

# Tracking Latent Dynamics

—潜在的構造変化検出の情報論的学習理論—

山西健司

東京大学情報理工学系研究科

2010年 6月16日

# 目次

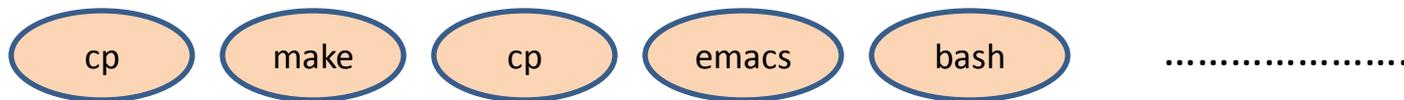
1. Latent Dynamicsの数学モデル
2. 動的モデル選択
3. 動的モデル選択の諸性質
4. データマイニング応用
5. おわりに

# 1. Latent Dynamicsの数学モデル

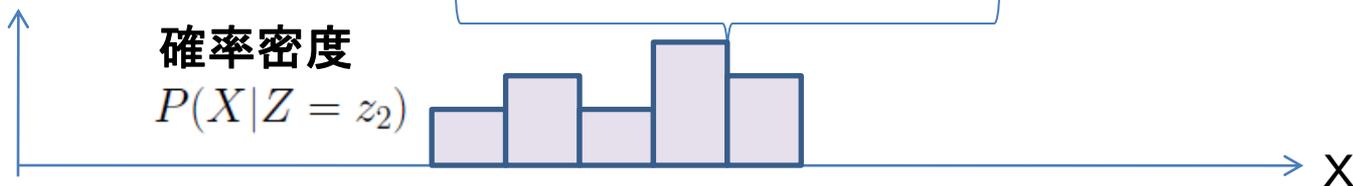
# 潜在情報とは

## コマンド系列の例: 潜在情報=意図

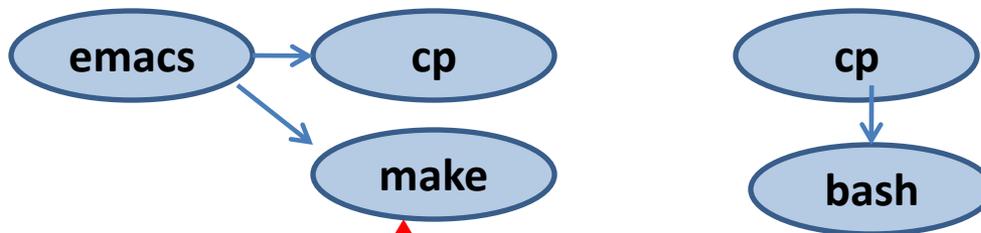
顕在変数  
(データ)



確率密度  
 $P(X|Z = z_2)$



確率モデル



潜在変数  
(クラスター)



確率密度



# 潜在情報の確率モデル

- $X$ : 顕在変数
- $Z$ : 潜在変数
- $Z = \{z_1, \dots, z_k\}$ : 潜在変数の集合
- $k$ : 潜在変数の数

$$P(X) = \sum_Z P(X|Z)P_k(Z).$$

問題: 潜在情報が動いていること (Latent Dynamics) をいかに扱うか?

# Latent Dynamicsの数学モデル

- 潜在変数の内容(数値、離散状態)が変わる  
⇒ 第一のLatent Dynamics

状態空間モデル

潜在状態のSwitching(レジュームSwitching)

潜在分布の変化検出

- 潜在変数を規定する世界の構造が変わる  
⇒ 第二のLatent Dynamics

動的モデル選択

Switching Distribution 理論

Tracking Best Expert 理論

# 状態空間モデル

## 潜在変数の値の時間的変動をモデリング

[赤池、北川93, 94]

$$z_t = Fz_{t-1} + Gx_t$$

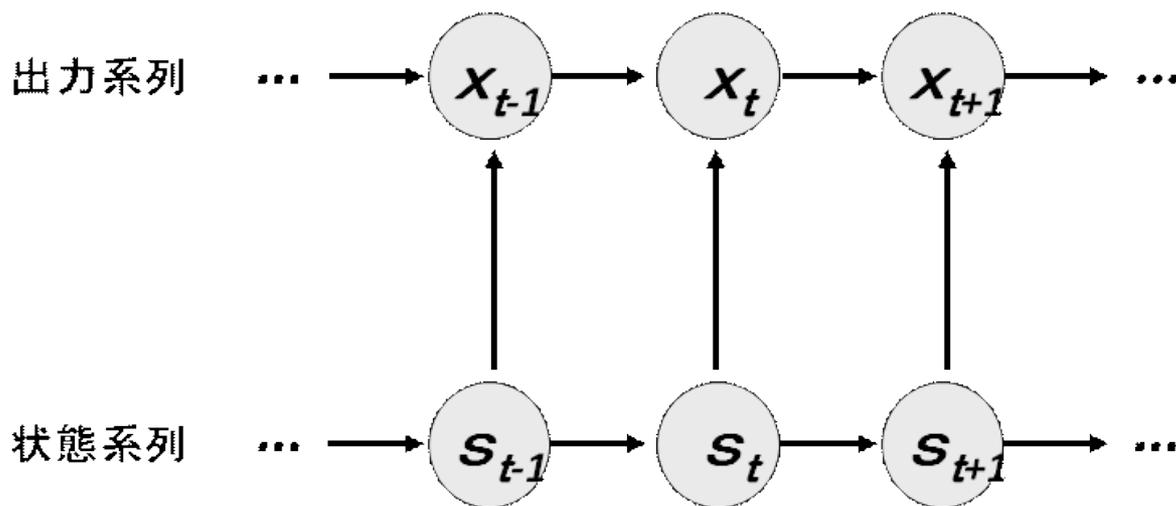
$$y_t = Hz_t$$

状態変数  $z_t$  が動的に変化する

⇒ 状態推定を通じて、時系列のトレンド、長周期、短周期成分などの潜在成分の変化を知ることができる

# 隠れマルコフモデルとノンパラメトリックベイズ

## 潜在状態のマルコフ遷移をモデリング



$$P(x^n) = \sum_{z^n} \prod_{t=1}^n P(x_t|z_t) \prod_{t=1}^{n-1} P(z_t|z_{t-1}) \cdot \gamma(z_1)$$

状態数  $K$  が無限に渡る事前分布 (Dirichlet Process)

⇒ ノンパラメトリックベイズ (Infinite HMM) [Beal et.al. NIPS2002]

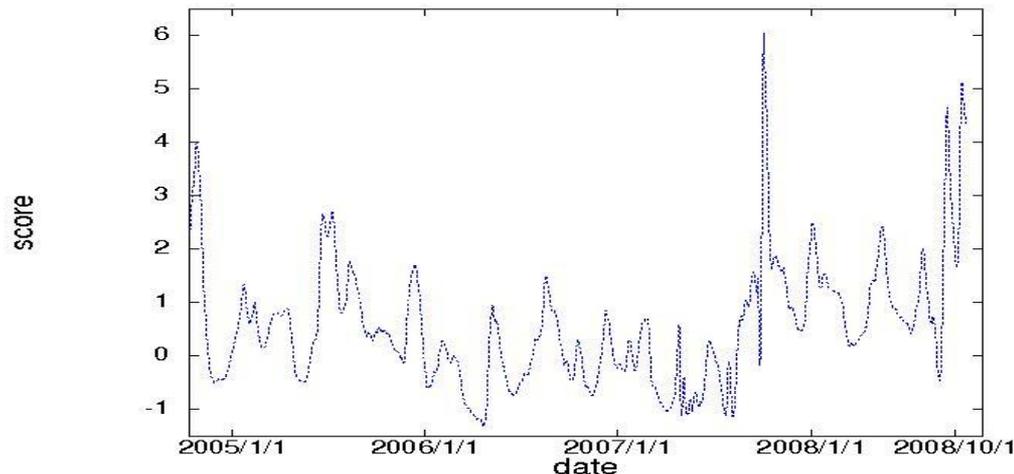
# 潜在状態のSwitchingを検出 —レジームスイッチング—

レジーム(=潜在変数) のスイッチングを伴う経済時系列モデリング

[Hamilton 1994]

$$X_t = \theta_1(s) + \theta_2(s)X_{t-1} + \sigma(s)\varepsilon_t$$

s: 離散状態をとるレジーム  
sが時間とともにスイッチングするダイナミクスを知る



S1

S2

S3

# 潜在空間の分布変動検出

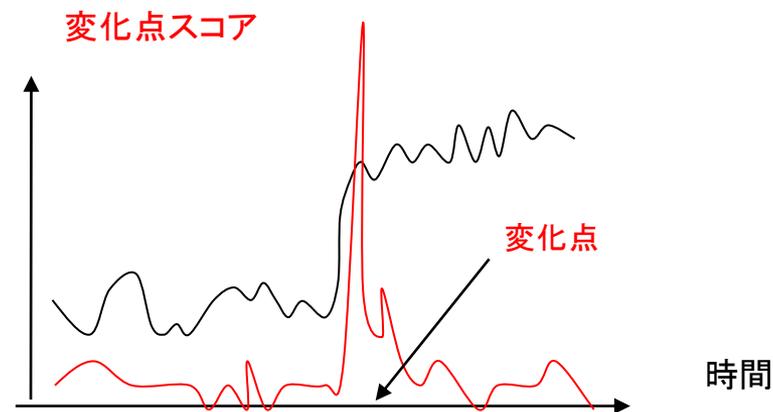
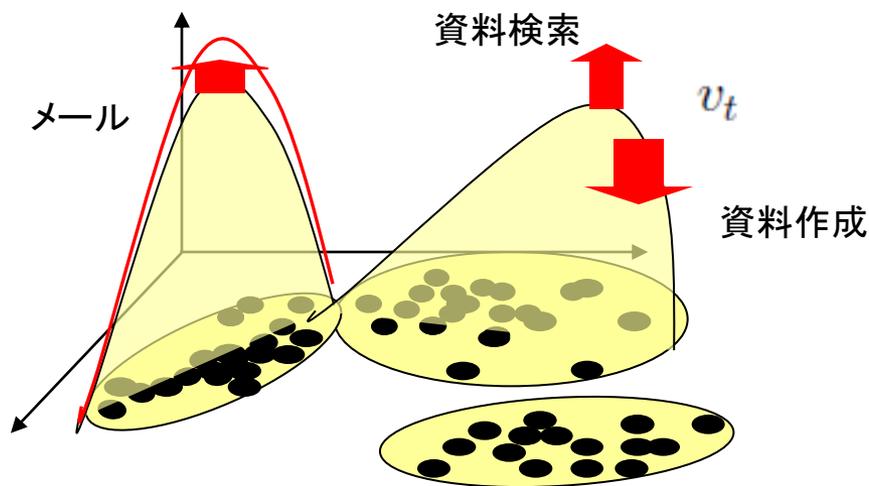
潜在空間の分布変動の変化からLatent Dynamicsを捉える

[Hirose , Yamanishi SDM08  
Stat. Sci. and DM 2009]

$$\begin{aligned} v_t &= d(P_t(z), P_{t-1}(z)) \\ &= \sum_z P_t(z|x^t) \log \frac{P_t(z|x^t)}{P_{t-1}(z|x^{t-1})} \end{aligned}$$



$\{v_t : t = 1, 2, \dots\}$  系列の変化点検出により  
潜在空間の変動を知る



# 潜在空間の構造的変化

潜在変数空間のモデル(変数の数、状態の数)が変化する

コンピュータ  
利用ログ  
(顕在情報)

```
1 cd
2 ls
3 emacs
4 platex
5 :
```

```
1 cd
2 ls
3 ema
4 plat
5 :
```

```
1 emacs
2 make
3 ./a.exe
4 :
5 :
```

```
1 emacs
2 cp
3 bash
4 perl
5 :
```

構造的  
変化

時間

利用パタン  
(潜在情報)

emacs → cp  
emacs → make

プログラム作成

cp → \*.exe

情報漏えい

emac → cp  
emac → make

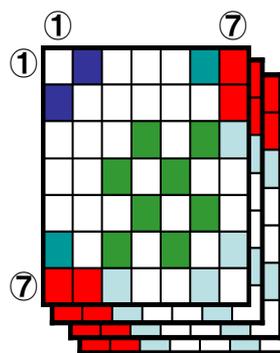
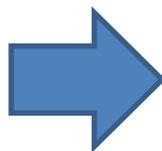
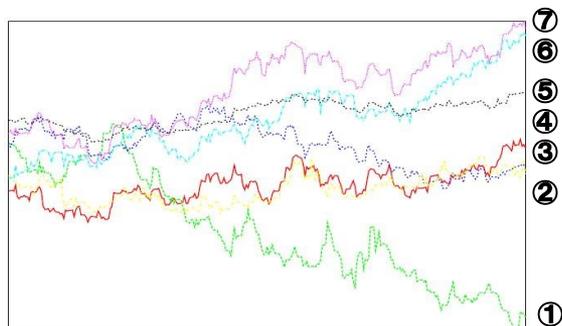
プログラム作成

$$Z = \{z_1\}$$

$$Z = \{z_1, z_2\}$$

# 潜在空間の構造変化検出の例

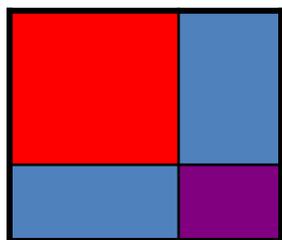
グラフ系列から、グラフ分割構造(コミュニティ構造)の変化を検出



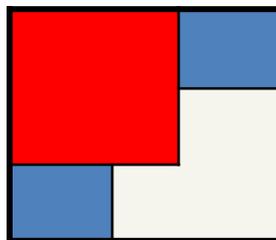
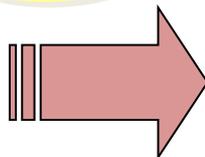
相関行列の時系列

...

e.g.  
テンソル解析

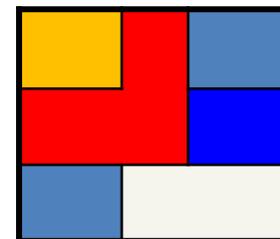
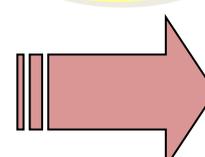


時間発展



分割構造の変化

時間発展

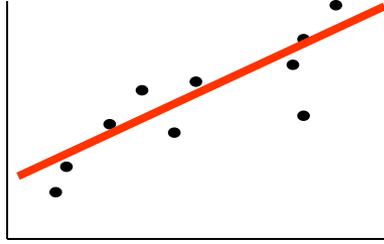


新しいクラス  
ターの出現

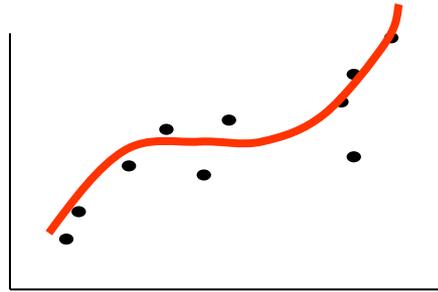
# 3. 動的モデル選択

# 動的モデル選択

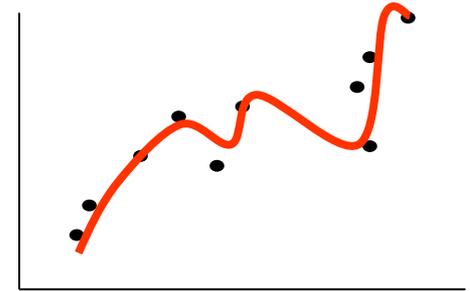
静的モデル  
選択



$$\theta_0 + \theta_1 x$$



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$



$$\theta_0 + \theta_1 x + \dots + \theta_d x^d$$

動的モデル  
選択

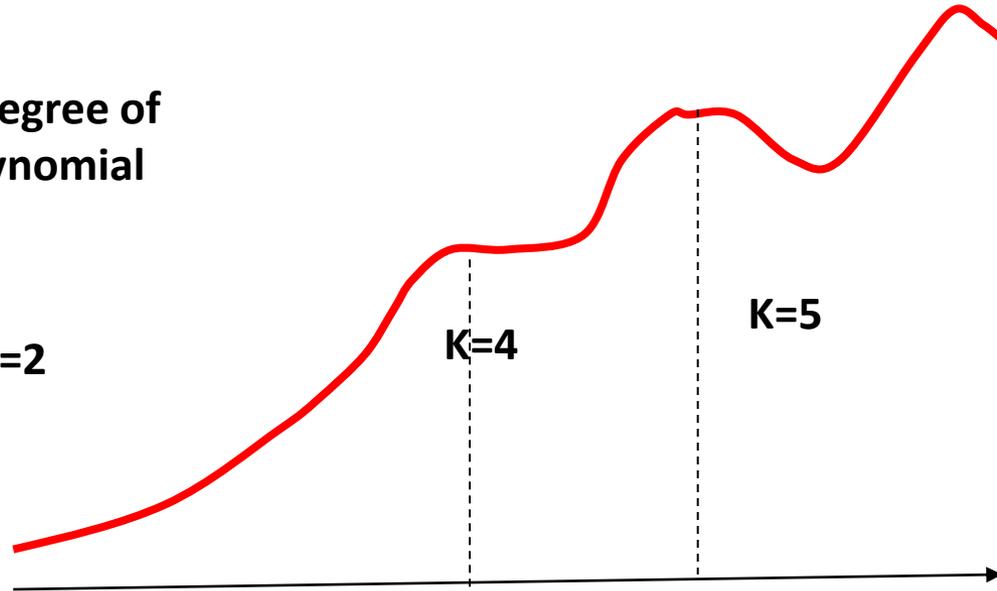
K: degree of  
polynomial

K=2

K=4

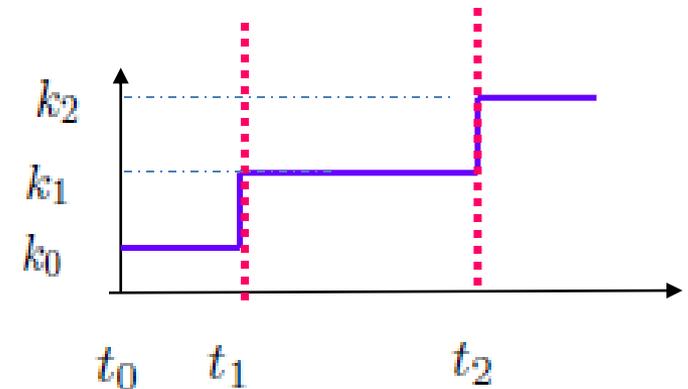
K=5

時間とともに  
次数が変化



# Switching分布

- $x^n = x_1 \cdots x_n$ : データ列
- $m$ : 変化点の総数
- $\mathbf{t} = (t_0 = 1, t_1, t_2, \dots, t_m)$ : 変化点系列
- $\mathbf{k} = (k_0, k_1, k_2, \dots, k_m)$ : モデル系列
- $\mathbf{s} = (m, \mathbf{t}, \mathbf{k})$ : latent dynamics (潜在変数)
- $P(\mathbf{s})$ :  $\mathbf{s}$  の事前分布



## Switching Distribution

$$P(x_i | x^{i-1} : \mathbf{s}) = \begin{cases} P(x_i | x^{i-1} : k_0) & t_0 \leq t \leq t_1 \\ P(x_i | x^{i-1} : k_1) & t_1 \leq t \leq t_2 \\ P(x_i | x^{i-1} : k_2) & t_2 \leq t \leq t_3 \\ \dots & \dots \end{cases} \quad m = 2$$

$$P(x^n | \mathbf{s}) = \prod_{i=1}^n P(x_i | x^{i-1} : \mathbf{s})$$

$$P(x^n) = \sum_{\mathbf{s}} P(x^n | \mathbf{s}) P(\mathbf{s})$$

**Latent Dynamicsを  
時間を含めた潜在変数  
として扱う**

# Latent Dynamicsの推定

$x^n$  のモデル系列に関する確率的コンプレキシティ

$$-\log \sum_s P(x^n | s) P(s) \leq \min_s \{-\log P(x^n | s) - \log P(s)\}$$

MDL (Minimum Description Length) 原理

$$\hat{s}_{opt} = \arg \min_s \{-\log P(x^n | s) - \log P(s)\}$$

データの符号長      LDの符号長

⇒ データ圧縮の意味で最適な Latent Dynamics

**動的モデル選択**  
**(Dynamic Model Selection)**

問題:

- 動的モデル選択のアルゴリズムをいかに設計するか？
- いかに上手く Latent Dynamicsを抽出できるか？

# 関連研究

- **従来モデル選択 (統計学)**  
AIC [Akaike 1973], BIC [Schwarz 1978], MDL [Rissanen 1978],  
– 定常性を仮定、最適なモデルを1つ選択
- **動的モデル選択 (情報論的学習理論)**  
[Yamanishi & Maruyama IEEE IT 2007]  
– 非定常なもとで、モデルが変化する場合のモデル系列選択
- **Switching 分布 (情報理論)**  
[Erven, Grunwald, Rooij, NIPS 2008]  
– 分布が時間とともにSwitchする場合の分布推定理論
- **Tracking Best Experts (計算論的学習理論)**  
[Herbster & Warmuth ML1998][Vovk COLT1997]  
– Best Expertが時間と共に変わる場合の予測アルゴリズム
- **Changing Dependencyの検出 (機械学習)**  
[Fearnhead & Liu 2007] [Xuan & Murphy ICML2007]

# 予測分布と予測符号長(1/2)

## 動的に変化する確率構造を記述する有効な手段

確率モデルのクラス

$$\mathcal{P}_k = \{P(x^n|\theta, k) : \theta \in \Theta_k\} \quad (n = 1, 2, \dots) \quad k: \text{モデル}$$

$$\dim \Theta_1 < \dots < \dim \Theta_k < \dim \Theta_{k+1} < \dots$$

予測分布:

$$P(x_t|x_{t-1} : k_t) = P(x_t|\hat{\theta}_{t-1} : k_t) \quad (\text{plug-in 分布})$$

$\hat{\theta}_{t-1}:x^{t-1} = x_1, \dots, x_{t-1}$ からの $\theta$ の最尤推定量

$$P(x_t|x_{t-1} : k_t) = \int P(x_t|\theta)P(\theta|x^{t-1} : k_t)d\theta \quad (\text{Bayes 予測分布})$$

$P(\theta|x^{t-1})$ :  $x^{t-1}$ からの $\theta$ の事後確率密度関数

$$P(x_t|x^{t-1} : k_t) = \frac{P(x_t \cdot x^{t-1}|\hat{\theta}(x_t \cdot x^{t-1}) : k_t)}{\sum_x P(x \cdot x^{t-1}|\hat{\theta}(x \cdot x^{t-1}) : k_t)} \quad (\text{SNML 分布})$$

SNML: Sequentially Normalized Maximum Likelihood

# 予測分布と予測符号長(2/2)

$x^{t-1}$  が与えられたもとでの  $x_t$  の予測符号長

$$-\log P(x_t | x^{t-1} : k_t)$$

$x^n = x_1, \dots, x_n$  にわたる予測符号長

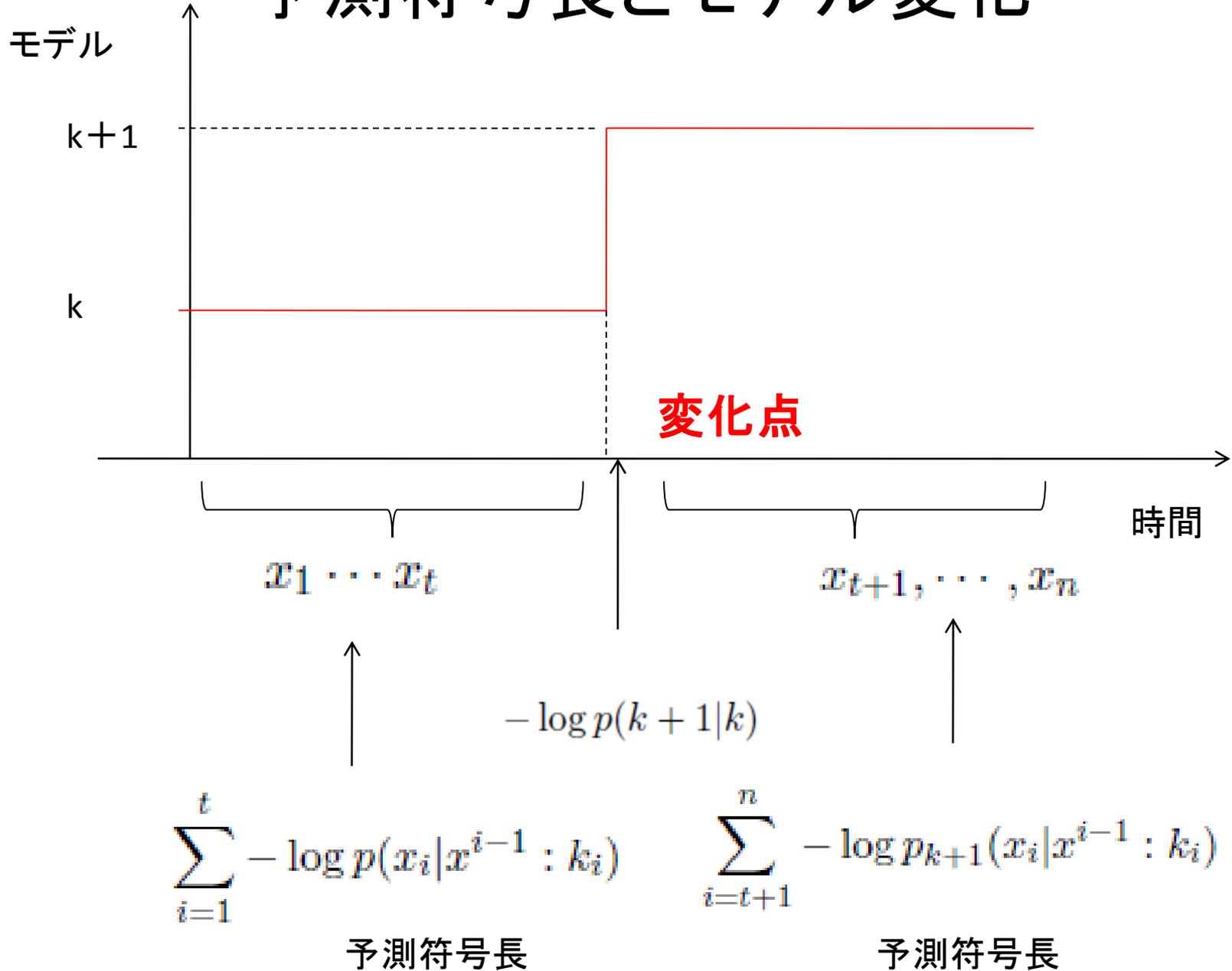
$$I(x^n) = \sum_{t=1}^n -\log P(x_t | x^{t-1} : k_t)$$

モデルが一定値  $k_t = k$  をとる場合には、どの予測分布を用いても

$$I(x^n) \approx -\log P(x^n | \hat{\theta}_n) + \frac{k}{2} \log n$$

**⇒モデルが時変のときには成り立たない**

# 予測符号長とモデル変化



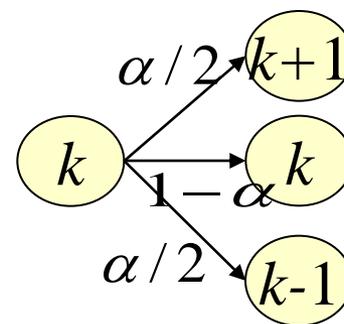
# 動的モデル選択規準

[Yamanishi Maruyama  
IEEE Trans. IT 2007]

■ 仮定: モデルは確率的に推移する

モデル遷移確率:  $P_t(k_t = k | k^{t-1}, \alpha)$ ,  $\alpha$  : パラメータ

$$P(k_t | k_{t-1} : \alpha) = \begin{cases} 1 - \alpha, & \text{if } k_t = k_{t-1} \text{ and } k_{t-1} \neq 1, K, \\ 1 - \frac{\alpha}{2}, & \text{if } k_t = k_{t-1} \text{ and } k_{t-1} = 1, K, \\ \frac{\alpha}{2}, & \text{if } k_t = k_{t-1} \pm 1 \end{cases}$$



■ DMS (Dynamic Model Selection) 規準:

$$\ell(x^n : k^n) = \underbrace{\sum_{t=1}^n -\log P(x_t | x^{t-1} : k_t)}_{\text{データ列の 予測符号長}} + \underbrace{\sum_{t=1}^n -\log P_t(k_t | k^{t-1}, \hat{\alpha})}_{\text{モデル列 } k^n \text{ の 予測符号長}}$$

Input:  $x^n = x_1, \dots, x_n$

$\hat{\alpha}$  : estimate of  $\alpha$

Output:  $k^n = k_1, \dots, k_n$  minimizing  $\ell(x^n : k^n)$

### 3. 動的モデル選択の諸性質

# Latent Dynamics推定の限界(1/3)

[Yamanishi Maruyama 2007]

**定理1:** Switching 分布に対して一括型でDMS規準を最小化するLatent Dynamicsを計算量  $O(n^2)$  で出力するアルゴリズムが存在し、その総記述長の上界は次式で抑えられる。

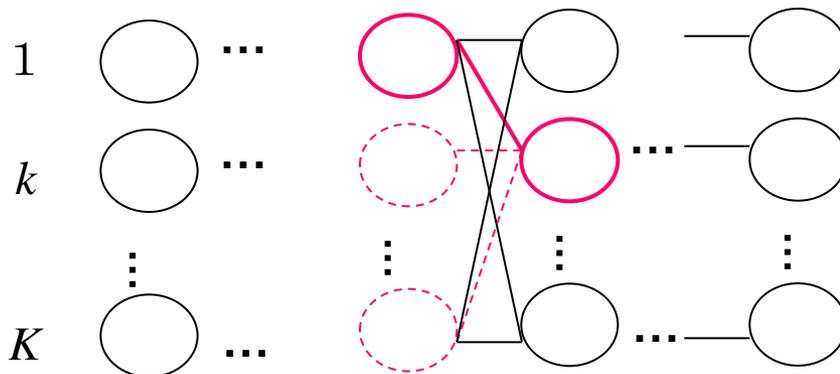
$$\star \min_m \min_{(t_0, k_0), \dots, (t_m, k_m)} \left\{ \sum_{j=0}^m \sum_{t_j+1}^{t_{j+1}-1} -\log P(x_t | x^{t-1} : k_i) + nH\left(\frac{m}{n}\right) + \frac{1}{2} \log n + m + o(\log n) \right\}$$

モデル系列の複雑さ

ポイント1: Dynamic Programming による最適パス探索

ポイント2: Krichevsky and Trofimov 推定法によるモデル遷移確率の推定

$$\hat{\alpha}_t = \frac{n_{k,t} + \frac{1}{2}}{t}$$



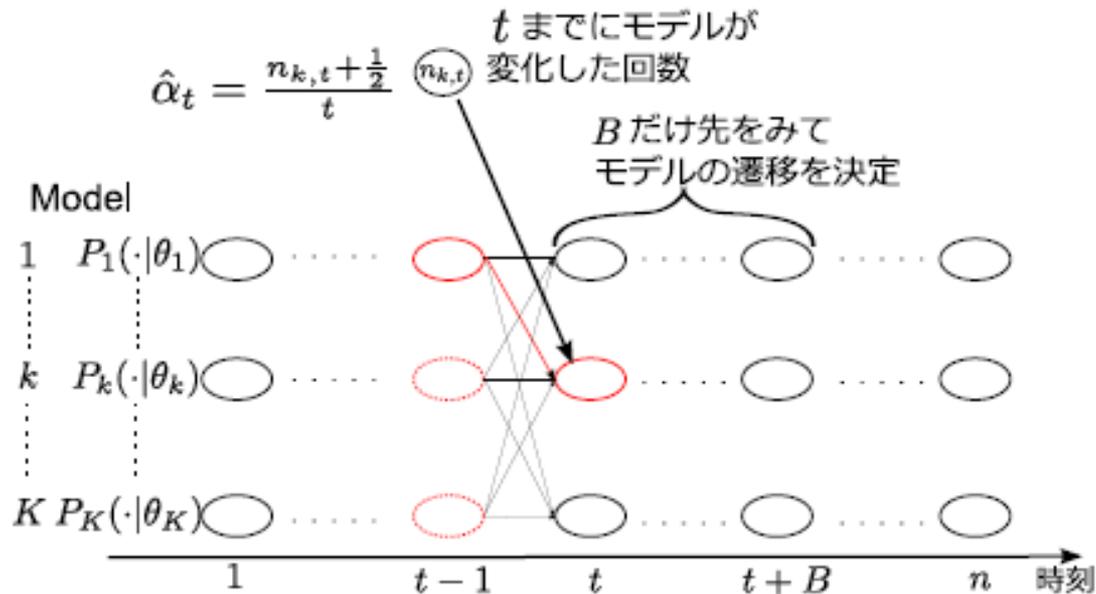
各時点、各状態に対し、  
前時点のパス候補から  
最適なパスを1つ選択

# Latent Dynamics推定の限界(2/3)

[ Sakurai Yamanishi 2010]

**定理2:** Switching 分布に対して**逐次的に**DMS規準を最小化する Latent Dynamicsを計算量  $O(n)$  で出力するアルゴリズムが存在し、そのときの総記述長の上界は★よりも大。

**ポイント:** 一定幅Bのウィンドウを設け、**B だけ先読み**をしてから、**変化点の位置を決定**  
**ウィンドウをスライディング**させてDynamic Programming で求めた最適系列を接続する



# Latent Dynamics推定の限界(3/3)

Resetting 分布.....モデルの変化点で予測分布をリセットさせる

注: Switching分布は各モデルで過去の予測分布を憶えている

**定理3:** Resetting 分布に対して(各変化点で初期化)一括型でDMS規準を最小化するLatent Dynamicsを計算量  $O(n^3)$  で出力するアルゴリズムが存在し、そのときの総記述長の上界は★よりも小。

LD1: 一括型Switching

LD2: 一括型Resetting

LD2: 逐次型Switching

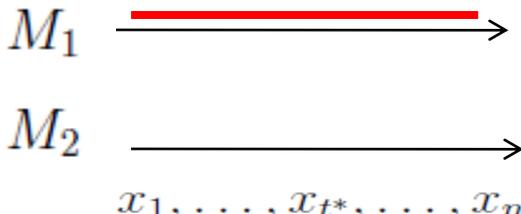
計算量: LD2  $O(n^3)$  > LD1  $O(n^2)$  > LD3  $O(n)$

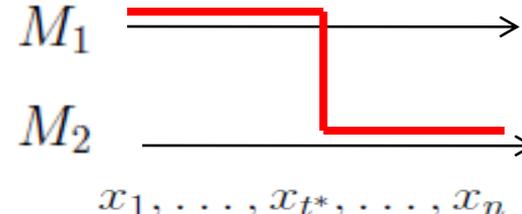
記述長: LD2 < LD1 < LD3

# 仮説検定からみた動的モデル選択(1/3)

モデルが2つ  $\{M_1, M_2\}$  の場合の仮説検定問題:

$t^*$ : モデルの変化点

仮説  $H_0$ :  $M_1$  for  $x_1^n = x_1 \cdots x_n$ ,  $M_2$  

仮説  $H_1$ :  $\begin{cases} M_1 & \text{for } x_1^{t^*} = x_1 \cdots x_{t^*}, \\ M_2 & \text{for } x_{t^*+1}^n = x_{t^*+1} \cdots x_n. \end{cases}$  

動的モデル選択規準により

$-\log P(x_{t^*+1}^n | x_1^{t^*} : M_1) + \log P(x_{t^*+1}^n | x_1^{t^*} : M_2) - \alpha < 0 \implies H_0$ : 採択  
otherwise  $\implies H_1$ : 採択

$$\alpha = \log \frac{\omega}{1-\omega}, \quad P(M_1|M_1) = P(M_2|M_2) = \omega, \quad P(M_2|M_1) = 1-\omega$$

# 仮説検定からみた動的モデル選択(2/3)

## 定理4

$$\begin{aligned} \text{第一種の誤り確率} &= \text{Prob}[\text{モデルが変化しないが } H_1 \text{ が採択}] \\ &\leq 2^{-\alpha} \end{aligned}$$

$$\begin{aligned} \text{第二種の誤り確率} &= \text{Prob}[\text{モデルが変化するが } H_0 \text{ が採択}] \\ &\leq 2 \exp(-Ch\beta^2) \end{aligned}$$

ここに

$$h \stackrel{\text{def}}{=} n - t^* \quad (\text{delay})$$

$$D_h(M_2 || M_1) \stackrel{\text{def}}{=} \sum_{x_{t^*+1}^n} P(x_{t^*+1}^n | x_1^{t^*} : M_2) \log \frac{P(x_{t^*+1}^n | x_1^{t^*} : M_2)}{P(x_{t^*+1}^n | x_1^{t^*} : M_1)}$$

$$\beta \stackrel{\text{def}}{=} \frac{1}{h} (D_h(M_2 || M_1) - \alpha)$$

# 仮説検定からみた動的モデル選択(3/3)

## 定理5

$$D_h(M_2||M_1) \stackrel{\text{def}}{=} \sum_{x_{t^*+1}^n} P(x_{t^*+1}^n | x_1^{t^*} : M_2) \log \frac{P(x_{t^*+1}^n | x_1^{t^*} : M_2)}{P(x_{t^*+1}^n | x_1^{t^*} : M_1)}$$

$$\lim_{h \rightarrow \infty} \frac{1}{h} (D_h(M_2||M_1) - \alpha) \geq \gamma$$

とするとき

変化点検出の遅延時間の期待値  $\leq O(1/\gamma^2)$

## 4. データマイニング応用

# 事例1: SyslogからのNovelty Detection(1/3)

Syslogとは

- BSD syslog プロトコルを用いて集められたイベント集合の系列
- 装置の状態に関するウォーニングメッセージ

ID	Time stamp	Event Severity	Att1	Att2	Message
##	Nov 13 00:06:23:	ERR	bridge:	!brdgursrv:	queue is full. discarding a message.
##	Nov 13 10:15:00:	WARN:	INTR:	ether2atm:	Ethernet Slot 2L/1 Lock-Up!!
##	Nov 13 10:15:10:	WARN:	INTR:	ether2atm:	Ethernet Slot 2L/2 Lock-Up!!
##	Nov 13 10:15:20:	WARN:	INTR:	ether2atm:	Ethernet Slot 2L/3 Lock-Up!!

# 事例1: SyslogからのNovelty Detection (2/3)

★セッション系列  $y_1, \dots, y_M$  を混合隠れマルコフモデルで表現

例. コマンド履歴データ  $y = (\text{ps}, \text{cd}, \text{ls}, \text{cp}, \dots)$

$$P(\mathbf{y}_j | \theta) = \sum_{k=1}^K \pi_k P_k(\mathbf{y}_j | \theta_k) \quad K: \text{混合数} = \text{行動パタンの数}$$

ここに

$$P_k(\mathbf{y}_j | \theta_k) = \sum_{(x_1, \dots, x_{T_j})} \gamma_k(x_1) \prod_{t=1}^{T_j-1} a_k(x_{t+1} | x_t) \prod_{t=1}^{T_j} b_k(y_t | x_t)$$

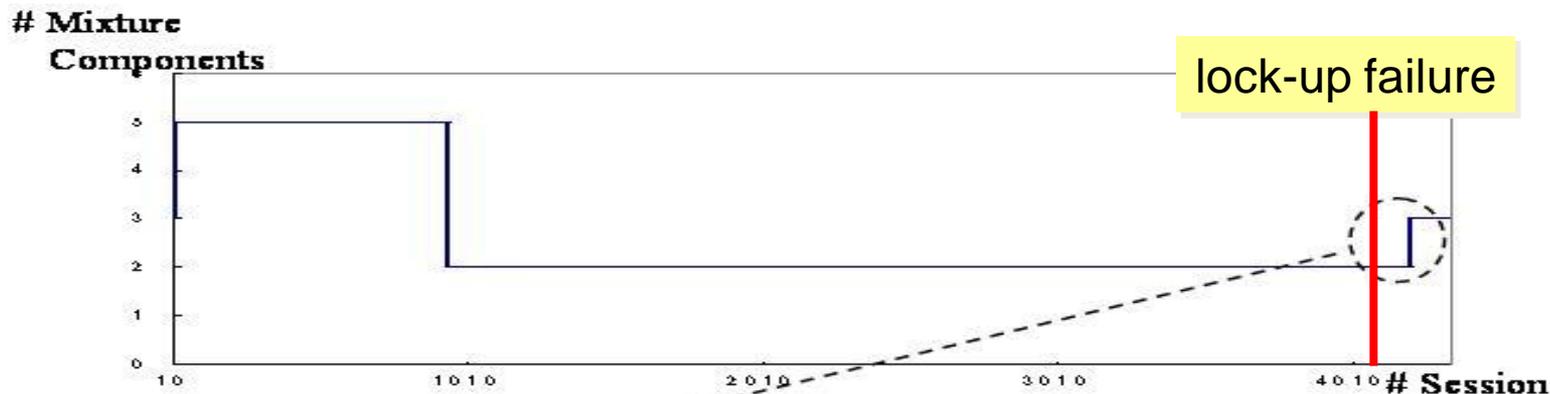
$(x_1, \dots, x_{T_j})$ : 状態変数

**行動パタンの数のLatent Dynamicsの推定により  
Novelty Detection(新しい行動パタンの発見)を行う**

# 事例1: Syslogからの障害パタンの検出(3/3)

[Yamanishi Maruyama KDD2005]

## Latent Dynamicsの推定により新しい重大障害のパターンを同定



Cluster1  
WARN:kern:!ATM: error (un→ WARN:kern:!ATM: error (un

Cluster2  
WARN:gated:rt\_add: interf→ WARN:gated:rt\_add: interf

Cluster3  
WARN:kern:!LEC: UNIT=0 commaE→ WARN:kern:!LEC: Called Pa

0.900
0.091
0.009

Message Trans.	Prob.
WARN:kern:! <u>LEC</u> : UNIT=0 comma→ WARN:kern:! <u>LEC</u> : Called Pa	0.954
ERR :bridge:! <u>brdgursrv</u> : q→ ERR :bridge:! <u>brdgursrv</u> : q	0.751
ERR :gated:krt_ifread: io→ ERR :bridge:! <u>brdgursrv</u> : q	0.734
WARN:kern:! <u>LEC</u> : Multicast→ WARN:kern:! <u>LEC</u> : Control D	0.691

New Pattern!

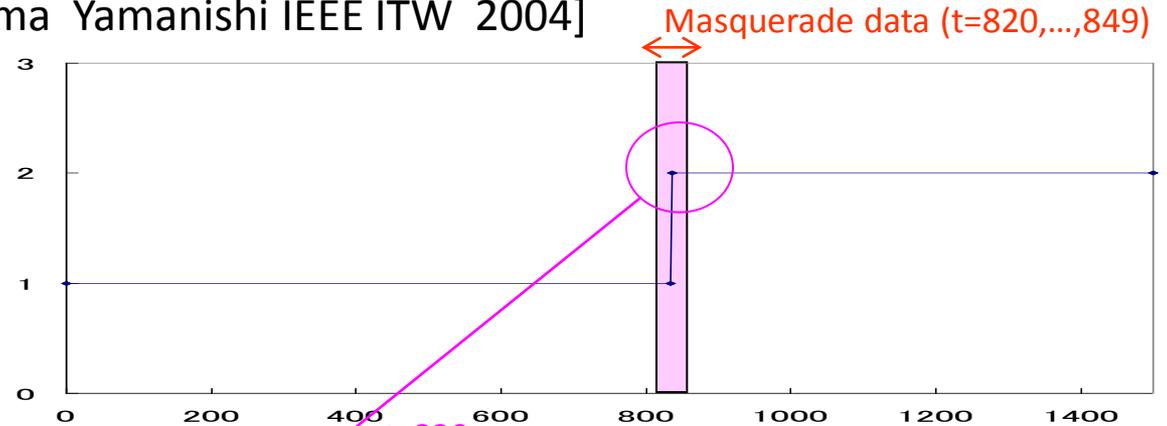
The component with smallest occurrence probability corresponds to a new pattern

# 事例2: アクセスログからのなりすまし検出

行動パタンの数のLatent Dynamicsからなりすましパターンを同定

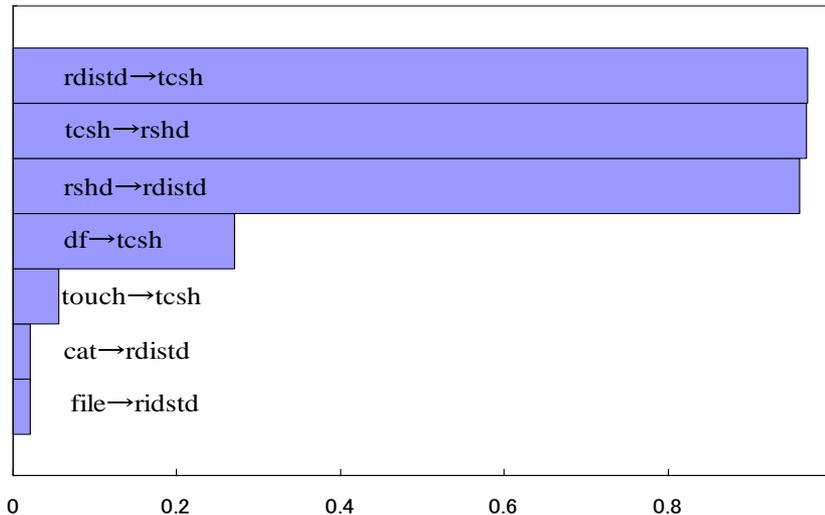
[Maruyama Yamanishi IEEE ITW 2004]

[Yamanishi Maruyama 2007]

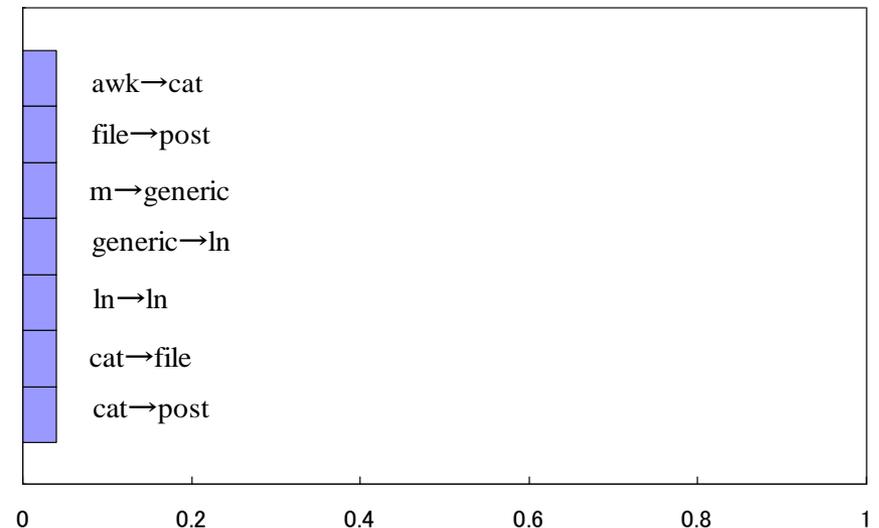


Top 7 transitions of each cluster (k=2 at t=820)

Pattern 1



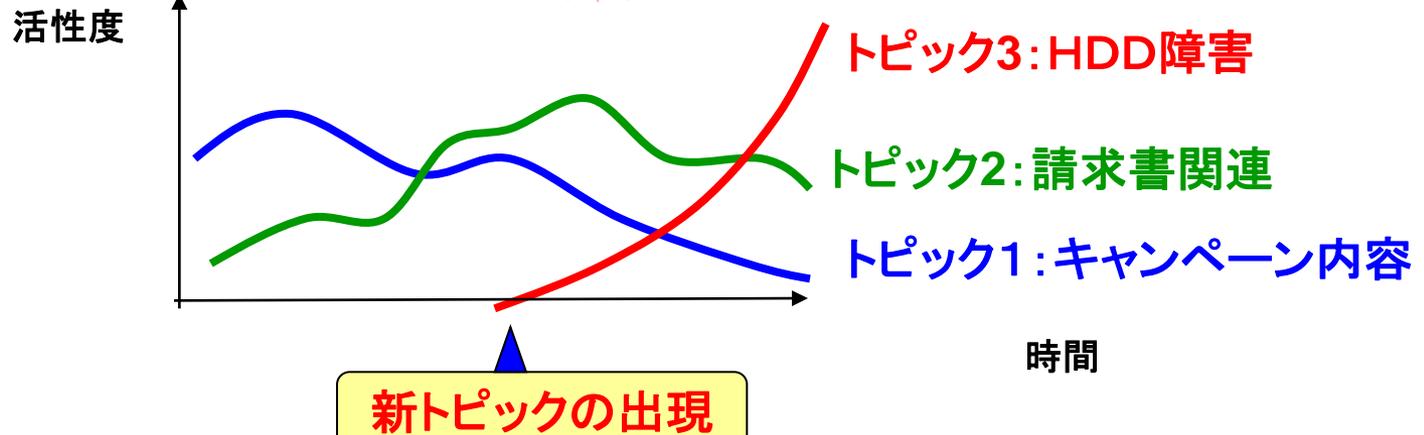
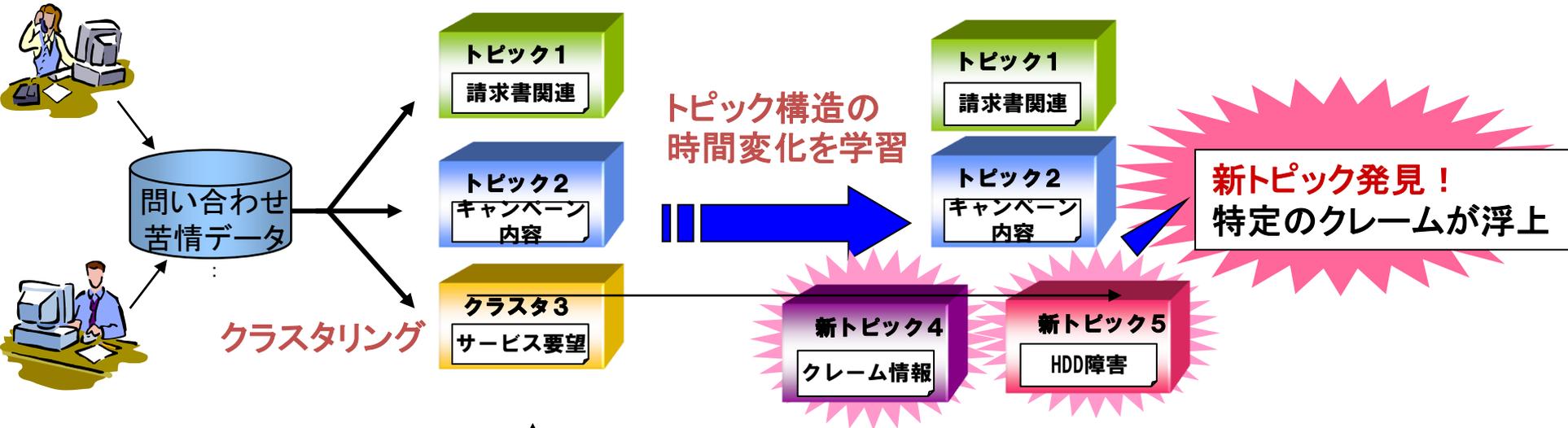
Pattern 2



# 事例3: トピック潮流の検出(1/2)

トピック混合モデルのトピック数の変化により新トピックの出現を同定

コンタクトセンターにおけるリアルタイム処理

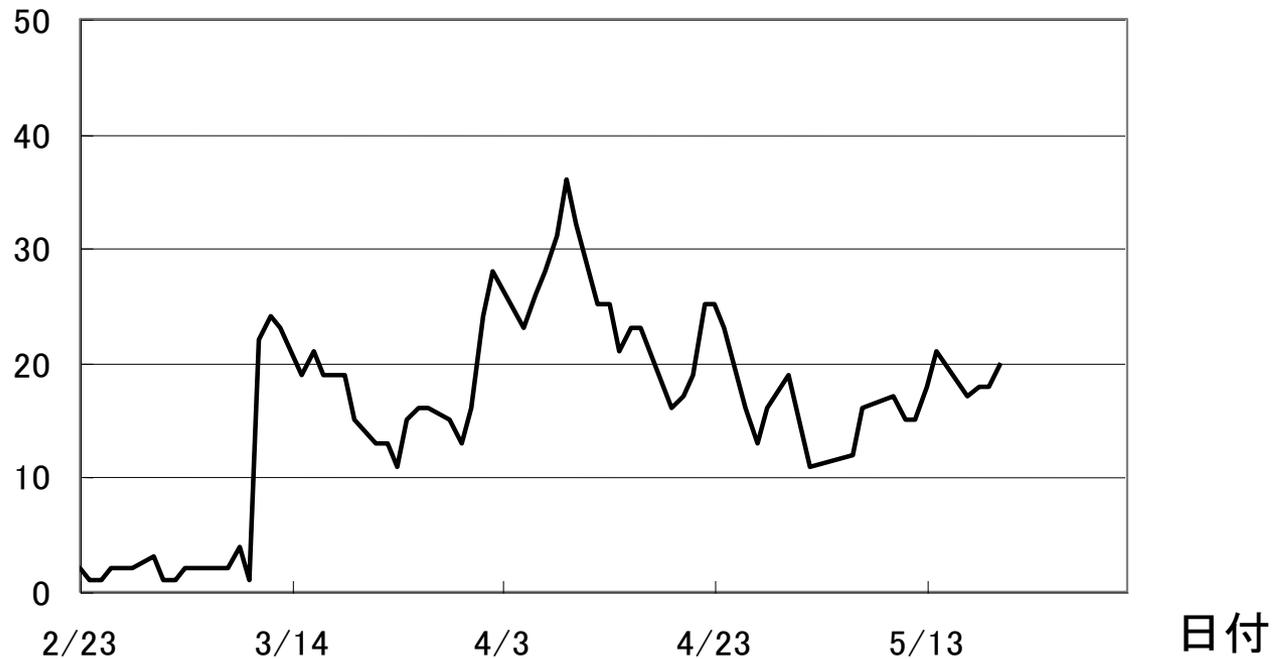


# 事例3：トピック潮流の検出(2/2)

[Morinaga Yamanishi KDD2004]

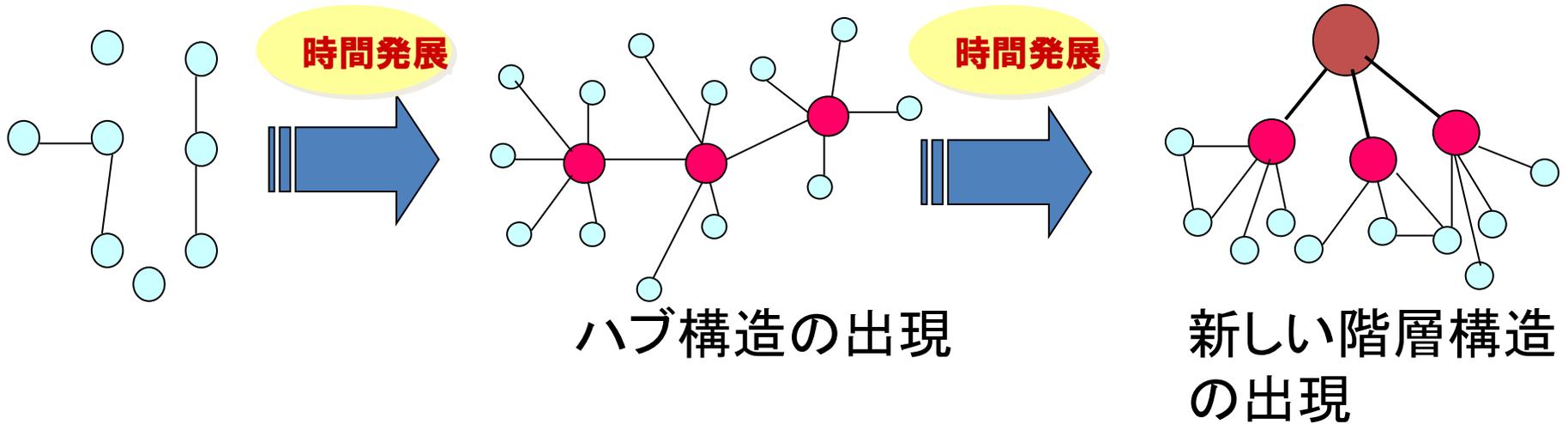
- ヘルプデスクにおけるコンタクトデータの分析（1202件）

主要トピック数の変化



# 事例4: ネットワーク構造変化検出

ネットワークの階層構造の変化を検知することでハブの存在を捉える



ex. SNS ネットワーク構造変化検知:

潜在的コミュニティの検知、犯罪集団の検知、Influencerの検知

Other examples:

物理ネットワーク、購買ネットワーク、共著者関係ネットワーク

# おわりに

1. Latent Dynamicsの問題としては、潜在変数の値の変化のみならず潜在空間を支配する構造の変化の研究こそがChallengeであろう。
2. そのような問題の数学的モデルとしてSwitching理論、動的モデル選択理論があり、それらは情報理論、学習理論の視点から1つの流れを形成しつつある。
3. それはNovelty Detection, 異常検出などの応用可能性をもち、ネットワークの進化的構造変化の発見などに対する有望な1アプローチとなる。