

大澤研究室

チャンス発見とデータ駆動イノベーションを実現する「データ市場」の創成

参照先・連絡先

Web: <http://www.panda.sys.t.u-tokyo.ac.jp>

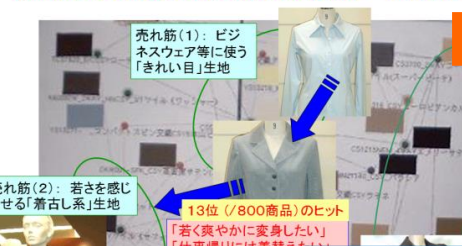
面談希望など: 大澤および秘書へ直通メール

info@panda.sys.t.u-tokyo.ac.jp

- 原理1 革新的モデルとデータ可視化により、微動の裏に潜む巨動を表出化
- 原理2 「気づき×表現」「会話×推論」により、実現性のある創造活動を促進
- 原理3 データ市場: 公開を強くないデータ駆動イノベーションの場を実現



戦略的シナリオ表出化を支援するためのデータ可視化



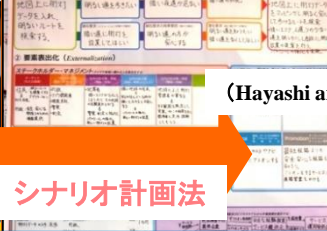
(大澤・白井2003)

データ駆動イノベーションとデータ駆動



(大澤研・2013~)

Action Planning



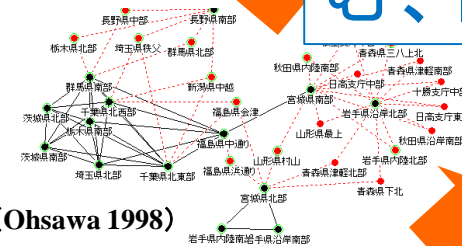
(Hayashi and Ohsawa 2016)



「発見」する筆跡

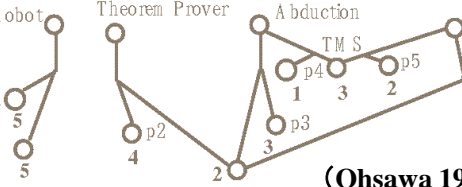
データ駆動ビジネスの全体設計に取り込む、世界に稀な研究室です。なぜなら、

基盤1: データ可視化

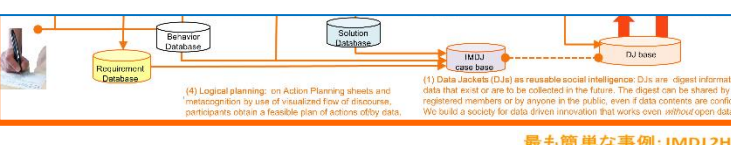


(Ohsawa 1998)

基盤2: 非単調推論

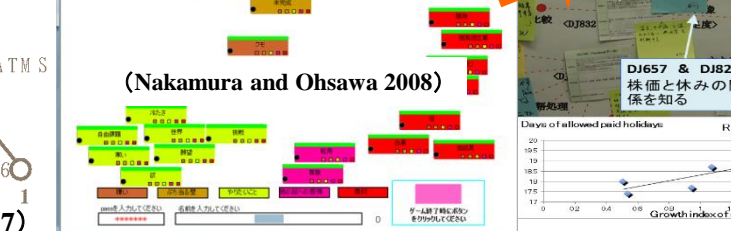


(Ohsawa 1997)



最も簡単な事例: IMDJ 2H, AP 1H, そして分析を1時間だけ

基盤3: 創造性の科学



(Nakamura and Ohsawa 2008)



DJ675: 各業界における株価

DJ825: 各業界における有給休暇の取得数

DJ657 & DJ825: 株価と休みの関係を知る

(DJ研究会・2016)



株価変動予兆の発見

(Ohsawa 2015)



地震予兆発見 (忙しくて未発表)

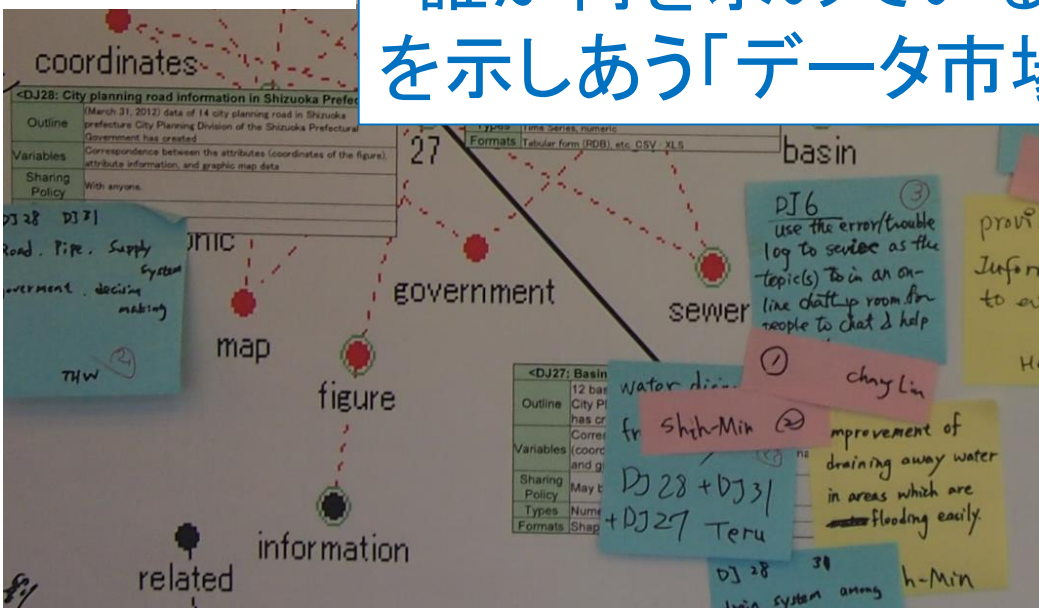
Innovators Marketplace on Data Jackets

(データジャケット市場ゲーム: Ohsawa, Kido, Hayashi, Liu, Data Jackets for Synthesizing Values in the Market of Data, *Procedia Computer Science* 22, 709-716 (2013))



- ・誰がどんなデータを持っているか
 - ・誰が何を求めているか
- を示しあう「データ市場」のデザインにより

(名古屋工大で)



(IEEE SPIN2015, 2016での実施, India)

(1) データ利活用者における着想を支援する手法の研究

潜在的な意識 (例: 患者の容態急変が怖い)

IMDJの入力情報と出力情報

IMDJ

提示される要求 (例: 「健康変化の未知の要因を知りたい」)

利用者
→提案者:
要求と評価

Requirement. Tell me the influence of weather on health!

中身を「公開」できないデータまで合意の上で「共有」し、

提案者→
利用者:

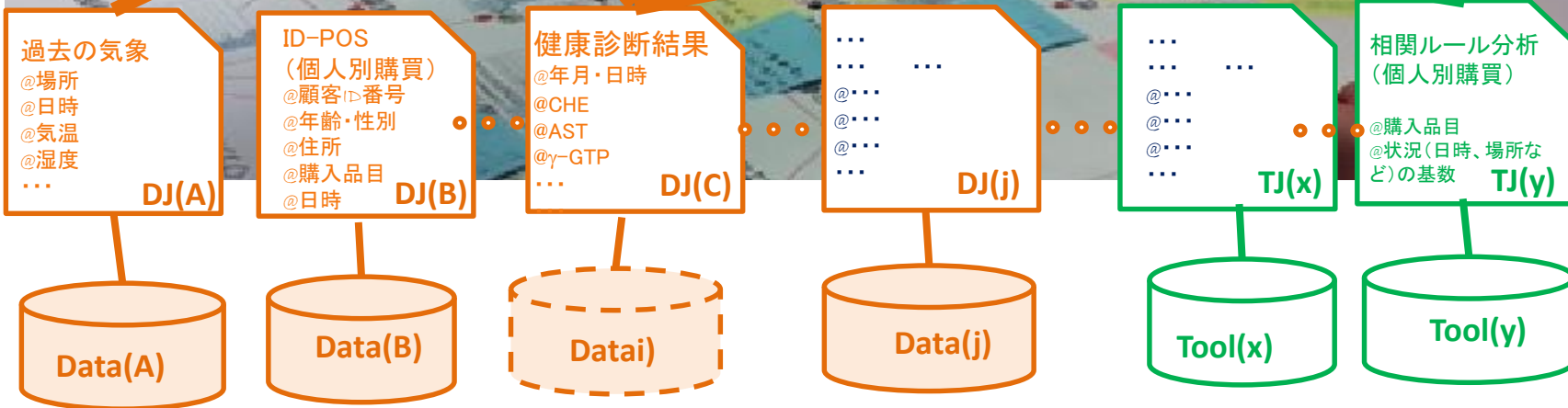
データ結合・
利用の方法を
提案(例: 提案1, 2)

Proposed plan: Combine data behind DJ(A), DJ(B), and DJ(C) to learn rules such as
(1) $temperature(t) > 25c \rightarrow consumption_beer(t) > 1L$
(2) $consumption_beer(t) > 0.7L \rightarrow fat(t) > \alpha \rightarrow \gamma\text{-GTP } \beta.$

Proposed plan: Analyze the data behind DJ(C) to grasp the latent condition of health.

DJ結合

DJ登録



(1) データ利活用者における着想を支援する手法の研究

DJ675: 各業界における株価

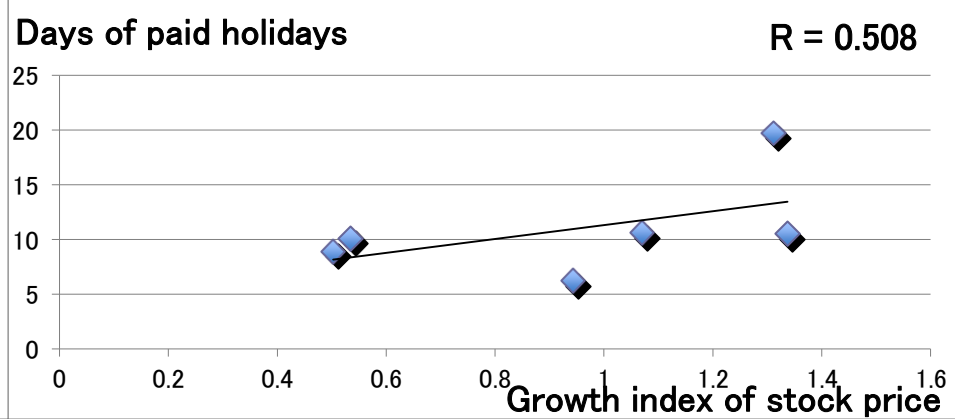
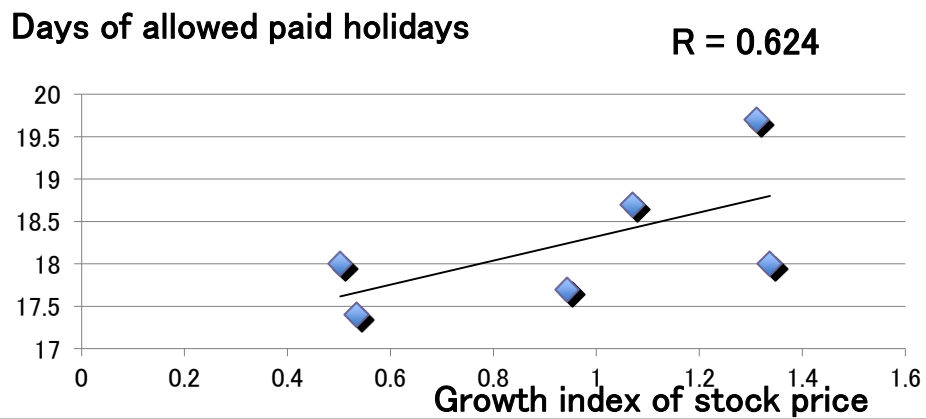
簡単な事例: IMDJ 2H, AP 1H, そして分析を1時間だけ

DJ825: 各業界における有給休暇の取得数

社会的要求にフォーカスを絞ってデータ分析する手法、

DJ657 & DJ825: 株価と休みの関係を知る

(IMDJワークショップから)



(1) データ利活用者における着想を支援する手法の研究

MDJの精緻化と思考過程データマイニング

Action Planning チーム番号: _____

① 要求分析 (Requirement Analysis)

② 要素表出化 (Externalization)

③ 要素系列化 (Serialization)

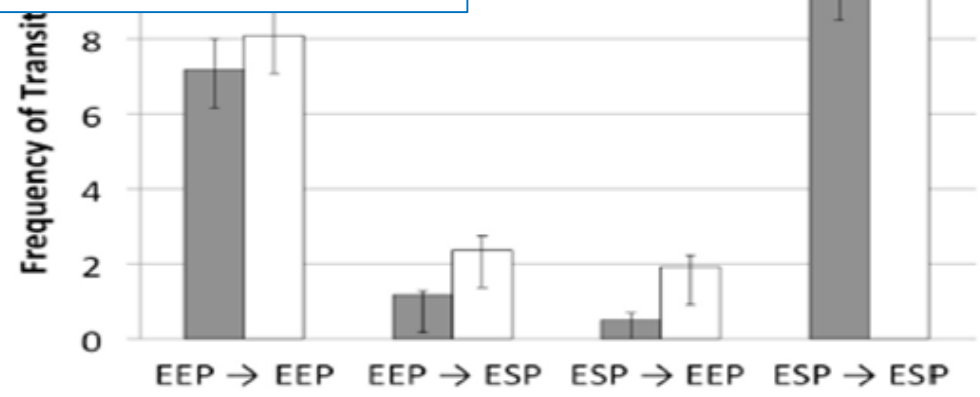


14

その結果に基づくビジネスシナリオ策定技術、



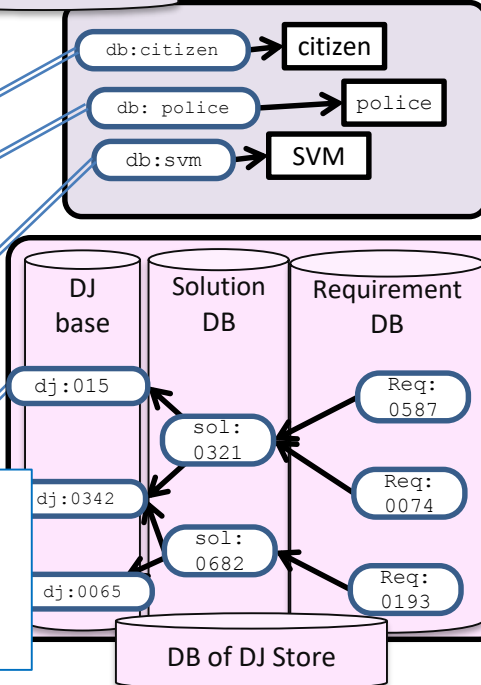
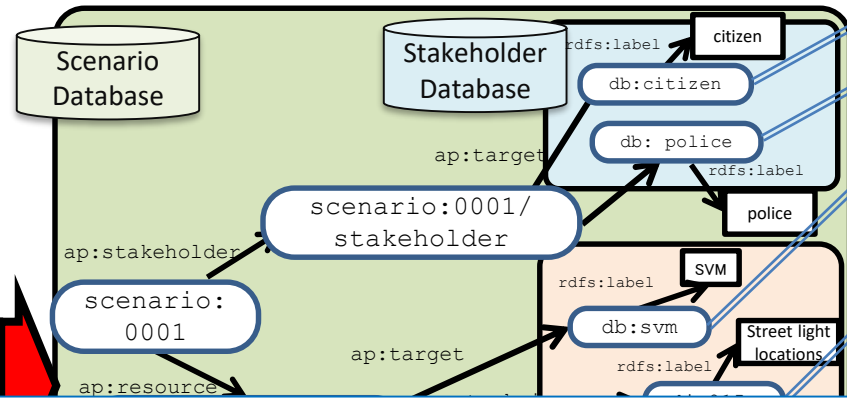
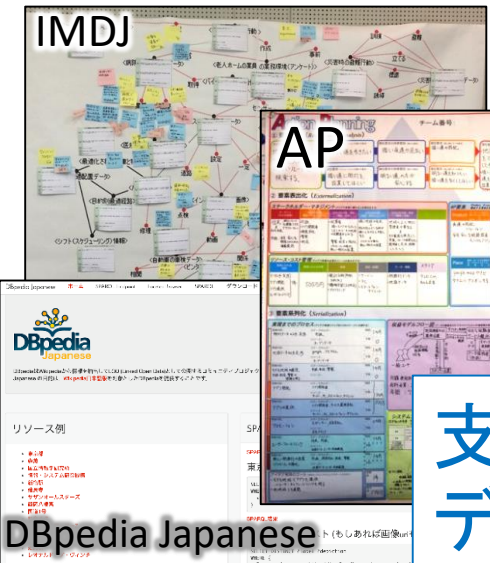
EFP: Element Externalization
ESP: Element Serialization



An Action Planning sheet

Hayashi and Ohsawa "Comparison of Conflict Resolution Behavior and Scenario Generating Process in Group and Individual by Handwriting Process Analysis," in Intelligent Decision Technologies (2016)

データ利活用知識の構造化と検索



支援ツールとして必要な潜在データの検索技術も実現し、

DJ store: 過去のIMDJで創出されたデータ利活用案を知識ベースとし、現在の興味に関連するDJを検索

Resource Finder: 過去のAPのシナリオに関連する要素(ステークホルダー、データ、分析手法など)を知識ベースとし、DBpediaと連結して新規シナリオに関連する要素を推薦



DJ検索結果: 39件

ID	DJタイトル	概要	項目	内容
79	知日外国人消費動向調査	知日旅行者の日本における消費行動データ	ID	79
31	ライフログ	居住者のライフログ(ライフイベント)データである。住居につづけたセンサにより、蓄積したライフログ、ライフイベント、家電情報操作履歴、電力値パラメータ(消費電力、電圧、電流、力率)および各家電機器が操作された際の消費電力波形の1年間のデータ	タイトル	知日外国人消費動向調査
32	スーパーマーケットにおける顧客の店内行動履歴データ	RFIDを付与したカードを利用して、スーパーマーケットにおける顧客の移動経路を記録し、データとして蓄積したもの。データにはカートの識別番号、タイムスタンプ、店内における位置情報(OX,Y座標)、移動状態(停止または移動)、エリア(どの売場か)などが含まれる。	概要	知日旅行者の日本における消費行動データ
101	シェルスヒート	生活に向けた具体的な事例に入ると発表した。2例目の探検試験を教団員ほかの協力で始めることも明らかにした。教団員が5月で、今年5-7月の試験で目標約40キロワットを生産した。JAP EXは「試験ごとに生産量が向上しており、期待でき	実数	知日旅行者の日本における消費行動データ

Knowledge Search	Category	Type	Name	Outline
Plotting the correlation between the brightness of the streets and crimes on the map, and identifying the area to strengthen patrol.	stakeholder	target	citizen	The status of a person recognized under the custom or ... click here
	stakeholder	target	government	The system by which a state or community is governed... click here
	stakeholder	external supporter	police	A police force is a constituted body of persons... click here
	resource	technique	machine learning (SVM)	In machine learning, support vector machines (SVMs, also... click here
	resource	data	man data	Desktop and mobile web mapping service
	search	clear		

経済産業省「データ駆動型イノベーション創出戦略協議会」の後継事業「先端課題に対応したベンチャー事業化支援等事業」、国土交通省「国土交通行政に資するビックデータの活用に関する調査業務」で採用

(2) 時系列データから未知の変化を捉える分析技術の研究

要求 R: 急な売れ行き変化の理由が知りたい

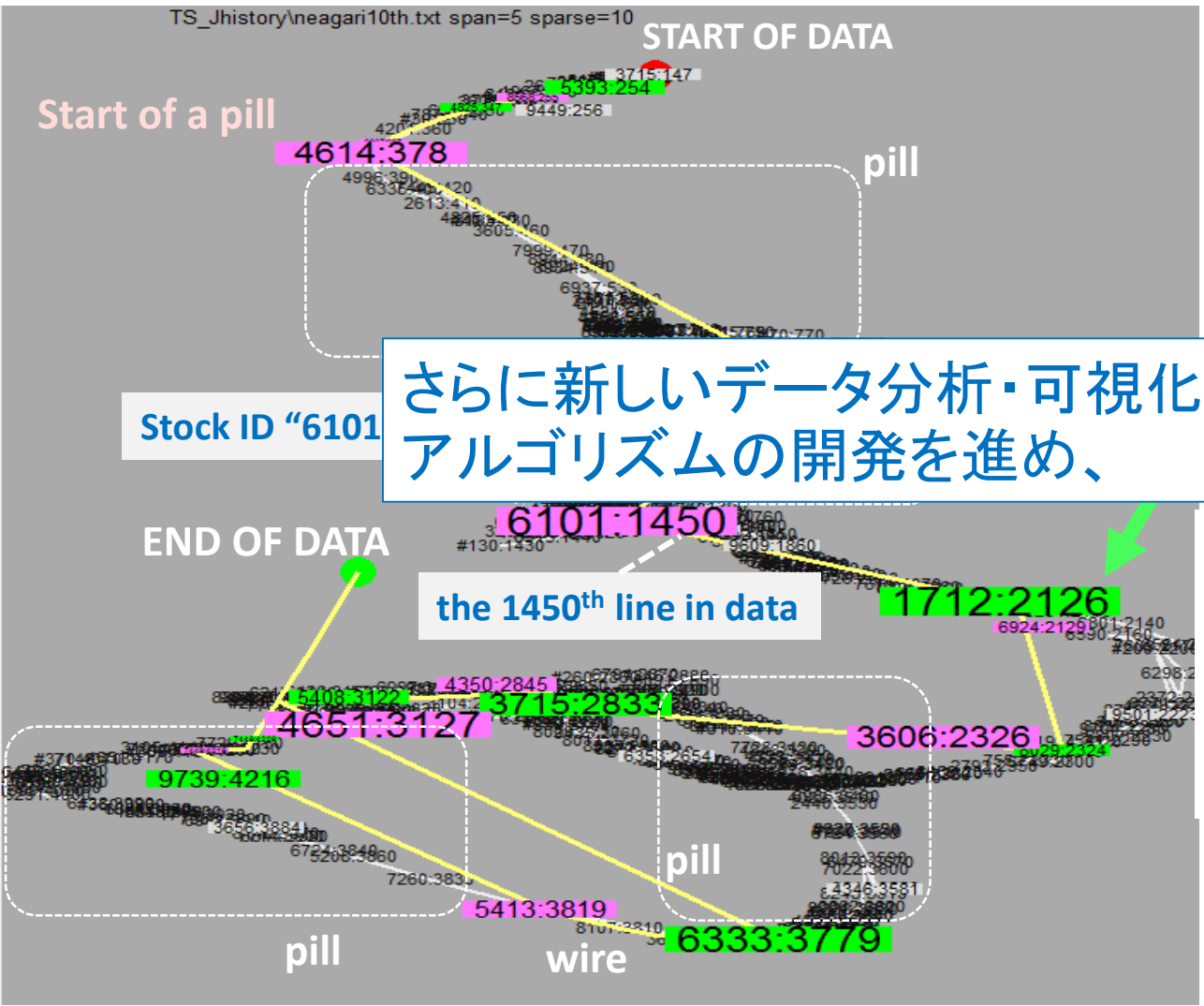
ソリューション(回答案) S: 変化のタイミングを把握し、外部情報と合わせて検討

Data for realizing S: $\{D\}_{567}$ 購買履歴データ $D(T)_{J_5}$: Tangled String

株価上昇上位株の履歴

IMDJから生まれた「絡まる糸 (Tangled String)」を多種データに適用

2011・3・12 (1712: 産業廃棄物処理大手: 土壌汚染の調査から浄化処理までを一貫して行っている)



さらに新しいデータ分析・可視化アルゴリズムの開発を進め、

既発表の結果: 流行商品の価格上昇は、Pillが終わると止まる → “株を売るタイミングを発見する手法は新しく、実務に使える画期的な成果だ(トップクオンツアナリストのコメント)

Yukio Ohsawa, "Tangled String Diverted for Evaluating Stock Risks - A by Product of Innovators Marketplace on Data Jackets" MoDAT2015 (Workshop on Designing Safe and Secure Life on the Market of Data), in IEEE International Conference on Data Mining, Atlantic City (2015)

(2) 時系列データから未知の変化を捉える分析技術の研究

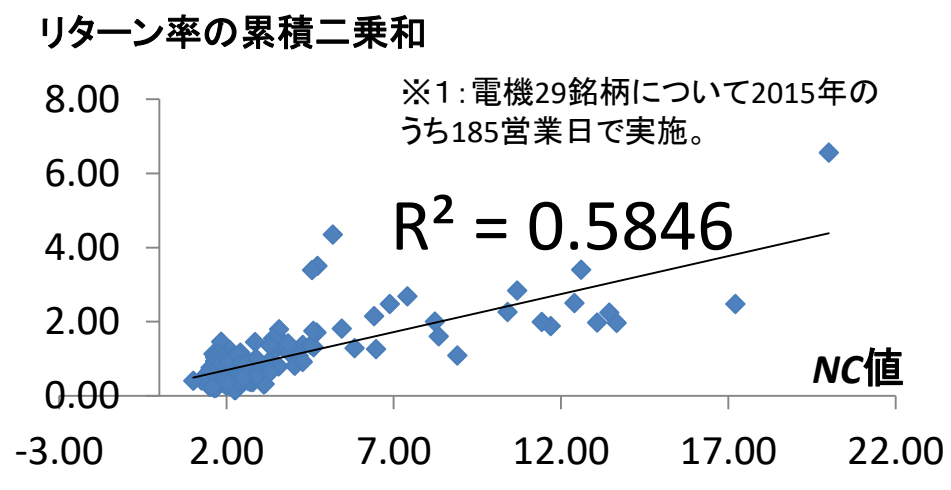
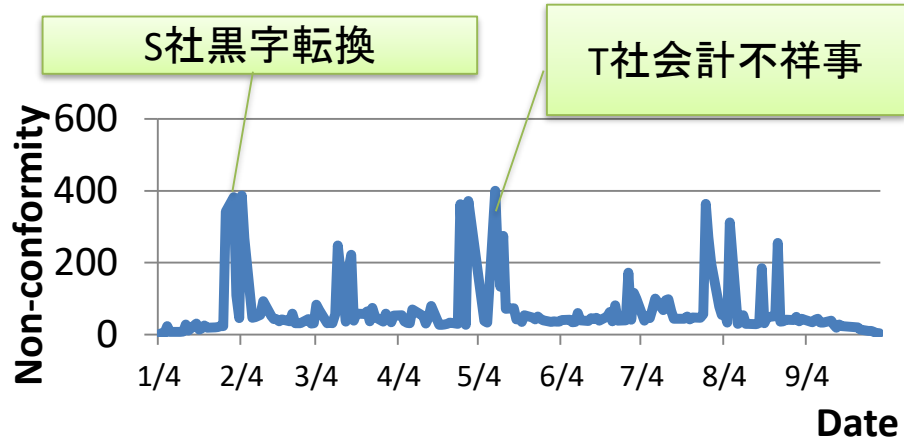
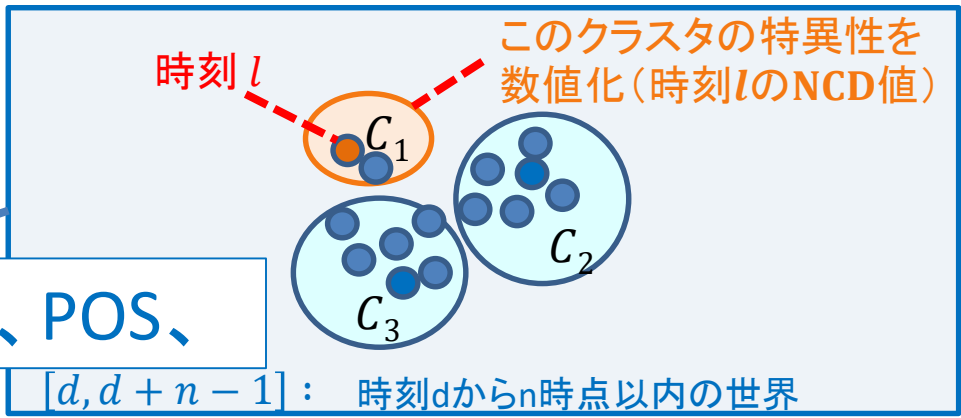
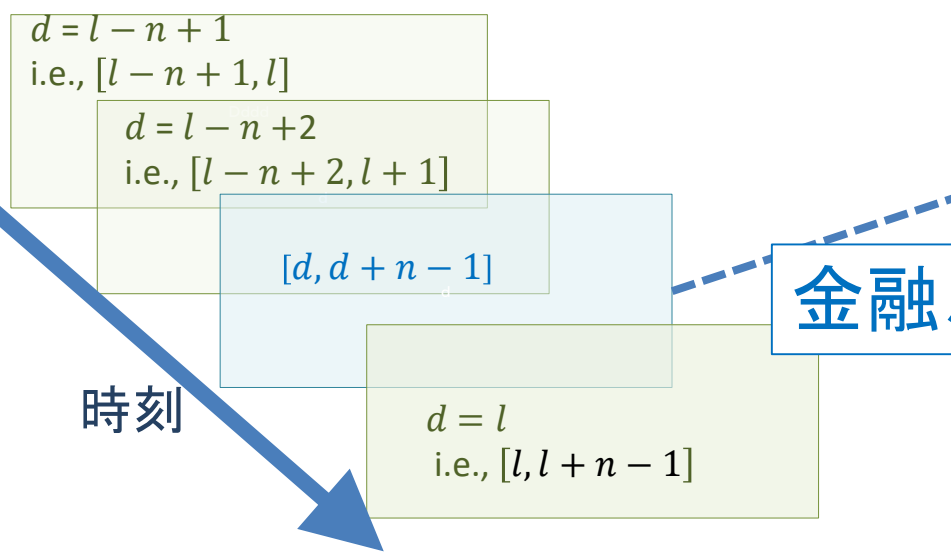
Non-Conformity Detection: マイナーなクラスタに属する時点(図の赤い粒)を抽出

サプライズ度即ち、

$$NC(t_l) = \sum_{d=l-n+1}^l \sum_{w=1}^K N_{w,d} / N_{w,l,d}$$

Kasuga, Ohsawa, Yoshino, Ashida (2016) Non-Conformity Detection in High-Dimensional Time Series of Stock Market Data, to be presented in IEA-AIE and published in Trends in Applied Knowledge-Based Systems and Data Science, Springer

※ w は、Affinity propagationで得た K クラスタのうち w 番目
 $N_{w,d}$ は、時区間 $[d, d+n-1]$ におけるクラスタ C_w 内の時点数
 $N_{w,l,d}$ は $[d, d+n-1]$ における「時点 t_l を含むクラスタ」内の時点数



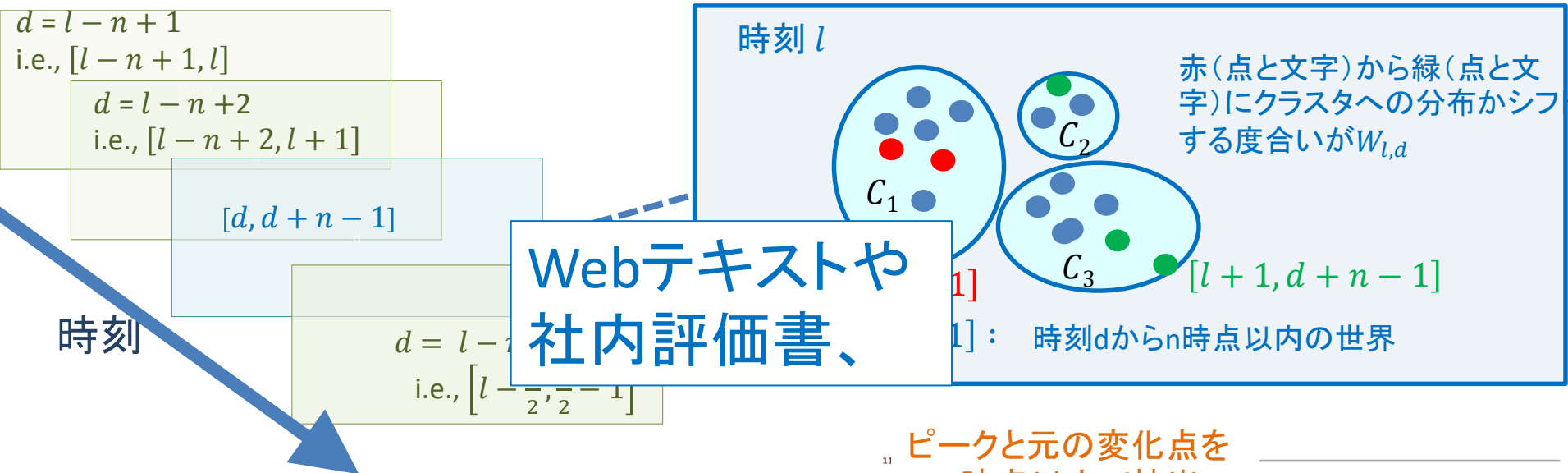
(2) 時系列データから未知の変化を捉える分析技術の研究

Conformity Change Detection: 各クラスタへの所属度が変わる時点を抽出

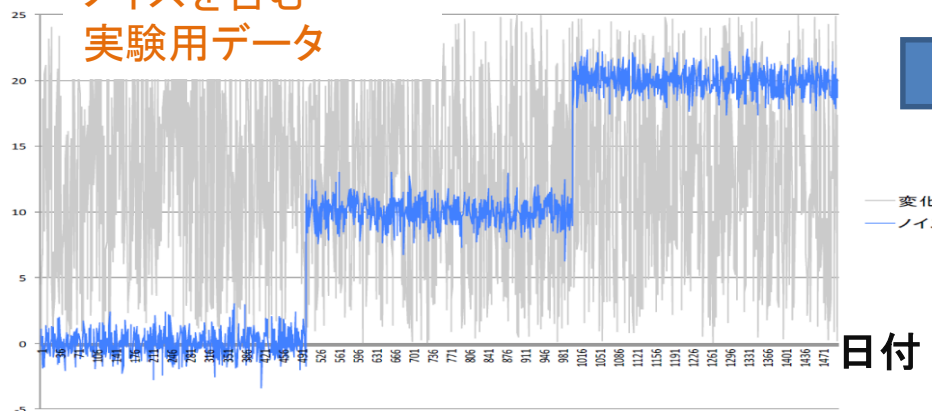
変化度即ち、

$$CC(t_l) = \sum_{w=1}^K \sum_{d=l-n+1}^{l-n/2} W_{l,d}$$

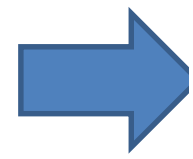
$$W_{l,d} = |\{t_q | t_q \in C_{w,d}, q \in [l+1, d+n-1]\}| - |\{t_q | t_q \in C_{w,d}, q \in [d, l-1]\}|$$



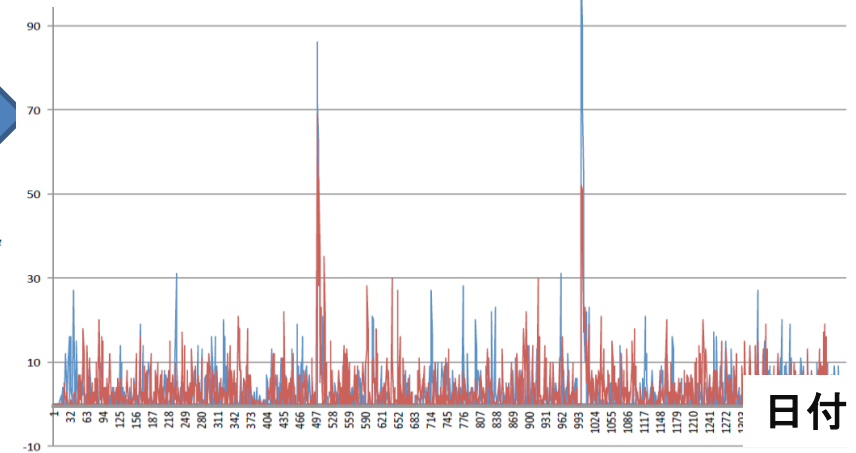
ノイズを含む
実験用データ



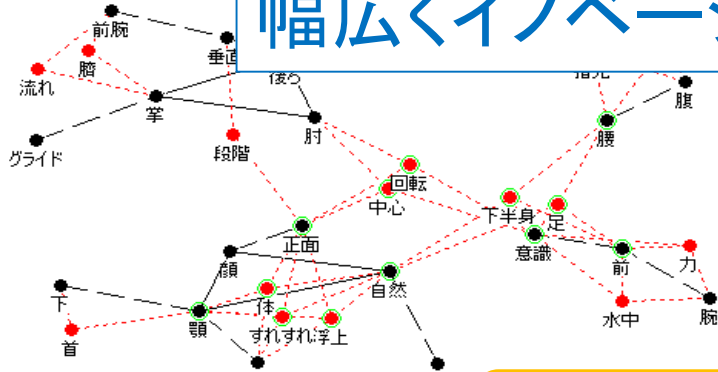
CCD値



ピークと元の変化点を
±1時点以内で抽出



地震から、スポーツデータまでを、 幅広くイノベーションに繋げているからです。



KeyGraph(上): 右の日記テキストの可視化



※写真は大澤本人



大澤幸生
工学系研究科・教授

第一キックは腰を水面に近く保ちながら・・・(中略)・・・ グライドの時は、肘を高くして、掌が下方の、やや後ろを・・・(中略)・・・ 第二キックを意識。この意識の中で、肘を回転中心として前腕を下方に・・・(中略)・・・ 自然に体が水面に浮上して顔が正面、ここで顎を突き出して水面すれすれに・・・(後略)

分析シナリオ①の実現

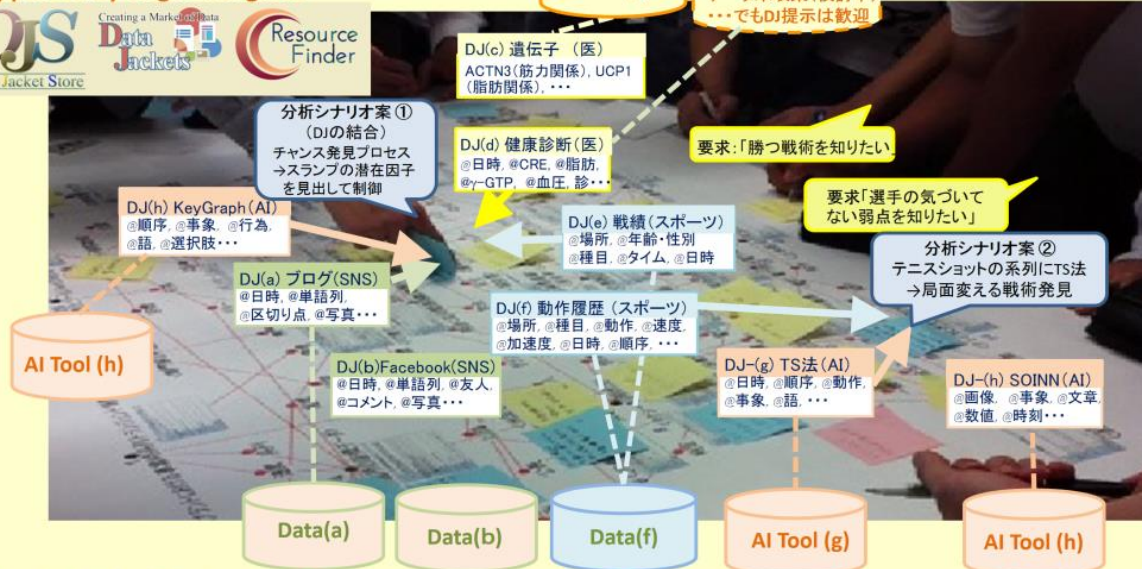
Innovators Marketplace on Data Jackets

(大澤研2013～)

supported by original engines



「人が読むメタデータ」(データジャケット:DJ)から、異分野データの結合と活用へ (写真はイメージ)



分析シナリオ②の実現

フォアハンドでバックハンド領域に打ち込んだ時、ネットにかかり試合展開が変わってしまう(あるプロの失敗パターンをTangled String法で発見)

