

因果方向推定に対するメタ学習

福水 健次

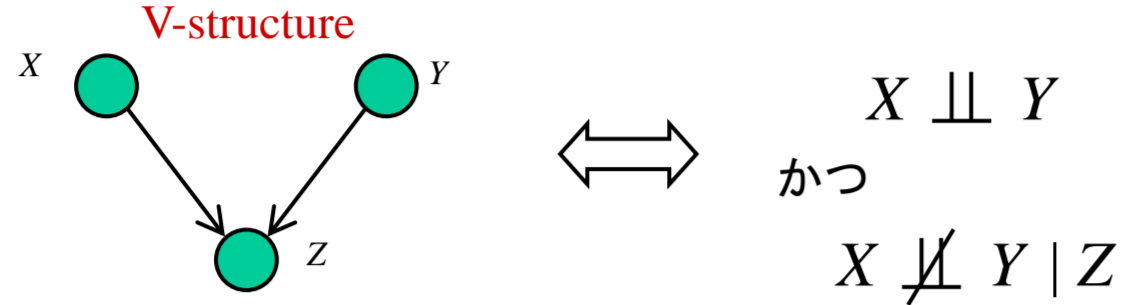
数理・推論研究系 教授

(Jean-Francois Ton (Oxford), Dino Sejdinovic (Oxford)との共同研究. AAAI 2021)

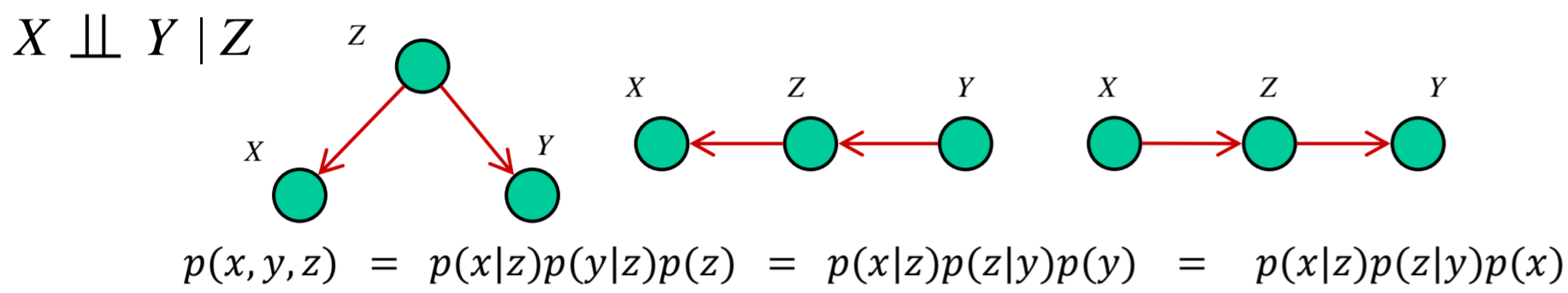
因果方向の推定問題

「非実験」データからの因果推論

- 非実験的, 非時系列データからの因果推論 (Judea Pearl; Spirtes & Glymore) → なぜ可能か?
- DAG : $X - Z - Y$



3変数に対して, この場合のみ同時分布から矢印の向きが決まる. 次のケースは, 区別できない



2変数データからの因果推論

- Markov性 (DAG) に基づく議論は機能しない
- 「因果」をどう定式化するか?
- Informal def.: Independence of cause and mechanism
- “Mechanism”? --- 写像? 条件付確率?
- “Independence”? --- 「統計的」な独立性は使えない



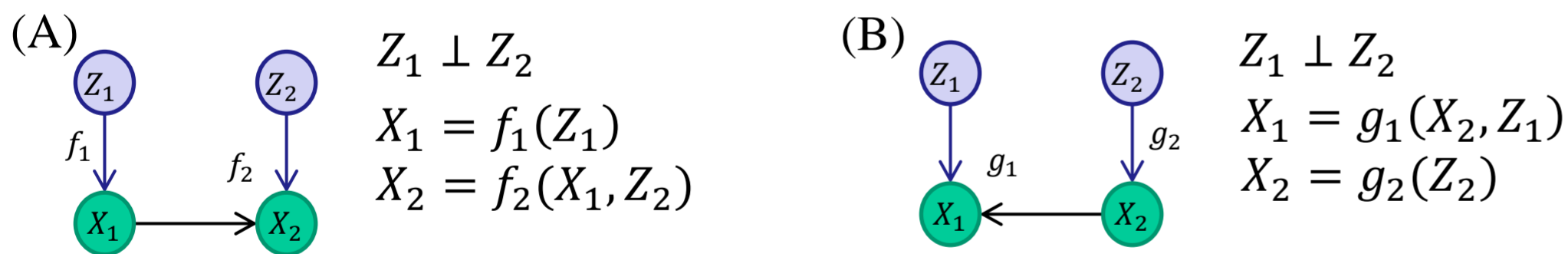
訓練データによる「識別」的なアプローチを考える.

- Additive Noise Model (ANM, Hoyer et al 2019) $Y=f(X)+N$, X と N の独立度合いを測る.
- Conditional Distribution Similarity (CDS, Fonollosa 2016) $X, Y, Y|X, X|Y$ の様々な統計量を特徴として, 識別問題を学習.
- Regression Error based Causal Inference (RECI, Blöbaum et al 2019) 非線形回帰の2乗誤差を比較
- Information Geometric Causal Inference (IGCI, Daniusis et al 2010) $Y=f(X)$ に対し, $\text{PearsonCorr}(X, |\log f'(X)|)$ を2方向に対して比較.
- Randomized Causation Coefficient (RCC, Lopez-Paz 2015) Kernel平均埋め込み $\Phi(X, Y) \rightarrow$ Cause-effect label の識別問題をNNで学習

構造方程式モデル (SEM) による因果方向学習

仮定:
Common hidden cause が無い.
Causal directionが存在する.

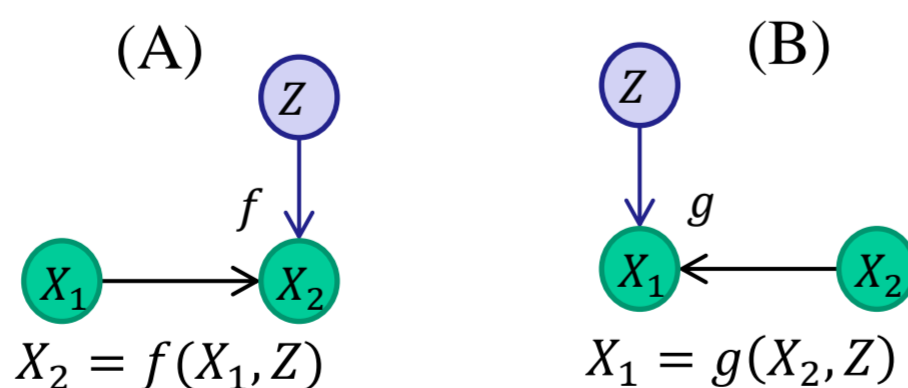
2変数の場合に取りえるSEMは以下のどちらか



Causal Generative Neural Networks (CGNN, Goudet et al 2017)

Data : $\mathbf{X} = \{(X_1^i, X_2^i)\}_{i=1}^m$

モデル(A), (B)の適合度により向き判定



Fitting:

(A) の場合 : $(X_1, \hat{X}_2^A) = (X_1, f(X_1, Z))$, $Z \sim N(0,1)$

f : NNによって (X_1, X_2) の生成モデルを学習 $\rightarrow \hat{\mathbf{X}}^A = \{(X_1^i, \hat{X}_2^{i,A})\}_{i=1}^m$

$MMD_{emp}(\mathbf{X}, \hat{\mathbf{X}}^A)$ を目的関数に用いる

向きの判定

If $MMD_{emp}(\mathbf{X}, \hat{\mathbf{X}}^A) \leq MMD_{emp}(\mathbf{X}, \hat{\mathbf{X}}^B)$, choose (A).

Else choose (B).

MMD_{emp} : 正定値カーネルを用いた分布間距離. サンプルからの計算が容易.

CGNNの問題点:

- よい正解率を出すには, 多くのデータ数が必要
- CGNN論文の人工データ: 訓練データ 1500 (1000 訓練, 500 validation)
- Hyper parameter tuning (MMDに使うカーネル, NNのアーキテクチャ) 多数のカーネル, アーキテクチャの結果 (32-64個) を平均.



Meta-learning を使う:

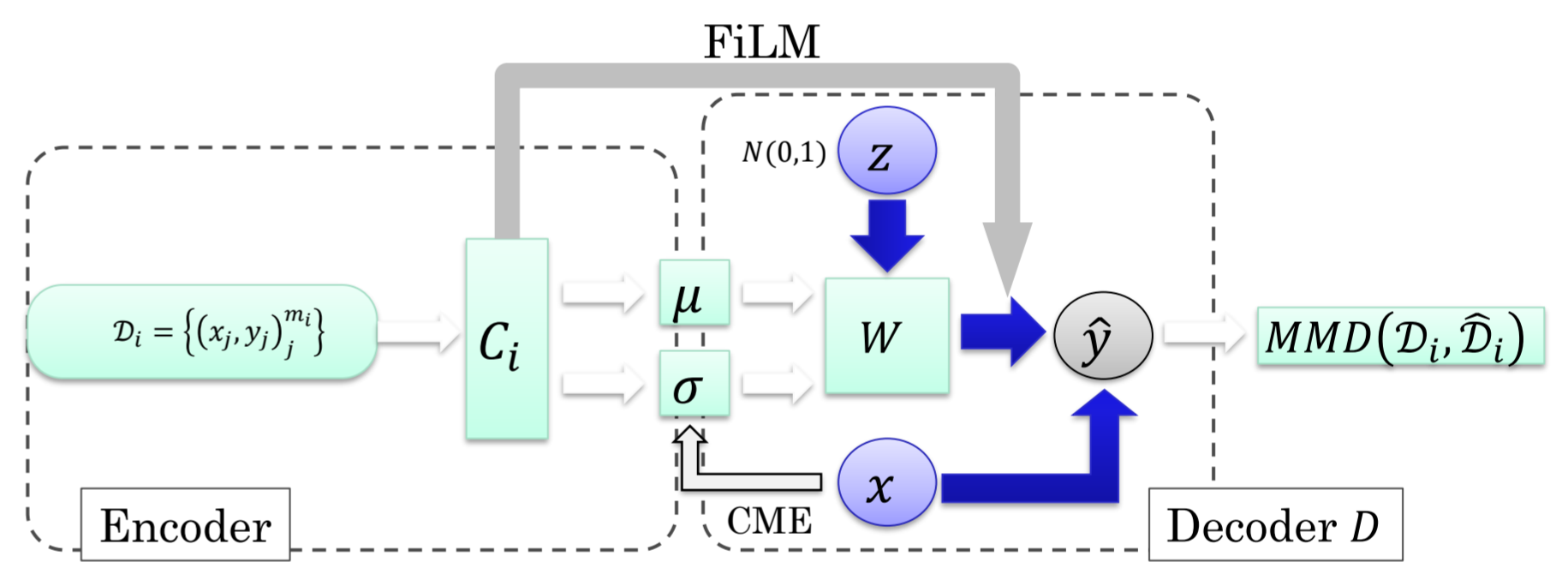
似たCause-effect pairの結果を利用する.
少ないデータ数でもcause-effectの向きを判定可能にする

Meta-CGNN (Tong, Sejdinovic, Fukumizu. AAAI 2021)

- N 個のcause-effect pairs: $\mathcal{D}_i = \{(X_j^i, Y_j^i)\}_{j=1}^{m_i}$ $i = 1, \dots, N$
- Direction $X^i \rightarrow Y^i$.
- Dataset feature: $C_i = \Phi(\mathcal{D}_i)$: Conditional kernel mean embedding (CME) または Deep set
- $\hat{Y}_j^i = F(X_j^i, Z_j^i, C_i)$ Datasetすべてに共通の写像 F を作る.
 $\rightarrow \hat{\mathcal{D}}_i := \{(X_j^i, \hat{Y}_j^i)\}_{j=1}^{m_i}$
- $F = \arg \min_F \sum_{i=1}^N MMD_{emp}^2(\mathcal{D}_i, \hat{\mathcal{D}}_i)$
- Test: $\mathcal{D} = \{(X_j, Y_j)\}_{j=1}^m$. $\hat{Y}_j = F(X_j, Z_j, C_i)$, $\hat{X}_j = F(Y_j, Z_j, C_i)$
- If $MMD_{emp}^2(\mathbf{X}, \hat{\mathbf{Y}}) < MMD_{emp}^2(\mathbf{Y}, \hat{\mathbf{X}})$, then $X \rightarrow Y$, else $Y \rightarrow X$.

構成

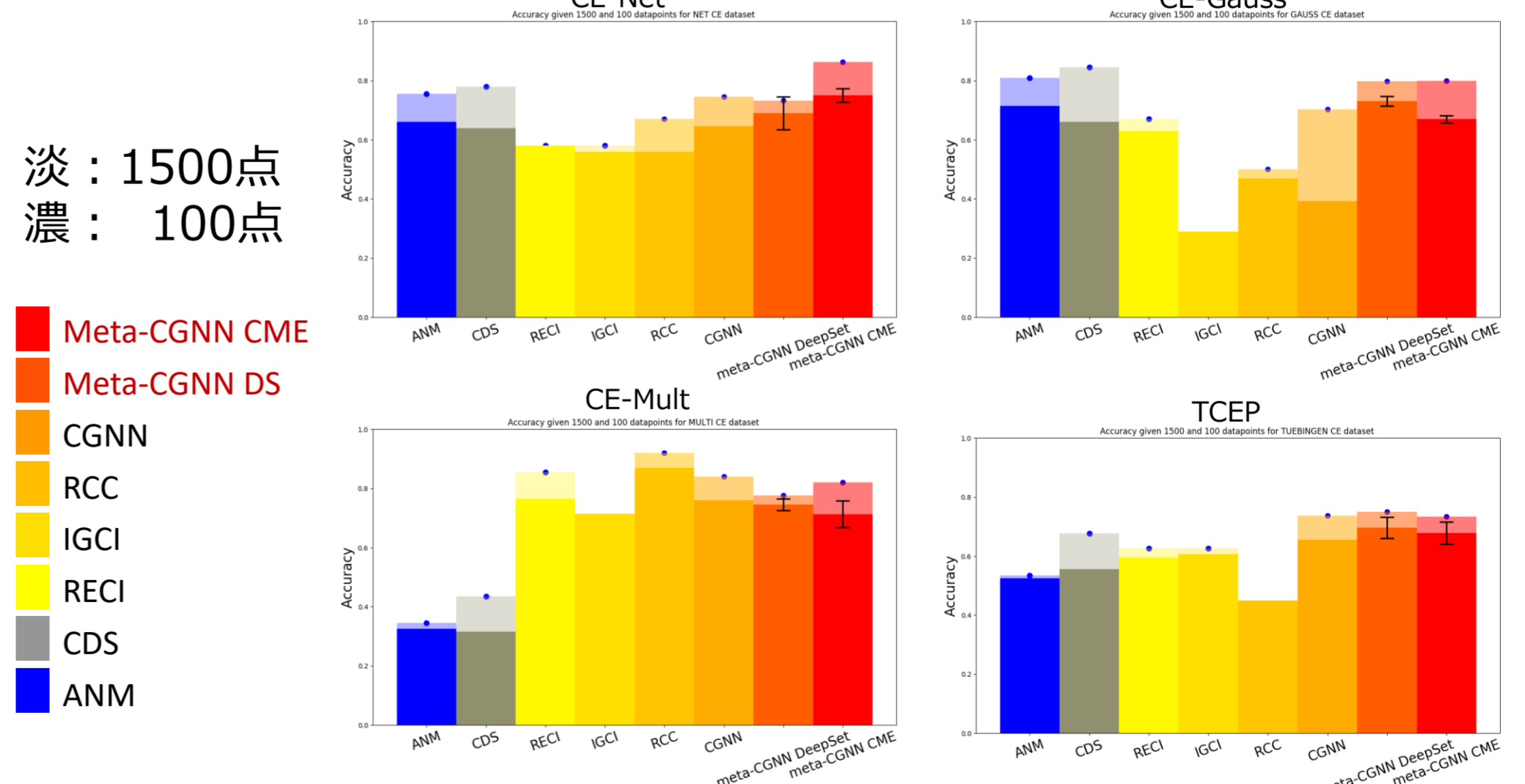
- Encoder: dataset feature $C_i = \Phi(\mathcal{D}_i)$ を計算.
- Dataset feature C_i は, Feature-wise Linear Modulation (FiLM, Perez et al AAAI 2018)に入力.
- Noise modulation: $\hat{y} = D(x, W)$, $W = \mu(C_i) + \sigma(C_i)Z$



数値実験

- データ
 - (A). CE-Net. 入力分布はランダムに生成, メカニズムはランダムNN.
 - (B). CE-Gauss. メカニズムをGPにより生成.
 - (C). CE-Mult. メカニズムはランダム linear/polyn. 加法/乗法ノイズ
- (A)-(C): 300組生成. 100で訓練, 200でテスト.
各データセット 1500点 or 100点
- (D) Tübingen Cause-Effect Pairs : 99組の実/人工データ
- 5-fold CV により評価.

正答率



参考文献: J.-F. Ton, D. Sejdinovic, K. Fukumizu Meta Learning for Causal Direction. AAAI 2021.