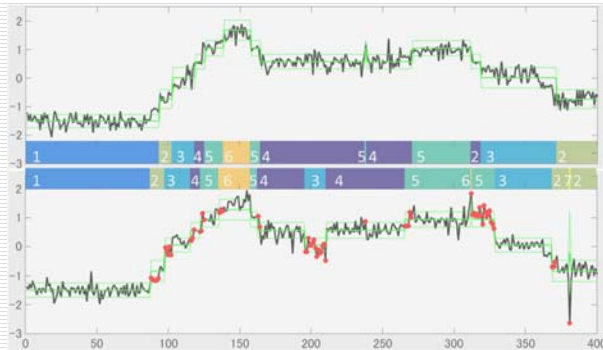


人物動態のモニタリングに向けた統計的異常検知



東京大学大学院
工学系研究科
社会基盤学専攻
布施孝志

背景

人物動態の把握の要請

多岐にわたる分野において重要
・交通モデリング ・マーケティング 等

位置情報取得の容易化

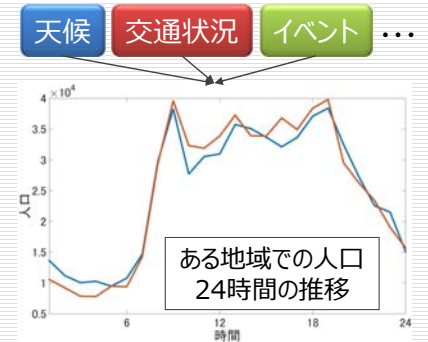
・GPSやWiFiを利用した測位技術の発達
・高分解能かつ低コスト

人物の位置情報をリアルタイムに集計し、時々刻々と変化する

人物動態のモニタリングへの期待

モニタリングでは異常状態の検知が重要

- 異常をもたらす要因全てを監視できない
- 天候, 交通状況, イベント...
- 観測値としての人物動態から異常を検知できれば大変有用である
- 人為による正常/異常の判定には限界



統計的異常検知手法の枠組みに着目

背景

3

人物動態の把握の要請

多岐にわたる分野において重要
・交通モデリング ・マーケティング 等

位置情報取得の容易化

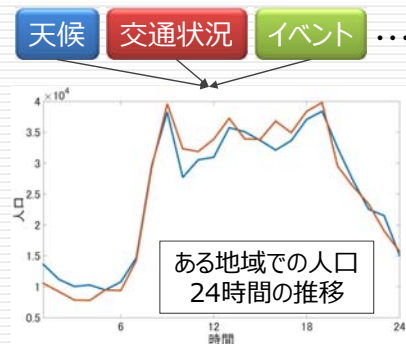
・GPSやWiFiを利用した測位技術の発達
・高分解能かつ低コスト

人物の位置情報をリアルタイムに集計し、時々刻々と変化する

人物動態のモニタリングへの期待

統計的異常検知手法の枠組みに着目

- 日常的に得られるデータから正常な状態を学習, 異常や変化を検知
- その上で, 最終的に対策が必要か人為による確認も可能となる
- 本研究における異常
= 学習した正常な状態と異なる状態

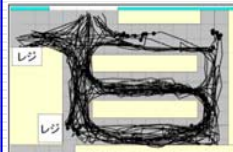


目的

4

統計的異常検知手法の関連研究

逸脱行動人物検出



交通事故等の検知



車両の混雑検知



ホットスポットの検出



動画中の動線を分析 (鈴木ら, 2007) NW上の航路を分析 (Pan, 2013) 集計QK曲線との乖離 (トラフィックスコープ) メッシュに内挿処理 (Horanont, 2010)

- 様々な手法の開発はアドホックに行われており, その整理も十分ではない
- 様々な対象への適用は限定的である

人物動態モニタリングの特徴を整理し, その上で適用可能な手法を探る

本研究の目的

- ✓ 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
- ✓ 整理結果に基づいた異常検知手法の構築および基本性能の検証

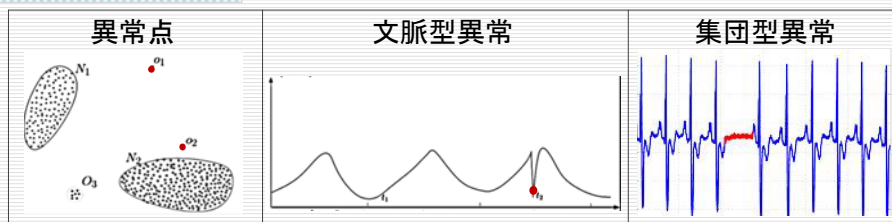
統計的異常検知問題の整理

5

- 多様な分野における統計的異常検知問題のレビューを行った
- 統計的異常検知問題は以下の4要素によって特徴付けられる

(Chandola et al, 2009)

データの性質	入力するデータの型やデータ間の関係性
出力方法	異常検知の出力方法 {異常スコア, 異常ラベル}
ラベル	学習用データの「正常」「異常」の事前情報の有無
異常の種類	検出する異常のタイプ



メッシュ人口データ

6

GPSログデータ

- 大規模かつ高時間分解能での人物移動データの取得が可能

ポイントデータ
個人の特定制が可能

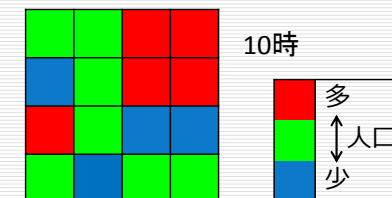
個人情報保護への意識の高まり

メッシュデータ
個人の特定制が不可能

活用が期待される

メッシュ人口データ

GPSログデータを基にメッシュの人口を推計したデータ



人口の多寡のみ表現

メッシュ人口データ

7

GPSログデータ

- 大規模かつ高時間分解能での人物移動データの取得が可能

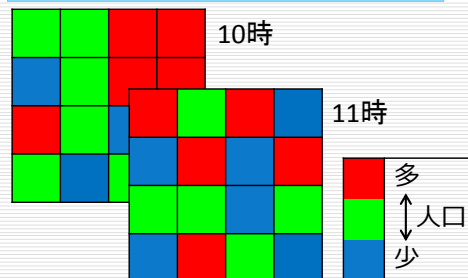
ポイントデータ
個人の特定制が可能

個人情報保護への意識の高まり

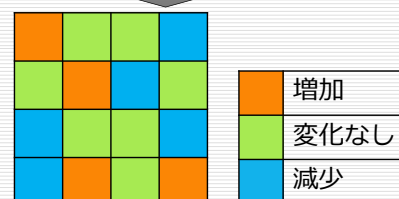
メッシュデータ
個人の特定制が不可能

活用が期待される

時系列メッシュ人口データ



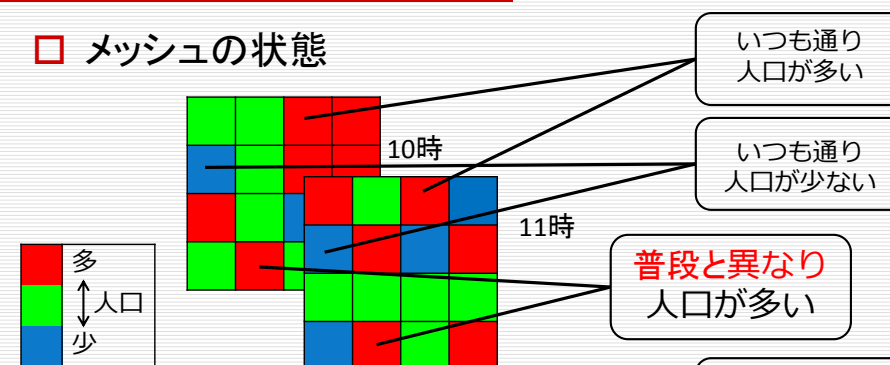
メッシュ人口の増減の視覚化



メッシュ人口データ

8

メッシュの状態



平常・異常等の状態は把握不可能

本研究での統計的異常検知問題の特徴

9

□ 人物動態モニタリングにおける統計的異常検知問題の4要素

▶ **メッシュ人口データ**を対象に以下の表に整理を行った

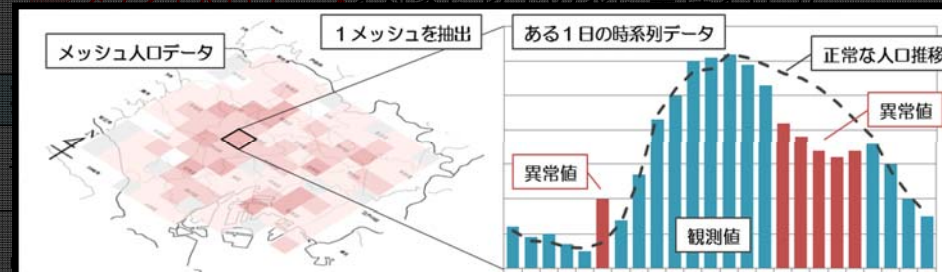
└ 時間帯別にメッシュ毎の人口を推計。匿名かつ入手可能性が高い

要素	人物動態モニタリングにおける特徴
データの性質	各メッシュから得られる メッシュ人口値の時系列データ とみなす また、時系列データの空間的な相関、影響も考えられる
出力方法	異常スコアの算出 / 「異常」「正常」メッシュの判断 (両方可)
ラベル	半教師付き異常検知 および 教師なし異常検知 日常的な交通状況より「正常」ラベルを持つデータが入手可能
異常の種類	文脈型異常 に集約される 事故による交通渋滞の影響で生じた過度な人口増加・減少 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

本研究での統計的異常検知問題の特徴

10

□ 人物動態モニタリングにおける統計的異常検知問題の4要素 一つのメッシュの時系列の観測データに着目



本研究では、1メッシュ中の**時系列データ**に潜む**異常**の検知を目指す
空間方向へはモデルを拡張することで対応を行う

異常の種類	文脈型異常 に集約される 事故による交通渋滞の影響で生じた過度な人口増加・減少 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
-------	---

→ 時系列データに潜む文脈型異常の検知が可能な手法が望まれる

メッシュ人口データと状態

13

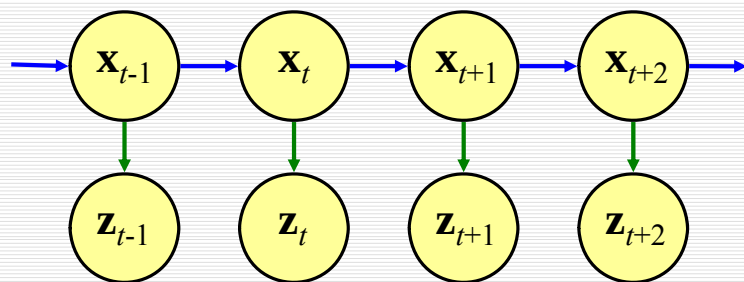
□ メッシュの状態

x_t : 時刻 t におけるメッシュの状態

視覚化の際の判読性を考慮し
離散的に表現

— 増減はないか多い etc...

異常: 普段とはかけ離れた状態



z_t : 時刻 t に観測されたメッシュ人口

メッシュ人口データ: 観測できない状態から出力されたデータ

メッシュの状態は時系列変化

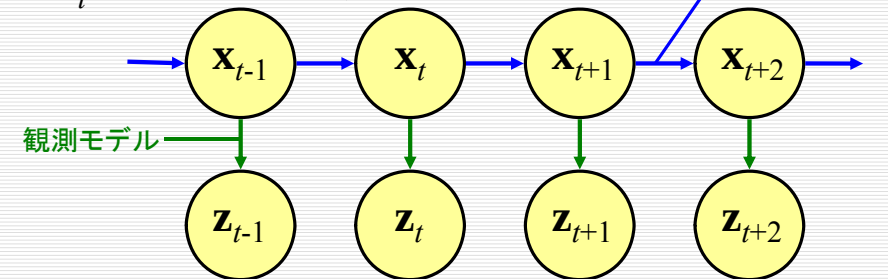
一般状態空間モデル

14

□ グラフィカル表現

x_t : 時刻 t における状態ベクトル: 観測不可

z_t : 時刻 t における観測ベクトル: 観測可



システムモデル $x_t = f_t(x_{t-1}, v_t)$ $x_t \sim p(x_t | x_{t-1})$

観測モデル $z_t = h_t(x_t, w_t)$ $z_t \sim p(z_t | x_t)$

事後確率最大化基準の下、最適な状態ベクトルを推定

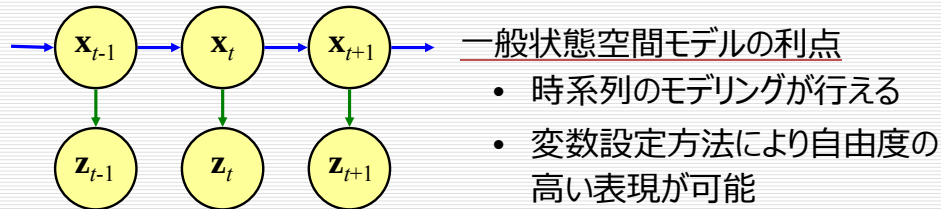
: MAP推定

$p(x_t | z_{1:t}) \rightarrow \max.$ → optimal x_t

一般状態空間モデルに基づく変化点検出

- 前述の要件を満たすと考えられる手法

▶ **一般状態空間モデル** を用いた異常検知手法が利用可能



一般状態空間モデルによる統計的異常検知手法

2つの方法に大別できる

- 観測値の尤度・出力確率計算による手法
KFやPFを用いて逐次状態推定を行い、観測値の尤度から異常を判断
- 異常状態の推定による手法
あらかじめ設計した異常状態に推定された時のデータを異常と判断

隠れマルコフモデル(HMM)

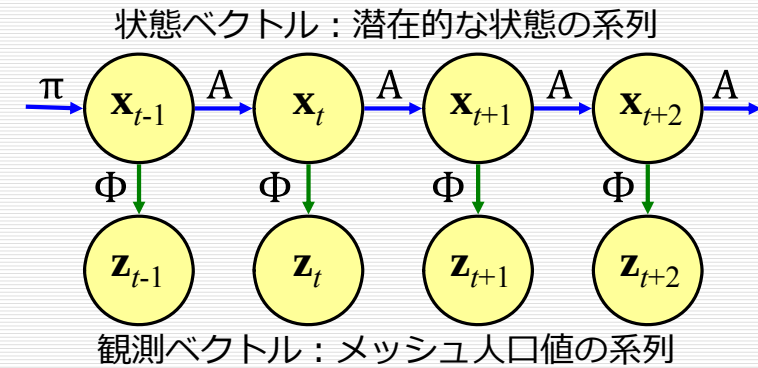
一般状態空間モデルにおいて、
離散的な潜在変数を導入したモデル
～自然言語処理などで発展

状態ベクトル

パラメータ群

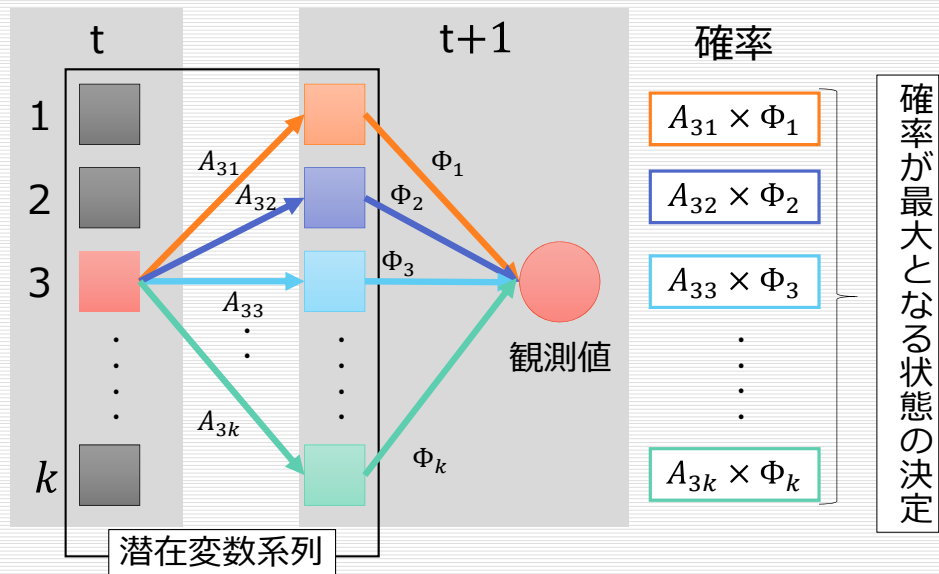
- π : 初期分布確率
- A : 遷移確率
- Φ : 出力確率

HMMのグラフィカル表現



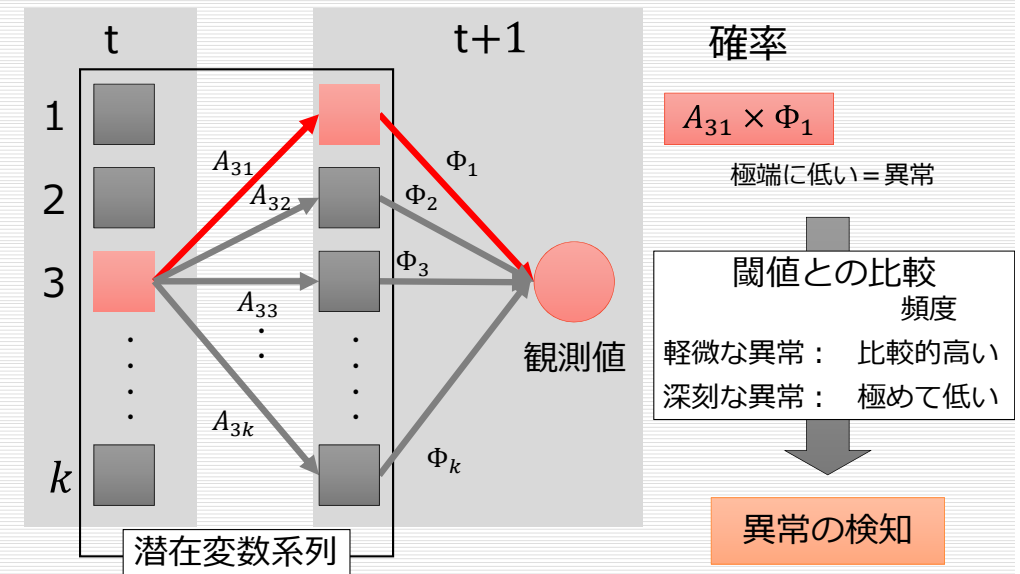
状態推定・異常検知

□ 状態遷移図



状態推定・異常検知

□ 状態遷移図

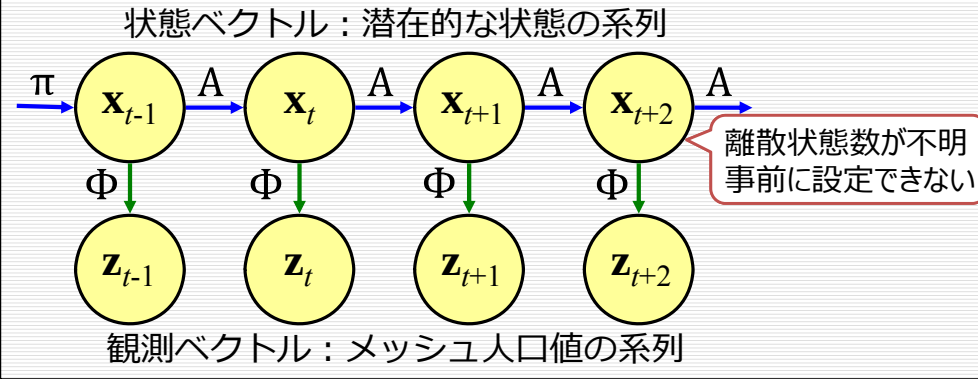


隠れマルコフモデル(HMM)

25

- 正常・異常状態の学習において、一般には状態数を事前に設定
 - しかし、メッシュ人口データの状態数は未知

HMMのグラフィカル表現



統計的異常検知

26

	観測値の尤度・出力確率の計算手法	異常な潜在状態の推定手法
長所	自由度の高いモデリングが可能	離散な状態を推定すればよく、異常の解釈が明確
短所	- データ生成モデル側に高い精度が必要 - 尤度の解釈が難しい	- 状態数の定義が必要 - 状態が離散的にしか取れず自由度の低いモデリングとなる

<観測値の尤度・出力確率の計算手法>

- 人物動態データの観測値に対し、観測モデルとシステムモデルの最適な組み合わせは未だ確立されていない
 - 尤度の低下がモデルの不具合か、異常の検出か判別が難しい

<異常な潜在状態の推定手法>

- 潜在状態に離散変数を事前に設定する必要がある
 - 人口メッシュデータの状態数は未知

そこで、データに応じて状態数を同時に推定可能な

階層ディリクレ過程隠れマルコフモデル(HDP-HMM)の枠組みに着目

階層ディリクレ過程隠れマルコフモデル

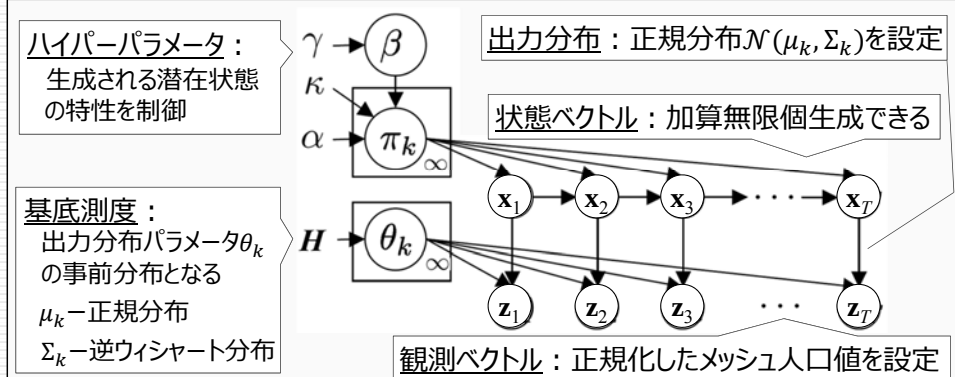
27

- 加算無限個の状態を潜在的に仮定する

階層ディリクレ過程に基づく隠れマルコフモデル (sticky HDP-HMM (Fox, 2008))

を基に異常検知手法を構築
- データに応じて状態数を同時に推定可能なモデル

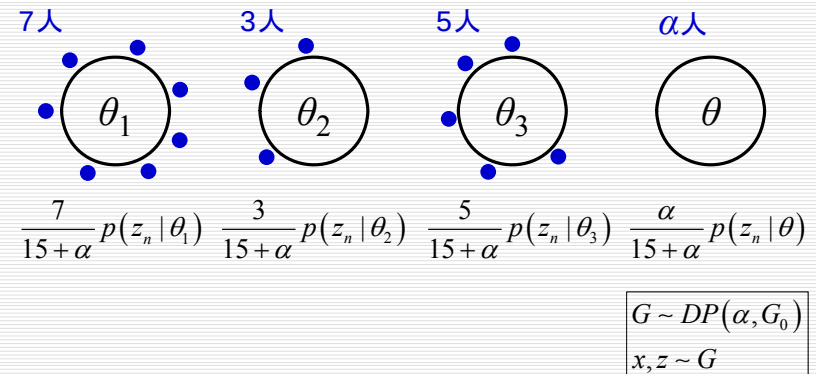
▼sHDP-HMMのグラフィカルモデル



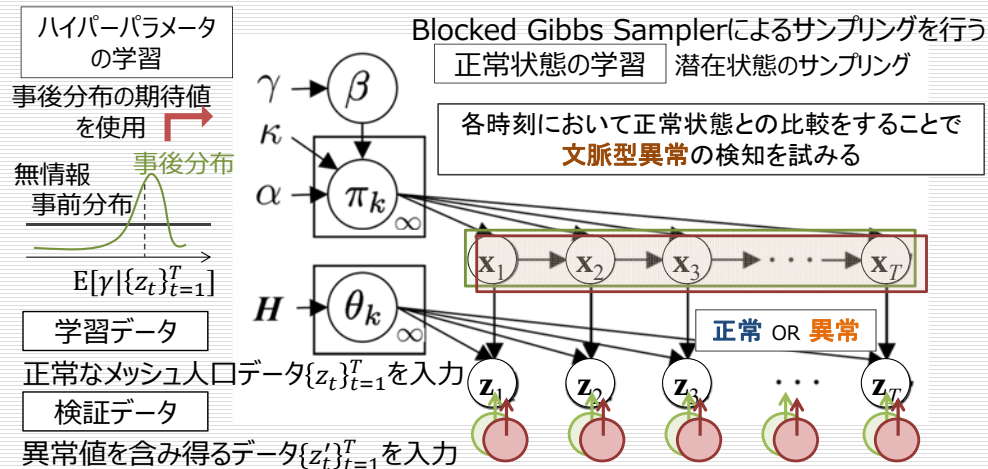
Chinese Restaurant Process: CRP

28

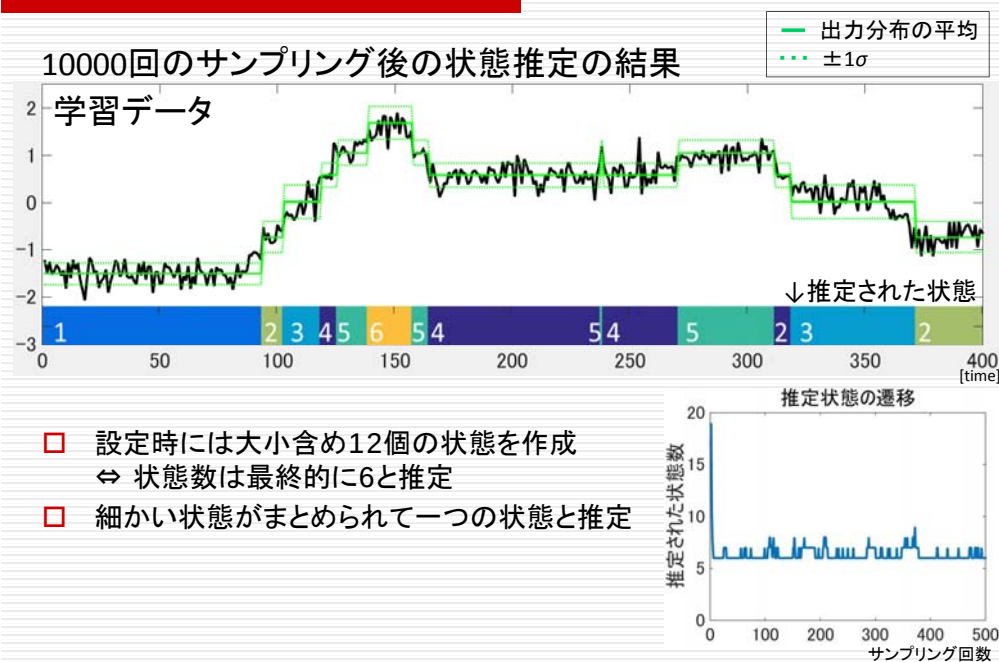
- 新しい客 x_n (状態) は、以前の客 $x_{1:n-1}$ のテーブル着席状態に従ってテーブルを決定
- テーブルが決まったらその料理 z_n (観測値) を θ_k に従って決定



1. 学習データによってハイパーパラメータおよび正常状態とする潜在状態を学習
2. 学習済みのsHDP-HMMに適用データを適用し、潜在状態を推定
3. 状態の対応付けを行い、各時刻において異なる状態となれば異常と判定



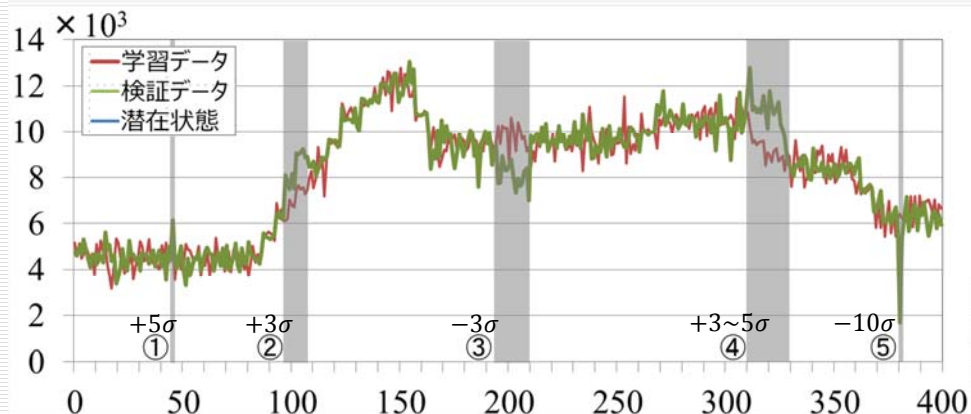
学習結果



- ❑ 設定時には大小含め12個の状態を作成
⇔ 状態数は最終的に6と推定
- ❑ 細かい状態がまとめられて一つの状態と推定

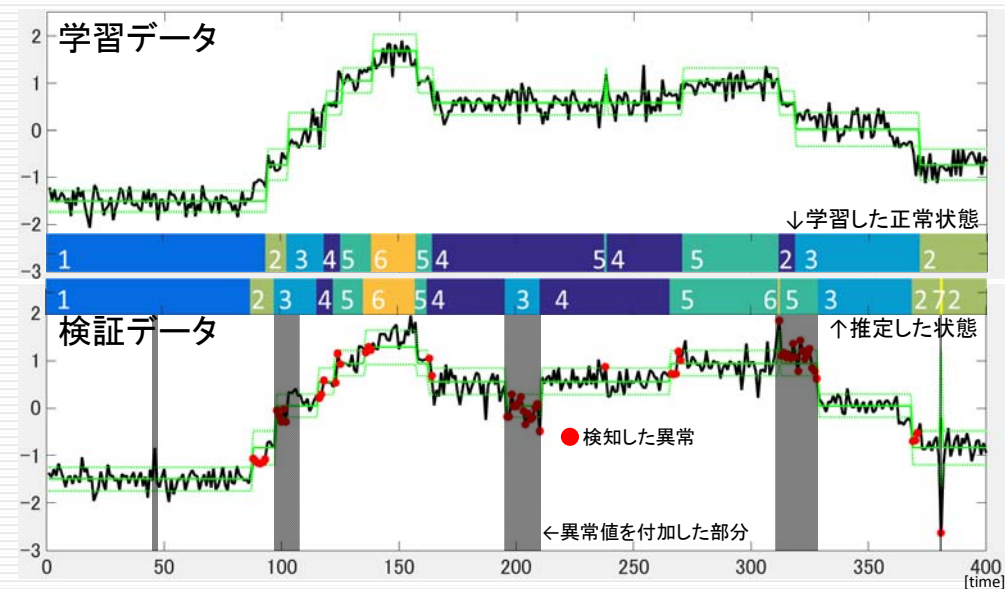
シミュレーションによる基本性能の検証

- ❑ 状態区分が細かく、値が滑らかに遷移する時系列データを作成
 - シミュレーションの状態の平均値は実データを参考
 - 各状態に対応した正規分布からのi.i.dサンプルを並べて時系列を生成
- ❑ 検証データには 5箇所①～⑤ に異常値をそれぞれ付加
- ❑ sHDP-HMMへの入力には時系列で正規化したものを利用

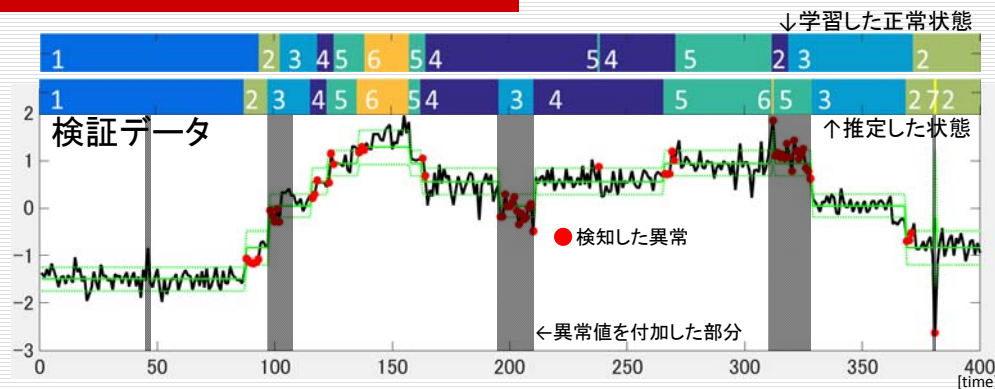


異常検知結果

- 学習したハイパーパラメータで設定したsHDP-HMMを用いて検証データを推定



-
- 検知した異常
- ←異常値を付加し



□ 全時系列長 $T=400$ 中、64点で異常を検知

■ 異常検出率: $TP/(TP+FN)=80\%$

■ 精度: $TP/(TP+FP)=50\%$

□ 状態の境界付近での検知が多見

□ 異常を含む状態の出力分布パラメータが変化し、状態区分も変化

□ 値の変動する箇所でも正しく検知

		異常検知結果	
		異常 64	正常 336
真の状態	異常 40	True Positive 32	False Negative 8
	正常 360	False Positive 32	True Negative 328

前実験の条件を変更し、
学習データ量の相違による影響、データの時間分解能の影響の検証

1 複数の学習データを用いた場合における実験

同じ状態を持つ学習データ5つを使用

– 異常検出率: $TP/(TP+FN)=92.5\%$

– 精度: $TP/(TP+FP)=51.4\%$

– 異常検出率の向上が確認できた
精度については大きな向上は無し

(参考)前実験結果
検出率80%, 精度50%

		異常検知結果	
		異常 72	正常 328
真の状態	異常 40	True Positive 37	False Negative 3
	正常 360	False Positive 35	True Negative 325

2 時間分解能の低いデータでの実験

時系列長 $T=24$ に圧縮したデータを使用

– 異常検出率: $TP/(TP+FN)=75\%$

– 精度: $TP/(TP+FP)=54.5\%$

– 大幅な検出率の低下を起こすこと
なく適用可能であることを確認

		異常検知結果	
		異常 11	正常 13
真の状態	異常 8	True Positive 6	False Negative 2
	正常 16	False Positive 5	True Negative 11

□ 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

■ 統計的異常検知に関する論文のレビュー

■ メッシュ人口データ中の文脈型異常を検知する問題として設定

□ 人物動態モニタリングにおける統計的異常検知手法の構築

■ 潜在状態数を自動推定可能なsHDP-HMMに基づく手法構築

□ シミュレーションによる提案手法の基本性能の検証

■ 状態区分数の相違による影響

■ 学習データ量の相違による影響

■ データの時間分解能の影響の検証

□ 実データによる適用可能性の検証

■ 列車の運転見合わせが発生した地点・時刻での異常検知

■ 人口メッシュデータにおける空間分解能への示唆

□ 精度改良および多様なデータへの適用

□ 推定状態・異常の交通ネットワーク上における解釈の深化

■ 交通NW上における交通パターンや実社会現象と照合し、
推定状態・検知異常の解釈可能性の検討

□ モデルの拡張

■ 空間方向へ拡張したモデルの構築

□ 周辺メッシュ値との空間相関の考慮

■ 多様なデータを統合した統計的異常検知手法の構築

□ 天候やイベント等の要素を変数に加えたモデル構築

■ オンライン型異常検知手法と制御手法との統合

□ 異常のリアルタイム検知 および
動的管制の影響を加味したモデル構築