



時系列ビッグデータのリアルタイム解析： 新技術と挑戦

大阪大学 産業科学研究所
産業科学AIセンター
櫻井保志



講義内容

- ① 概要
- ② 時系列解析の基礎
- ③ リアルタイム解析技術
- ④ 社会実装への取り組み



研究の理念

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ための技術開発

【対象データ】

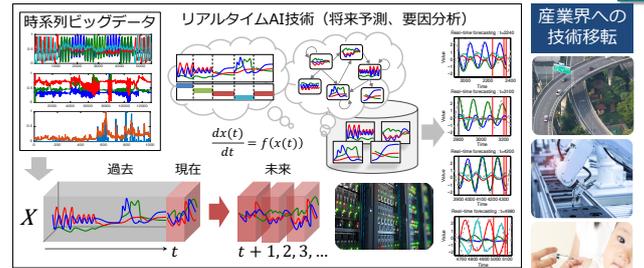
- IoT (スマート工場、車両走行データ)
- 医療、Web

【応用】

- 製造業、交通、医療、Web
- 産業界への技術移転：
「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」



研究開発の取り組み



時系列ビッグデータのリアルタイム解析・学習と
将来予測・要因分析に基づく知的支援サービス

講義内容

- ① 概要
- ② 時系列解析の基礎
- ③ リアルタイム解析技術
- ④ 社会実装への取り組み

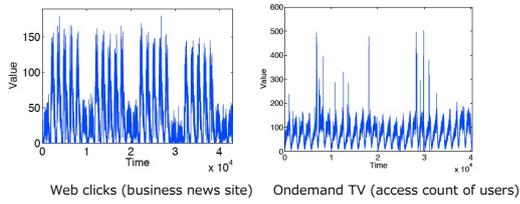


時系列データ

- Webオンラインアクティビティ
 - Webアクセスログ
 - Web検索数
 - オンラインレビュー
- IoTデバイスデータ
- 医療、ヘルスケア

時系列データ

- Webアクセスログ

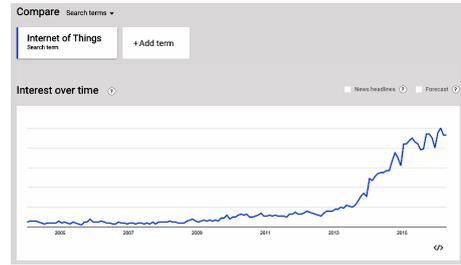


7

時系列データ

- Web検索数

from Google trends

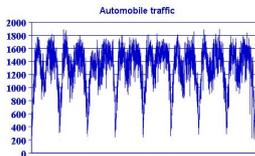


8

時系列データ

- IoTデバイスデータ

- インフラ (建物、橋梁など)
- 道路、交通
- 環境

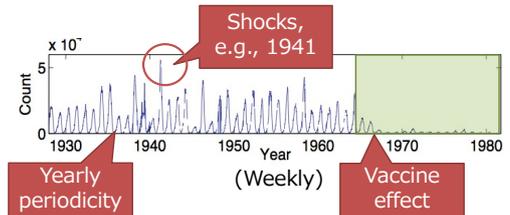


9

時系列データ

- 医療情報

例：アメリカ、はしかの患者数

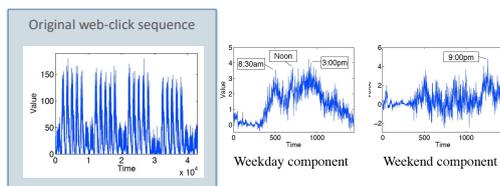


10

問題1

Given: 時系列データ (例、時刻毎のクリック数)

Find: パターンや周期性、圧縮

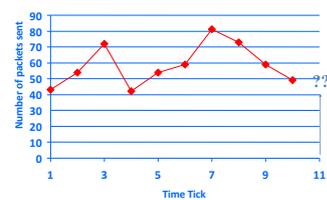


11

問題2

時系列予測

Given x_t, x_{t-1}, \dots , forecast x_{t+1}

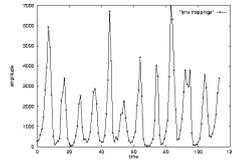


12

様々なタスク

パターン、外れ値、モデル推定、予測は密接に関連

- 予測のためには
 - パターン、ルール、モデルが必要
- 外れ値のためには
 - 予測が必要
 - (外れ値とは予測を大きく外れたもの)



13

講義内容

- 概要
- 特徴抽出
- 線形予測
- ストリーム処理とパターン発見

14

DFT

- シーケンス x_0, x_1, \dots, x_{n-1}
- Discrete Fourier Transform
- X_0, X_1, \dots, X_{n-1} :

$$X_f = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} x_t * \exp(-j2\pi tf / n) \quad f = 0, \dots, n-1$$

$$(j = \sqrt{-1})$$

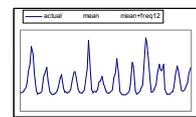
$$x_t = \frac{1}{\sqrt{n}} \sum_{f=0}^{n-1} X_f * \exp(+j2\pi tf / n) \quad \leftarrow \text{inverse DFT}$$

15

DFT: 振幅スペクトラム

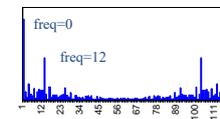
$$A_f^2 = \text{Re}^2(X_f) + \text{Im}^2(X_f)$$

count



year

Ampl.

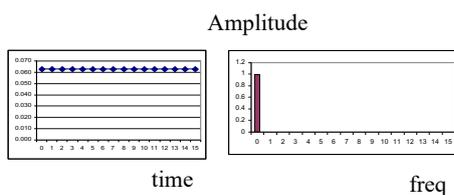


Freq.

16

DFT: 例

- Flat



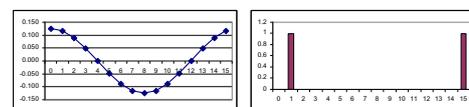
time

freq

17

DFT: 例

- Low frequency sinusoid



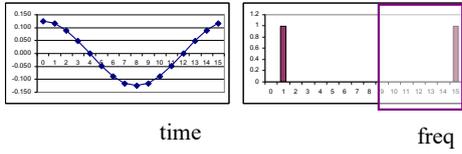
time

freq

18

DFT: 例

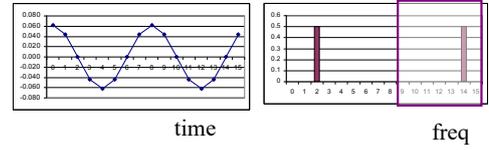
- Sinusoid - symmetry property: $X_f = X_{n-f}^*$
- 複素共役



19

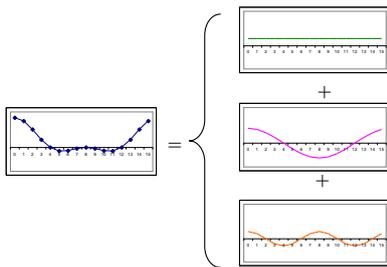
DFT: 例

- Higher frequency sinusoid



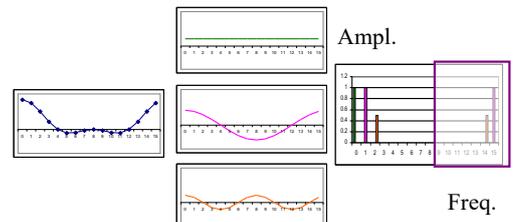
20

DFT: 例



21

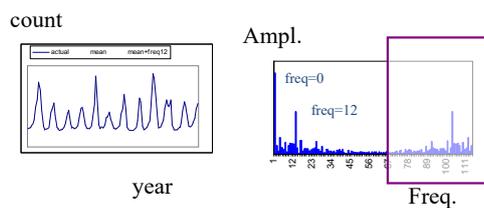
DFT: 例



22

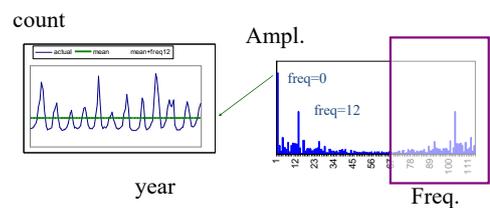
DFT: 振幅スペクトラム

Amplitude: $A_f^2 = \text{Re}^2(X_f) + \text{Im}^2(X_f)$



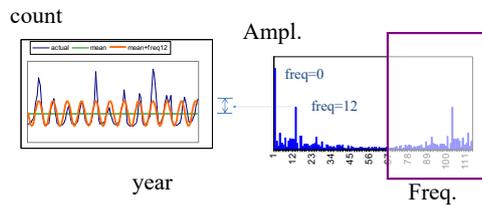
23

DFT: 振幅スペクトラム



24

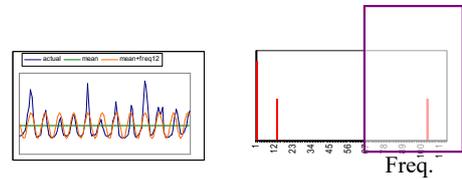
DFT: 振幅スペクトラム



25

DFT: 振幅スペクトラム

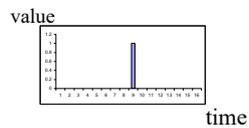
- 2つの周波数で優れた近似が可能
 - 圧縮
 - パターン発見



26

ウェーブレット DWT

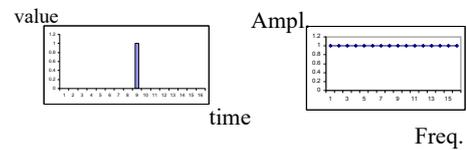
- DFTは有用
- しかし、以下のようなスパイクがある場合は？



27

ウェーブレット DWT

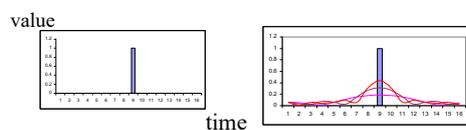
- DFTは有用
- しかし、以下のようなスパイクがある場合は？
- 全てのDFT係数が必要



28

ウェーブレット DWT

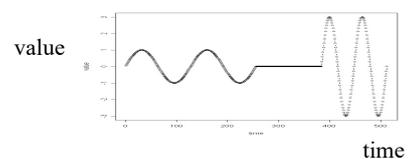
- DFTは有用
- しかし、以下のようなスパイクがある場合は？
- 全てのDFT係数が必要



29

ウェーブレット DWT

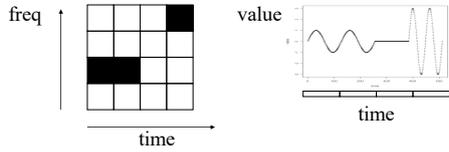
- 同様に、DFTは短い波の扱いも不得意
- 例: バリトン、静寂、ソプラノ



30

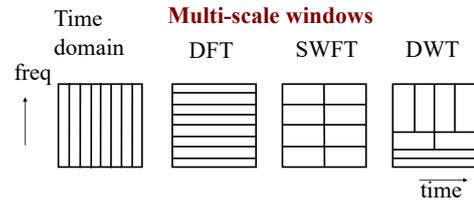
ウェーブレット DWT

- 解決策? : Short window Fourier transform (SWFT)
- しかし、どのようにウィンドウサイズを決めるのか?



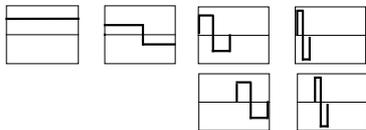
ウェーブレット DWT

- 解決策! : 複数のウィンドウサイズ
- DWT



Haar Wavelets

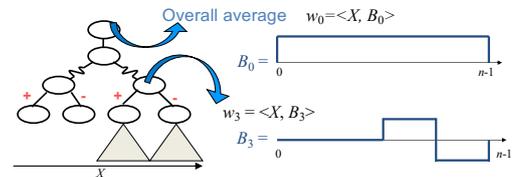
- 右半分の合計値から左半分の合計値を引く
- 再帰的に繰り返す、1/4、1/8、...



ウェーブレット変換

- 正規直交ウェーブレット基底
- ウェーブレット係数 w_i : シーケンス X と基底 B_i の内積

$$w_i = \langle X, B_i \rangle$$



ウェーブレット変換

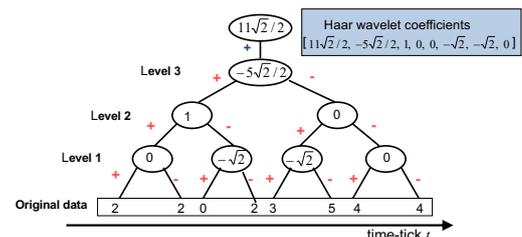
- シーケンスを階層的に表現
 - Haar wavelets: 最もシンプルな基底、簡単に実装
 - レベル毎に平均と差だけで表現
 - 小さい係数を無視することによってデータ圧縮

Level	Averages	Detail Coefficients
3	[2.75]	[-1.25]
2	[1.5, 4]	[0.5, 0]
1	[2, 1, 4, 4]	[0, -1, -1, 0]
0	[2, 2, 0, 2, 3, 5, 4, 4]	----

Haar wavelet coefficients: $[11\sqrt{2}/2, -5\sqrt{2}/2, 1, 0, 0, -\sqrt{2}, -\sqrt{2}, 0]$

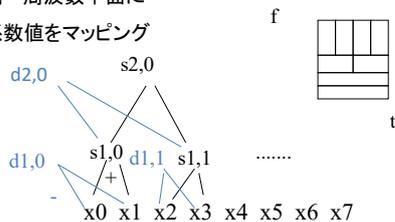
ウェーブレット変換

- シーケンスを階層的に表現



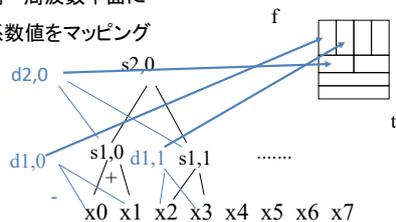
Wavelets – 計算方法

時間一周波数平面に
各係数値をマッピング



Wavelets – 計算方法

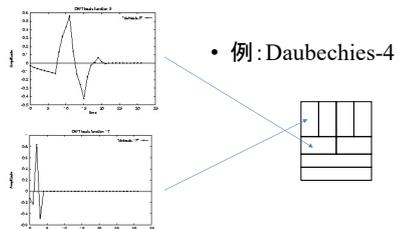
時間一周波数平面に
各係数値をマッピング



Wavelets – 計算方法

- 特徴1
 - '+' は重み付け加算
 - '-' は対応する重み付け差分(直交フィルタ)
- 特徴2
 - DFT/DCTと異なり数多くのウェーブレット基底
 - Haar, Daubechies-4, Daubechies-6, Coifman, Morlet, Gabor, ...

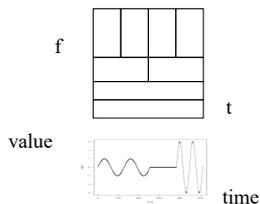
Wavelets – その他



- 例: Daubechies-4

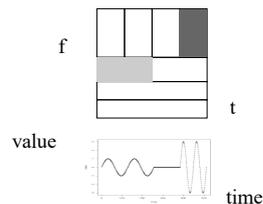
Wavelets – 例題1

- Q: baritone/silence/soprano - DWT?



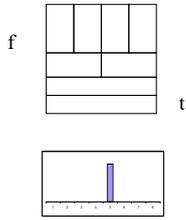
Wavelets – 例題1

- Q: baritone/silence/soprano - DWT?



Wavelets – 例題2

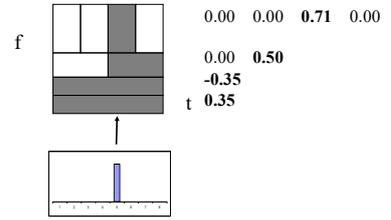
- Q: spike - DWT?



43

Wavelets – 例題2

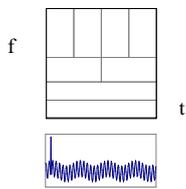
- Q: spike - DWT?



44

Wavelets – 例題3

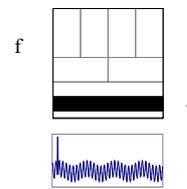
- Q: weekly + daily periodicity, + spike - DWT?



45

Wavelets – 例題3

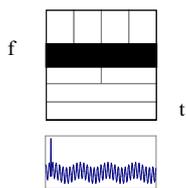
- Q: **weekly** + daily periodicity, + spike - DWT?



46

Wavelets – 例題3

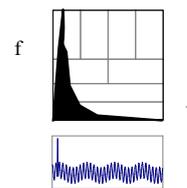
- Q: weekly + **daily** periodicity, + spike - DWT?



47

Wavelets – 例題3

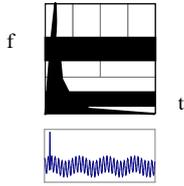
- Q: weekly + daily periodicity, + **spike** - DWT?



48

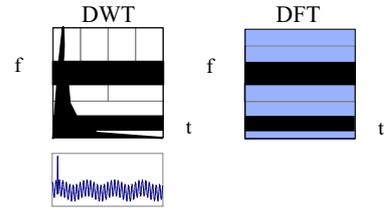
Wavelets – 例題3

- Q: weekly + daily periodicity, + spike - DWT?



Wavelets – 例題3

- Q: DFT?

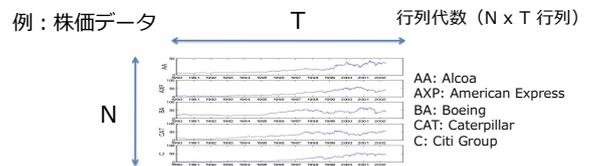


Waveletの優位性

- 圧縮効率の良さ(同じ係数の個数で少ない誤差、JPEG-2000で用いられている)
- 少ない計算コスト — $O(n)$
- スパイクを効率的に扱うことが可能

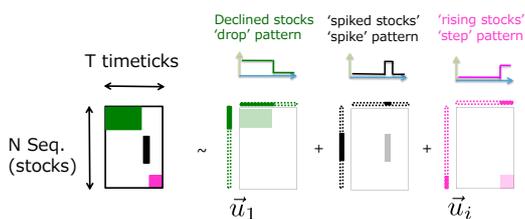
特異値分解 (SVD)

- 特異値分解 Singular Value Decomposition
 - 次元圧縮において最適 (ユークリッド距離基準において)
- Given: 数多くの時系列シーケンス
- Find: 潜在的な (隠れ) 変数



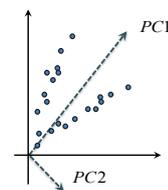
特異値分解 (SVD)

- 行列代数の解釈
- SVD -> 行列分解: ブロックを見つける



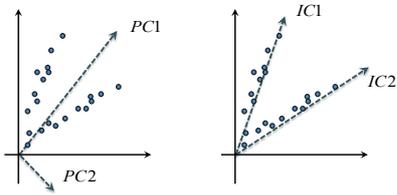
SVDとその改良

- Q: もっと上手く特徴をとらえたい?
- A: 独立成分分析 ICA
 - ICA: Independent Component Analysis



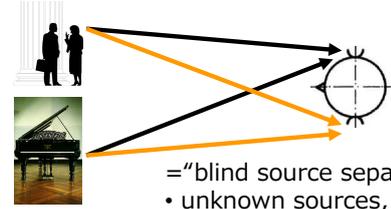
独立成分分析 ICA

- 主成分分析PCA (or SVD) は時々本質的な特徴を逃す
PCA vs. ICA



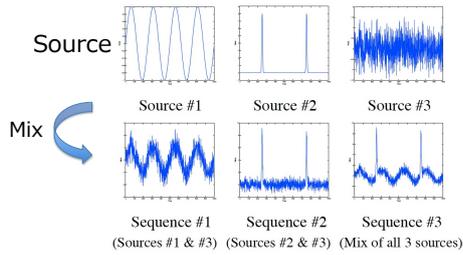
BSS : カクテルパーティ問題

- 二つの音源のもつれを解く



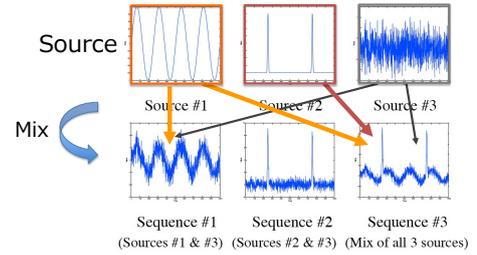
独立成分分析 ICA

- Why not PCA



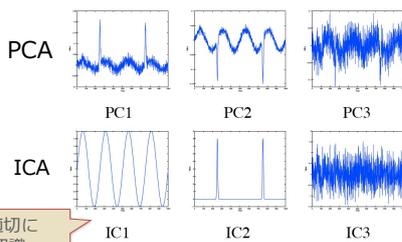
独立成分分析 ICA

- Why not PCA

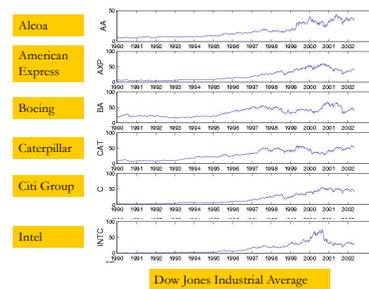


独立成分分析 ICA

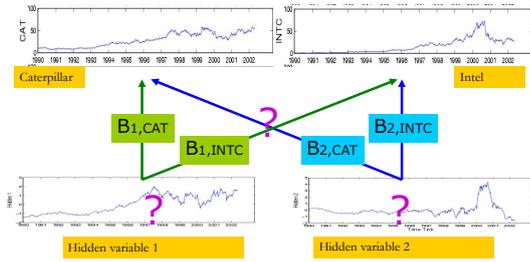
- Why not PCA



隠れ変数

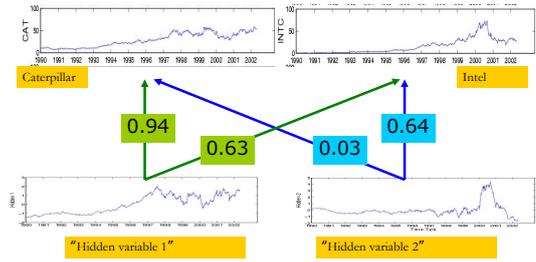


隠れ変数



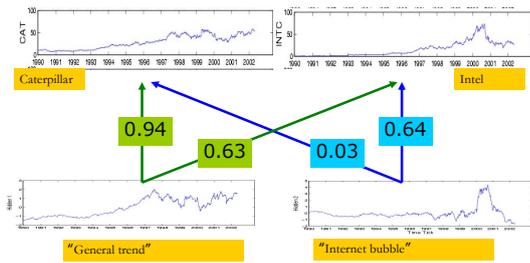
61

隠れ変数



62

隠れ変数



63

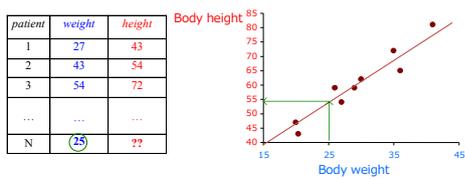
時系列予測

- 過去の時系列データを線形関数により表現
 - 過去の時系列ウィンドウ w 中のデータ
 - x_{t-2}, x_{t-2}, \dots
- 時系列データ x_t を表現

$$x_t \approx a_1 x_{t-1} + \dots + a_w x_{t-w} + noise$$

64

線形回帰



- 既知のデータに基づく線形関数を用いて (独立変数)
- 未知のデータを表現 (従属変数)

65

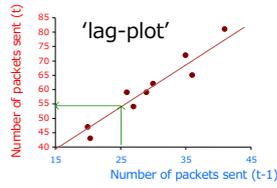
線形自己回帰

Time	Packets Sent(t)
1	43
2	54
3	72
...	...
N	??

66

線形自己回帰

Time	Packets Sent (t-1)	Packets Sent(t)
1	-	43
2	43	54
3	54	72
...
N	25	??

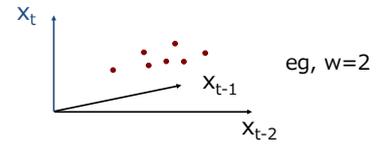


- lag $w=1$
- 従属変数 = # of packets sent ($S[t]$)
- 独立変数 = # of packets sent ($S[t-1]$)

67

線形自己回帰

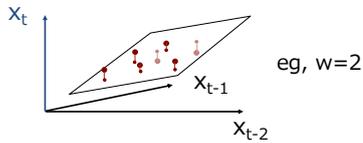
- 2以上の長さのウィンドウ $w>1$ の場合



68

線形自己回帰

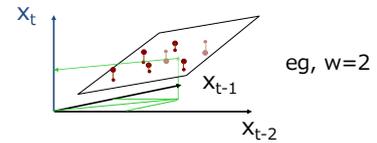
- 2以上の長さのウィンドウ $w>1$ の場合
- 超平面によってフィットさせる



69

線形自己回帰

- 2以上の長さのウィンドウ $w>1$ の場合
- 超平面によってフィットさせる



70

講義内容

- ✓ 概要
- ✓ 時系列解析の基礎
- リアルタイム解析技術
- 社会実装への取り組み



71

研究の理念

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ための技術開発

【対象データ】

- IoT (スマート工場、車両走行データ)
- 医療、Web

【応用】

- 製造業、交通、医療、Web、材料開発
- 産業界への技術移転：
「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」



72

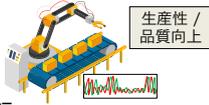
研究課題と目標



時系列ビッグデータ解析に基づく社会行動支援技術の開発

【要素技術】リアルタイムAI技術

- リアルタイム将来**イベント予測**
- 社会行動最適化のための**動的要因分析**



生産性 / 品質向上

【技術開発目標】産業貢献のための技術移転

- **スマート工場**、設計製造過程の高度化 (生産性、品質向上)
- **車両走行データ解析** (運転支援システム、自動運転)
- 医療・ヘルスケア分野におけるIoTデータ解析
- Web情報解析と市場分析



運転支援 / 事故回避

学術貢献



【トップ国際会議】最先端技術開発

- ACM **KDD**, IEEE ICDM, SIAM SDM (データマイニング)
 - **WWW**, ACM WSDM (Web)
 - ACM **SIGMOD**, VLDB, IEEE ICDE (データベース)
 - 難関トップ会議 (KDD, WWW, SIGMOD) の継続的な採択
- 【国際的な研究成果と受賞】
- KDD Best research paper awards (2008, 2010) 受賞
 - 3時間のチュートリアル講演 (SIGMOD2015, WWW2016, KDD2017、世界の研究コミュニティのリーダーとして講演)
 - 日本人初、日本人唯一の受賞、採択
 - IPSJ/ACM Award for Early Career Contributions to Global Research (2018)受賞



社会実装に向けた産学連携



トヨタ自動車 (2014~)、TTDC (2020~)

- 車両センサデータ解析、運転支援サービス

ソニー-SCK (2017~)

- スマート工場 (CMOSイメージセンサー)

三菱重工MHJET (2017~)

- スマート工場 (ターボチャージャー)

住友電工 (2021~)

- 送電線AI保守

SCREEN (2020~)

- AI装置制御システム

ローム (2020~)

- パワーデバイス制御

小松製作所 (2021~)

- 建設機械AI

トッパン・フォームズ (2021~)

- ヘルスケア

富士通研究所 (2016~)

- 人工知能ソフトウェア

電通デジタル (2019~)

- デジタル広告

自動運転・運転支援 スマート工場・設備管理 エッジデバイスAI その他AI



理念



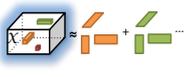
未来の予測によって社会を変革する

様々な企業・法人との共同研究、技術移転、産業・社会貢献
IoTビッグデータ解析の世界デファクトスタンダード
次世代のイノベーションのための若手スター研究者の育成

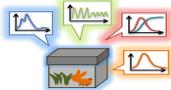


開発技術：新たなビッグデータ解析

テンソル解析



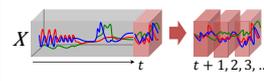
非線形モデリング



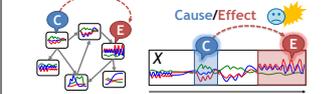
特徴自動抽出



リアルタイム予測



動的要因分析



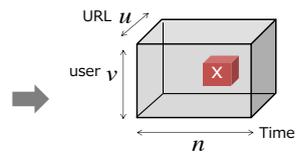
(R1) 大規模テンソル解析



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
...



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

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

Element x (# of events)
e.g., 'Smith', 'CNN.com', 'Aug 1, 10pm'; 21 times

TriMine @KDD 2012

大規模テンソル解析

Fast Mining and Forecasting of Complex Time-Stamped Events

Given
時刻付きイベントデータ
(例: Webアクセス履歴)

Goal
複数の観点からの潜在トピックの抽出、将来イベント予測

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
...

Original sequences (web-click events)

URL Actor Time

User

Drive

Media

Business

79

TriMine @KDD 2012

大規模テンソル解析

x を3つのトピックに分解

Object Actor Time

Webクリックデータ x

Object/URL

Actor/user

Time

Topic A (business)

Topic B (news)

Topic C (media)

e.g., business topic vectors

Object / URL

Actor / user

Time

Smith Johnson

Mon-Fri Sat-Sun

高い値: トピックとの強い関連性

90

TriMine @KDD 2012

大規模テンソル解析

時刻付きイベント

{time, URL, user ID, access devices, http referrer,...}

Timestamp	URL	User	Device
2012-08-01-12:00	CNN.com	Smith	iphone
2012-08-02-15:00	YouTube.com	Brown	iphone
2012-08-02-19:00	CNET.com	Smith	mac
2012-08-03-11:00	CNN.com	Johnson	ipad
...
2012-08-05-12:00	CNN.com	Smith	iphone
2012-08-05-19:00	CNET.com	Smith	iphone

Forecast

81

FUNNEL @KDD 2014

大規模テンソル解析

FUNNEL: Automatic Mining of Spatially Coevolving Epidemics

Original data

Time	Disease	Location	Cases
2021-04-01	measles	PA	4740
2021-04-01	measles	NY	5310
2021-04-02	rubella	CA	1923
...

Fitting results

Potential population of susceptibles (measles)

Vaccination

Seasonality strength vs. peak season

82

SpikeM @KDD 2012, TWEB 17

EcoWeb @WWW 2015

CompCube @WWW 2016

(R2) 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)
- 非線形社会現象の解析
 - ビッグデータの非線形解析
 - Web, IoTなど

83

SpikeM @KDD 2012, TWEB 2017

非線形モデリング

Rise and Fall Patterns of Information Diffusion: Model and Implications

"you can put lipstick on a pig"

Breaking news

News spread

Decay (power law)

"yes we can"

(per hour, 1 week)

84

SpikeM @KDD 2012, TWEB 2017 **非線形モデリング** **AIRC-ISIR** **ISIR**

Rise and Fall Patterns of Information Diffusion: Model and Implications

Power law function
 $f(n) = \beta * n^{-1.5}$

Linear scale $f(n)$ Log scale $f(n)$ -1.5

$\Delta B(n+1) = \frac{p(n+1)}{B(n)} \left[U(n) \cdot \sum_{t=n_0}^n (\Delta B(t) + S(t)) \cdot f(n+1-t) + \epsilon \right]$

Blogged **Periodicity**

$U(n+1) = U(n) - \Delta B(n+1)$

Un-informed

External shock (breaking news) S

Infection (word-of-mouth) β

Uninformed users

Time $n=0$ Time $n=n_b$ Time $n=n_b+1$

85

SpikeM @KDD 2012, TWEB 2017 **非線形モデリング** **AIRC-ISIR** **ISIR**

Rise and Fall Patterns of Information Diffusion: Model and Implications

(1) First spike (2) Release date (3) Two weeks before release

November 19, 2010 "Deathly Hallows part 1"

July 15, 2009 "Harry Potter and the Half-Blood Prince"

July 15, 2011 "Deathly Hallows part 2"

"harry potter"

Google search counts

Real-time modeling of information diffusion patterns **Forecast future spikes**

86

EcoWeb @WWW 2015 **非線形モデリング** **AIRC-ISIR** **ISIR**

The Web as a Jungle: Non-linear Dynamical Systems for Co-evolving Online Activities

1. Squirrel monkeys 2. Spider monkeys 3. Macaws 4. Capybaras

生物種

食料資源

ジャンルの生態系

Ecosystem in the Jungle

オンライン活動 (キーワード)

1. Xbox 2. PlayStation 3. Wii 4. Android

ユーザ資源

Web上の生態系

Ecosystem on the Web

Fitting result - RMSE=0.058817

PlayStation Wii Android Xbox

Modeling power of EcoWeb

Interactions between keywords

Seasonality

87

EcoWeb @WWW 2015 **非線形モデリング** **AIRC-ISIR** **ISIR**

生物種

食料資源

個体数

気候・季節

ジャンクル

キーワード

ユーザ資源

注目度

季節イベント (Xmasなど)

Web

88

EcoWeb @WWW 2015 **非線形モデリング** **AIRC-ISIR** **ISIR**

The Web as a Jungle !

Fitting result - RMSE=0.058817

PlayStation Wii Android Xbox

Model estimation

Interactions between keywords

Seasonality

89

CompCube @WWW 2016 **非線形テンソル解析** **AIRC-ISIR** **ISIR**

Non-Linear Mining of Competing Local Activities

e.g., Google search volumes for Kindle, Nexus (for 236 countries, from 2004 to 2015)

Global US CA IT AU ZA BR CN JP

Location m

Time-stamped data {activity, location, time}

Keyword d

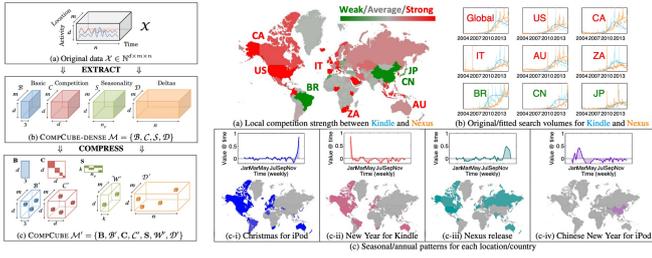
Time (weekly) n

M^{th} order tensor ($M=3$)
 $\mathcal{X} \in \mathbb{N}^{d \times m \times n}$

90

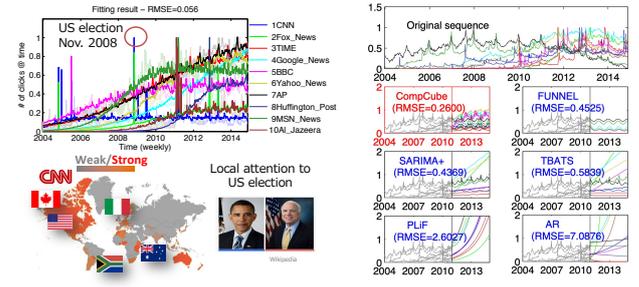
非線形テンソル解析

Non-Linear Mining of Competing Local Activities



非線形テンソル解析

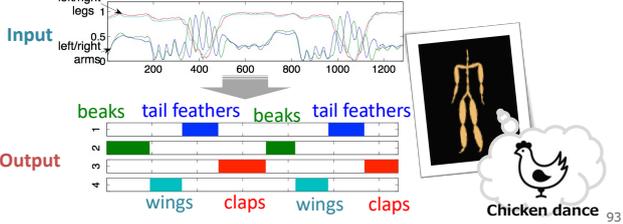
Non-Linear Mining of Competing Local Activities



(R3) 特徴自動抽出

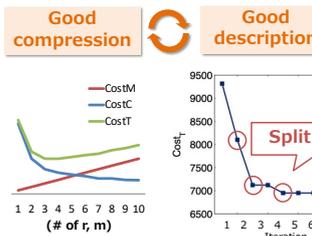
ビッグデータから「完全自動」で重要な特徴を抽出

- 技術者・専門家のチューニング・パラメータ調整が不要
- 「人の手を介さない」ビッグデータの学習手法



特徴自動抽出

$$\min (\text{Model cost } \text{Cost}_M(M) + \text{Coding cost } \text{Cost}_C(X|M))$$



特徴自動抽出

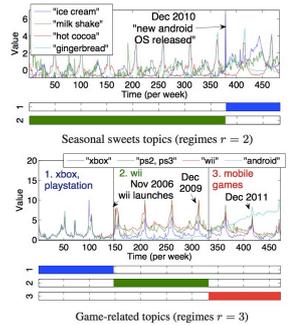
AutoPlait at work

(1) Turning point detection

- seasonal sweets
- Trend suddenly changed in 2010 i.e., release of android OS ("Ginger bread", "Ice Cream Sandwich")

(2) Trend discovery

- game-related topics
- It discovers 3 phases of "game console war" (Xbox&PlayStation/Wii/Mobile social games)

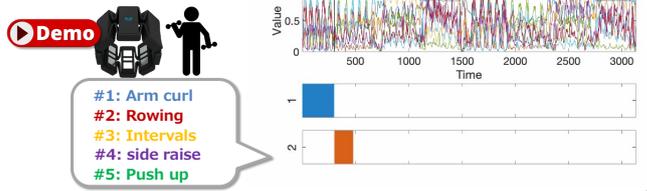


特徴自動抽出

ビッグデータから「完全自動」で重要な特徴を抽出

- 技術者・専門家のチューニング・パラメータ調整が不要
- 「人の手を介さない」ビッグデータの学習手法

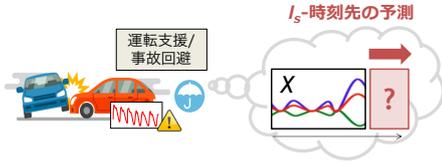
"Real-time" time-series mining



(R4) リアルタイム予測

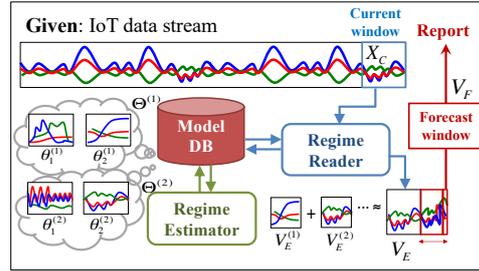
リアルタイム、継続的、かつ適応力のある学習・予測

- IoTデータストリーム (スマート工場、車両走行など)
- 時系列パターンとトレンドを継続的に把握・モデリング
- 突発的な変化にもリアルタイムに対応、予測



リアルタイム予測

RegimeCast: システム概要

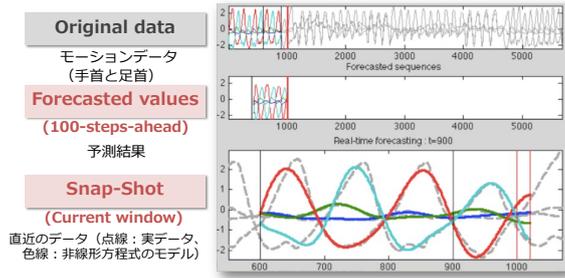


Point 1:
データストリームから学習した様々な非線形方程式のモデルをDBに格納

Point 2:
現在の時系列パターンに合うモデルをDBから探索し予測

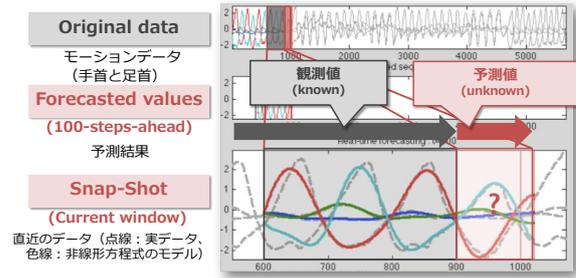
リアルタイム予測

デモ動画 (RegimeCast)



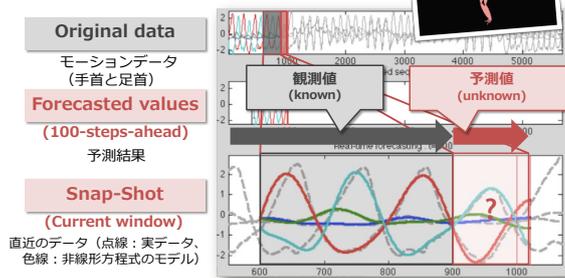
リアルタイム予測

デモ動画 (RegimeCast)



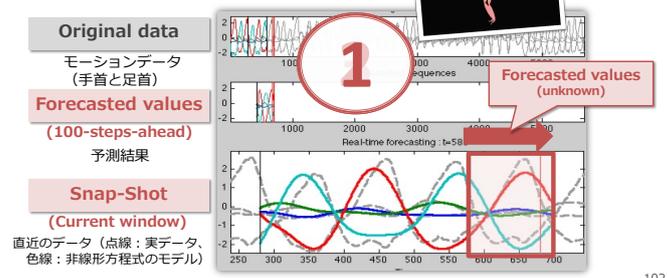
リアルタイム予測

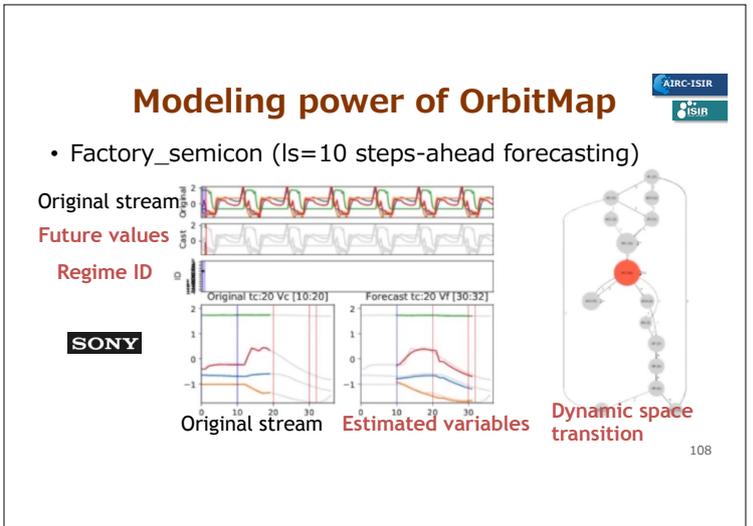
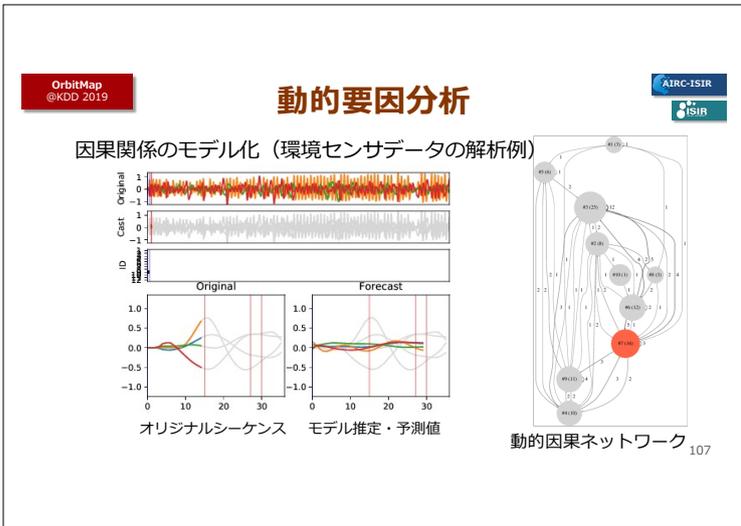
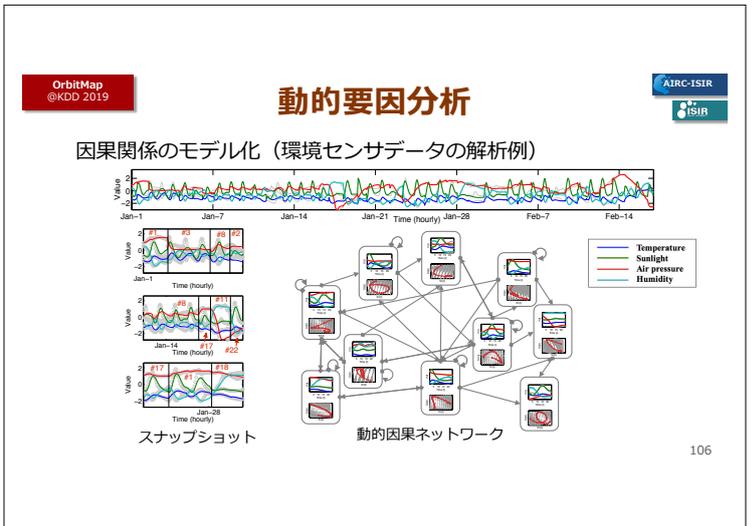
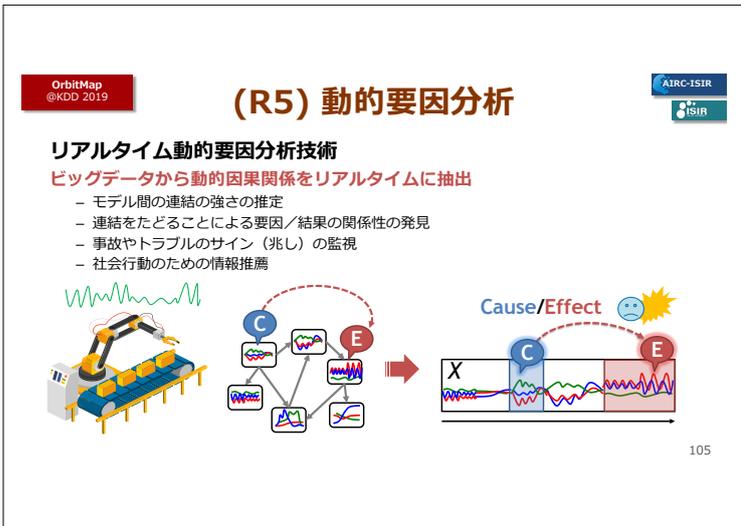
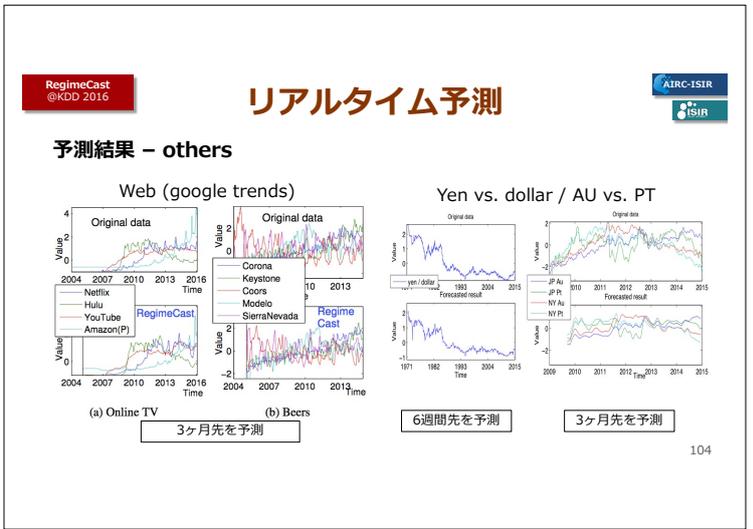
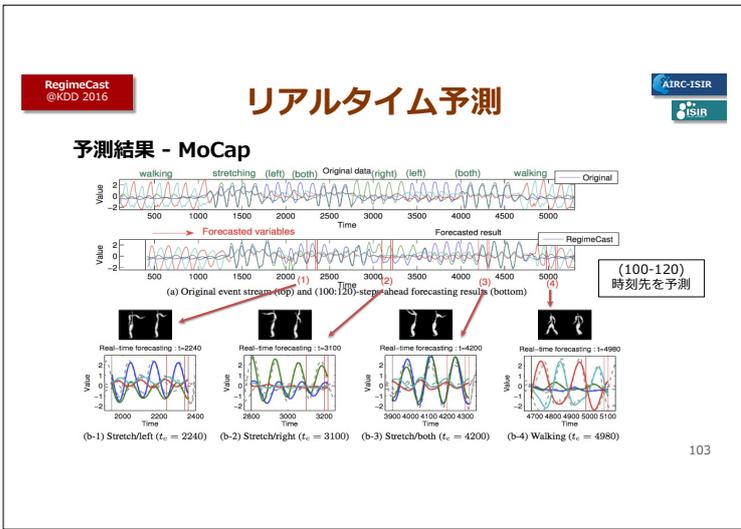
デモ動画 (RegimeCast)



リアルタイム予測

デモ動画 (RegimeCast)





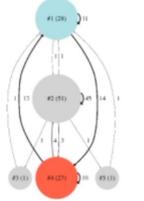
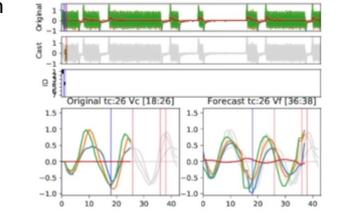
OrbitMap at work

- Factory_engine (ls=10 steps-ahead forecasting)

Original stream

Future values

Regime ID



Original stream Estimated variables Dynamic space transition 109

技術の優位性

- 関連技術
 - 情報推薦 (例: 協調フィルタリング)
 - 時間発展を考慮していない
 - 刻々と変化する状況に対応するような情報推薦は難しい
 - 予測、時系列解析 (例: カルマンフィルター、統計分野の SARIMA や TBATS など)
 - 高い精度の予測をリアルタイム処理で行うことは難しい
 - 深層学習
 - 解析結果の要因や判断基準が利用者には分からず、ブラックボックスになっている
 - 時系列解析に弱い (LSTM や GRU など、深層学習を用いた時系列解析の仕組みもあるが、予測精度が低い)
 - 大量の学習データを必要とする
 - 学習に多くの時間を要する
 - メモリ使用量が大い

VS

技術の優位性

- 本研究の技術的な強み
 - リアルタイム予測
 - 世界の全手法、全技術の中で最高の予測精度
 - 最小の計算コスト
 - 深層学習と比べ10万倍の高速化、10倍の高精度化
 - 省メモリ
 - ダイナミック要因分析
 - 世界における唯一の技術
 - 時系列解析における新しい取り組み
 - 利用者が直感的に理解できる予測/解析内容
 - スマート工場、エッジなどIoT関連産業に有効な技術

報道記事

OrbitMap @KDD 2019

2019年11月25日
 日刊工業新聞 2019/11/25 05:00配信
 「自然・社会現象変化、深層学習の10万倍速で予測 阪大がAI」
<https://www.nikkan.co.jp/articles/view/00539314>

日刊工業新聞 2019年11月25日17面
 「現象変化を即時予測 阪大がAI技術・深層学習の10万倍速」
 産業科学AIセンター
 櫻井保志、松原靖子

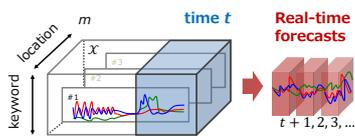


テンソルデータのリアルタイム予測

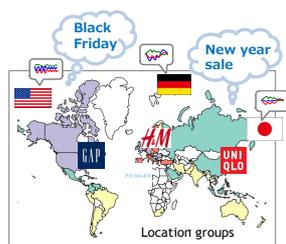
CubeCast @KDD 2020

テンソルデータストリームの非線形解析

例: Webにおけるキーワード検索回数データ



トレンドの変化、季節性、地域の局所パターン
 リアルタイム学習とリアルタイム予測

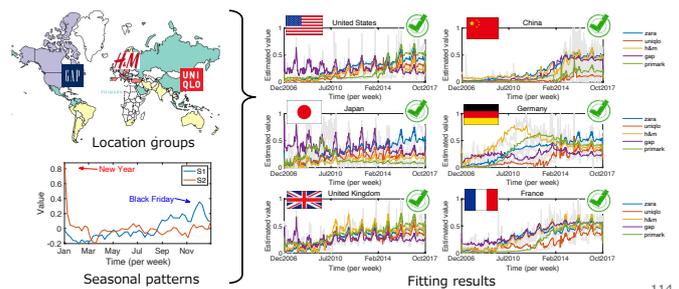


非線形モデリングと周期性の取得

テンソルデータのリアルタイム予測

CubeCast @KDD 2020

テンソルデータストリームの非線形解析



リアルタイムAI技術の特徴

【技術の概要】

- 非線形微分方程式に基づくリアルタイム学習技術を応用し、一つ一つの対象（部品、人々等）ごとに、来歴情報に合わせた予測モデルを、リアルタイムにパッチワーク的に合成し続ける
- 個性が高く、変化のパターンが突然大きく切り替わるような複雑な物理現象、そこから発生する事象（部品・デバイス故障等）にも瞬時に追従し、その未来を高い精度で予測することを可能とする

【技術の特長】

- ① 非定型データ群
- ② エッジAI化（クロスデータ処理）
- ③ 省電力データ処理
- ④ 高速データ処理



リアルタイムAI技術の特徴

	本研究（リアルタイムAI）の技術	既存手法（深層学習）の技術
予測モデル	多数の非線形方程式からリアルタイムに予測モデルを生成	ビッグデータから固定的な予測モデルを生成
利用シーン例	各々が個別の予測モデルを持つ	全員が単一の予測モデルを共有する
モデルの説明性	高い	低い
モデル生成のコスト	計算コスト：小 メモリ消費量：小	計算コスト：大 メモリ消費量：大
モデル修正のコスト	修正が容易 → 予測モデルを常時更新（秒1回以上）	修正が困難 → 予測モデルを間欠的に更新（定期更新）
適用対象	環境変化や外的要因の影響が大きい 個別的で不安定な現象 <small>(例) 環境に応じて動作や故障パターンが変化するシステムの新規および継続的学習、風入風の強い風車・ヘルメット等</small>	環境変化や外的要因の影響が少ない 普遍的で安定な現象 <small>(例) 識別性が少ないシステムや現象の認識および予測、画像の翻訳、画像の認識等</small>

日常における変動要素が大きなモノづくり現場
 低計算コストによる日常運用
 変動予測のフィードバックをすぐに行うべき場合
 新製品開発における最適化
 製品の大規模な性能向上、機能付与
 新材料、プロセス探索
 AIの「適材適所」活用

講義内容

- ✓ 概要
- ✓ 時系列解析の基礎
- ✓ リアルタイム解析技術
- 社会実装への取り組み



産業界への技術移転

第4次産業革命のためのAI・ビッグデータ解析

- デジタルツイン・DX（製造業）
- コネクティッドカー・サービス（自動車）など



【課題】従来とは異なる新たなAIテクノロジーの開発が急務

- 深層学習：技術的な問題点、限定された使用用途
 - ・ 使用用途：文書解析や対話サービス（自然言語処理）、不良品検知（画像処理）
 - ・ 問題点・課題：膨大な学習データと計算時間、ブラックボックス、低い予測精度
 - ・ 複雑現象の学習に不向き（個性が高く突然パターンが変化する非定常性の高いデータの学習が困難）

生産工程における設備稼働データを統合的、高速に解析し、製造工程全体を最適化、リアルタイムに情報提供するようなAIソフトウェアが必要

産業界への技術移転

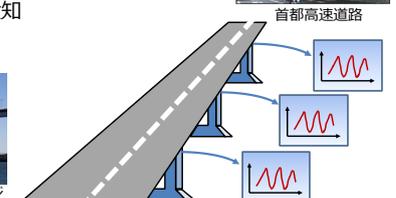
リアルタイムAI技術の特徴を生かした未来のモノづくり社会の実現

大阪大学・産業科学研究所・産業科学AIセンター
 連携企業：ソニー、三菱重工、小松製作所、住友電工、SCREEN、トッパン・フォームズ、富士通研、トヨタ自動車、トヨタTTDC

複雑事象のモデリングによる知的支援システムの産業IoTへの展開

産業界への技術移転

- 橋梁モニタリングと異常検知
 - 大量の振動センサーデータを解析
 - リアルタイムに橋梁の状態を監視
 - 災害時の異常検知
 - 早期異常把握



産業界への技術移転



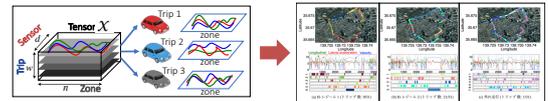
- 橋梁モニタリングと異常検知
 - 道路橋に振動を計測するセンサを設置
 - 数百のセンサから毎秒数万のセンサデータ
 - リアルタイムかつ長期継続的に橋梁の状態を監視
 - 災害時のリアルタイム異常検知、早期異常把握



- 橋梁の異常監視
- 振動、衝撃計測
- 構造物異常監視システム
- 首都高速道路・三軒茶屋高架橋など4箇所以上に導入
- 東京ゲートブリッジに導入

産業界への技術移転

- トヨタ自動車との共同研究
 - トップ国際会議ACM SIGMOD2014、WWW2015、WWW2016の研究成果を実用化、車両走行パターンの自動検出に成功
 - 車両走行センサデータに対し走行パターンを表現する要約情報を自動抽出



(a) 標準的な走行
(複数車線道路)



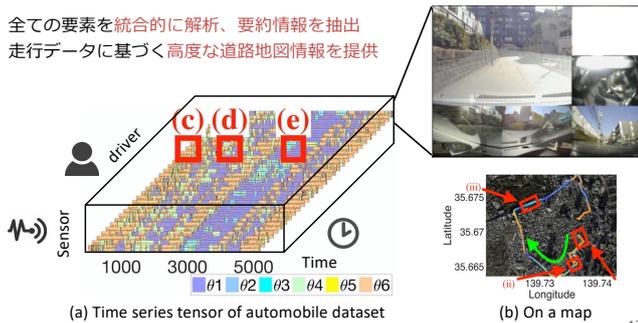
(b) 慢重な走行
(路地、横断歩道)



(c) 複雑な走行
(交差点)

時系列テンソル解析技術

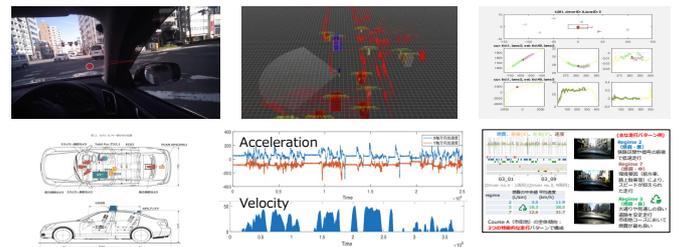
- 全ての要素を統合的に解析、要約情報を抽出
- 走行データに基づく高度な道路地図情報を提供



IoTビッグデータを用いた自動運転技術

高度な自動運転・運転支援技術（運転快適性の向上、省エネルギー）

- IoTビッグデータ：車両走行・周辺車・位置情報等、燃費・電力消費情報、運転者評価データ



自律飛行への応用イメージ

- オンデマンド型自律飛行ドローンのスマート工場への展開
 - 製造システムの指示によって動作する工場内ドローン
 - 3Dプリンタとの連携



リアルタイムAI技術の製造業への導入

製造業への最先端AI技術の導入、その重要性

技術導入の社会的インパクト

- 日本のGDP（533兆円）に占める製造業比率は22%
- 日本経済を支える大きな産業

製造業の国内回帰

- 新興国の人件費上昇、サプライチェーンの見直し
- 最先端AI技術の導入によるさらなる生産性向上が必須
- 製品の高付加価値化につながるAI技術

先端技術導入のニーズ

- 多くの製造業企業は高度なAI技術の活用に積極的（DX、デジタルツイン）
- 製造業では世界シェアが高い高収益企業ほどAI技術導入に積極的
- 企業の生声：企業独自の強みとしてHWとSWの総合技術を開発、ノウハウとして蓄積したい
- 革新的AI生産プラットフォームの開発と完全自律製造の実現



リアルタイムAI技術の製造業への導入

製造業DXのためのリアルタイムAI技術

- スマート工場：設備から稼働状態などの情報をリアルタイムに収集、ビッグデータを学習
- 複雑事象の把握、リアルタイム学習、行動・オペレーションの最適化
- リアルタイム情報提供のためのAI技術基盤

基礎技術の開発と産業への応用

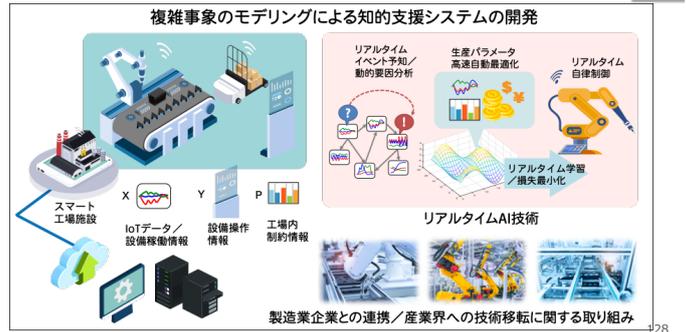
- 予測、要因分析、トラブル予知、行動最適化のためのAI 技術を開発
- 産学連携と異分野融合により産業界における社会実装



127

リアルタイムAI技術の製造業への導入

AIRC-ISIR
ISIR



128

製造業IoT・設備管理向けAI

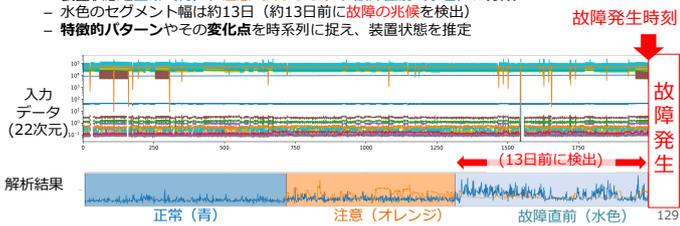
AIRC-ISIR
ISIR

ターボ分子ポンプの故障発生予測

- ・ 下図上段：オリジナルの入力情報
- 設備管理および稼働データ

・ 下図下段：解析結果

- 装置状態を正常(青)、注意(オレンジ)、故障直前(水色)に分類
- 水色のセグメント幅は約13日(約13日前に故障の兆候を検出)
- 特徴的パターンやその変化点を時系列に捉え、装置状態を推定



129

製造業IoT・設備管理向けAI

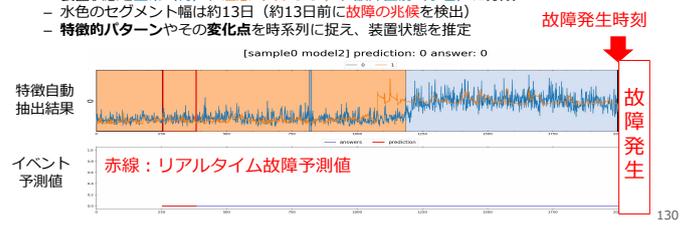
AIRC-ISIR
ISIR

ターボ分子ポンプの故障発生予測

- ・ 下図上段：オリジナルの入力情報
- 設備管理および稼働データ

・ 下図下段：解析結果

- 装置状態を正常(青)、注意(オレンジ)、故障直前(水色)に分類
- 水色のセグメント幅は約13日(約13日前に故障の兆候を検出)
- 特徴的パターンやその変化点を時系列に捉え、装置状態を推定



130

製造業IoT・設備管理向けAI

AIRC-ISIR
ISIR

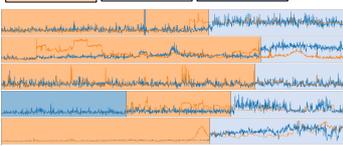
設備	regime	StartDate	EndDate
#1	0	2020-01-30 14:38:56	2020-01-30 15:41:04
	1	2020-02-05 10:03:31	2020-02-15 01:46:28
	2	2020-01-23 03:39:22	2020-02-05 09:58:35
#2	1	2020-06-14 10:21:28	2020-06-22 15:18:01
	2	2020-05-10 14:29:12	2020-06-14 09:55:35
#3	1	2020-08-31 18:17:57	2020-09-15 05:59:22
	2	2020-07-28 05:55:25	2020-08-31 17:51:16
#4	0	2021-02-06 06:25:24	2021-02-14 17:13:49
	1	2021-02-24 09:14:52	2021-03-07 15:42:19
#5	1	2021-02-14 17:22:33	2021-06-05 09:06:05
	2	2021-05-17 18:00:45	2021-06-07 14:47:04
#5	1	2021-05-06 06:46:12	2021-05-27 17:39:32
	2		

【予備実験】

- 5/5件で故障兆候の事前検知に成功
- ・ 8日~15日前に故障発生を予測
- ・ Regime #1: 故障の予兆

本技術の事業導入に向けた取り組みを開始

Regime #0 Regime #1 Regime #2



131

設備管理向けAI (電力設備保全)

AIRC-ISIR
ISIR

住友電気工業との共同研究

- ・ 都内大規模停電
 - 2016年10月12日14時49分
 - 埼玉県新座市内の送電設備の火災
 - 最大36.7万戸、のべ約58.6万戸の停電が発生
 - 完全復旧に1時間
 - 経年と地震によりケーブルの老朽化、絶縁破壊が発生、銅管破裂により着火、火災発生
- ・ 保守監視の高度化が急務(電力業界の緊急課題)
 - 保守人員の減少に直面
 - 様々な場所に敷設している送電線のデータを全て送ることは困難



エッジでの保守監視AIモジュールの開発
ケーブルの絶縁破壊を監視、寿命を予測、改修の優先度を推定
安定供給と国民負担抑制のためのケーブル改修工事の最適化

132

エッジデバイスAI



小型デバイス、組み込み機器のためのリアルタイム学習機構

- AI実装による小型デバイスや組み込み機器の高度化
 - 小松製作所：建設機械における異常や部品故障の予測
 - 住友電工：送電線設備における自律AI監視
 - SCREEN：AI装置制御システム
 - ローム：パワーデバイス制御
 - トッパン・フォームズ：バイタルモニタリングによる女性（妊婦）向け健康管理、各種イベント予測



【実用化のための課題】

- モデル学習の高速化、省メモリ化
 - 組み込み機器や小型デバイスでは計算環境（メモリ量やCPUパワーなど）に制約
 - モデルは固定ではなく、変化させていかなければならない
 - 個体差、使用環境、顧客先での使われ方、経年変化によって傾向変化



【開発技術の強み】

- メモリ量やCPUパワーに制約がある中でもリアルタイムかつ継続的にモデル学習
- 状況に合わせて、デバイス、機器の内部でモデル更新/追記学習

小型デバイスに実装可能な高速・軽量リアルタイムAIソフトウェアを開発

- Raspberry PiにエッジAIソフトウェアとして搭載、連携企業とともにFPGAにも実装

133

若手研究者の育成



産学連携と国際連携による人材育成

– 基礎研究と産学連携

- 若手研究者が企業の開発現場から学ぶ
- 新しい研究課題を創造、インパクトのある課題に挑戦
- 社会の技術革新をリードする研究者の育成
- 企業からの社会人ドクターの受け入れ

– 国際連携

- カーネギーメロン大学への海外派遣プログラム
- 若手スター研究者の育成



Christos Faloutsos

134

若手研究者の育成



IPSJ/ACM Award for Early Career Contributions to Global Research

- 国際間の共同研究で成果をあげた若手研究者を表彰
- IPSJ とACMとの共同表彰、ACM・チューリング賞授賞式への参加
- 松原靖子准教授が第1回目の受賞者（2018年）



チューリング賞授賞式



情報処理学会 西尾章治郎会長
(阪大・総長)



ACM Vicki L. Hanson会長

135

Acknowledgements



- JST未来社会創造事業「複雑事象のモデリングによる知的支援システムの開発」（研究代表者：櫻井保志）[2019-2022]
- 文部科学省科研費基盤（A）「ビッグデータからの材料特性の高速モデル学習と最適化」（研究代表者：櫻井保志）[2020-2023]
- NICT Beyond 5G研究開発促進事業「Beyond5Gの高速通信・低遅延等に適したエッジAIソフトウェアの開発と動作実証に関する研究開発」（研究代表者：櫻井保志）[2021-2023]
- 総務省（SCOPE）ICT基礎・育成型研究開発「複合ビッグデータストリームの動的空間モデリングと最適化に関する研究」（研究代表者：松原靖子）[2019-2022]
- 環境研究総合推進費（ERCA）革新型研究開発「リアルタイムAI技術に基づく省エネルギー化に資する高度自動運転支援技術に関する研究開発」（研究代表者：松原靖子）[2020-2023]
- JST AIP加速課題「リアルタイム将来予測に基づく自律型オペレーション最適化に関する研究開発」（研究代表者：松原靖子）[2021-2024]
- 文部科学省科研費基盤（B）「時系列ビッグデータストリームの複合モデリングに関する研究」（研究代表者：松原靖子）[2021-2024]
- トヨタ自動車（2014～）、富士通研究所（2016～）、ソニーSCK（2017～）、三菱重工MHJET（2017～）、三菱重工MAT（2017～）、電通デジタル（2019～）、SCREEN（2020～）、TTDC（2020～）、ローム（2020～）
- 小松製作所（2021～）、住友電工（2021～）、トッパン・フォームズ（2021～）

136