真値同士を近づける．$\alpha < \beta$ により，1 項目の平滑化項から低精度センサーの観測位置での真値 z_i が高精度センサーの観測位置での z_j に影響され，ドリフト μ_i が推定される．

2.2　Spalogh（環境内確率モデル）

少数の高精度センサーと多数の低精度センサーを用い，多数の低精度センサーのドリフトと空間分布を推定する手法を Spalogh と呼ぶ．まず，式 (1) のエネルギー関数をギブス・ボルツマン分布で定義し，エネルギー関数の各項ごとに分布をガウス分布に変形させる．その後，パラメータに対して事前確率を置き，同時確率を求めていく．尤度関数は, 式 (1) の 2 項と 3 項目をまとめてガウス分布を用いて以下の式で定義される．

$$p(\mathbf{Y}^m | \mathbf{z}^m, \boldsymbol{\mu}^m, \alpha, \beta) = \prod_{n=1}^{N_m} \mathcal{N}(\mathbf{y}_n^m | \mathbf{z}^m + \boldsymbol{\mu}^m, \mathbf{I}_{S_m}), \qquad (2)$$

ただし，$\mathbf{I}_{S_m}$ は対角行列でその対角要素は最初の L_m 次元は α^{-1} であり，残りの H_m 次元が β^{-1} である．　系統誤差 $\boldsymbol{\mu}^m$ ベクトルは最後の高精度センサー数分の次元は 0 で与えられている．事前分布 $p(\boldsymbol{\mu}^m)$ はある品番の低精度センサーの系統誤差は母平均があると仮定し, 以下の式で与えらえる．

$$p(\boldsymbol{\mu}^m) = \prod_{i=1}^{L_m} \mathcal{N}(\mu_i^m | m_\mu, \sigma_\mu^2), \qquad (3)$$

m_μ には無情報事前分布を置く．σ_μ^2 は超パラメータになる．空間分布の真値 $\mathbf{z}^m$ の事前分布 $p(\mathbf{z}^m | \mathbf{w}_m)$ は式 (1) の 1 項目を変形すると MRF の構造を持つ多次元ガウス分布で以下のように書き下される．

$$p(\mathbf{z}^m | \mathbf{w}_m) = \mathcal{N}(\mathbf{z}^m | \mathbf{0}, \mathbf{A}^{-1}(\mathbf{w}_m)), \qquad (4)$$

ただし，

$$[\mathbf{A}(\mathbf{w}_m)]_{ij} = \begin{cases} \sum_{k \in \mathcal{N}(i)} k(\mathbf{x}_i, \mathbf{x}_k) & (\mathrm{i = j}) \\ -k(\mathbf{x}_i, \mathbf{x}_j) & (j \in \mathcal{N}(i)) \\ 0 & (\text{otherwise}), \end{cases}$$

ここで, i の近傍は半径 d とした．α, β は共役事前分布で $\alpha \sim \mathrm{Gam}(a_{L0}, b_{L0}), \beta \sim \mathrm{Gam}(a_{H0}, b_{H0})$ である．

2.3　Spalogh - iWMM（多環境マルチタスク学習）

環境にも類似性があると仮定し，複数のクラスが存在するとする．ただし，事前知識ではクラス数も環境間の類似度もわからない．そこで，環境内の $\mathbf{z}^m$ を近傍で平滑化させる重み $k(\mathbf{x}_i, \mathbf{x}_j)$ がもつパラメータ $\mathbf{w}_m$ に着目する．$\mathbf{w}_m, m = 1, ..., M$，をクラスタリングし，環境の類似性を表現することで，同じクラスでは似た $\mathbf{w}_m$ を持つように正則化するマルチタスク学習を行う．このクラスタリングは，環境内の近傍の補助データ $\mathbf{x}_i^m, \mathbf{x}_j^m$ と空間分布の真値 z_i^m, z_j^m の関係性が環境間で似ていると同じクラスに属するようになる．

$\mathbf{w}_m$ を全環境分まとめた行列 $\mathbf{W} \in \mathbb{R}^{M \times Dh}$ を GPLVM で非線形に低次元圧縮し，低次元空間内で iGMM により正則化させる（iWMM）．また，RBF カーネルを用いた GP の性質上，入力が似ると出力も似るため，環境間でパラメータ $\mathbf{w}_m$ も似るようになる．通常，$\mathbf{w}_m$ は，エネルギー関数最小化を行

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

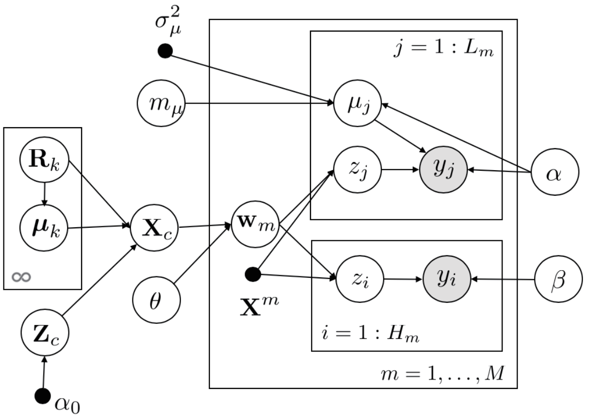

図 2: 提案モデルのベイジアンネットワーク．大きなプレートが各環境を表し，小さいプレートの上は低精度センサーを表し，下のプレートは高精度センサーを表す．

うと，大きな値をとり，近傍で平滑化させる効果がなくなる．したがって正則化が必要である．ここでは，GPLVM の共分散行列 $\mathbf{K}$ の各要素が小さくなるように RBF カーネルの尺度パラメータ θ_0 とスケールパラメータ σ_0 にガンマ分布を仮定し，超パラメータをうまく設計することで，$\mathbf{w}_m$ の正則化を実現する．
まず，$p(\mathbf{W})$ の分布として GPLVM を用いる．

$$p(\mathbf{W}|\mathbf{X}_c, \boldsymbol{\theta}) = \prod_{j=1}^{D_h} \mathcal{N}(\mathbf{w}_j|\mathbf{0}, \mathbf{K}). \tag{5}$$

本研究ではパラメータ $\{\theta_0, \sigma_0, \beta_0\}$ を持つ RBF カーネルを使用した．カーネルパラメータの分布として $\boldsymbol{\theta} = \{\theta_0, \sigma_0\}$ は $\theta_0, \sigma_0 \sim \mathrm{Gam}(a_{k0}, b_{k0})$ で与える．次に，GPLVM の入力になる確率変数 $\mathbf{X}_c$ の分布に iGMM を用いる．

$$p(\mathbf{x}_c|\{\lambda_k, \boldsymbol{\mu}_k, \mathbf{R}_k\}) = \sum_{k=1}^{\infty} \lambda_k \mathcal{N}(\mathbf{x}_c|\boldsymbol{\mu}_k, \mathbf{R}_k^{-1}), \tag{6}$$

この位置 $\mathbf{x}_c$ が潜在的なクラスを決める要因である．$\{\boldsymbol{\mu}_k, \mathbf{R}_k\}$ に事前分布として超パラメータ $\{\mathbf{u}, r, \mathbf{S}, \nu\}$ を持つガウス・ウィシャート分布を使用する．クラス分割数を決める $\mathbf{Z}_c$ に超パラメータ α_0 を持つ中華料理店過程を導入した．クラス数を決めるのに重要な α_0 は 0.1 に設定し，クラス数が少なくなるようにした．iWMM の詳細は [Iwata 13] を参照．提案モデルのベイジアンネットワークを図 2 に示す．
　式 (2) から (6) と超パラメータの分布の積により提案モデルの同時分布が以下の式で与えられる．

$$\begin{aligned} p(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, \mathbf{W}, \mathbf{Z}_c, \mathbf{X}_c, \boldsymbol{\theta}, \alpha, \beta, m_\mu) &= p(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\mu}, \alpha, \beta)p(\boldsymbol{\mu}|m_\mu) \\ &\times p(\mathbf{Z}|\mathbf{W})p(\mathbf{W}|\mathbf{X}_c, \boldsymbol{\theta}) \\ &\times p(\mathbf{X}_c|\mathbf{Z}_c)p(\mathbf{Z}_c) \\ &\times p(\alpha)p(\beta)p(\boldsymbol{\theta})p(m_\mu). \end{aligned}$$

事後確率 $p(\mathbf{Z}, \boldsymbol{\mu}, \mathbf{W}, \mathbf{Z}_c, \mathbf{X}_c, \boldsymbol{\theta}, \alpha, \beta, m_\mu|\,\mathbf{Y})$，をギブスサンプリングとハミルトンモンテカルロ・サンプリング（HMC）により推定する．

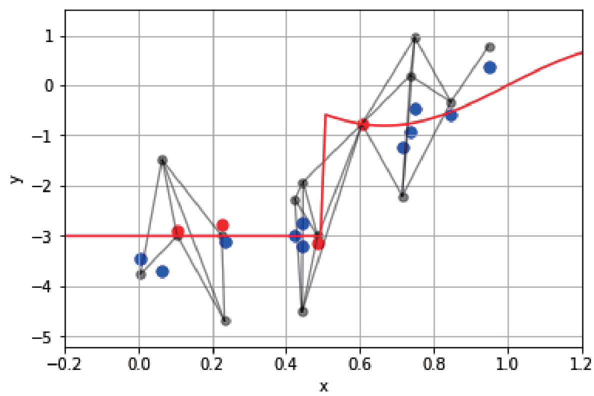

図 3: ドリフト推定による低精度センサー値の補正結果．青丸は低精度，赤丸は高精度センサーでの空間分布の推定値を表している．黒丸はドリフトを含んだ観測データである．赤線は真値，黒線は近傍のエッジを表す．元々のドリフトの平均絶対値誤差は 1.13, 補正後の誤差は 0.29.

3. 実験

　以下 2 つを検証する．
1）環境内で高精度センサーを頼りに低精度センサーが補正されるか．
2）マルチタスク学習により多環境ならびに低精度だけの環境も補正の推定精度が上がるのか．

3.1 データセット

　本実験では，1）人工データ 2）アメリカの点観測の気象データ *1*2 を用いた．1）1 次元の環境を仮定し，真値は x 軸が 0.5 までは $y = -3$，それ以上は $y = \cos(1.5\pi x)$ の曲線である．x の生成は $x \sim \mathrm{Uni}(0, 1)$ で 11 点は低精度としてドリフト $\mu_i \sim \mathrm{Uni}(-2, 2)$ とノイズ $\mathcal{N}(0, 0.5)$ を真値に加えた．高精度は 4 点でノイズ $\mathcal{N}(0, 0.1)$ を加えて用いた．近傍の距離を $d = 0.18$ とした．2）については環境は全 10 環境で Frolida，NewYork，Washington，California，Texas，Indiana，Wyoming 州で，後ろ 3 州だけ 2 環境ずつ作成した．環境内の観測点が 1 5 地点で全環境同数になるように用意した．補助データとして緯度，経度，標高，雨量，湿度など 12 種類を全環境において正規化し用いた．目的変数は日中平均気温とする．15 地点中 4 地点を高精度センサーとし，11 地点を低精度センサーとする．低精度センサーにはドリフトとして $\mathrm{Uni}(0, 2)$ を加えて実験を行なった．高精度センサーの位置は全体的に散らばるように選び，その観測値には人工データ同様ノイズを加えた．

3.2 比較手法

　単純な Lasso 回帰と [Obozinski 07] が提案したマルチタスク学習の Group Lasso 及びグラフラプラシアンによる正則化を用いた Lasso 回帰 [Xu 16]，ただし，空間的近接性に基づく距離の逆数のタスク関係行列を用いた．さらに，Spalogh-L^2 ノルムで行なった．Lasso の正則化パラメータは共通して $\lambda = 1.3$ を使用した．Spalogh-L^2 ノルムは，$\mathbf{W}$ の事前分布に平均 0 の等方ガウス分布を用いたモデルである．環境間で重み $\mathbf{w}_m$ を独立にした手法になる．

*1 https://weather.us/observations/california/total-cloud-coverage/20180201-1800z.html
*2 http://climod.unl.edu

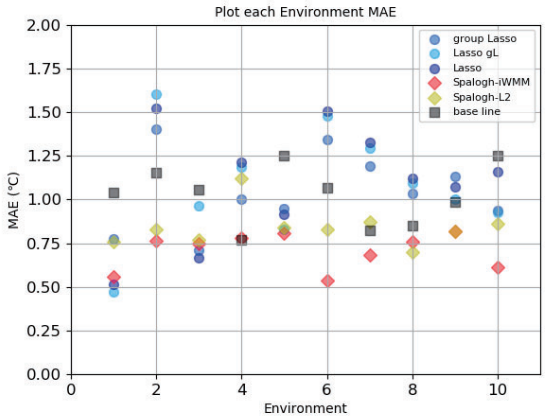

図 4: 各環境で生成したドリフト量と推定したドリフト量 $\boldsymbol{\mu}_{\mathrm{MAP}}$ の MAE を比較した図．横軸は環境の番号であり，縦軸は摂氏の MAE である．10 番が低精度センサーのみの環境になる．

3.3 結果・考察

1）について図 3 に，人工データでのドリフト推定による低精度センサー値の補正結果を示す．赤丸の高精度センサー値によって全体的に青丸の低精度センサー値が補正されていることが分かる．しかし，どの観測位置が高精度センサーになるかで推定結果が大きく異なることが分かる．一般に高精度センサーの位置は空間代表性を持っていることが望まれる．高精度センサーが空間分布の局地的な場所に配置されている場合，低精度センサーもその局地的な場所に強く影響される．

2）の実験の結果である，各環境におけるドリフトの推定誤差を図 4 に示した．Lasso はドリフトを表す確率変数がモデルにないため，λ の大きさに依存して，環境によって推定結果が良い場合もあれば悪い場合もある．環境の 10 番目は低精度だけの環境であり，Spalogh-iWMM は比較手法より高い精度を示している．本実験では，$\mathbf{w}_m$ は L^2 ノルムにあたる事前分布より多環境の影響を受けた方が精度が良いことが分かる．他の環境についても同様に推定結果が Spalogh-L^2 と同等かそれ以上の結果が示されている．各環境の潜在的なクラスと重みパラメータ $\mathbf{W}$ を図 5 に示す．クラスタリングにおいて，3 州だけ同じ環境が 2 つ ($\{4,6\},\{5,7\},\{9,10\}$) あったが，どの 2 つの環境も同じクラスに属することはなかった．これは，クラスが景観類似性や，空間的近接性によって決まるのではなく，各環境の観測地点近傍の補助データの類似性と観測データに依存して決まることがわかる．さらに，今回のデータセットは環境内で観測位置の間隔が広く，高精度センサーを頼りすぎ，誤差が逆に大きくなる場合があった．この問題の解決策に空間分布の真値間の差が大きい場合，不連続とみなす画像の分野で有名なラインプロセスを用いることが期待される．

4. おわりに

本稿では，観測データの信頼性をもとに，少数の高精度センサーと多数の低精度センサーを用いたセンサー値の補正と観測地点における空間分布を推定した．さらに環境をクラスタリングしつつ，マルチタスク学習を行い，比較手法より精度が高いことを示した．また，低精度だけの環境においても高い精度を示した．本手法の課題として，補助データに欠損値があってはならない．補助データに欠損値があっても対応できるようなテンソル分解などのモデルを検討する．さらに超パラメータの決定がヒューリスティックになってる部分をより正確に決定する方法が課題である．今後の発展として，環境内の不連続性に対応するラインプロセスの導入がある．また欠損値に強いモデルへの拡張を行う．環境モニタリングの応用を見据えた時空間データにおける長期計測に有効なモデルを構築することが挙げられる．

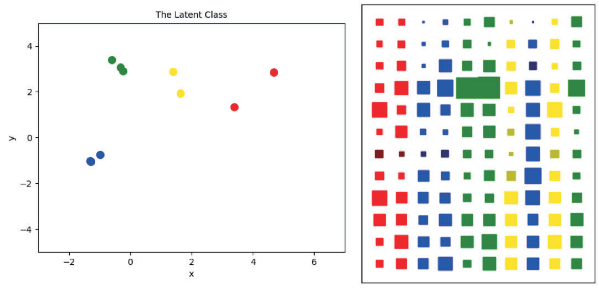

図 5: [左] iWMM による各環境の潜在的な二次元空間の位置．色はクラス．[右] 各環境内のエッジの重みを決めるパラメータのヒントン図．横に各環境が並び，縦が各環境内で補助データにかかるパラメータを表す．四角形の大きさがパラメータの大きさを示す．暗い色は負の値．特に緑色や黄色をみると，パラメータベクトルが似ていることが分かる．

参考文献

[Xu 17] Xu, J., "Multi-Task Learning and Its Application to Geospatio-Temporal Data," *Michigan State University*, 2017.

[Atluri 18] Atluri, G., A. Karpatne, and V. Kumar, "Spatio-temporal data mining: A survey of problems and methods," *ACM Computing Surveys* (*CSUR*), vol. 51, no. 4, p. 83, 2018.

[Iwata 13] Iwata, T., Duvenaud, D., and Ghahramani, Z., "Warped mixtures for nonparametric cluster shapes," *In Uncertainty in Artificial Intelligence*, 2013.

[Geman 84] Geman, S. and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on pattern analysis and machine intelligence*, 6, pp. 721-741, 1984.

[Stakhovych 09] Stakhovych, S. and B. H. Tammo, "Specification of spatial models: A simulation study on weights matrices." *Papers in Regional Science*, 88.2, pp. 389-408, 2009.

[Obozinski 07] Obozinski, G., B. Taskar, and M. I. Jordan, "Joint covariate selection for grouped classification," *Technical report, UC Berkeley*, 2007.

[Xu 16] Xu, J., et al., "Gspartan: a geospatio-temporal multi-task learning framework for multi-location prediction," *Proceedings of the 2016 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics*, 2016.