



時系列ビッグデータ解析と リアルタイム予測

熊本大学大学院先端科学研究部
科学技術振興機構さきがけ研究員
松原靖子



Kumamoto University



研究領域「新しい社会システムデザインに向けた情報
基盤技術の創出」（研究総括：黒橋 禎夫先生）

研究理念

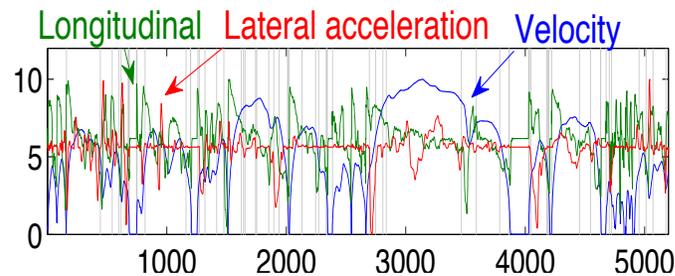
未来の予測によって社会を変革する

大規模データを用いて
自然現象や社会現象の時間発展を
リアルタイムに予測し、
社会活動を最適化する

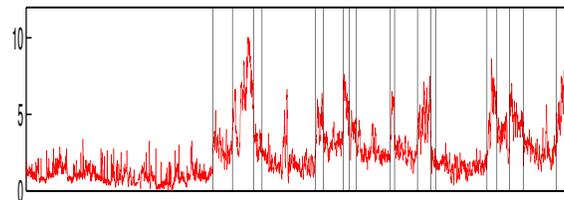
時系列ビッグデータ

• IoTビッグデータ

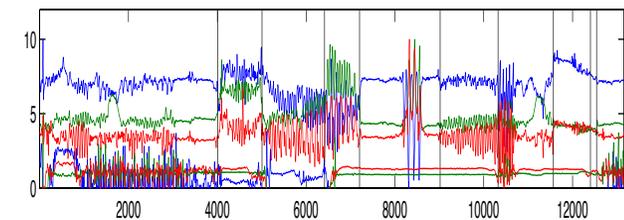
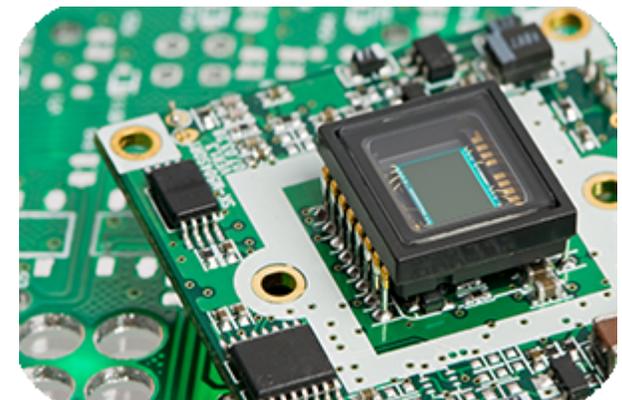
– 各種センサデータストリーム



車両走行センサ



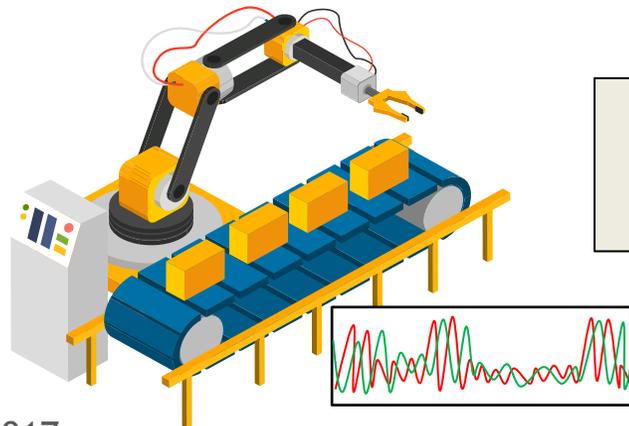
生体センサ



モバイルセンサ

時系列ビッグデータ解析

- **情報化社会におけるデータ量の飛躍的な増大**
 - IoTビッグデータ（環境、交通、生体）
 - Web、ソーシャルメディア、医療情報
- **社会を変革する情報工学**
 - IoTビッグデータ解析による
製造業の変革、付加価値の向上



生産性 /
品質向上

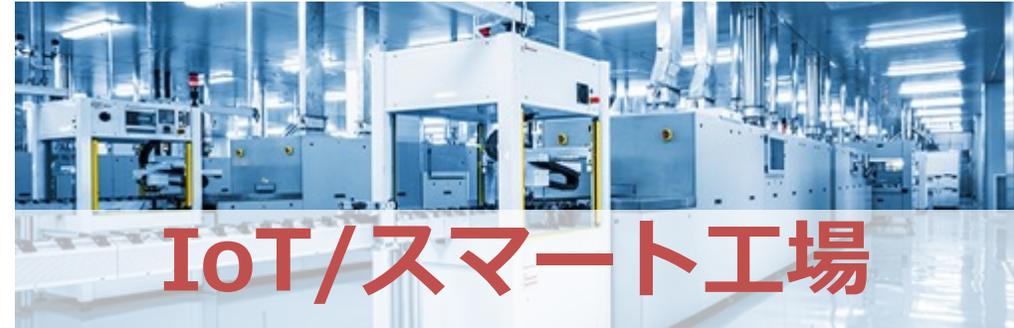


運転支援 /
事故回避



産業貢献のための共同研究

- より良いサービスのために
 - トヨタIT開発センター様 [2014, 2015, 2016年度]
 - 富士通研究所様 [2016年度]
 - 2017年度：契約手続き中（本年度は8社程度を予定）

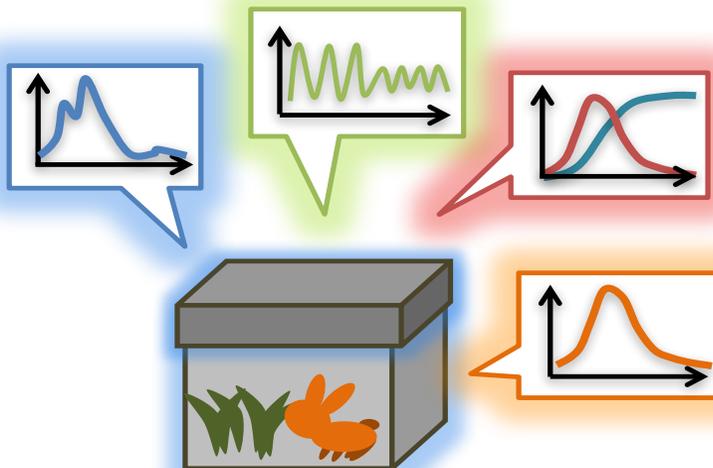


Roadmap

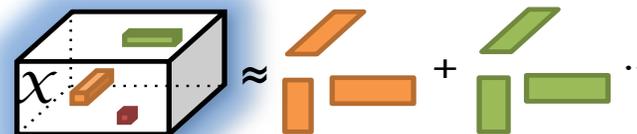


ベースとなる技術

非線形モデリング



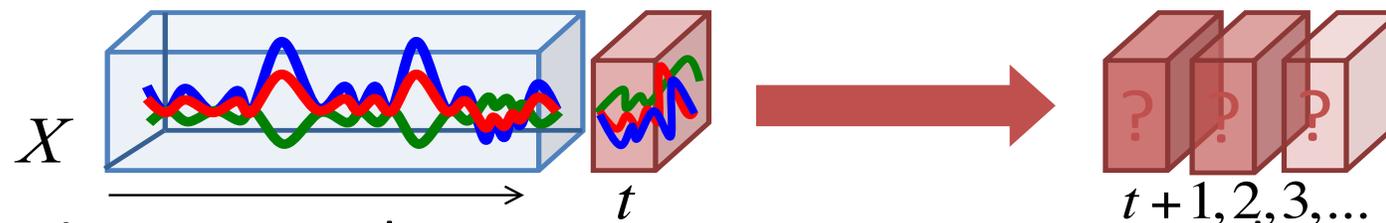
テンソル解析



特徴自動抽出



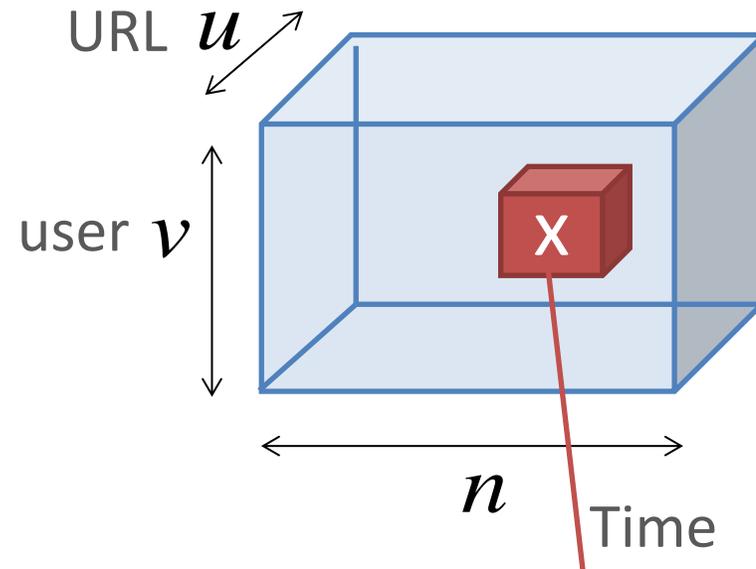
リアルタイム処理



(T1) 大規模テンソル解析

- 時刻付きイベント
 – 例：Webアクセスログ

Time	URL	User
08-01-12:00	CNN.com	Smith
08-02-15:00	YouTube.com	Brown
08-02-19:00	CNET.com	Smith
08-03-11:00	CNN.com	Johnson
...



Element x : # of events

e.g., 'Smith', 'CNN.com',
 'Aug 1, 10pm'; 21 times

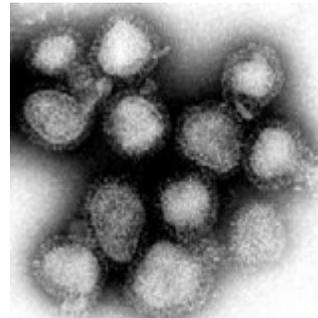
M^{th} order tensor ($M=3$)

$$\mathcal{X} \in \mathbb{N}^{u \times v \times n}$$

(T2) 非線形モデリング

• 非線形方程式 (non-linear equations)

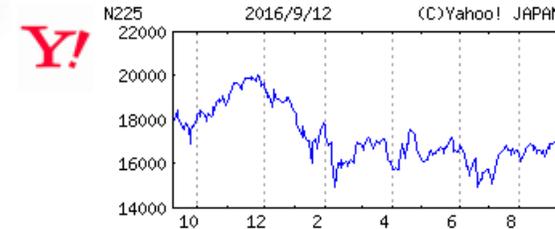
- 疫学 (epidemiology)
- 生物学 (biology)
- 物理学 (physics)
- 経済学 (economics)



Influenza@Wikipedia



FreeDigitalPhotos.net



• 非線形社会現象の解析

- 時系列ビッグデータの非線形解析
- Web, IoTなど

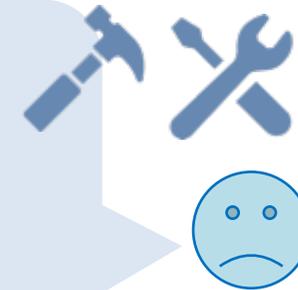


(T3) 特徴自動抽出

- 特徴自動抽出の重要性

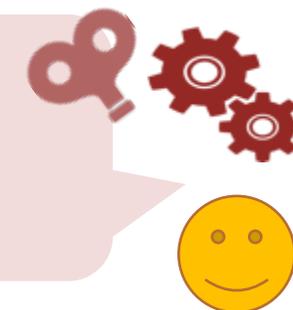
手動

- パタメータ調整がセンシティブ
- 調整作業に長い時間（数時間、数日…）



自動

- 技術者、専門家のチューニングが不要



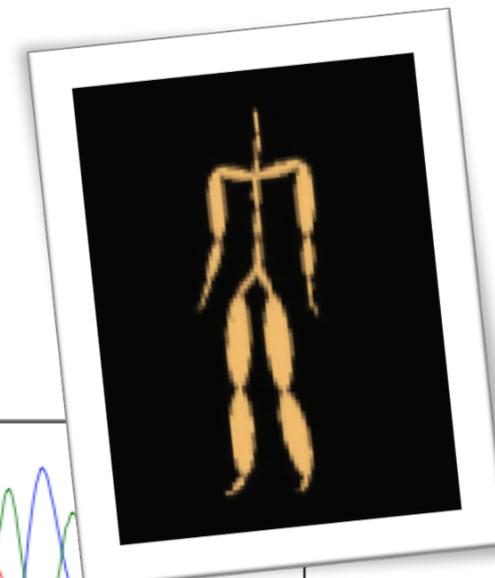
- ビッグデータマイニング：
-> 人手を介さない処理が必要

特徴自動抽出

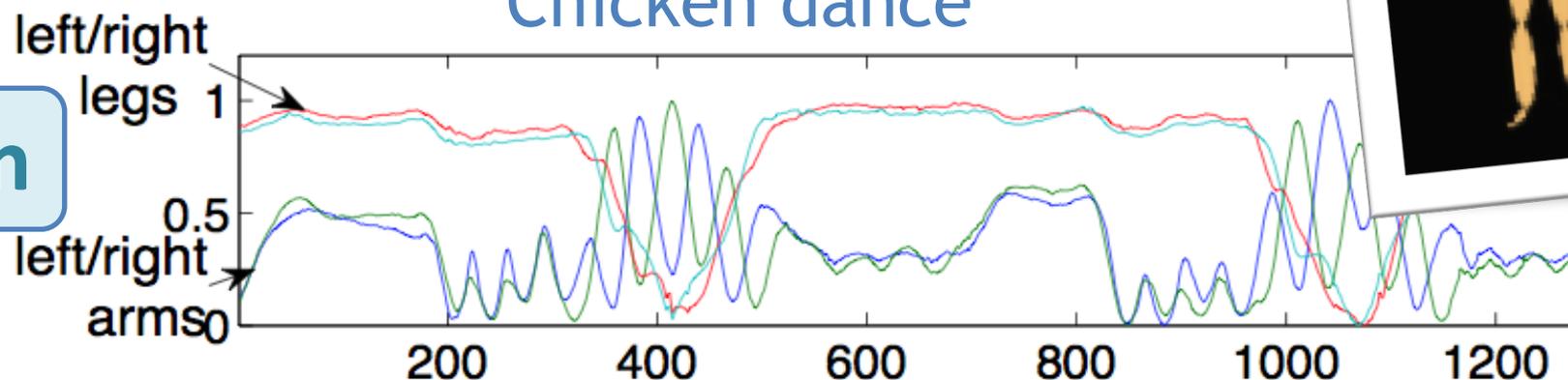
“Automatic” mining algorithm

Find: compact description of data X

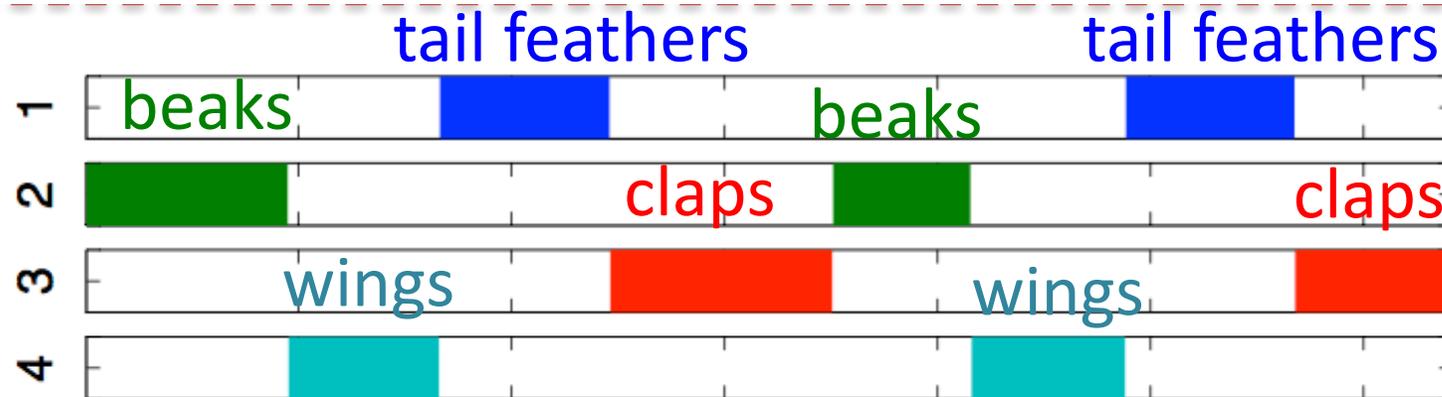
Chicken dance



Given

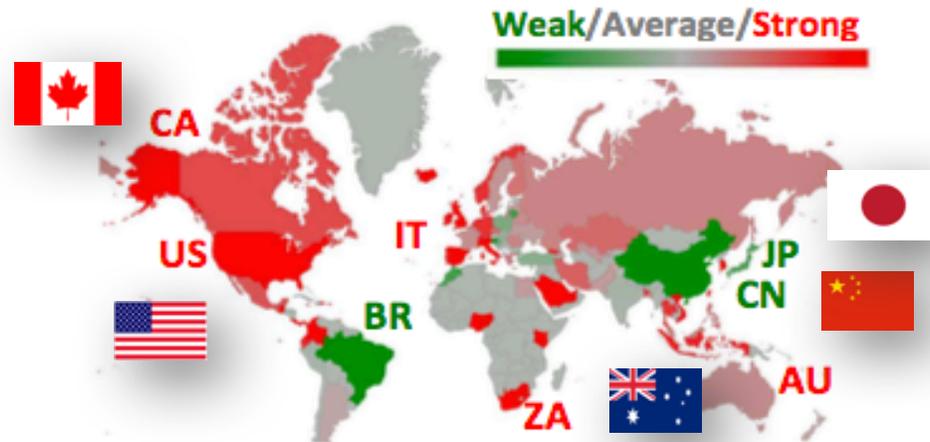


Find

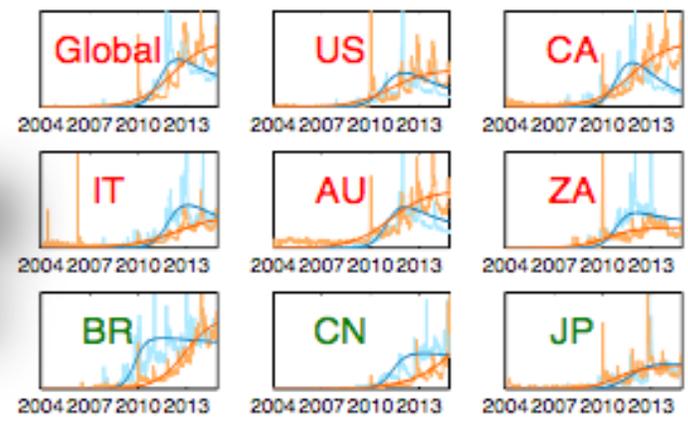


非線形テンソル解析

- Non-Linear Mining of Competing Local Activities (WWW2016)



(a) Local competition strength between Kindle and Nexus



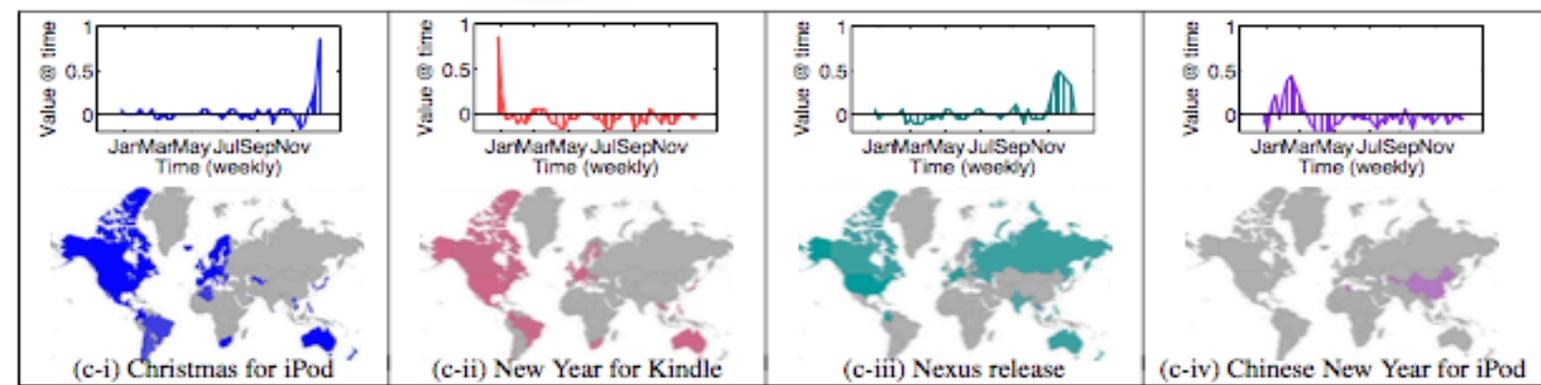
(b) Original/fitted search volumes for Kindle and Nexus



Nexus VS. Kindle



VS.



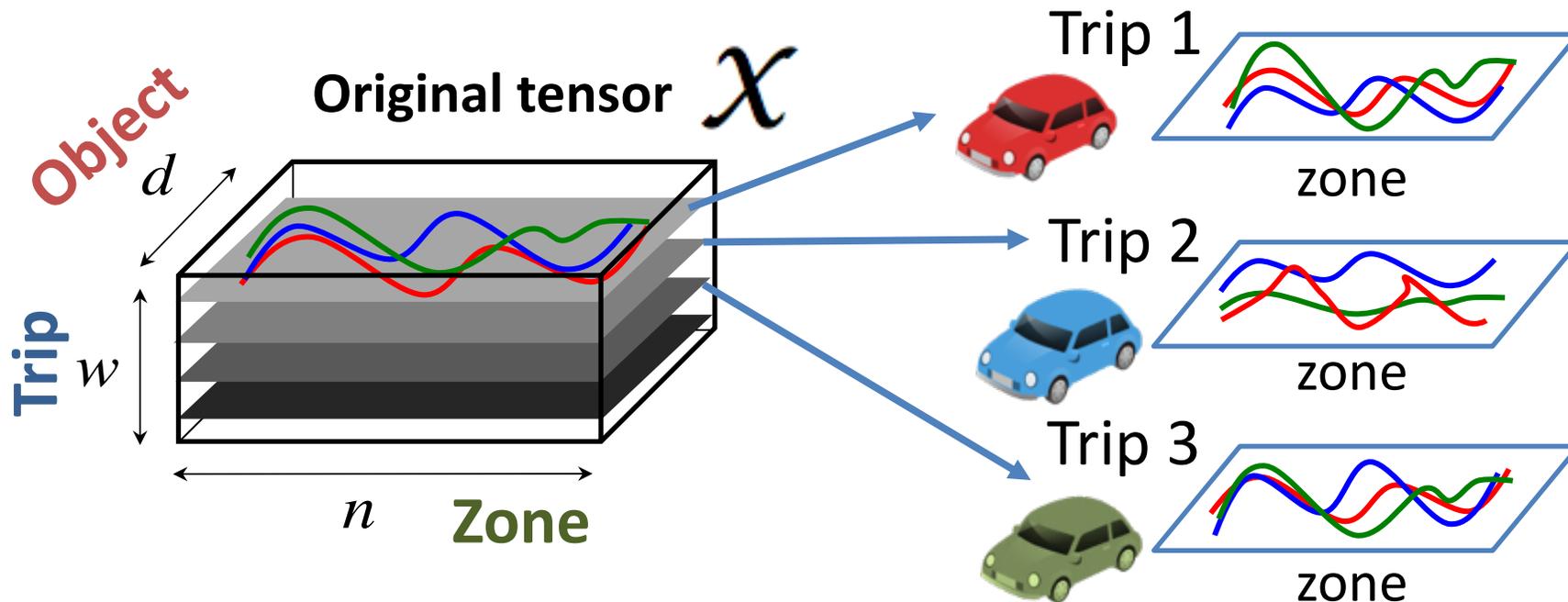
(c) Seasonal/annual patterns for each location/country



VS.

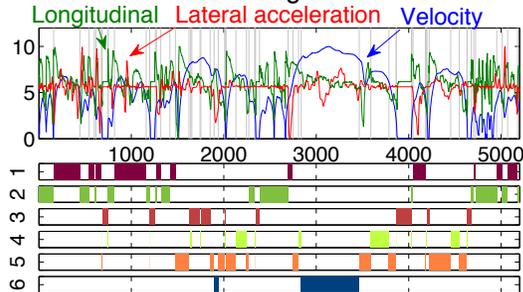
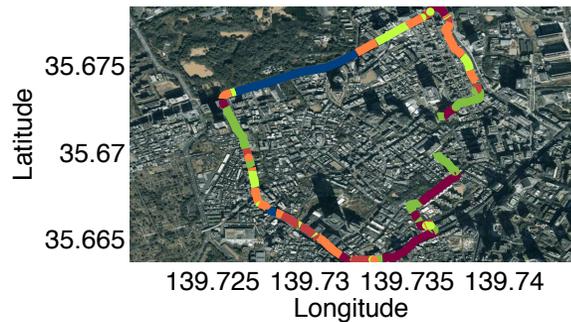
複合時系列データ解析の実用化

- 車両センサデータのテンソル解析
地理情報テンソル \mathcal{X} {**trip**, **zone**, **object**}

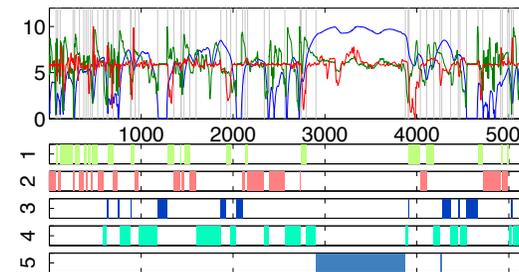
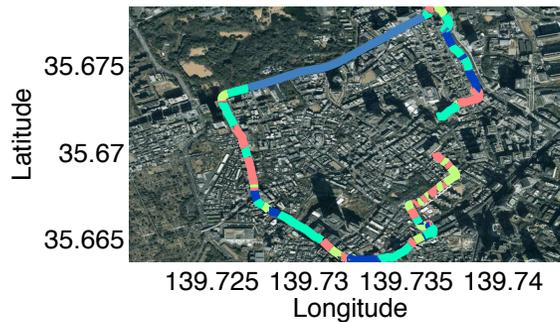


複合時系列データ解析の実用化

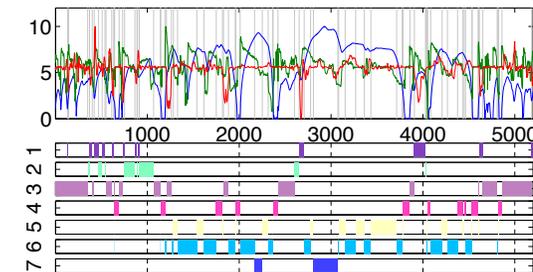
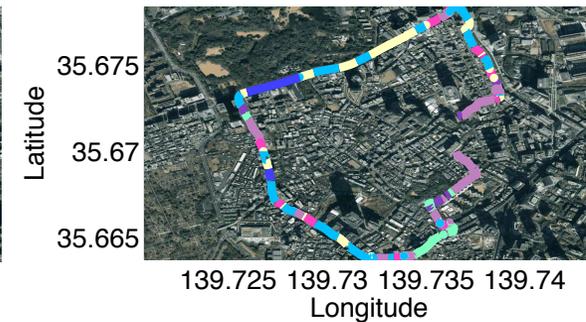
- 全ての要素を統合的に解析
- データ全体を表現する要約情報を抽出
- 走行データに基づく高度な道路地図情報を提供



(a) H-regime 1 (9/31)



(b) H-regime 2 (21/31)



(c) H-regime 3 (1/31)

Roadmap

- ✓ Research issues
- ✓ Technologies
- Ongoing projects



Smart assistant service

技術的課題

1. リアルタイム予測と自律的モデル学習
2. 社会行動支援サービスのための動的要因分析

応用

- 建設業、製造業、交通サービス、Web、環境



リアルタイム予測

リアルタイム予測に必要なこと

(a) l_s -時刻先の予測

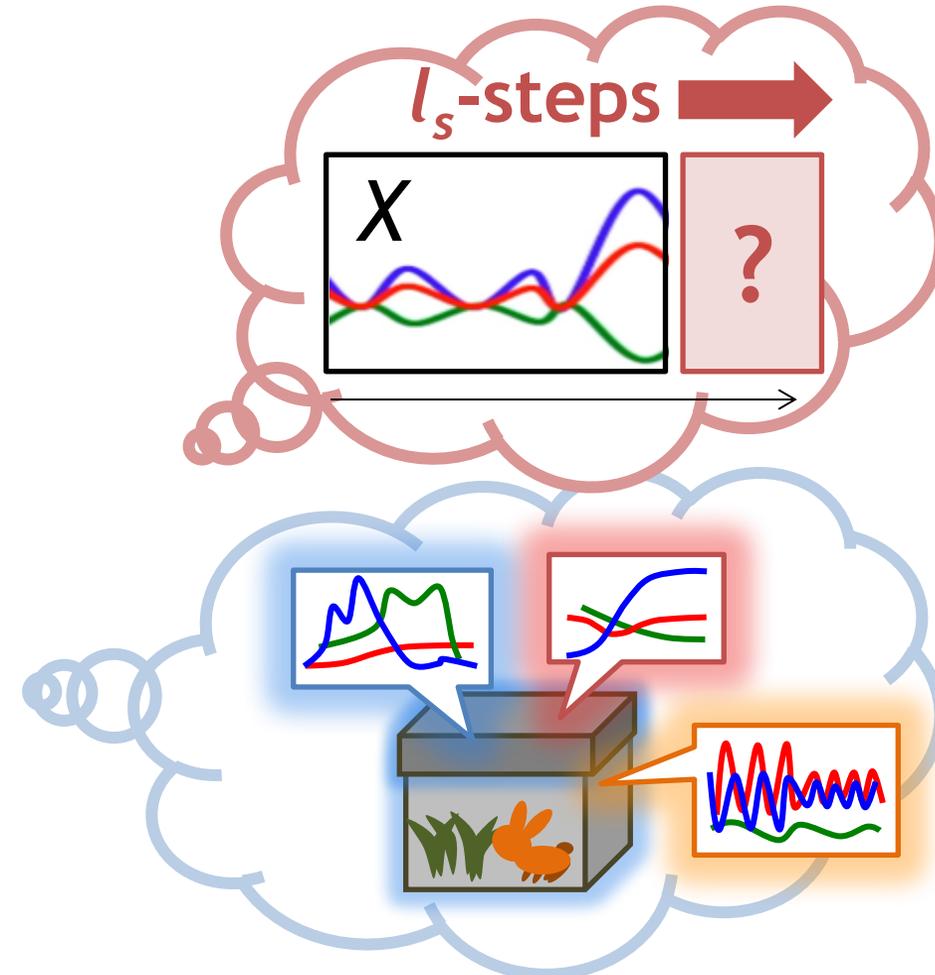
Long-term

(b) 継続的にパターン検出

Continuous

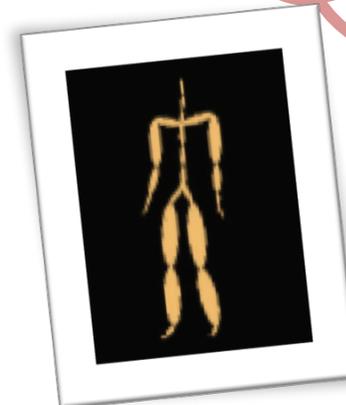
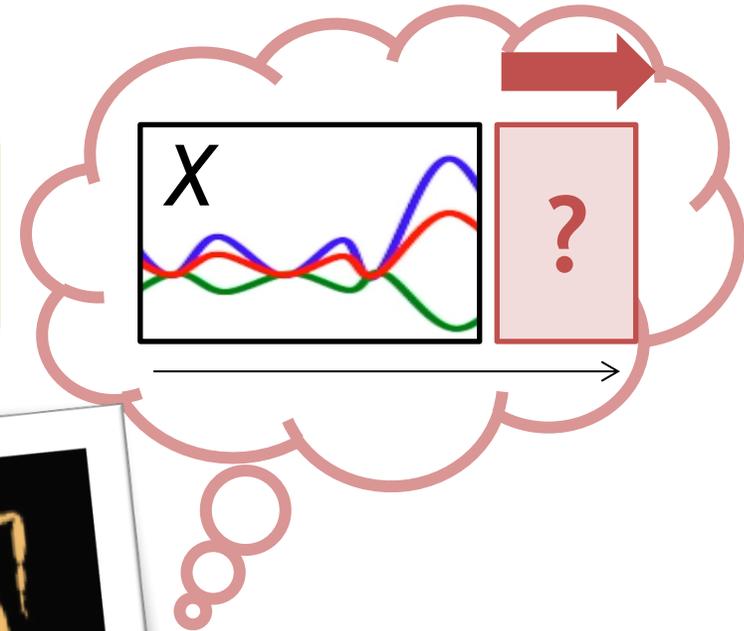
(c) 適応力のある予測

Adaptive



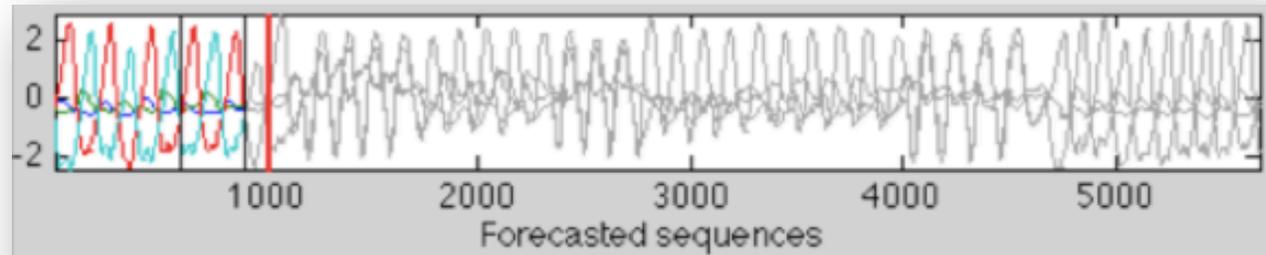
- Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences (KDD2016)

未来を予測し続けるには？



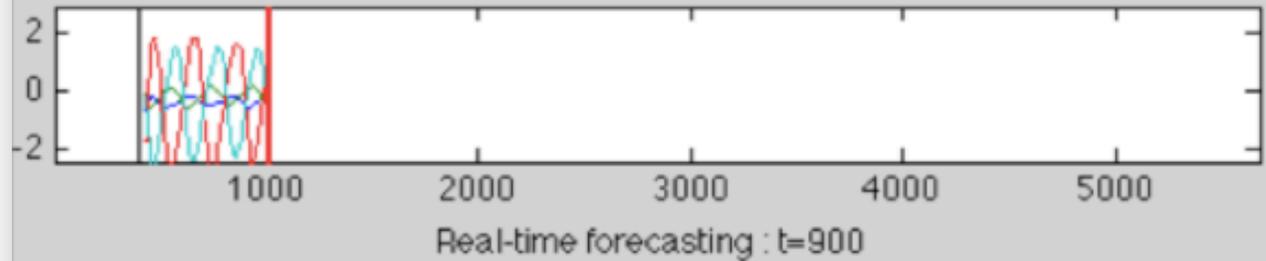
リアルタイム予測

Original



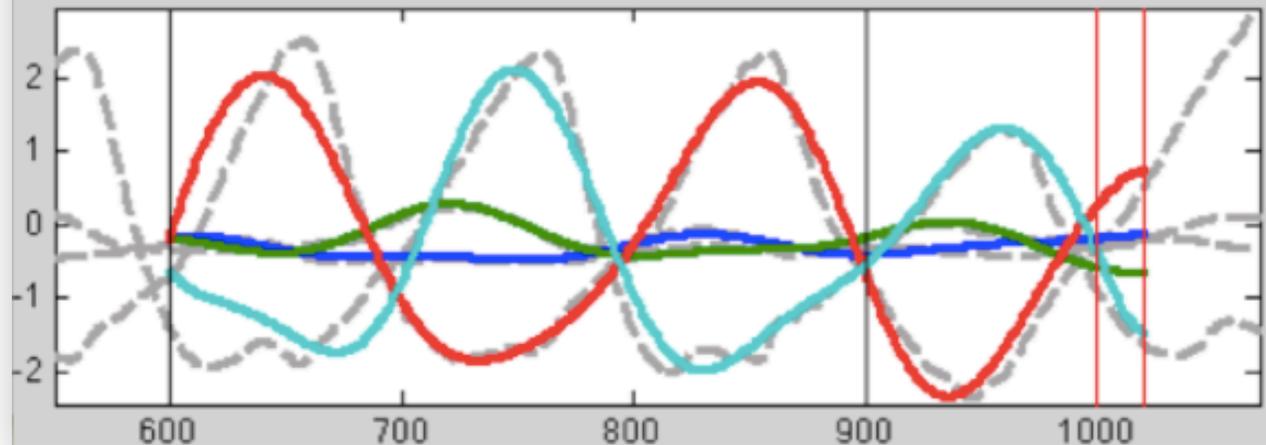
Forecast

(100-steps
-ahead)

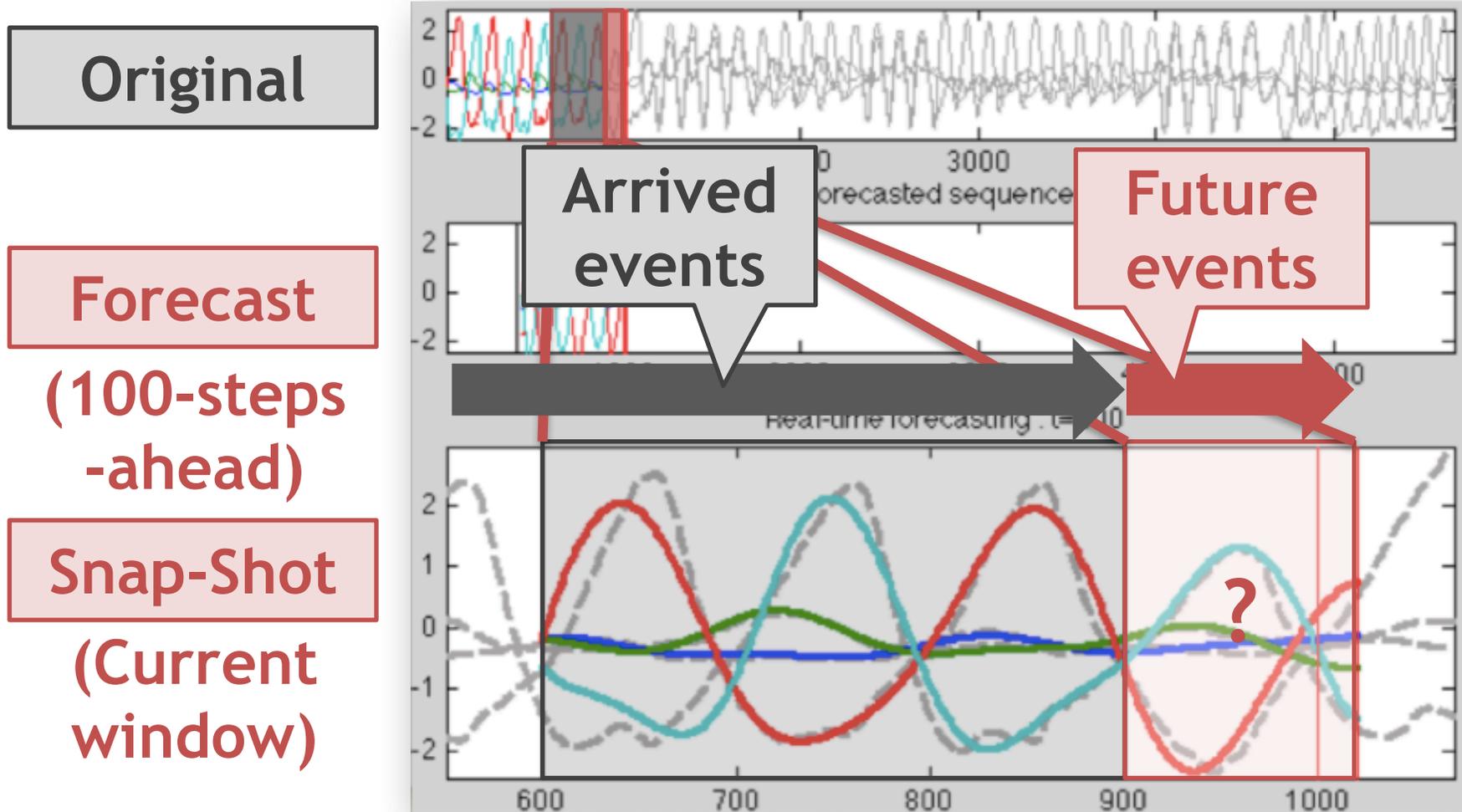


Snap-Shot

(Current
window)



リアルタイム予測



リアルタイム予測



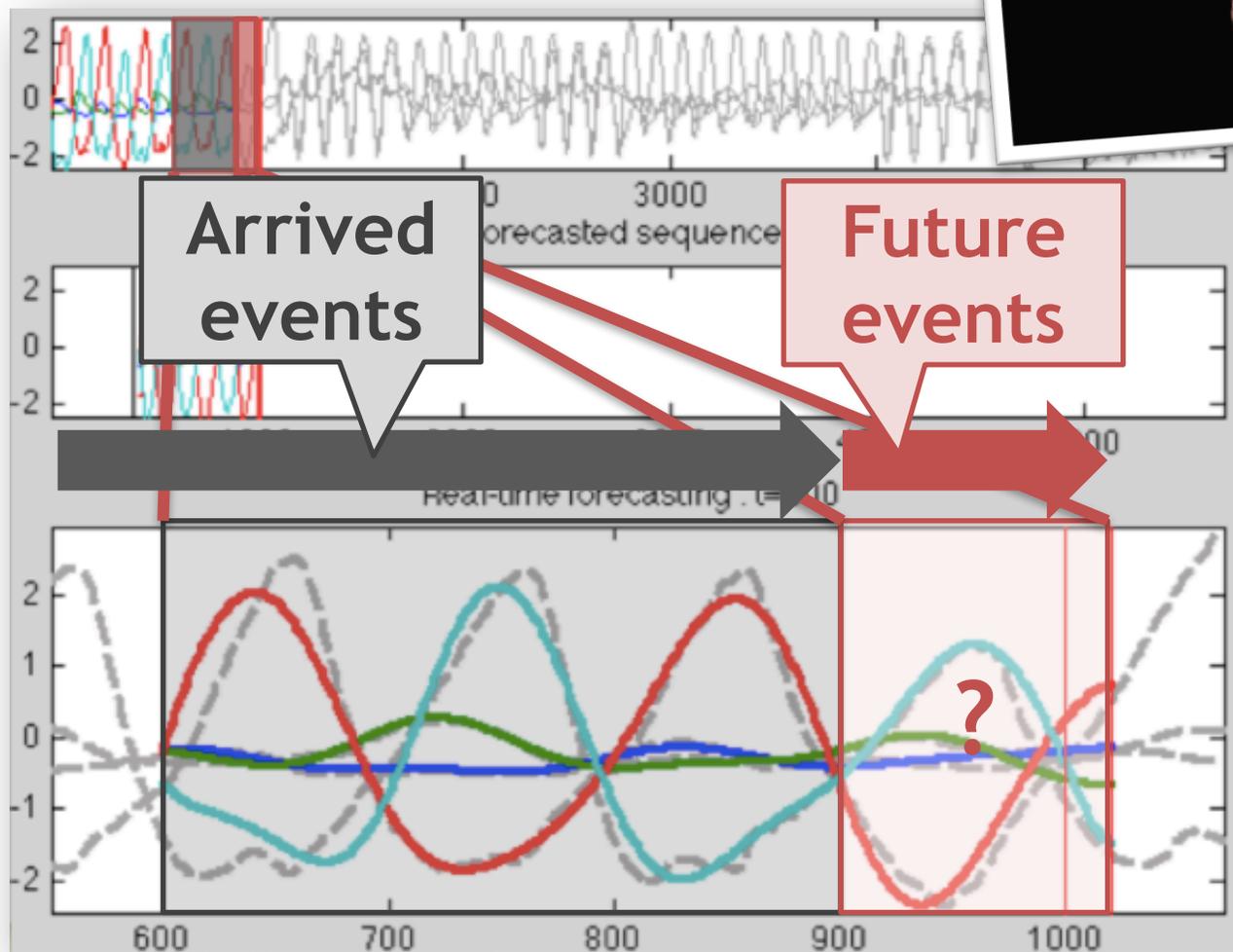
Original

Forecast

(100-steps-ahead)

Snap-Shot

(Current window)



リアルタイム予測

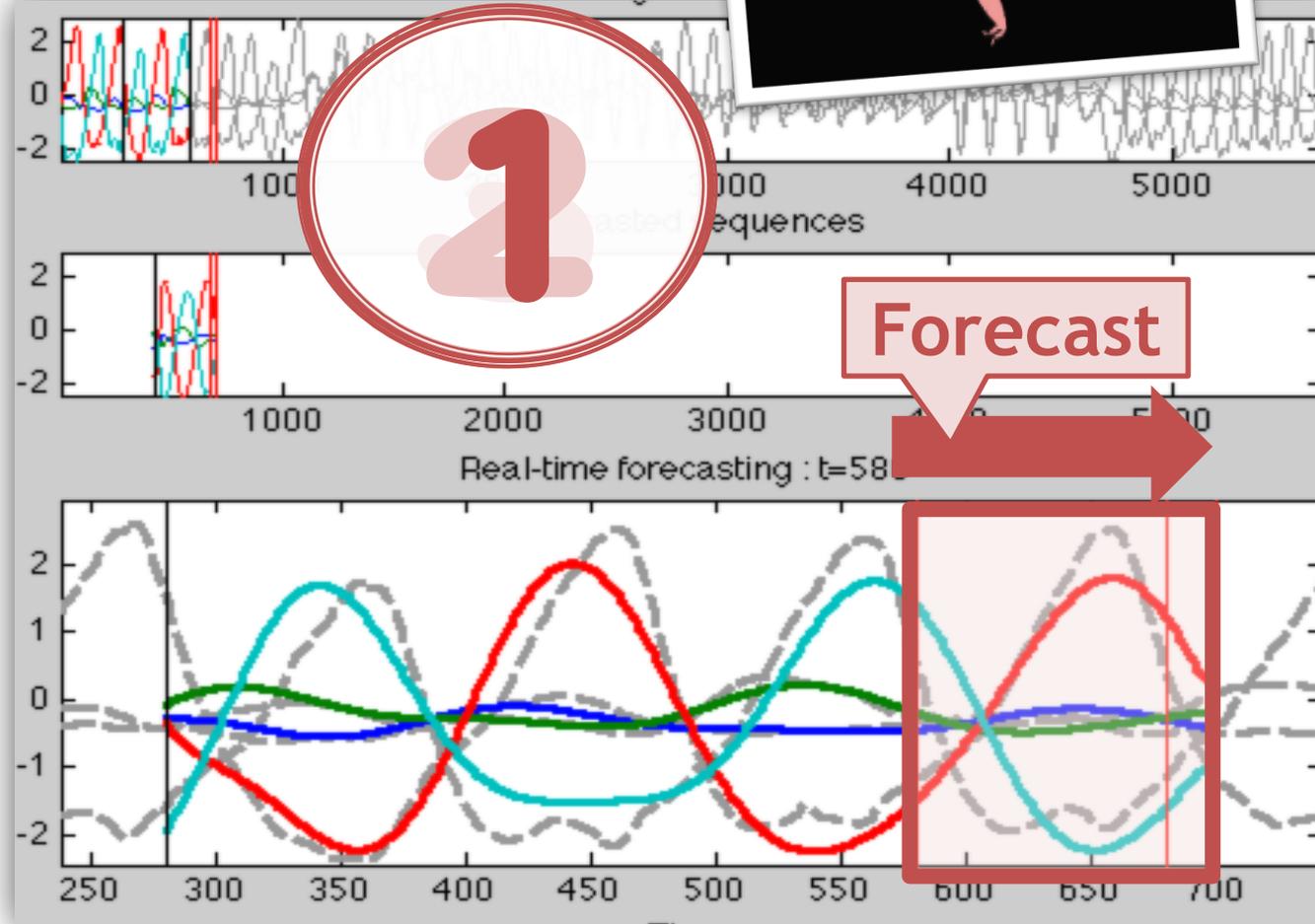
Original

Forecast

(100-steps
-ahead)

Snap-Shot

(Current
window)

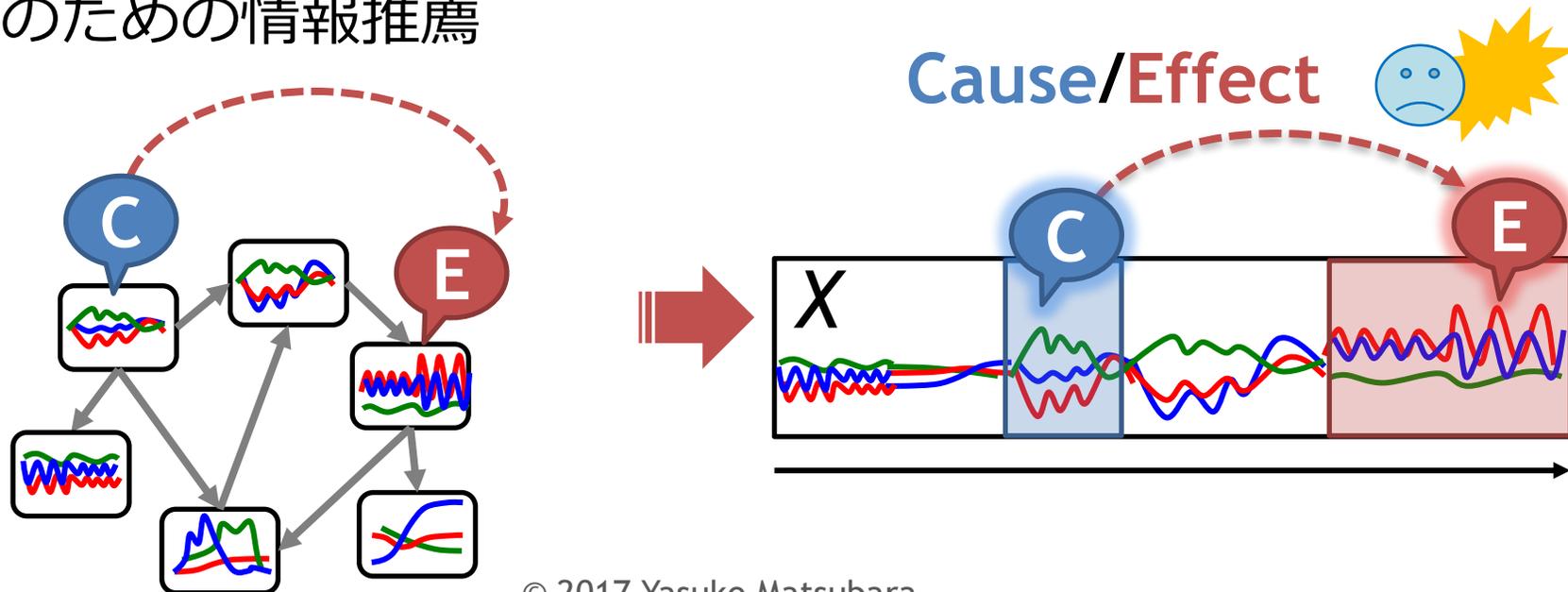


動的的要因分析

技術的課題

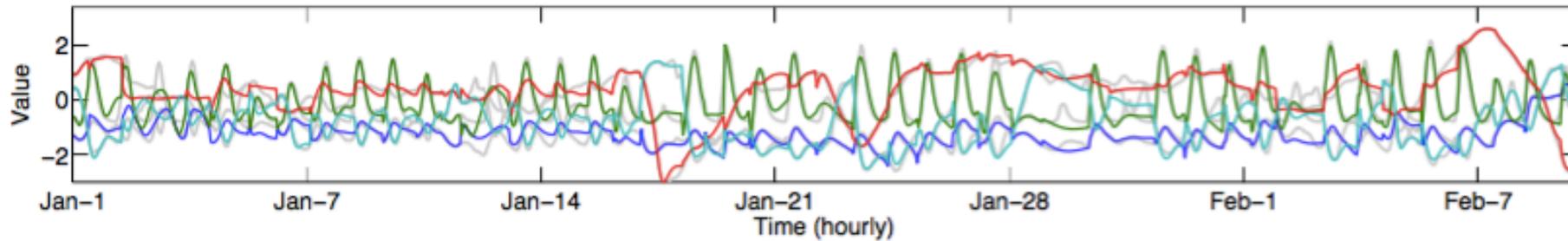
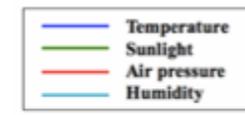
2. 社会行動支援サービスのための動的的要因分析

- モデル間の連結の強さの推定
- 連結をたどることによる要因／結果の関係性の発見
- 事故やトラブルのサイン（兆し）の監視
- 社会行動のための情報推薦

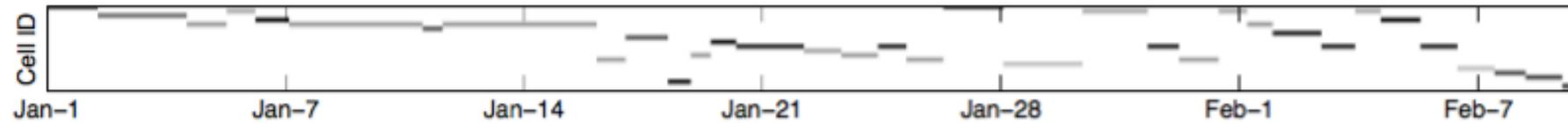


動的要因分析

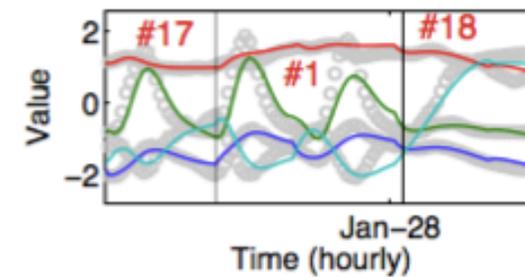
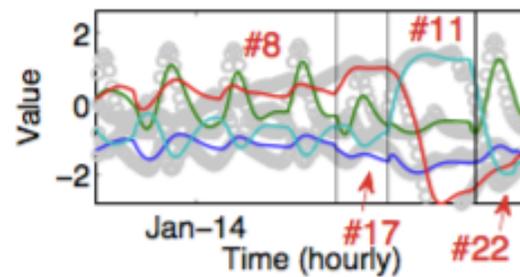
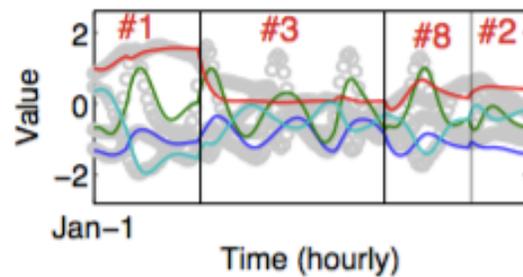
環境センサデータの解析例



(a) Fitting result (solid lines) vs. original data (gray lines)

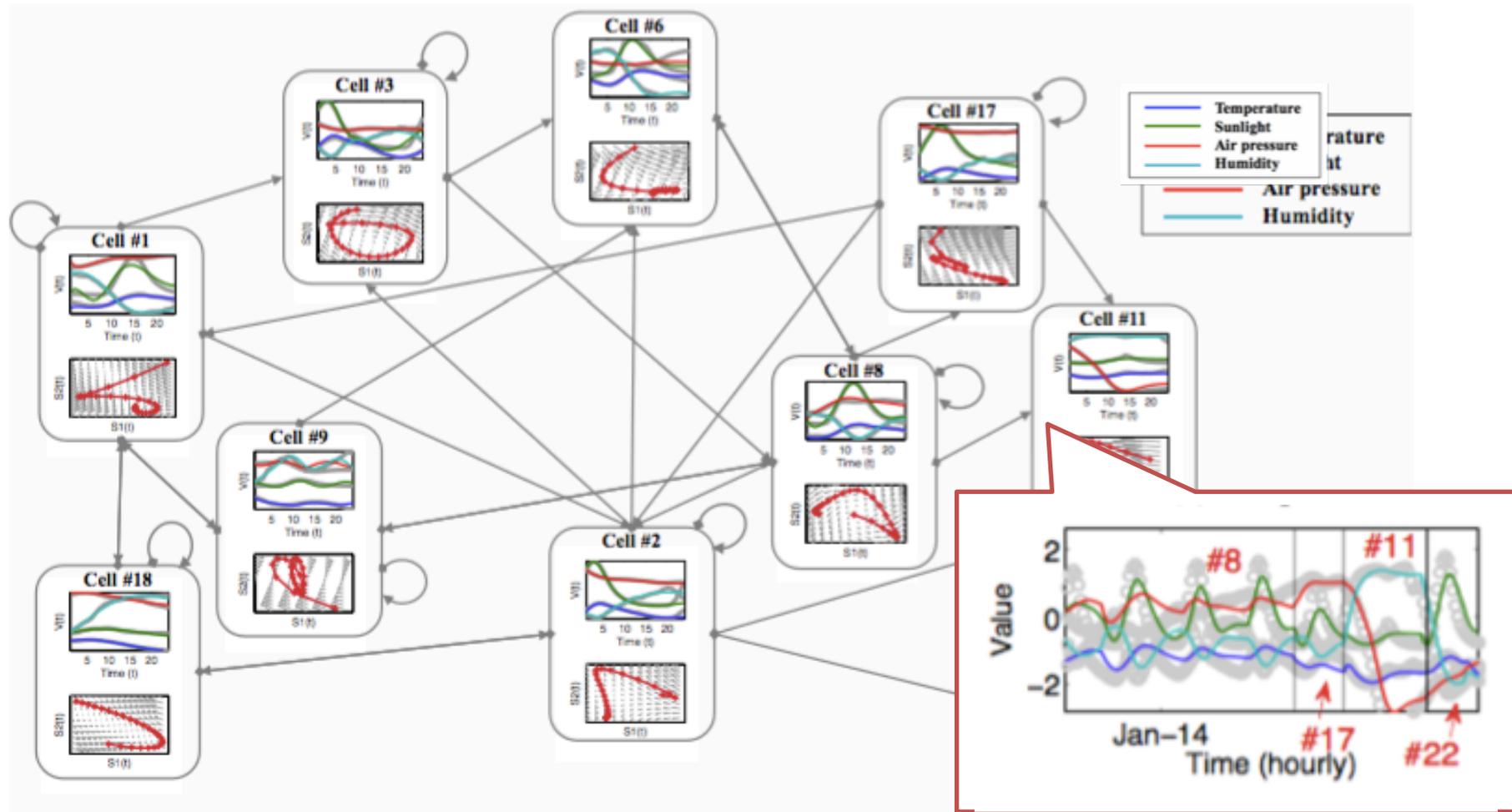


(b) Footprint



動的要因分析

環境センサデータの解析例



Roadmap

- ✓ Research issues
- ✓ Technologies
- ✓ Ongoing projects



将来の社会構造への貢献

- リアルタイム予測に基づく高度な社会サービスの実現



交通システム

(渋滞緩和/事故防止)



製造/流通/開発

(作業ストレス緩和/事故防止)



ヘルスケア

(健康維持)

将来の社会構造への貢献

- ・ **リアルタイム予測**に基づく高度な社会サービスの実現

展示（ポスター・デモ）

「時系列ビッグデータ解析とリアルタイム予測」

松原靖子 [J4] さきがけ