時系列データからの因果関係の発見

NTTコミュニケーション科学基礎研究所
研究員
近原　鷹一

自己紹介

- Name: 近原 鷹一（ちかはらよういち）
  - ✓ 2013.03: 慶應義塾大学理工学部生命情報学科（舟橋研）
  - ✓ 2015.03: 東京大学大学院情報理工学系研究科コンピュータ科学専攻（宮野研）
  - ➢ システム生物学、バイオインフォマティクス
  - ✓ 2015.04–: NTT コミュニケーション科学基礎研究所（CS研）
  - ✓ 2019.10-: 京都大学情報学研究科知能情報学専攻（鹿島研; 社会人博士課程)
  - ➢ 機械学習・因果推論(因果探索・因果効果推定)

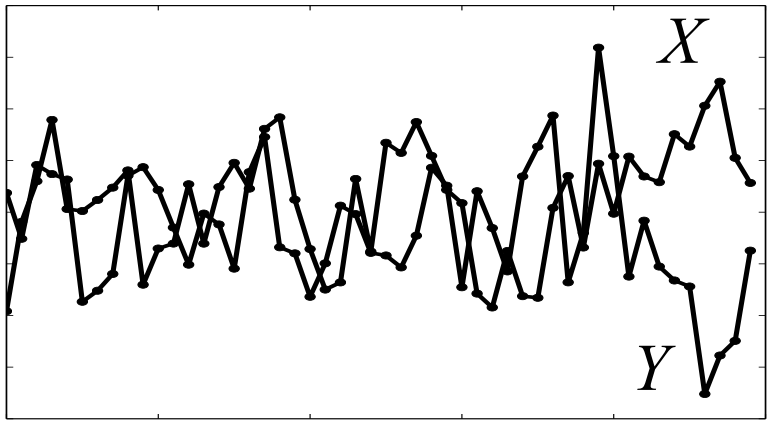
本日の内容

- 時系列データからの因果発見とは
  - ✓応用例
- Granger causalityの推定
  - ✓因果関係とは　– Granger causalityの定義 –
  - ✓教師あり学習に基づくGranger causalityの推定
    [Chikahara+; IJCAI2018(**AI分野最難関会議**)]
- Granger causality研究の最前線
  - ✓イベント間の因果関係 – 連続時間系のGranger –

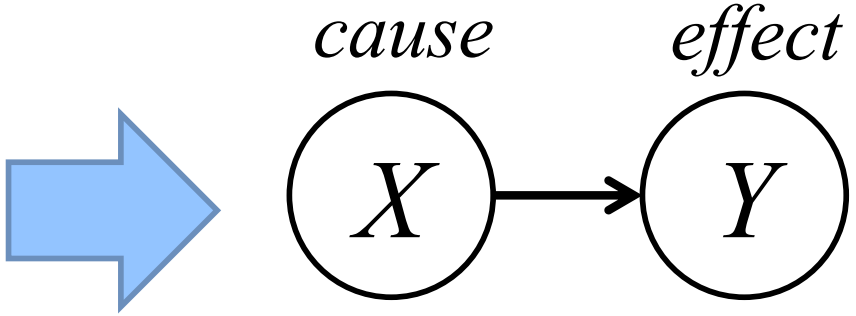
時系列データからの因果関係の発見

NTT

- 与えられた時系列データから
  変数間の**因果関係**を推定するタスク



入力: 時系列データ

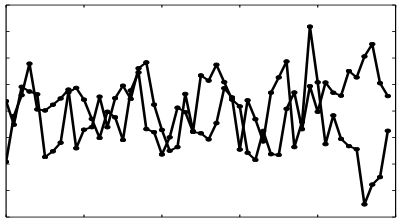


出力: **因果関係**

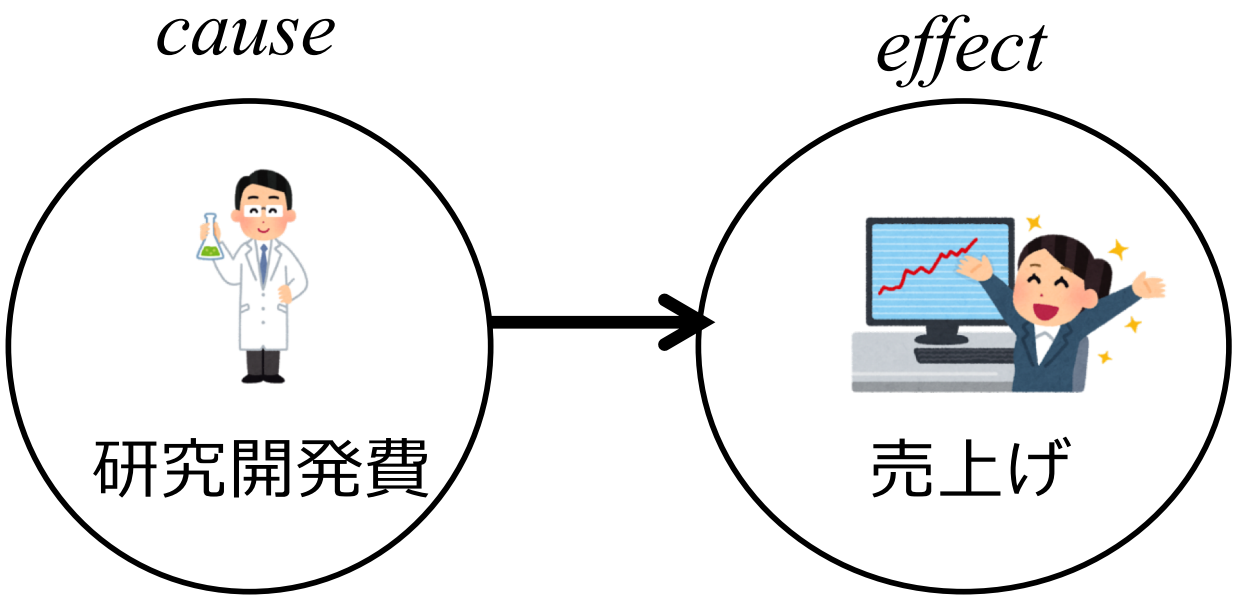
目的：因果関係に関する知識の発見

Application 1: Economics

- 「研究開発費は売上金に影響を与える」という知識は企業にとって有用



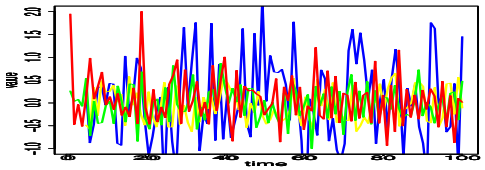
研究開発費と売上げ
の時系列データ



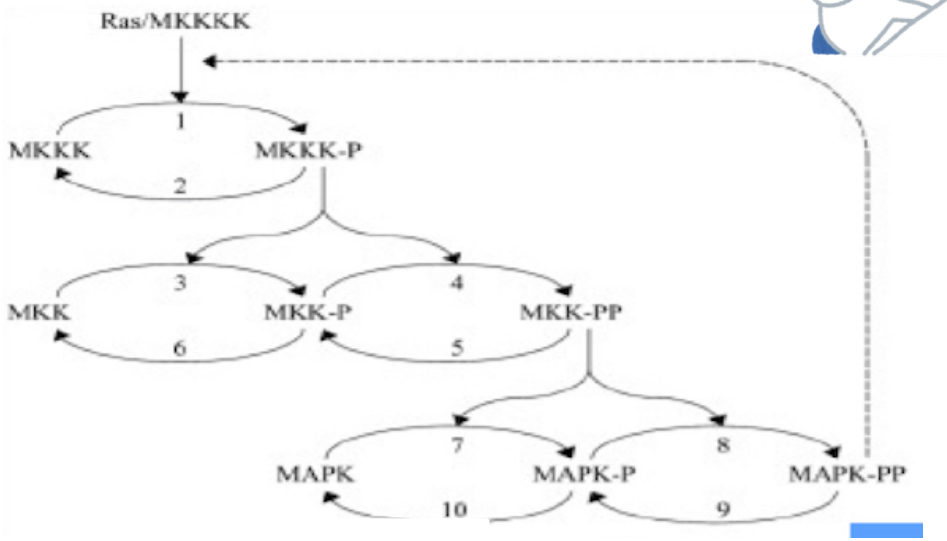
※逆ではない
（売上げは研究開発費に影響を与えない）

Application 2: Bioinformatics

- 遺伝子の発現量の時系列データから「遺伝子の制御関係」がわかると新規薬剤の探索に役立つ



遺伝子発現量に関するデータ



遺伝子の制御関係

“因果関係”とは何か？

どう定義されるのか？

時系列における因果関係の定義

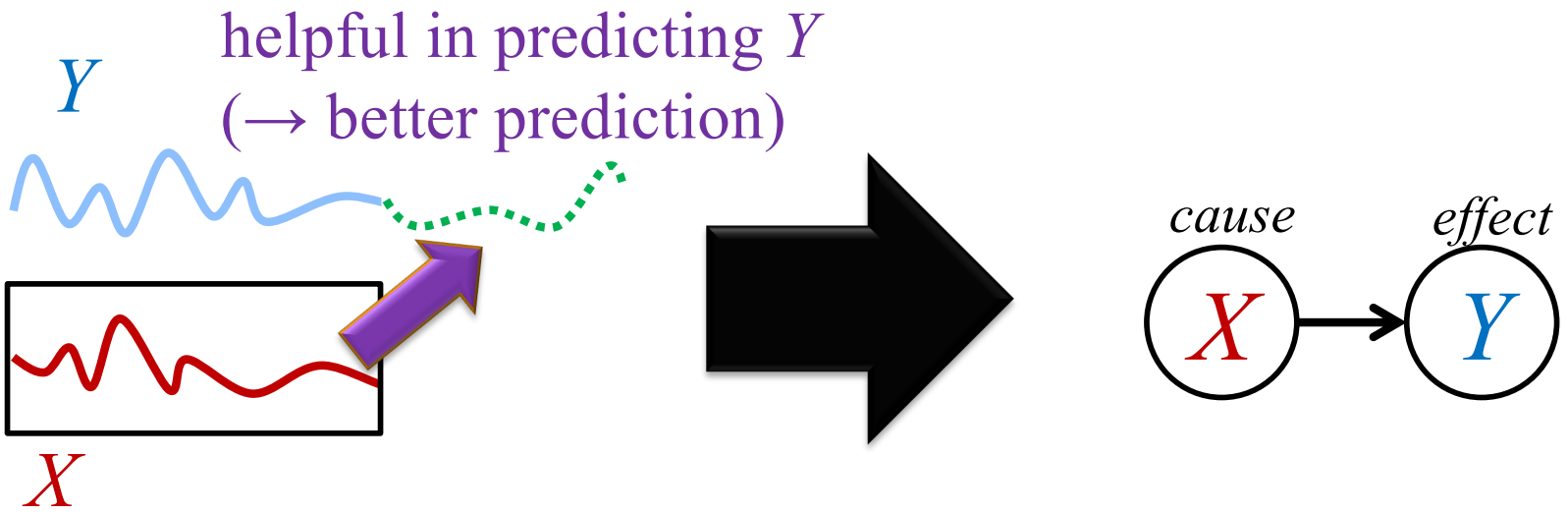
NTT

Granger causality [Granger1969]

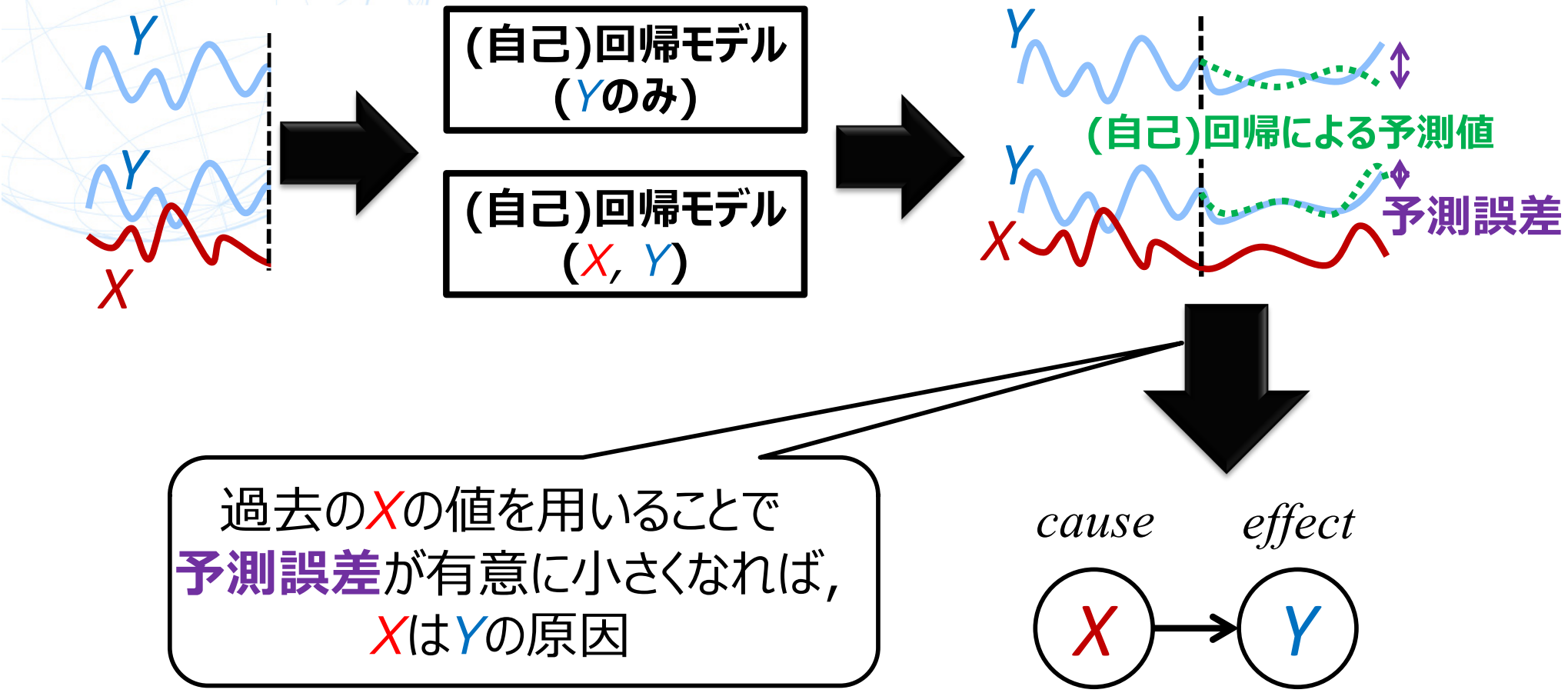
X is the cause of *Y*

if the past values of *X* are **helpful in predicting**

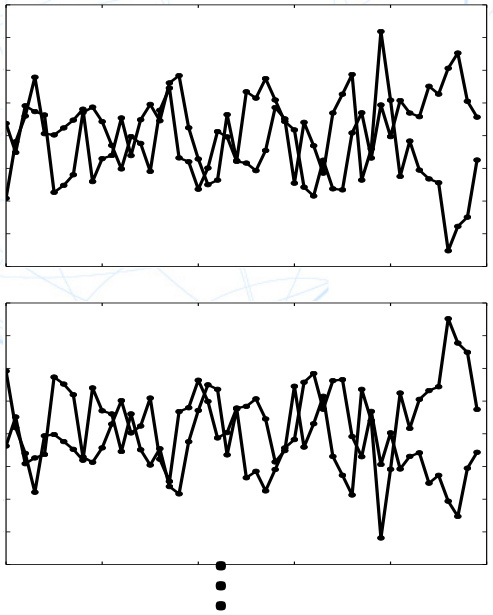
the future values of *Y*



既存手法によるGranger causalityの推定



既存手法の問題点



(自己)回帰モデル

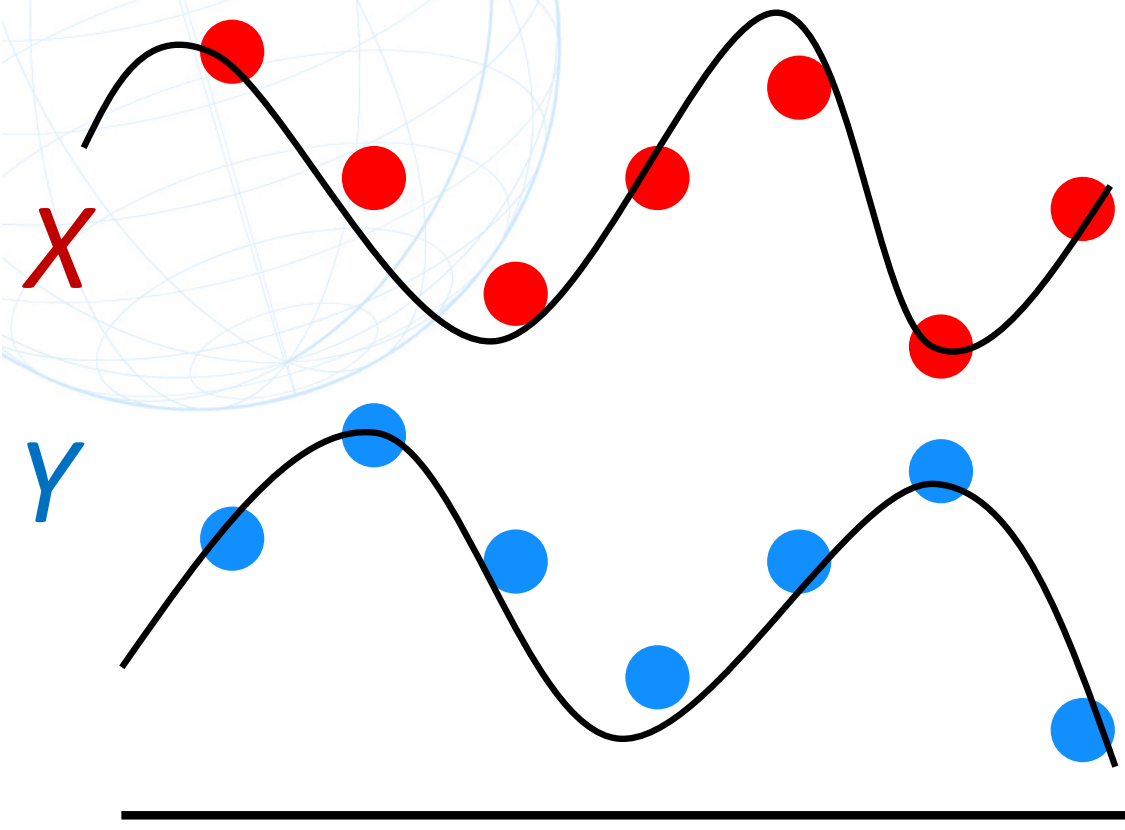
どの回帰モデルを
使えばいいの？

- VARモデル
  $$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \frac{1}{P}\sum_{\tau=1}^{P} A_\tau \begin{bmatrix} X_{t-\tau} \\ Y_{t-\tau} \end{bmatrix} + \begin{bmatrix} E_{X_t} \\ E_{Y_t} \end{bmatrix}$$
- GAM
- ガウス過程
- ⋮

**(自己)回帰モデルを各データに対して適切に選ばなければ
正しくGranger causalityの方向・有無を推定できない**

本研究のゴール：回帰モデルの選択が不要なアプローチの提案

（補足）既存のGranger causality推定



VARモデルの場合

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \frac{1}{P}\sum_{\tau=1}^{P} A_\tau \begin{bmatrix} X_{t-\tau} \\ Y_{t-\tau} \end{bmatrix} + \begin{bmatrix} E_{X_t} \\ E_{Y_t} \end{bmatrix}$$

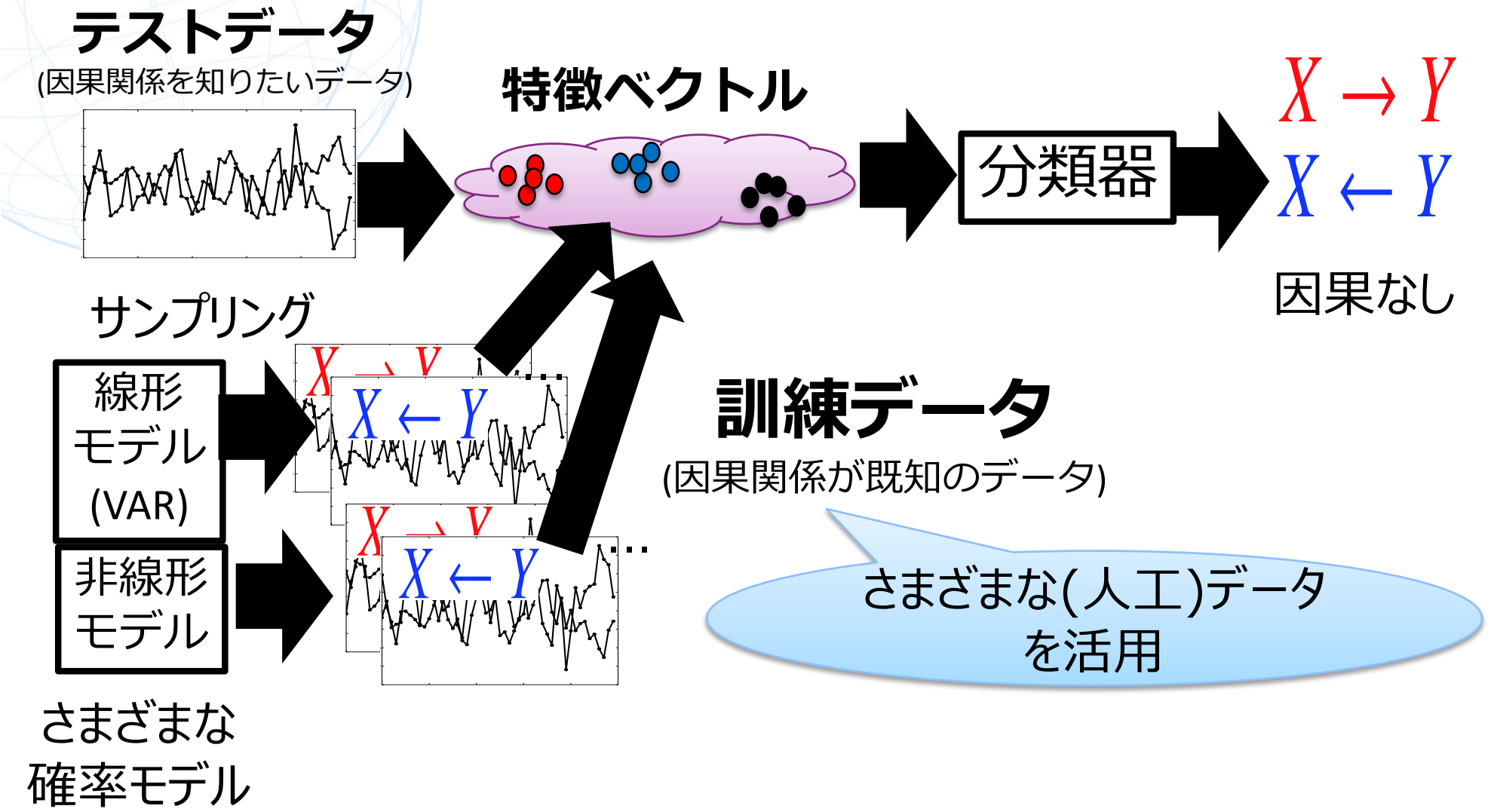
モデルの係数

データ点に合わせて回帰モデルを学習する（最小二乗法）

モデルの係数のゼロ/非ゼロで判定

予測誤差に関する統計量で有意差判定(仮説検定)

提案手法：
教師あり学習に基づくGranger causalityの推定



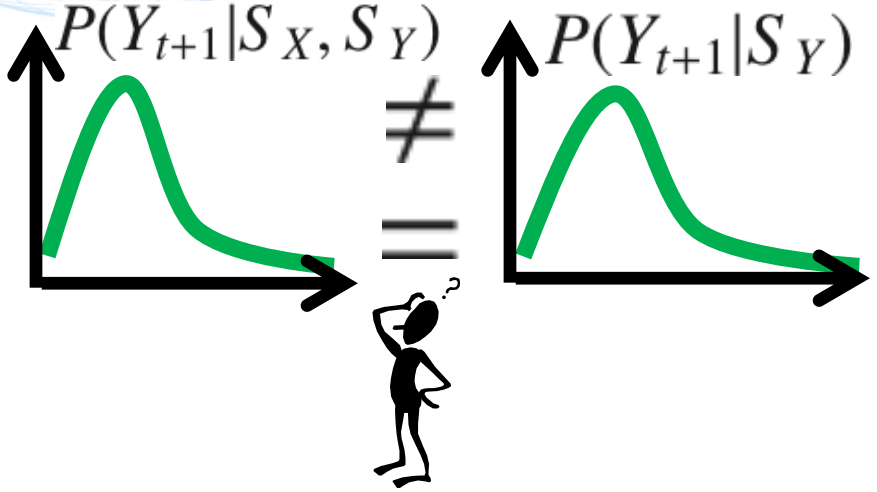
どのように特徴ベクトルを得るか？

– Granger causalityの定義から考える –

特徴量の設計方法

Granger因果性の有無・方向は
２つの(条件付き)分布が等しいか否かで決まる

Granger因果性の
フォーマルな定義

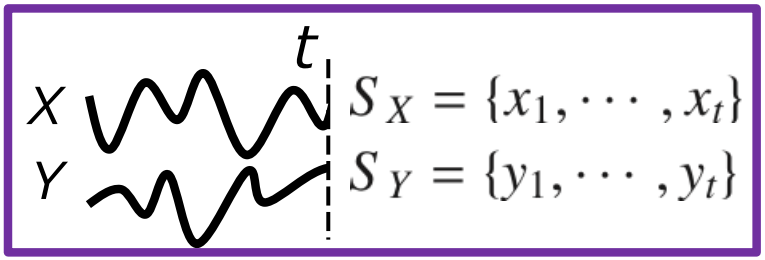


$X \rightarrow Y$
if $P(Y_{t+1}|S_X, S_Y) \neq P(Y_{t+1}|S_Y)$

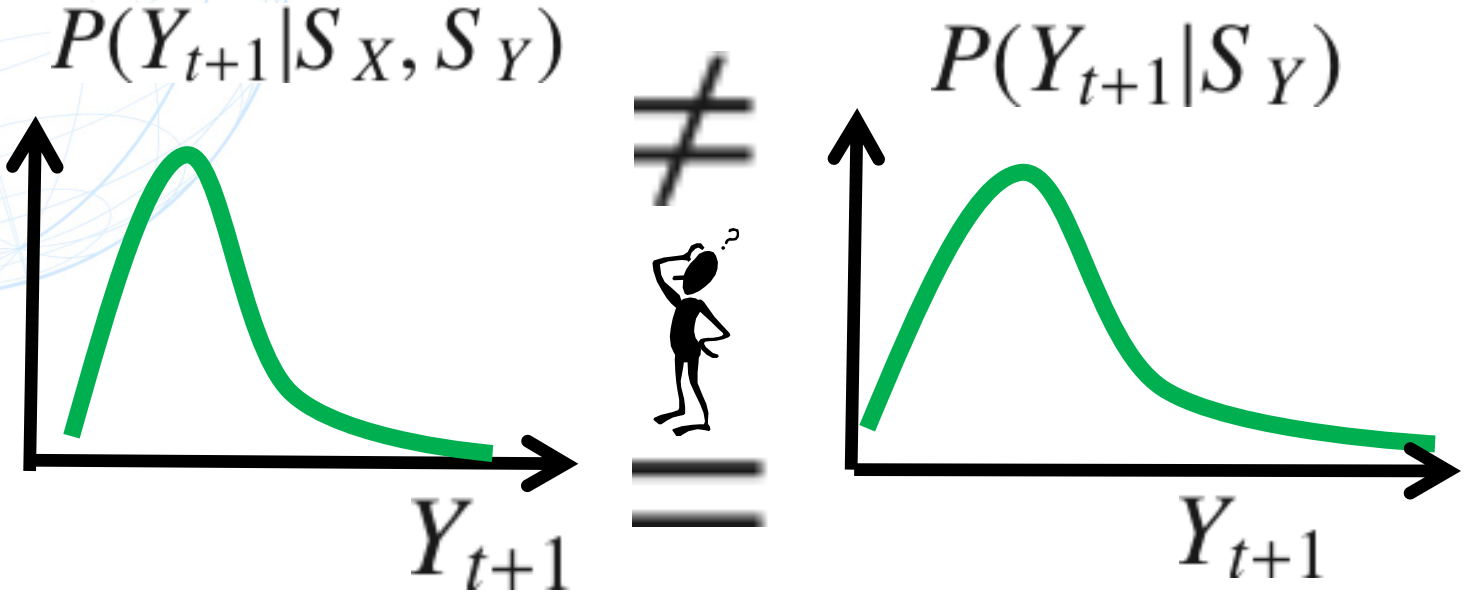
Xを用いると
Yの予測が異なる

$X \quad Y$
if $P(Y_{t+1}|S_X, S_Y) = P(Y_{t+1}|S_Y)$

Xがあってもなくても
Yの予測は変わらない



２つの(条件付き)分布が等しいか否かでGranger因果性の有無はわかる



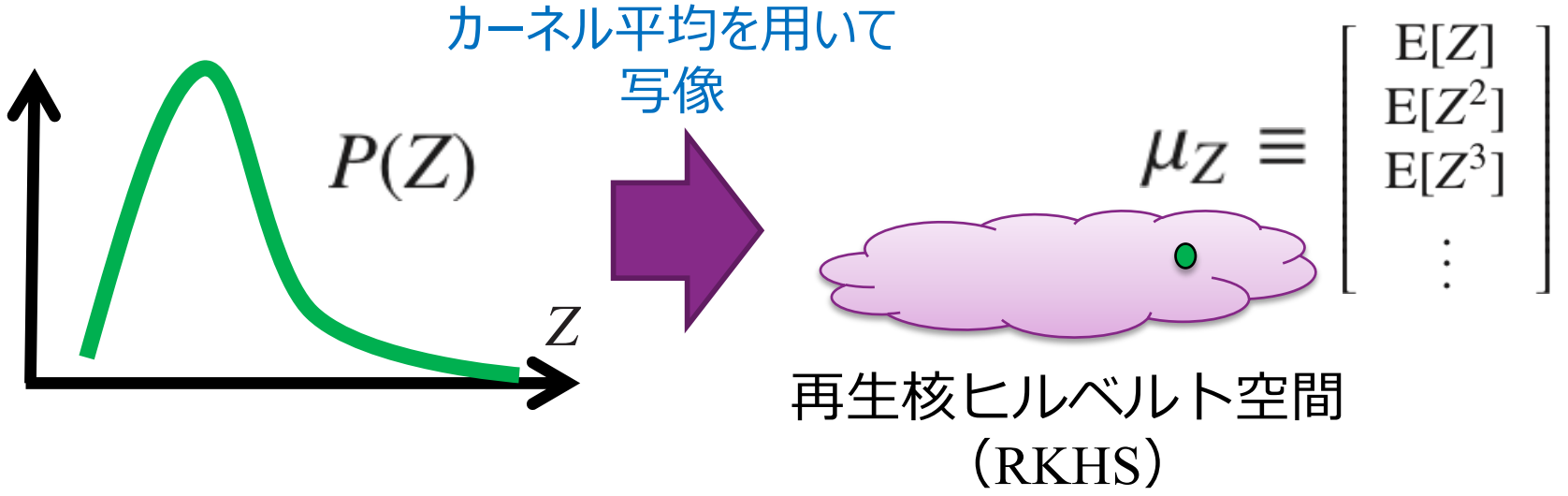
２つの分布が等しいか否かを
判定する特徴量が必要

分布間距離尺度MMDを使う

Kernel meanを用いて確率分布の特徴を抽出

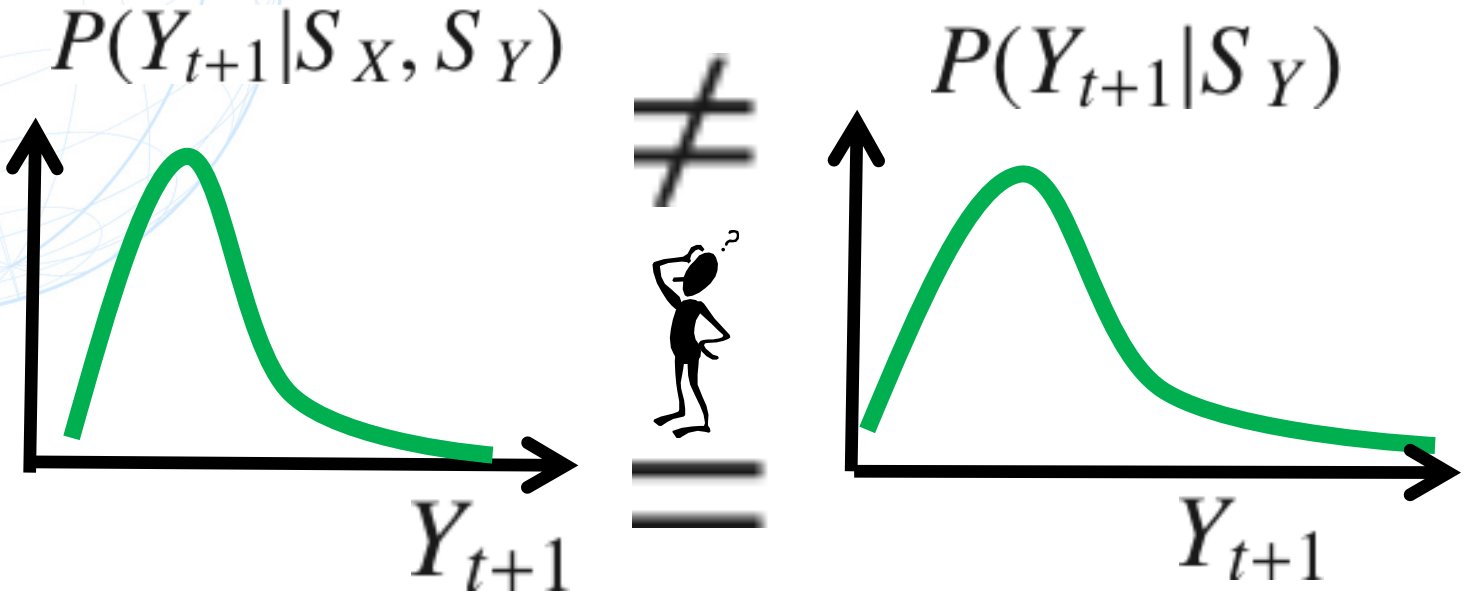
MMD: 条件付き確率分布の違いを**kernel mean間の距離として定量**

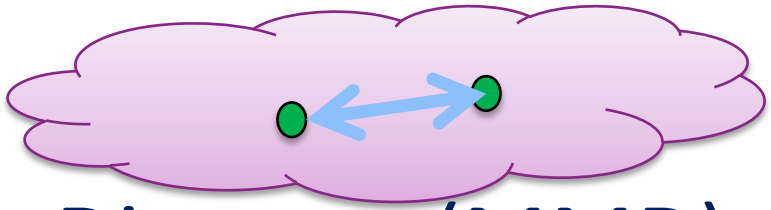
- **kernel mean**とは：分布の平均, 分散, …を返す関数



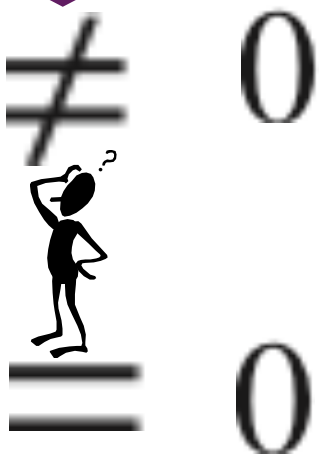
※特殊なカーネル(Gaussian kernel等)を用いれば
異なる２つの分布は同一の点に写像されない(**写像による情報損失がない**)

２つの(条件付き)分布が等しいか否かでGranger因果性の有無はわかる

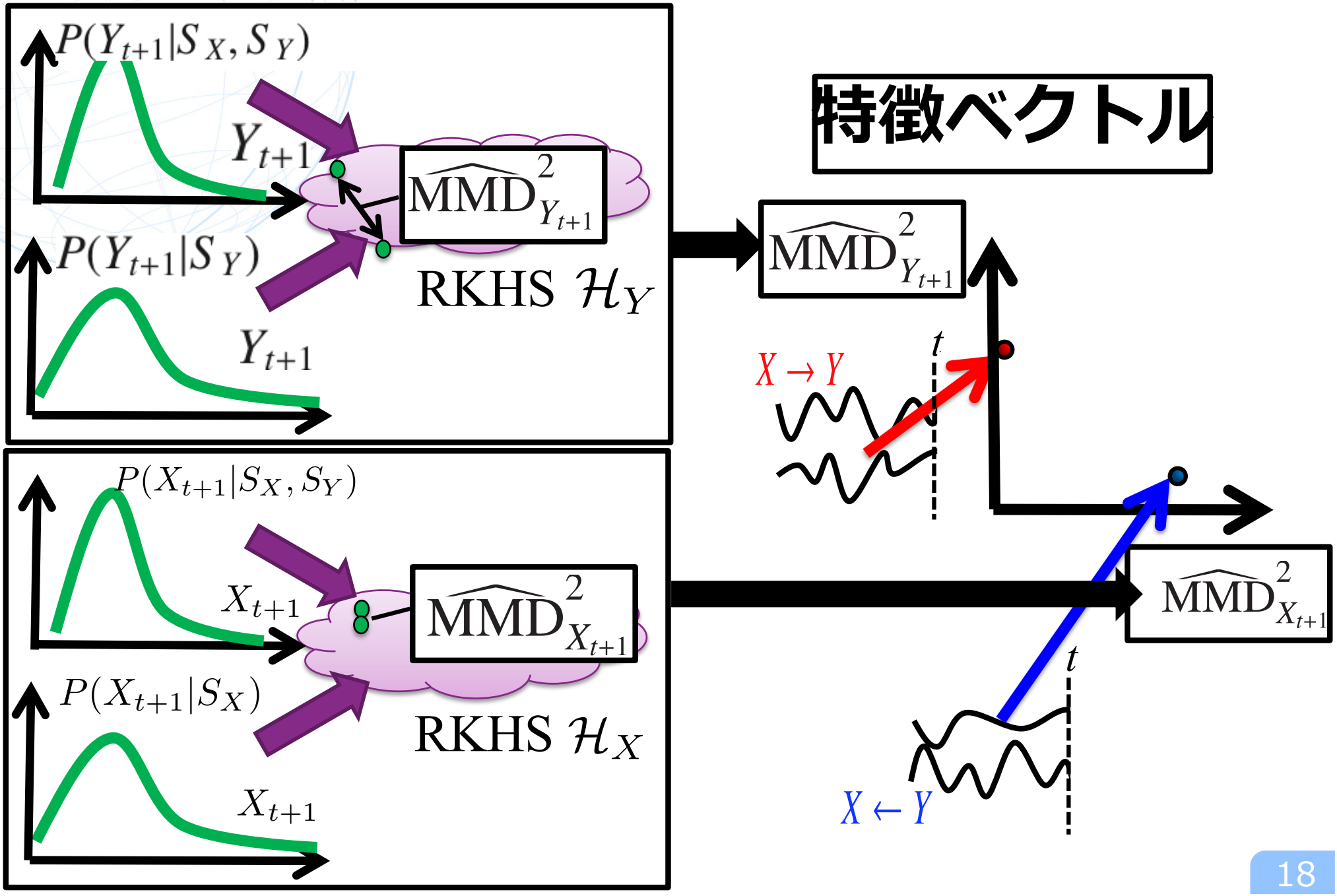




Distance (MMD)
(maximum mean discrepancy
[Gretton+; NIPS2007])



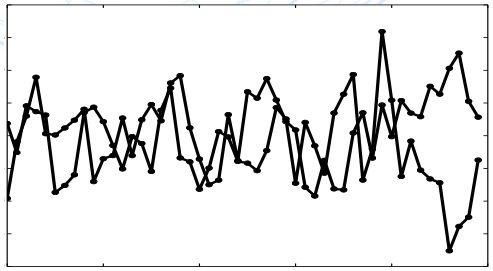
Kernel meanを用いて確率分布の特徴を抽出



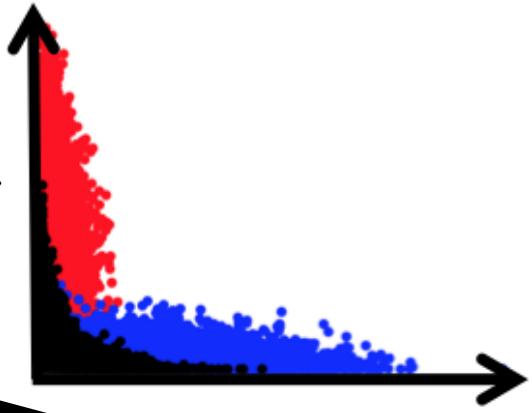
実験

NTT

テストデータ



分類器
（e.g., Random Forest）

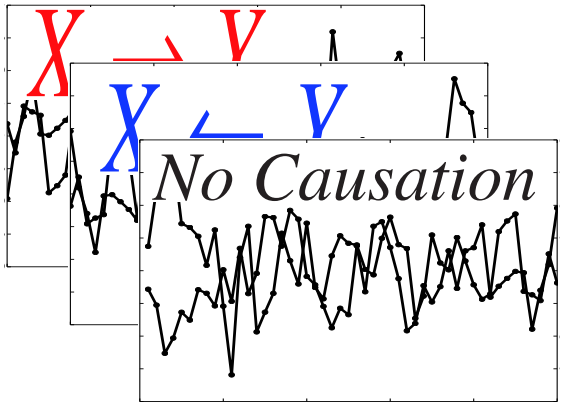


$X \rightarrow Y$

$X \leftarrow Y$

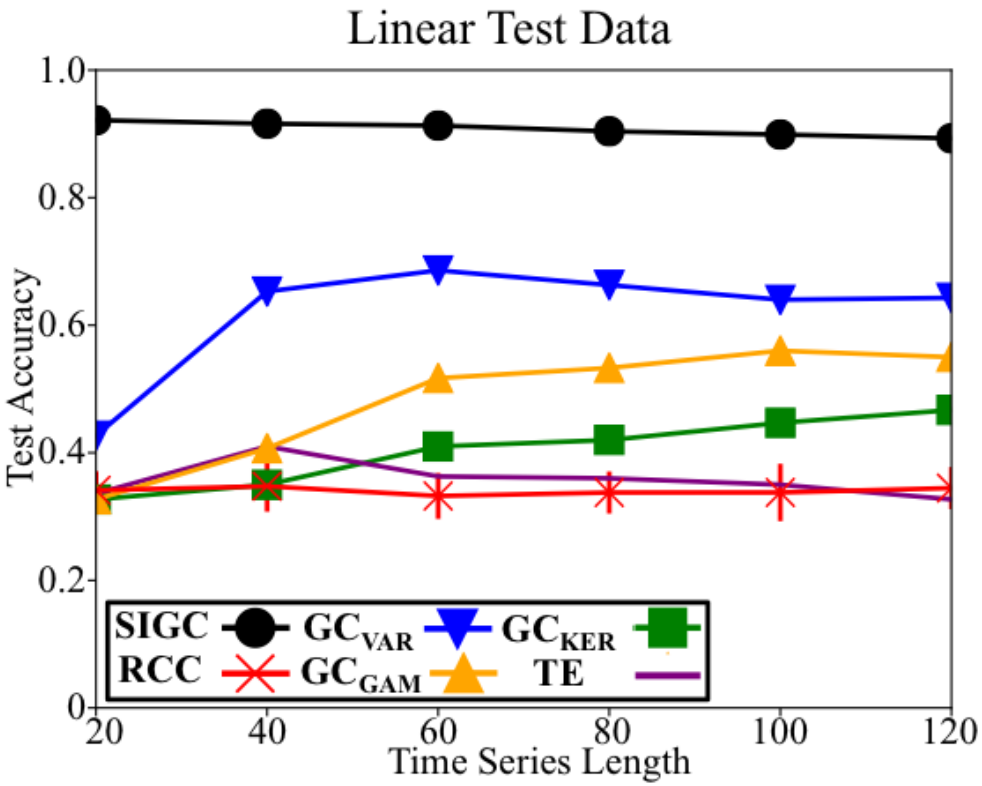
No Causation

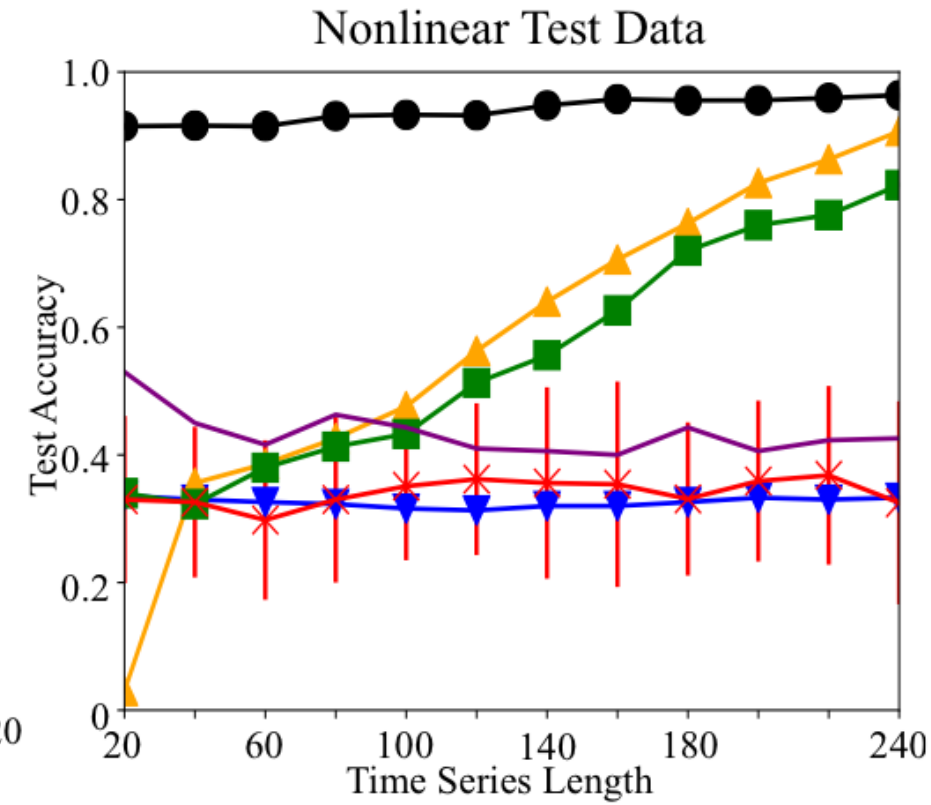
訓練データ
(人工データ)



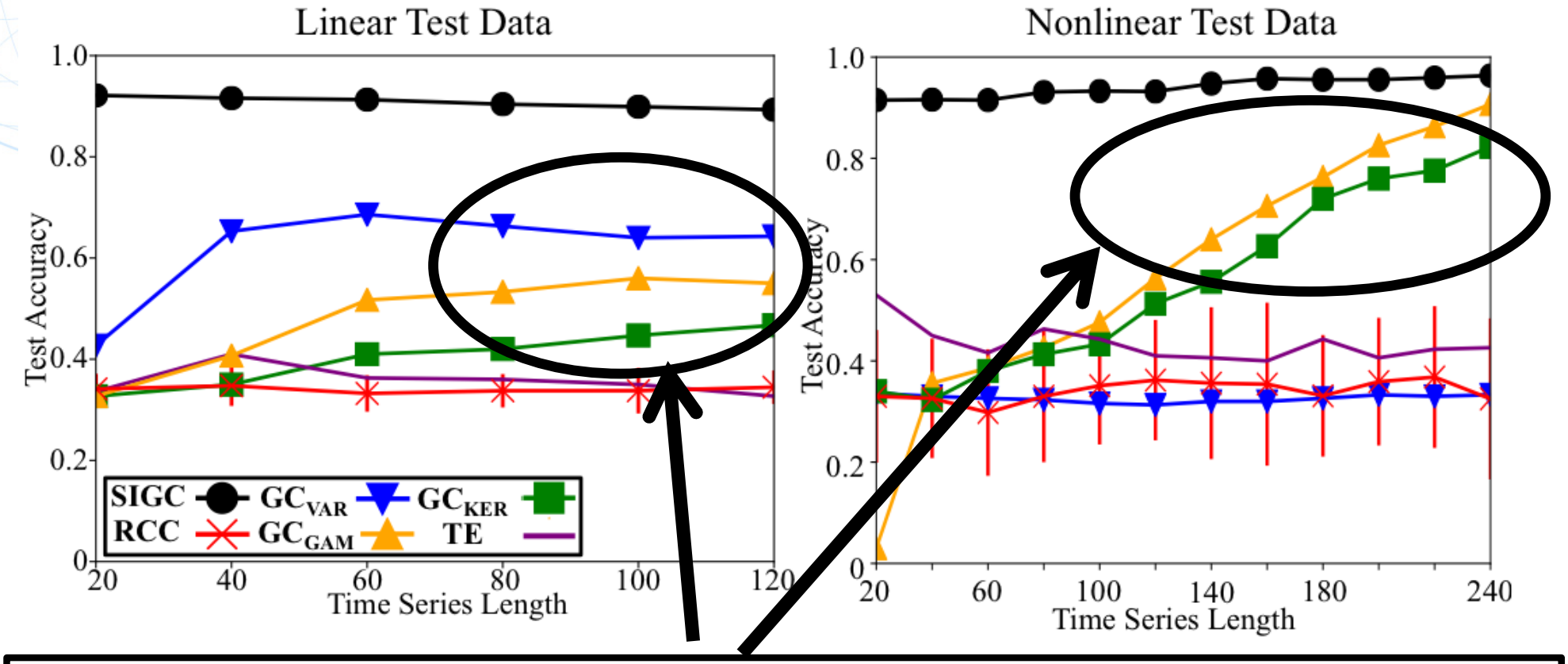
…

人工データ実験


Linear Test Data
Test Accuracy
Time Series Length
1.0
0.8
0.6
0.4
0.2
0
20
40
60
80
100
120
SIGC
GC_VAR
GC_KER
RCC
GC_GAM
TE


Nonlinear Test Data
Test Accuracy
Time Series Length
1.0
0.8
0.6
0.4
0.2
0
20
60
100
140
180
240

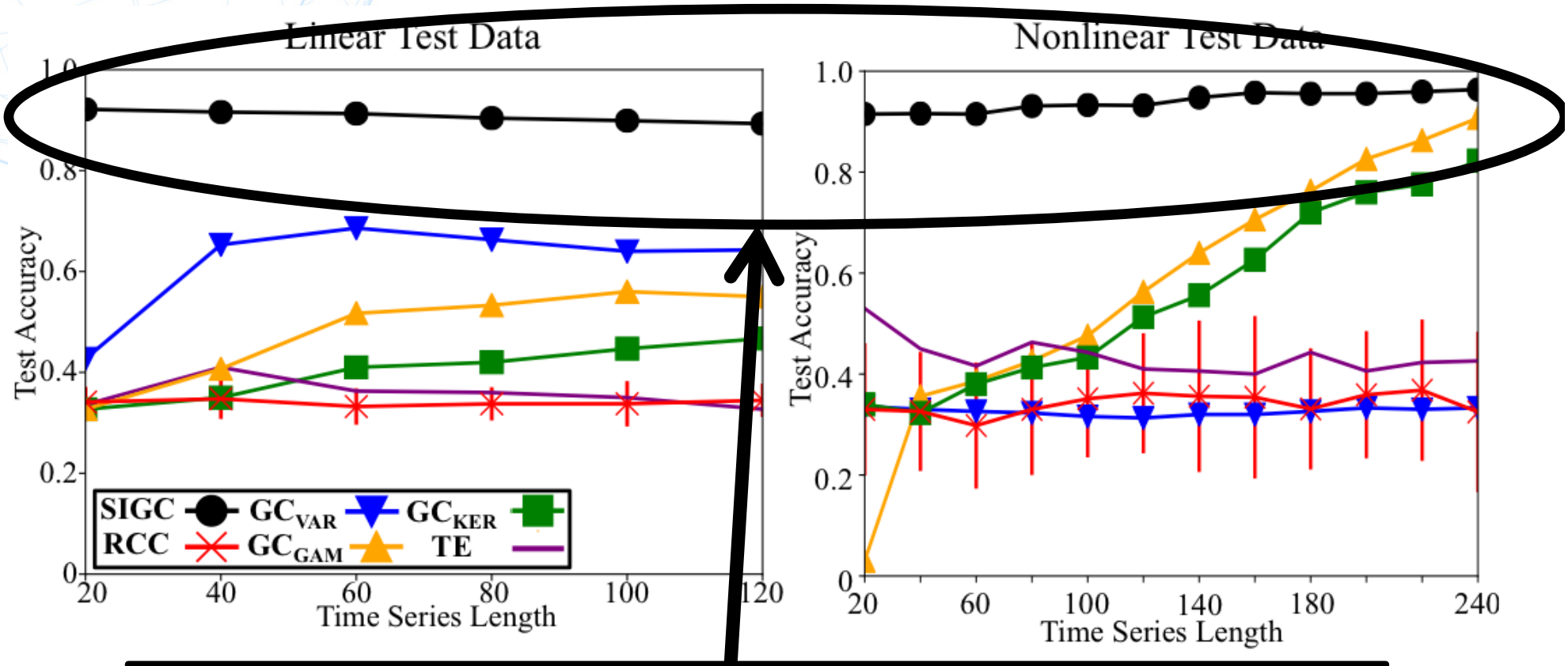
人工データ実験



既存手法

推定精度は選んだ回帰モデルの当てはまり度合いに強く依存

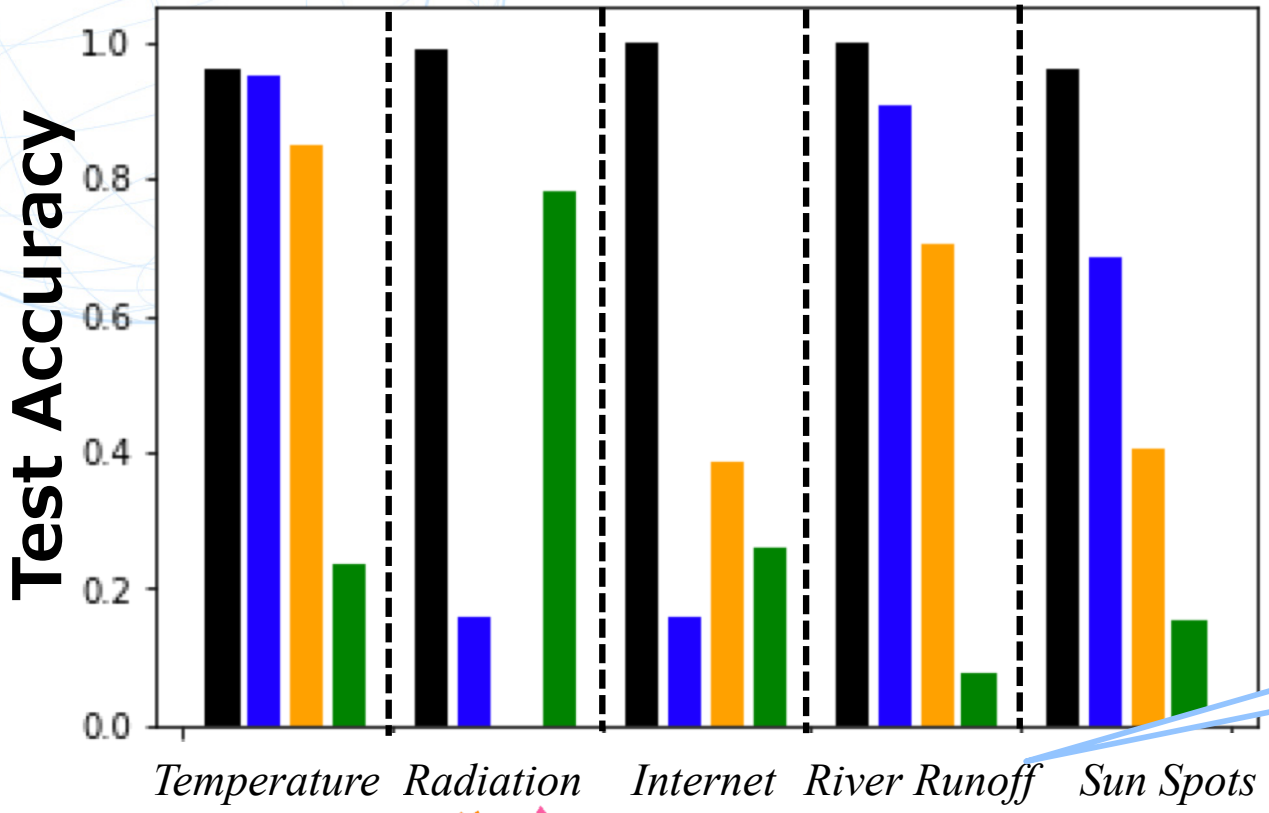
人工データ実験



提案手法

線形・非線形データの両方で高い推定精度

実データ実験



Proposed
Linear (VAR)
Nonlinear (GAM)
Nonlinear (kernel)

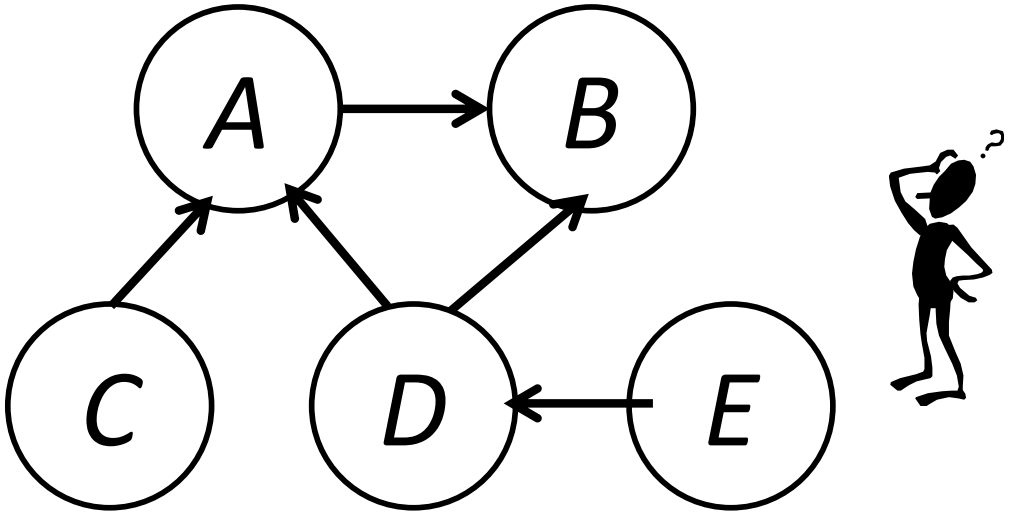
e.g., *River Runoff*
X: 降水量; *Y*: 河川流量
(※真の因果関係: $X \rightarrow Y$)

５つの実データセット

多変数(n変数)への拡張

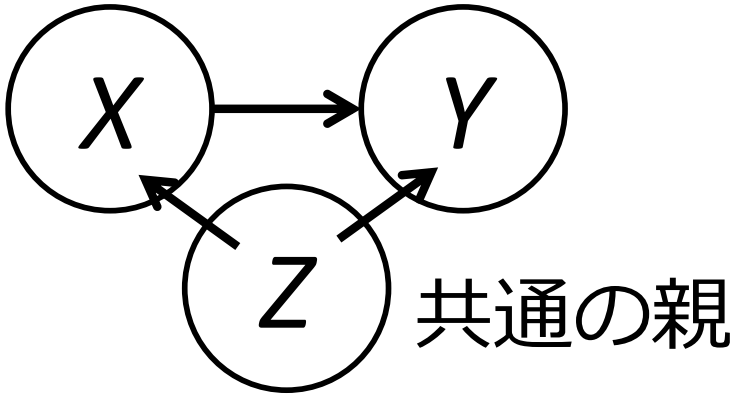
- ${}_nC_2$ 種類の各ペアに因果関係のラベルを割り当てる

$X \to Y$　　$X \leftarrow Y$　　*No Causation*



（補足）多変数(n変数)への拡張

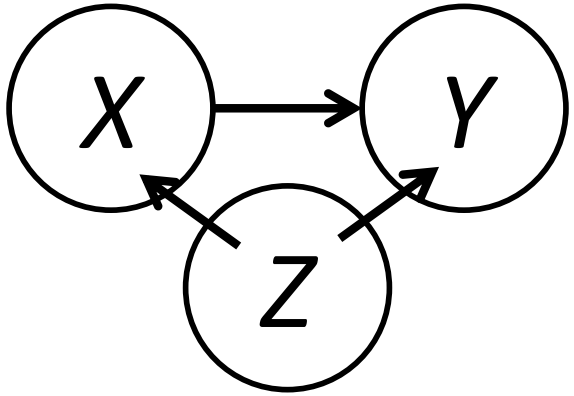
- 仮定：各ペアの共通の親は**高々１つ**
  → **第３の変数の影響のみ考えて３つ組をとる**



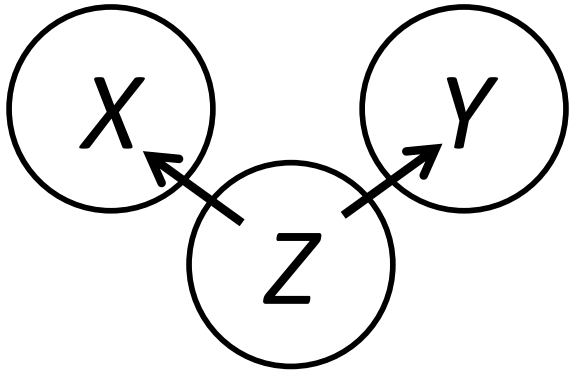
- ペア(X, Y)に対するラベルの割り当て方：
  ✓(n-2)種類の**各３つ組**$(X, Y, Z_v)\ (v \in \{1, \cdots, n-2\})$ **に対し**
  1. 特徴ベクトルを計算
  2. ラベル確率を計算
  ✓ラベル確率の**平均が最大のラベルを出力**

特徴ベクトルの計算

- **Conditional Granger causality** [Geweke JASA1984]:
  第３の変数Zで条件付けた２つの条件付き分布を比較



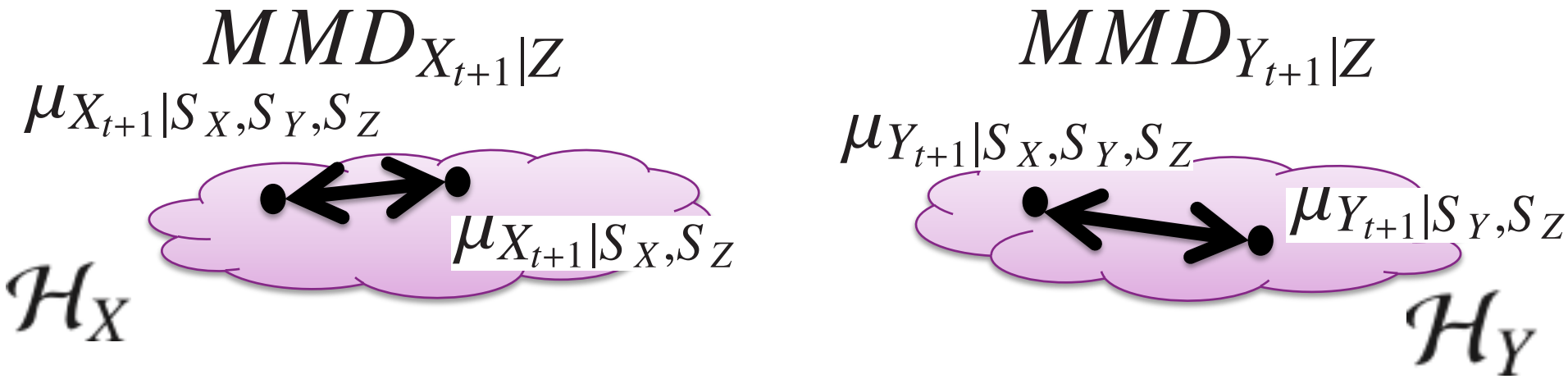
$$\text{if} \quad P(Y_{t+1}|S_X, S_Y, \underline{S_Z}) \neq P(Y_{t+1}|S_Y, \underline{S_Z})$$



$$\text{if} \quad P(Y_{t+1}|S_X, S_Y, \underline{S_Z}) = P(Y_{t+1}|S_Y, \underline{S_Z})$$

特徴ベクトルの計算

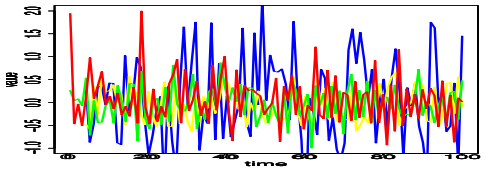
- 同様に求めた条件付き分布間の距離を使って分類の特徴ベクトルを得る



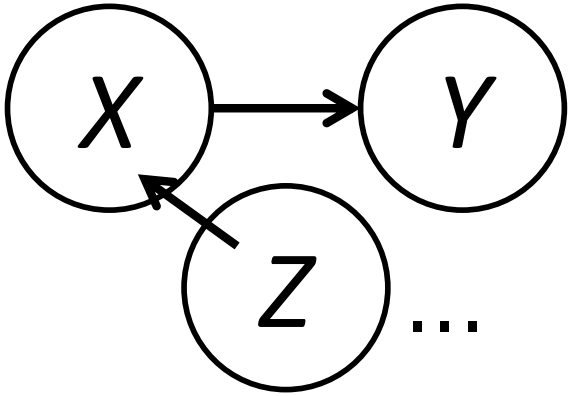
多変数実データで実験

テストデータ

Yeast cell cycle gene expression data
[Spellman+ 1998]
14 variables (genes)



分類器



訓練データ
（大量の３つ組人工時系列）

※真の因果関係：パスウェイデータベースKEGGを参照

結果(macro F1 score, micro F1 scoreで評価)

	$\mathbf{SIGC}_{tri}$	$\mathbf{GC}_{VAR}$	$\mathbf{GC}_{KER}$	$\mathbf{SIGC}_{bi}$	$\mathbf{GC}_{GAM}$	**TE**	**RCC**
macro-averaged F1	**0.483** (0.0)	0.351	0.437	0.431 (0.007)	0.457	0.430	0.407 (0.096)
micro-averaged F1	**0.637** (0.0)	0.436	0.513	0.578 (0.011)	0.567	0.449	0.567 (0.161)

※Higher is better

既存手法よりは高いが**十分な推定精度ではない**

1. 仮定（共通の親が高々１つ）が強すぎる
2. **時系列が短すぎる**($T = 57$), etc.

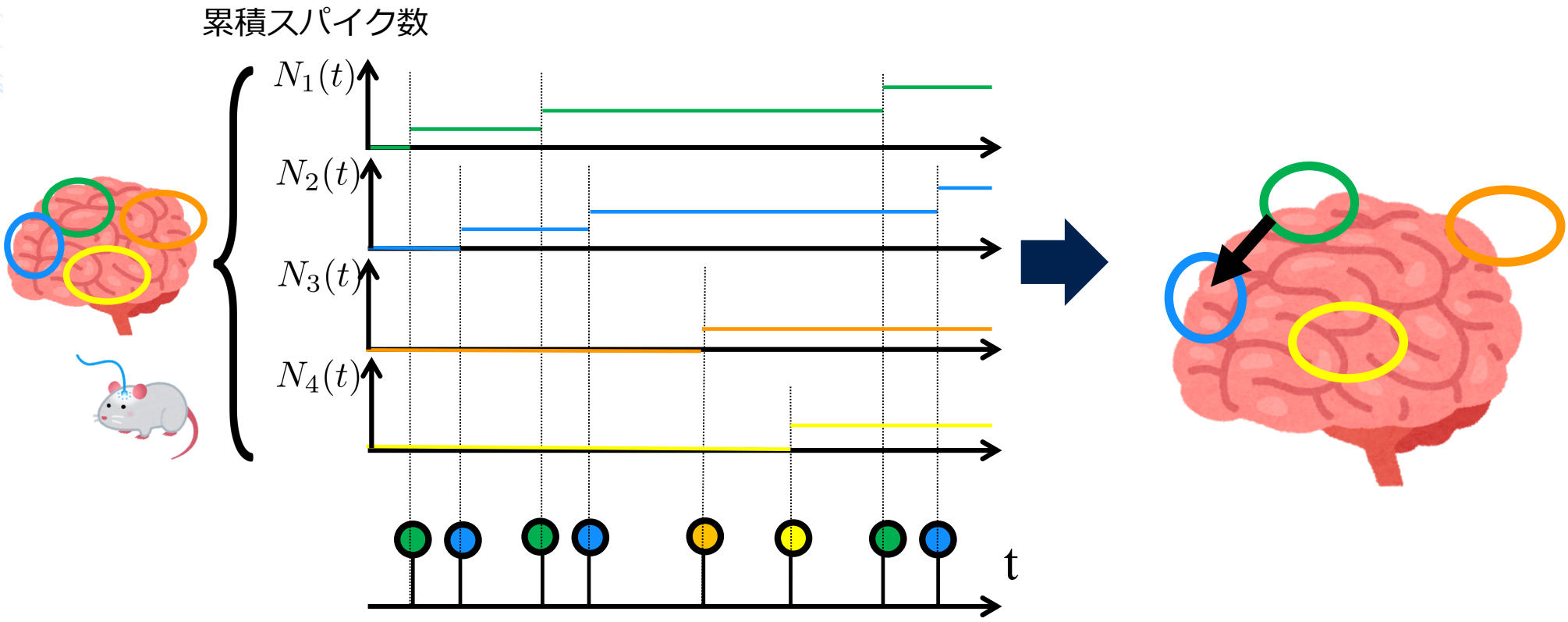
データ取得コストが大きいライフサイエンス分野において
高精度に因果関係の有無・方向を推定するのは未だ難しい

本日の内容

- 時系列データからの因果発見とは
  - ✓応用例
- Granger causalityの推定
  - ✓因果関係とは　– Granger causalityの定義 –
  - ✓教師あり学習に基づくGranger causalityの推定
    [Chikahara+; IJCAI2018(**AI分野最難関会議**)]
- Granger causality研究の最前線
  - ✓イベント間の因果関係 – 連続時間系のGranger –

イベント時系列間のGranger causality

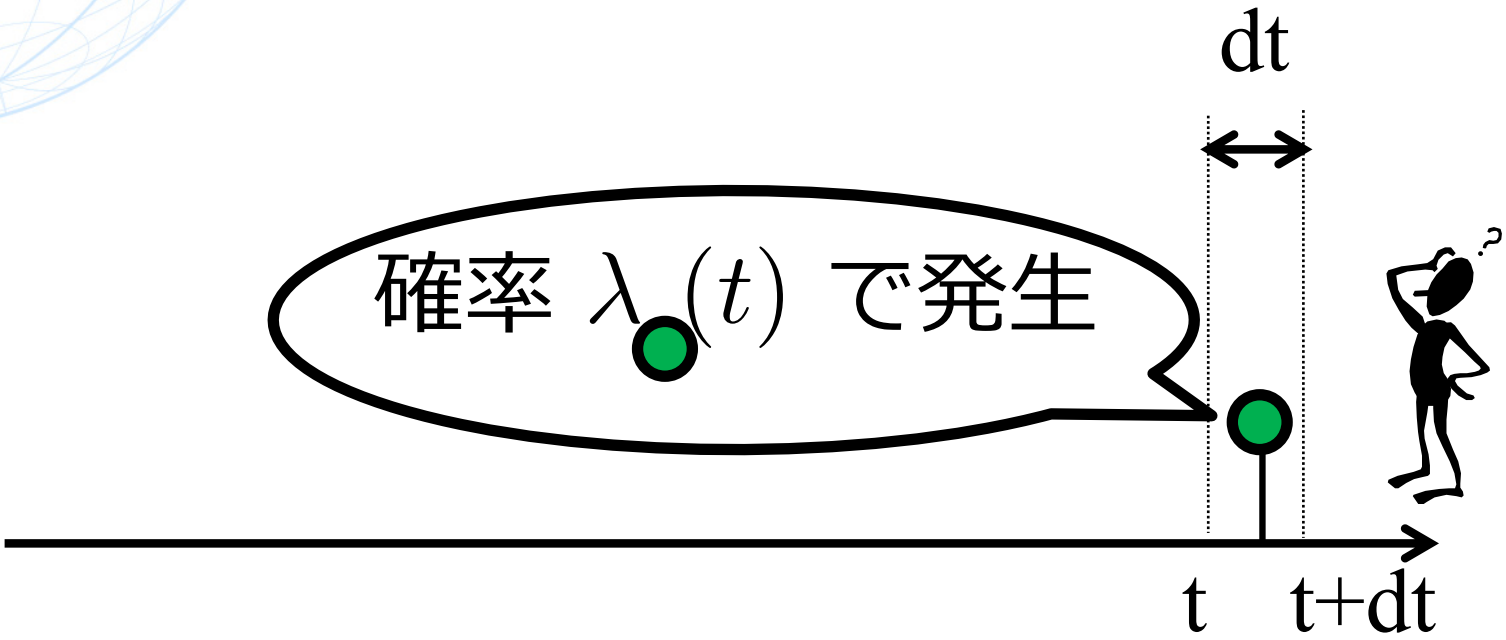
- 不均一な時間間隔で生じる離散的なイベントを考える
  - ✓（例）イベント＝特定の脳領域での脳波スパイク



イベント間のGranger causalityは
点過程モデルを用いて表せる

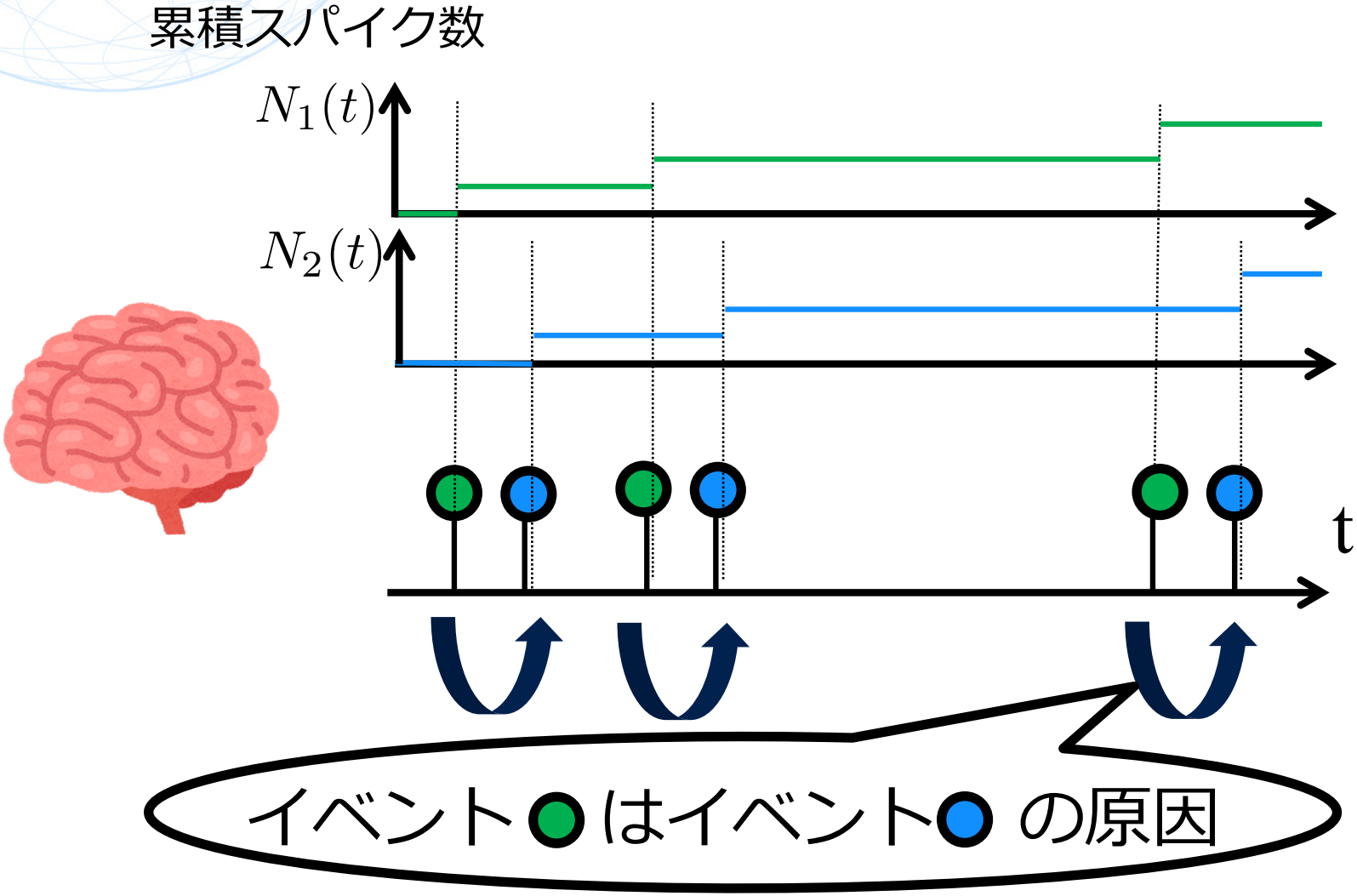
点過程モデル(Point Processes)

- 点過程モデル：微小時間 dt の間にイベントが生じる確率 $\lambda(t)$をモデル化



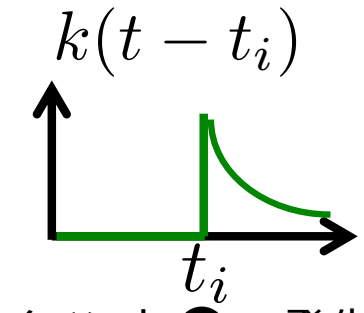
Hawkes過程

- Hawkes過程：イベント発生確率が他のイベントの発生に影響される点過程モデル



Hawkes過程

- イベント● の発生確率：



$$\lambda_{\bullet}(t) = \mu_{\bullet} + \alpha_{\bullet} \sum_{t_i \in \mathcal{H}_{\bullet}(t)} k(t-t_i) + \alpha_{\bullet} \sum_{t_i \in \mathcal{H}_{\bullet}(t)} k(t-t_i)$$

時間によらず
一定確率で発生

過去の自分自身の
イベント発生に影響

過去のイベント● の
発生に影響

Hawkes過程によるGranger causalityの定義

NTT

- イベント● の発生確率：

$$\lambda_{\bullet}(t) = \mu_{\bullet} + \alpha_{\bullet\bullet} \sum_{t_i \in \mathcal{H}_{\bullet}(t)} k(t - t_i) + \alpha_{\bullet\color{green}\bullet} \sum_{t_i \in \mathcal{H}_{\color{green}\bullet}(t)} k(t - t_i)$$

$\alpha_{\bullet\color{green}\bullet} \neq 0$ ならば
イベント● はイベント● の原因
[Eichler+; 2016]

- **推定**：最尤推定でパラメータ μ, α を推定

結論

- 時系列の因果関係の定義の一つとして Granger causalityをご紹介
  - ✓予測の「有用性」で定義
  - ✓離散的なイベントどうしの因果関係も定義可能

- 多種多様な応用が期待される一方、依然として多くの課題がある
  - ✓予測モデルの与え方
  - ✓**非観測変数が影響を及ぼす場合（非常に難しいが重要）**