

# 人物動態のモニタリングに向けた 統計的異常検知に関する研究

神谷啓太<sup>1</sup>・布施孝志<sup>2</sup>

<sup>1</sup>学生会員 工修 東京大学大学院 工学系研究科社会基盤学専攻 (〒113-8656 東京都文京区本郷七丁目 3-1)  
E-mail: kamiya@trip.t.u-tokyo.ac.jp

<sup>2</sup>正会員 工博 東京大学准教授 工学系研究科社会基盤学専攻 (〒113-8656 東京都文京区本郷七丁目 3-1)  
E-mail: fuse@civil.t.u-tokyo.ac.jp

人物の位置情報を高分解能で収集し、時々刻々と変化する人物動態のモニタリングを行うことで、動的な交通管制の実現が期待されている。モニタリングでは異常状態の把握が重要となるが、人物動態に関する統計的異常検知手法の開発はアドホックに行われており、整理も十分でない。それに対し本研究では、多様な分野におけるレビューに基づき、手法の要件を整理し、時系列データ中の文脈型異常を検知する問題として設定した。そして、データに応じて状態及び状態数を推定可能な階層ディリクレ過程に基づく隠れマルコフモデルを利用した異常検知手法を構築した。シミュレーション実験では優れた検出率を、実データへの適用では列車運転見合わせが発生した地点・時刻での異常検知を確認した。

**Key Words :** *Anomaly detection, Human dynamics monitoring, Dirichlet process, Hidden markov model, time-series data*

## 1. はじめに

都市における人々の移動や分布等の大局的な人物動態の把握は、都市計画やマーケティング等多岐にわたる分野から要請されている。また、近年のGPSや公共無線LAN等を利用した測位技術の発達により、時間的・空間的に高分解能かつ低コストで位置情報の入手が容易となっている。そのため、個々人の位置情報をリアルタイムに収集することによって、時々刻々と変化する人物動態をモニタリングすることが可能である。将来的には交通渋滞や交通パターンの変化等を検知することで、動的な交通制御の実現が期待されている。

モニタリングでは人物動態の異常状態の検知が重要となる。異常状態を引き起こす要因は交通事故、自然災害・天候、イベントの有無等、無数に考えられ、それら全てをモニタリング対象に置くことは不可能である。そのためメッシュ人口データやGPS軌跡データ等の人物動態データを直接分析することによって異常状態が検出できれば、それは有用である。ただし、モニタリングによって得られる人物動態データは、地域・時間によって大きく異なる挙動を示し、かつ膨大なデータ量であるため、人為による異常検知は困難かつ高コストとなる。そこで、日常的に得られるデータから正常な状態を学習し、そこからデータに潜む異常を自動的に検知できる統計的異常検知の枠組みに着目する。

大局的な人物動態を対象にした統計的異常検知に関し

て様々な研究が行われている。Candiaら<sup>1)</sup>は、メッシュ毎に集計したCDRデータに対し浸透理論を用い、異常イベントの発生場所・日時の分析を行った。Horanont<sup>2)</sup>は同様のデータに対しメッシュに内挿処理を施し、ホットスポットの検出・可視化を容易にした。車両プローブデータを利用した例として、Panら<sup>3)</sup>が交通ネットワーク上の運転航路に着目して交通事故等の異常検知を行った。Horiguchiら<sup>4)</sup>は、エリア集計QK曲線と実測値との解離度を測ることで非日常的な混雑状況を持つメッシュを検出、可視化した。

このように人物動態の統計的異常検知に関する研究はこれまでに複数行われているが、手法の開発は現状アドホックに行われており、その整理も十分でない。さらには、ある適用領域で精度良い手法であっても、別の適用領域で同様の精度を保つことは難しいとされている。そのため、人物動態モニタリングにおける統計的異常検知手法の開発に際しては、使用するデータの特性やモニタリングで着目する異常の特徴を考慮した上で、汎用的な異常検知手法の開発が望まれる。

以上の背景に基づき、本研究の目的は、人物動態モニタリングにおける統計的異常検知問題の整理、及び、その整理結果に基づく異常検知手法の構築と、基本性能の検証である。なお、本研究ではデータの匿名性、および入手可能性の高さから、携帯電話のGPSログデータを集計しメッシュ毎・時間帯別の人口を推定したメッシュ人口データを用いる。本研究ではまず、統計的異

常検知に関する論文をレビューし、メッシュ人口データを用いた人物動態モニタリングにおける統計的異常検知手法の要件の整理を行う。その整理結果に基づいた問題設定の下で、異常検知手法の開発を行う。詳しくは後述するが、データに応じて潜在状態数を自動的に決定可能な階層ディリクレ過程に基づく隠れマルコフモデルを利用する。そして、提案手法の基本性能をシミュレーション及び実データを通じて確認する。

本稿の構成は以下の通りである。まず第2章で、異常検知手法を特徴付ける要因に着目し、人物動態モニタリングにおける異常検知問題の設定を行う。同時に、空間状態モデルが適用可能であることを説明する。第3章ではディリクレ過程他、ノンパラメトリックベイジアンモデルを概説し、その後提案手法の詳細を述べる。第4章ではシミュレーション及び実データを用いた実験結果と考察を行い、本研究の成果と今後の課題を第5章で述べる。

## 2. 人物動態モニタリングに向けた異常検知問題の設定

### (1) 異常検知手法の要件の整理

統計的異常検知問題とは、大概のデータが従うパターンからある程度の尺度を以って外れたデータを検知する問題である<sup>5)6)7)</sup>。異常検知手法によって検知されるデータは、大量のデータの中に潜む傾向であったり、大部分のデータとは異なる珍しいデータであったりする<sup>8)</sup>。近年では、データ計測手法や計算機処理能力の発展により、異常検知はネットワーク侵入検知や画像処理、構造物ヘルスマニタリング等の多彩な分野で盛んに研究が行われている。

本研究では多様な分野における統計的異常検知のレビューを行った。その結果、異常検知問題の特徴付ける要因は、対象とするデータの性質、異常検知の出力方法、学習時のラベル、そして異常の種類<sup>4)</sup>の4要素であることが明らかとなった<sup>5)</sup>。以下では、メッシュ人口データを対象に、上述の4要素の視点から、人物動態モニタリングにおける異常検知手法の要件を整理する。

#### a) データの性質

異常検知問題の特徴付ける第一の要因は入力データの性質である。データの性質は二値、離散値、連続値に大別される。同時に、時系列やグラフデータ等、入力データ間の関係性も重要な要因となる。

本研究では、メッシュ人口データ中のある一つのメッシュに着目し、入力データはメッシュ人口値がなす時系列データであるとする。さらに、都市における人々の移動や分布は空間的な相関や相互影響を持っていると考えるのが一般的であるため、この時系列データが

空間的な相関、依存関係も有すると考える。本研究では時系列のみに着目することとし、時系列データに対応可能で、空間方向への拡張が可能な手法が望まれる。

#### b) 異常検知の出力方法

異常検知の出力方法には、各データに異常スコアを算出する方法と、「異常」か「正常」かを表すラベルを付加する方法がある。前者の出力方法によりメッシュ毎に異常スコアが算出されれば、交通管制対象の優先度の指標になり得る点で有用である。後者の場合、どのような正常・異常状態かを表すラベル作成ができれば、有用性は同様に高いと考えられる。従って、異常スコア及びラベルの両方が適用可能であると考えられる。

#### c) 学習時のラベル

第三の要因は、異常検知モデルを学習する際の学習データに含まれる「正常」「異常」ラベルの事前情報の有無である。学習データにどこまで教師ラベルが含まれているかによって、教師付き異常検知、半教師付き異常検知、教師無し異常検知の3手法に分類される。

正常な教師データの取得は、大規模な事故や自然災害が発生していない日時・地点の観測データを用いることである程度可能である。一方異常な教師データの十分数の確保は、異常状態とみなされる状態には無数のパターンが存在することより、現実的に不可能である。そのため、本研究では「正常」ラベルを用いた半教師付き異常検知、もしくは教師無し異常検知が望まれる。

#### d) 異常の種類

Chandolaら<sup>5)</sup>に依ると、統計的異常検知問題における異常は、異常点(Point Anomaly)、文脈型異常(Contextual Anomaly)、集団型異常(Collective Anomaly)の3種類に分類される。それぞれ、データ単独で異常なデータ、ある文脈の上で異常なデータ、異常な振る舞いをするデータ集合と解釈される。

本研究で対象とするメッシュ人口データ中から検出される異常の例として、事故・災害による交通渋滞の影響で生じた人口増加・減少や交通需要の変化等から生じるメッシュ人口値の推移パターンの変動等が挙げられる。図-1はそれらを模式的に表したものである。従って、時系列中の異常、すなわち、文脈型異常の検知が可能な手法が望まれる。

以上、4つの要因について議論を行った。表-1はそれらを簡潔にまとめたものである。この整理結果に基づき、本研究では、人物動態モニタリングにおける異常検知問題を、教師データにより異常検知モデルを学習し、時系列データ中の文脈型異常を検知する問題として設定する。

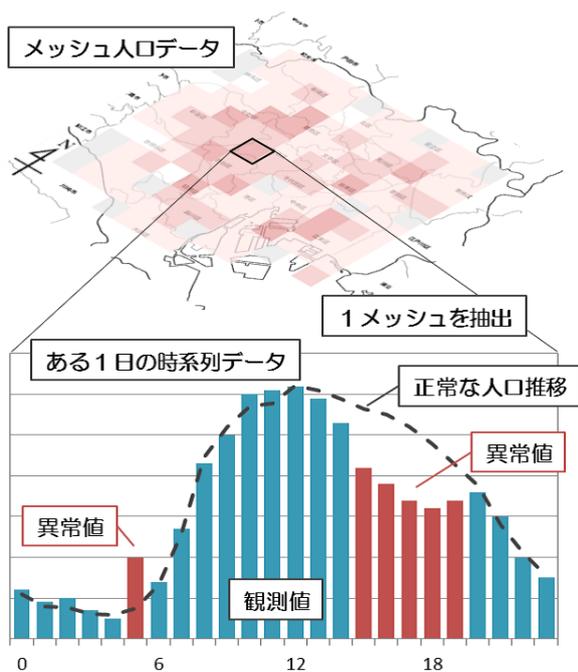


図-1 メッシュ人口データに含まれる異常の例

表-1 人物動態モニタリングにおける統計的異常検知手法の要件の整理結果

要因	人物動態モニタリングでの特徴
データの性質	各メッシュの時系列データ データ間の空間的な相関, 影響
出力方法	異常スコアまたは異常ラベル
学習ラベル	半教師または教師無し異常検知 日常データより正常ラベル取得可
異常の種類	文脈型異常 渋滞による人口変化 需要変化による推移パターン変化

## (2) 状態空間モデルによる統計的異常検知

上述の問題設定に基づく、状態空間モデルを使用した異常検知手法が適用可能と考えられる。状態空間モデルでは時間変化する隠れ状態から観測値が出力される。この表現により複雑に時系列変化するデータのモデリングが可能となるため、時系列データの異常検知に状態空間モデルが用いられることが多い<sup>6)</sup>。

状態空間モデルを利用した異常検知手法は主に2種類存在する。一つは各観測値に対して尤度や出力確率を計算し、その値を異常スコアとして算出する方法であり、もう一つは隠れマルコフモデルにおいて異常ラベルを持つ潜在状態を定義しておき、その状態に推定された時の値を異常とみなす方法である。前者の手法を本研究へ適用する際、異常スコアの信頼性を保つ

に十分な精度を持つシステムモデルや観測モデルの構築が困難であると予測される。一方で後者の手法を用いる場合、状態空間モデルの潜在状態数の設定が大きな問題となる。人物動態データが取りうる適切な状態数は未知であり、また、異常状態数は無数に存在するため、事前の設定は実質上不可能になってしまう。しかしながら、データから適切な状態数を推定することができれば、その問題が克服できると考えられる。

そこで本研究では、ノンパラメトリックベイジアンモデルの一つであり、加算無限個の状態数を事前に設定することで入力データに応じて適切な状態数を推定することができる階層ディリクレ過程に基づく隠れマルコフモデルに着目し、これを異常検知手法に用いることとする。特に、連続的に状態が遷移するようなデータの効率的なモデリングが可能である sticky Hierarchical Dirichlet Process - Hidden Markov Model (sHDP-HMM) を利用する。次章にて、その詳細を述べる。

## 3. 階層ディリクレ過程隠れマルコフモデルを利用した異常検知手法

### (1) ノンパラメトリックベイジアンモデル

本節では、ディリクレ過程をはじめとするノンパラメトリックベイズ法について概説した後、本研究で用いる sHDP-HMM について説明する。

#### a) ディリクレ過程

ディリクレ過程 (Dirichlet Process; DP) は「分布に対する分布」と呼ばれており、パラメータ空間  $\Theta$  上の基底測度  $H$  とパラメータ  $\alpha > 0$  によって一意に定義される。ここでは、 $DP(\alpha, H)$  と記す。端的には、基底測度  $H$  を、それに似た無限次元の離散分布によって近似した分布  $G_0 \sim DP(H, \alpha)$  を生成する確率過程であると換言でき<sup>9)</sup>、具体的には以下のように記述される。

$$G_0 = \sum_{k=1}^{\infty} \pi_k \delta(\theta_k), \quad \theta_k \sim H \quad (1)$$

ここで、 $\delta(\cdot)$  はディラックデルタである。また、重み  $\pi_k$  は以下の棒折り過程 (stick-breaking process; SBP) と呼ばれる確率過程から生成される。

$$\begin{aligned} \nu_k &= \text{Beta}(1, \alpha) \\ \pi_k &= \nu_k \prod_{l=1}^{k-1} (1 - \nu_l) \end{aligned} \quad (2)$$

ここで、これらの重みをまとめて  $\pi = (\pi_1, \pi_2, \dots, \pi_{\infty}) \sim \text{GEM}(\alpha)$  と記す。SBP では確率の総和である長さ 1 の棒を左から切っていくことで  $G_0$  を生成している。また、 $\delta(\theta_k)$  は  $\theta = \theta_k$  に位置するアトムと呼ばれる。

DP は混合数が未知もしくは加算無限個の混合分布モデルの事前分布として用いられる。このようなモデル

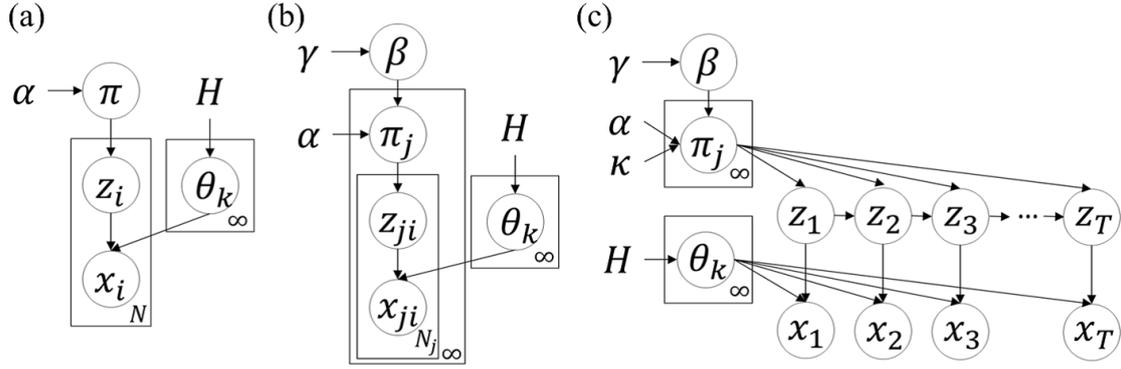


図-2 (a) ディリクレ過程, (b) 階層ディリクレ過程, (c) sticky HDP-HMM のグラフィカルモデル

はディリクレ過程混合モデルと呼ばれ, そのグラフィカルモデルは図-2 (a) に示す通りである. このモデルでは,  $GEM(\alpha)$  より生成された無限次元の多項分布  $\pi$  から状態  $z_i$  が生成される. そして, その状態  $k = z_i$  に対応するパラメータ  $\theta_k$  によって一意に定義された出力分布  $F$  によって観測値  $x_i$  が出力される.

#### b) 階層ディリクレ過程

階層ディリクレ過程 (Hierarchical Dirichlet Process; HDP) では, ディリクレ過程  $DP(\gamma, H)$  により生成された確率分布  $G_0$  が, さらにディリクレ過程  $DP(\alpha, G_0)$  の基底測度となり, 確率測度  $G_j$  が生成される. 上述の通り  $G_0$  は離散分布になるため,  $G_j$  はインデックス  $j$  が異なっても同様のアトム  $\delta(\theta_k)$  を共有する.

図-2 (b) には HDP の SBP 表現によるグラフィカルモデルを示す. ここでは,  $GEM(\gamma)$  よりパラメータ  $\beta$  が生成され,  $\pi_j$  が  $DP(\alpha, \beta)$  から生成される. そして, 状態  $z_{ji}$  と観測値  $x_{ji}$  が DP と同様の機構で生成される.

#### c) 階層ディリクレ過程隠れマルコフモデル

HDP を隠れマルコフモデルに拡張したものが, 階層ディリクレ過程隠れマルコフモデル (Hierarchical Dirichlet Process - Hidden Markov Model; HDP-HMM) <sup>10)</sup> である. HDP がアトムを共有する性質により, HDP-HMM における  $\pi$  は加算無限個の行, 列を持つ遷移行列として働く. 全ての加算無限個の状態が用いられるわけではないため, データに応じた潜在状態数を決定する柔軟なモデルであると解釈可能である. この特性により, 状態数が未知である場合の異常検知問題に適用されている <sup>11)</sup>.

HDP-HMM の課題として潜在状態が頻繁に切り替わる傾向にある点が挙げられる. この問題に対し Fox ら <sup>12)</sup> は自己遷移確率を制御するハイパーパラメータ  $\kappa$  を新たに導入した sticky Hierarchical Dirichlet Process - Hidden Markov Model (sHDP-HMM) を開発した. これにより滑らかな状態遷移を可能とするモデルの構築

が可能となる. sHDP-HMM によるデータ生成機構は次式で与えられ, グラフィカルモデルを図-2 (c) に示す.

$$\begin{aligned}
 \beta | \gamma &\sim GEM(\gamma) \\
 \theta_k | H &\sim H \\
 \pi_j | \alpha, \beta, \kappa &\sim DP(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_j}{\alpha + \kappa}) \\
 z_t | z_{t-1}, \{\pi_j\}_{j=1}^{\infty} &\sim \pi_{z_{t-1}} \\
 x_t | z_t, \{\theta_j\}_{j=1}^{\infty} &\sim F(x_t | \theta_{z_t})
 \end{aligned} \quad (3)$$

#### (2) sHDP-HMM を利用した異常検知手法の開発

本研究では, sHDP-HMM の枠組みを利用することで学習用の人物動態データの正常状態とその数を推定できると考える. そして, 学習済みモデルに検証データを適用し, 先ほど推定した正常状態と異なる状態が推定されていれば, 異常と見なす.

提案手法は, 主に以下の3段階で構成される. ①学習データの潜在状態列及び sHDP-HMM のハイパーパラメータを学習する. ②学習済みの sHDP-HMM によって検証データの潜在状態を推定する. ③推定された状態列を時刻毎に比較し, 異なる状態が推定されている場合に異常を検知する.

まず手順①において, 正常な学習データとして,  $d$  次元の時系列データ  $\{x_t\}_{t=1}^T$  を sHDP-HMM の観測値系列として設定する. ただし,  $T$  は時系列長さである. また, 潜在状態  $k$  に対応した出力分布にはパラメータ  $\theta_k = \{\mu_k, \Sigma_k\}$  で定義される  $d$  次元ガウス分布  $N(\mu_k, \Sigma_k)$  を設定する. この時, パラメータ  $\theta_k$  の事前分布  $H$  として, 平均  $\mu_k$  の事前分布にガウス分布  $N(\mu_0, \Sigma_0)$  を, 分散共分散行列  $\Sigma_k$  の事前分布に逆ウィンシャート分布  $IW(\nu, \Delta)$  をそれぞれ設定する. また, ハイパーパラメータは未知であるため, 無情報事前分布を与え, ハイパーパラメータもサンプリングによって推定する. 以上のように設定した sHDP-HMM に対し, Blocked Gibbs Sampler によって潜在状態  $\{z_t\}_{t=1}^T$  を含む全変数を推定する.

手順②では, 検証データ時系列  $\{x'_t\}_{t=1}^T$  の潜在状態

列  $\{z_t'\}_{t=1}^T$  を推定する。ハイパーパラメータには、手順①でサンプリングしたハイパーパラメータの事後分布の期待値を利用する。また、 $\beta, \pi, \theta$  のサンプリング初期値には手順①の最終サンプリング結果を用いる。

最後に手順③にて正常状態列  $\{z_t\}_{t=1}^T$  と検証用状態列  $\{z_t'\}_{t=1}^T$  を比較する。ただし、 $\beta, \pi, \theta$  の再推定によって各状態に割り当てられた番号が最終的に変わり得るため、ハミング距離をコストとする割当問題をハンガリー法を用いて、解く。その後、割り当てられた状態を時刻  $t$  毎に比較し、 $z_t \neq z_t'$  ならば異常と判断する。

2. で整理した結果に基づくと、提案手法は正常なデータを利用した半教師付き学習を行い、時系列データ中の文脈型異常を異常ラベルによって出力するものである。ある時刻では正常でも他の時刻では異常である、という文脈型異常の検知を試みる。

#### 4. 提案手法の適用と精度検証

この章ではシミュレーションデータ及び実データに提案手法を適用し、その基本性能を検証する。なお本研究では sHDP-HMM の推定プログラムの実装に、Foxら<sup>12)</sup> が公開している MATLAB HDP-HMM Toolbox を使用している。

##### (1) シミュレーションデータによる実験

本実験にあたり、図-3(a) に示す時系列長  $T = 400$  のデータを作成した。図中の番号で示されている計 12 個の状態毎に手で設定した正規分布から i.i.d サンプルを得ることで時系列を生成した。メッシュ人口データの 1 日の挙動に倣うように、正規分布の平均と分散を調整している。また、検証データとして同じ状態から別サンプルを並べた時系列データを作成し、計 5 箇所に異常値を手動で加えた。加えた異常値の大きさと時刻はそれぞれ、(i)  $5\sigma(t = 46)$ , (ii)  $3\sigma(t = 98 \sim 108)$ , (iii)  $3\sigma(t = 196 \sim 210)$ , (iv)  $3\sigma(t = 311 \sim 328)$ , (v)  $10\sigma(t = 381)$  である。ただし  $\sigma$  は正規分布の標準偏差を示す。

また、観測データの平均値と分散をそれぞれ  $\mu_0, \Sigma_0$  とし、自由度  $\nu = 4$  及び  $\Delta$  は単位行列  $I$  とした。なお、推定の安定化のため、入力データ時系列の平均が 0、分散が 1 となるように正規化したものを用いた。

図-3(b) に学習データの潜在状態の推定結果を示す。図中の黒色の実線が入力データ、緑色の実線及び破線が推定された各状態の平均と標準偏差を表しており、その下に推定された各状態を色毎に示す。データ作成時には計 12 個の状態を設定したが、10000 回のサンプリング後には状態数は 6 と推定された。詳細は省くが、出力分布パラメータの推定結果が、事前に設定したパラメータと大きく相違することはなかった。手順②にお

表-2 シミュレーション実験結果の混合行列

		異常検知結果	
		Anomaly	Normal
真値	Anomaly	64	336
	True Positive	32	8
	False Negative	32	328
Normal	False Positive	32	328
True Negative	360	328	



図-4 対象メッシュの位置

ける検証データの推定結果を図-3(c) に示す。検証データの推定によって状態数は 7 と推定された。また、異常検知の結果、図中に赤い円で示した箇所に計 64 個の異常が検知された。表-2 に、本実験における異常検知結果をまとめた混合行列を示す。異常検出率が 80.0% であり、提案手法が良好な異常検出率の下で適用可能であると考えられる。一方で、状態の境界付近で検知された誤検出が主な原因となり、精度は 50.0% であった。

##### (2) 実データへの適用

本実験ではメッシュ人口データとしてゼンリンデータコムが販売している、500m メッシュ・1 時間間隔の「流動人口データ」を用いた。対象期間として、2012 年 10 月 30 日(火) 及び 2012 年 11 月 6 日(火) のデータを使用する。10 月 30 日には大きな災害やイベント等が確認されず、一方で 11 月 6 日には午前 6 時頃東武伊勢崎線内で発生した人身事故により北千住-浅草間で約 1 時間半の運転見合わせが確認されているため、前者を学習データ、後者を検証データとして使用する。対象メッシュとして、運転見合わせの影響を受けていると考えられる東武浅草駅を含む図-4 に示す 500m メッシュに着目する。なお、先程と同様に、入力データには観測値を正規化したデータを用いる。

図-5 上部に手順①による学習結果を示す。朝ラッシュ

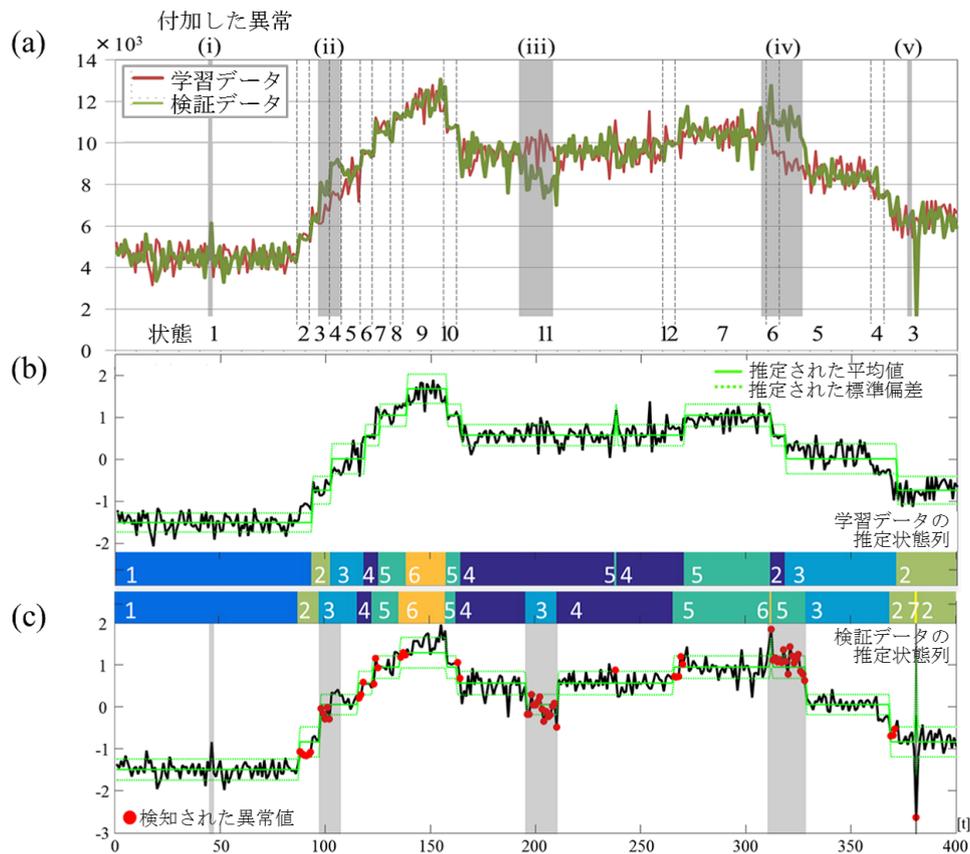


図-3 (a) 入力データ, (b) 学習データの推定結果, (c) 検証データの推定結果と異常検知結果

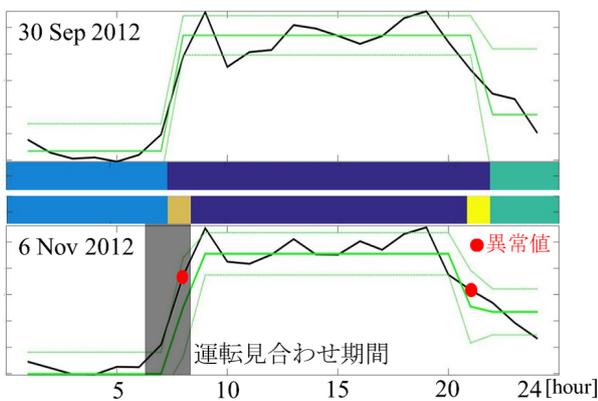


図-5 (上) 学習データの推定結果, (下) 検証データの推定結果と異常検知結果

が始まった7時と夜の人口減少が発生した21時の2箇所を境に、計3つの状態が推定された。異常検知の結果、図-5下部に赤い円で示した通り、8時および21時の2時点で正常状態とは異なる新たな状態が割り当てられ、異常が検知された。この内、8時に検知された異常については運転見合わせ時間である6-8時帯と一致した。一方で21時の異常について、同時間帯における

事故やイベント等は確認されなかった。この異常が潜在的異常の検知結果なのか、提案手法の誤検知なのか、今後詳しい分析を進める必要がある。

## 5. おわりに

本研究ではメッシュ人口データを対象に、人物動態モニタリングにおける統計的異常検知問題の整理を行い、その整理結果に基づき異常検知手法の構築を行った。具体的には、異常検知に関する論文のレビュー結果に基づき、入力データの性質、異常検知の出力方法、学習時のラベル、異常の種類、4要素の視点から、人物動態モニタリングにおける統計的異常検知手法の要件を整理し、時系列データ中の文脈型異常を検知する問題として設定した。さらに、状態空間モデルが適用可能であることを示し、状態数を同時に推定可能なsHDP-HMMに基づく異常検知手法を構築した。

また、提案手法をシミュレーションデータ及び実データに適用し、その基本性能の検証を行った。その結果、提案手法が良好な異常検出率の下で適用可能であることが確認され、実際に列車の運転見合わせが発生した時刻・メッシュにおける異常の検出を確認した。

今後、提案手法の精度改良および多様なデータへの適用を行い、推定された状態・検知された状態や異常がどのように解釈可能か分析を進める必要がある。さらに、モデルの空間方向への拡張、天候やイベント等の多様なデータとの統合、また制御手法との統合が今後の課題として挙げられる。

#### 参考文献

- 1) Candia, J. and González, M.C., Wang, P., Schoenharl, T., Madey, G. and Barabási, a.L.: Uncovering individual and collective human dynamics from mobile phone records, *Journal of Physics A: Mathematical and Theoretical*, Vol.41, No.22, pp.1-11, 2008.
- 2) Horanont, T.: A study on urban mobility and dynamic population estimation by using aggregate mobile phone sources, *CSIS Discussion Paper*, 2010.
- 3) Pan, B., Zheng, Y., Wilkie, D. and Shahabi, C.: Crowd sensing of traffic anomalies based on human mobility and social media, *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp.344-353, 2013.
- 4) Horiguchi, R., Iijima, M., Kobayashi, M. and Hanabusa, H.: Traffic anomaly detection for surface street networks with the mesh-wised traffic indices on macroscopic fundamental diagram, *OPTIMUM 2013 International Symposium on Recent Advances in Transport Modelling*, pp.1-7, 2013.
- 5) Chandola, V., Banerjee, A. and Kumar, V.: Anomaly detection: a survey, *ACM Computing Surveys*, Vol.41, No.3, pp.15:1-15:58, 2009.
- 6) Pimentel, M.A.F., Clifton, D.A., Clifton, L. and Tarassenko, L.: A review of novelty detection, *Signal Processing*, Vol.99, pp.215-249, 2014.
- 7) Hodge, V.J. and Austin, J.: A survey of outlier detection methodologies, *Artificial Intelligence Review*, Vol.22, No.2, pp.85-126, 2004.
- 8) 山西健司: データマイニングによる異常検知, 共立出版, 2009.
- 9) 持橋大地: 最近のベイズ理論の進展と応用 [III] ノンパラメトリックベイズ, 電子情報通信学会誌, Vol.93, No.1, pp.73-79, 2010.
- 10) Teh, Y.W., Jordan, M.I., Beal, M.J. and Blei, D.M.: Hierarchical dirichlet processes, *Journal of the american statistical association*, Vol.101, No.476, pp.1-30, 2006.
- 11) Lello, E.D., Laet, T.D. and Bruyninckx, H.: Hierarchical dirichlet process hidden markov models for abnormality detection in robotic assembly, *Workshop on Bayesian Nonparametric Models For Reliable Planning And Decision-Making Under Uncertainty*, pp.1-8, 2012.
- 12) Fox, E.B., Sudderth, E.B., Jordan, M.I. and Willsky, A.S.: Nonparametric bayesian learning of switching dynamical systems, *Advances in Neural Information Processing Systems*, pp.457-464, 2009.

## STATISTICAL ANOMALY DETECTION FOR MONITORING OF HUMAN DYNAMICS

Keita KAMIYA and Takashi FUSE

Human dynamics monitoring has recently drawn attention. Although this monitoring focuses on anomalous states, anomaly detection methods are developed ad hoc and not fully systematized. We discussed the characteristics of the anomaly detection of human dynamics monitoring with gridded population data and assumed a semi-supervised anomaly detection for contextual anomalies behind time-series data. We developed a technique based on a sticky HDP-HMM, which estimated the number of hidden states according to input data. Through the experiments with synthetic and real data, we found our method with high detection rate and detected an anomaly when and where an actual social event had occurred.