FY2023/2023年度
Causal Inference Team
Shohei Shimizu
因果推論チーム　清水 昌平

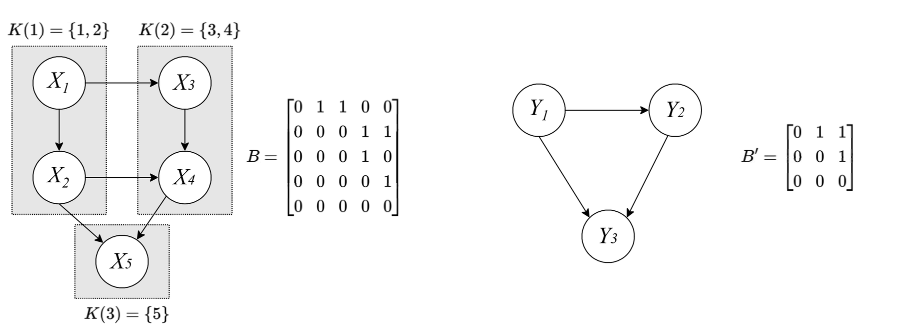
Causal Discovery
因果探索

1. Structure Learning for Groups of Variables in Nonlinear Time-Series Data with Location-Scale Noise (Kikuchi & Shimizu, 2023)

Model:

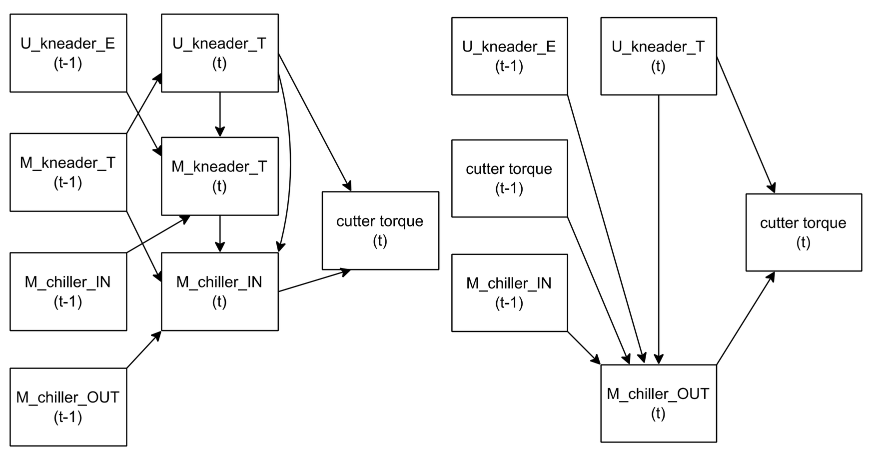
$$X_j^t = f_j\left(\mathrm{PA}_j^t, \ldots, \mathrm{PA}_j^{t-L}\right) + s_j\left(\mathrm{PA}_j^t, \ldots, \mathrm{PA}_j^{t-L}\right) N_j^t, \quad j = 1, \ldots, P$$

Capable of incorporating prior knowledge on grouping of variables



(a) Variable DAG (b) Group DAG

Figure 1: Example of a variable DAG and a group DAG



(a) TS-LSNM-ACY (b) NTS-NOTEARS

Figure 5: Estimated group DAGs for ceramic substrate manufacturing process data

2. Case studies of causal discovery (Jiang & Shimizu, 2023)

- Studied causal relations among financial markets in Japan and USA
- Found a hypothesis that previous day's US market influences the following day's Japanese market for both stocks and bonds, and the bond markets of the previous day impact the following day's FX market directly and the following day's Japanese market indirectly.

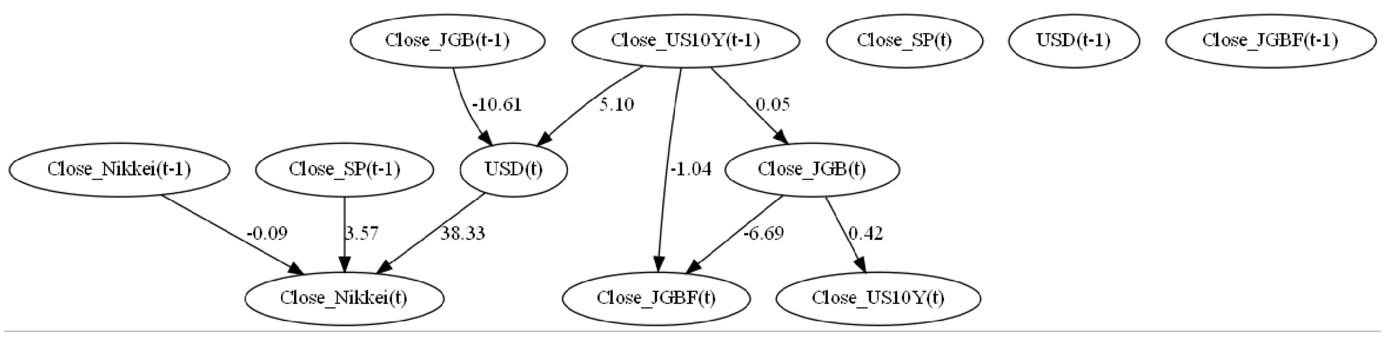


Figure 5: Output DAG of VAR-LiNGAM with Domain Knowledge

Causal Effect Estimation
因果効果推定

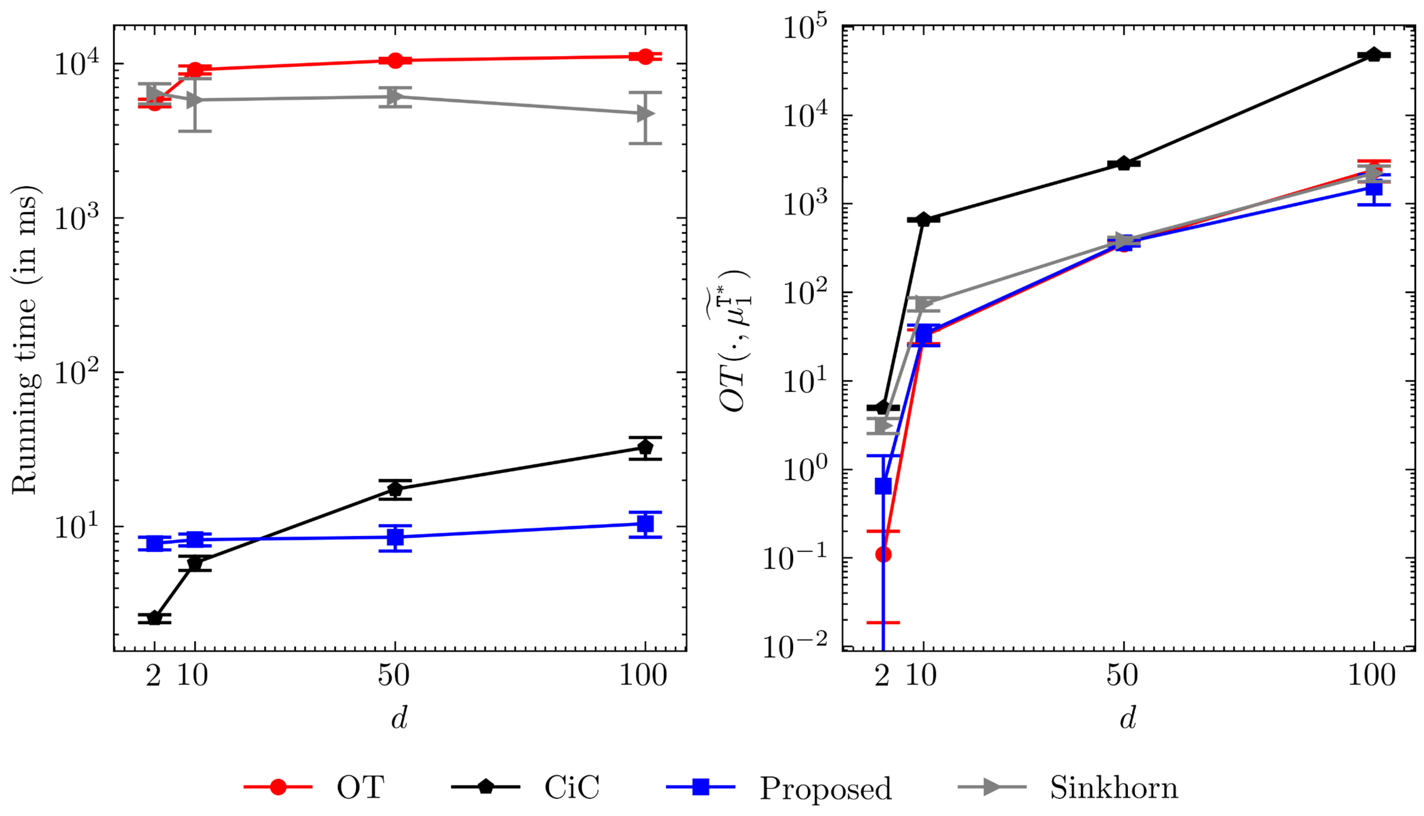
Thong Pham, Shohei Shimizu, Hideitsu Hino, and Tam Le. "Scalable Counterfactual Distribution Estimation in Multivariate Causal Models". To appear in *CLeaR 2024*

- Estimating the multivariate counterfactual distribution in Differences-in-Differences setting
- Existing methods are either accurate but slow and prone to overfitting in high-dimensions, e.g., optimal transport (OT) and Sinkhorn, or fast but inaccurate, e.g., CiC (Changes-in-Changes)
- Our method is based on robust OT and is the best in both running time and accuracy in high dimensions

$$\max_{\omega \in \Omega} \inf_{\pi(x,y) \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} d(\langle \omega, x \rangle, \langle \omega, y \rangle) \mathrm{d}\pi(x,y)$$

Ω: set of directions for finding the robust direction ω

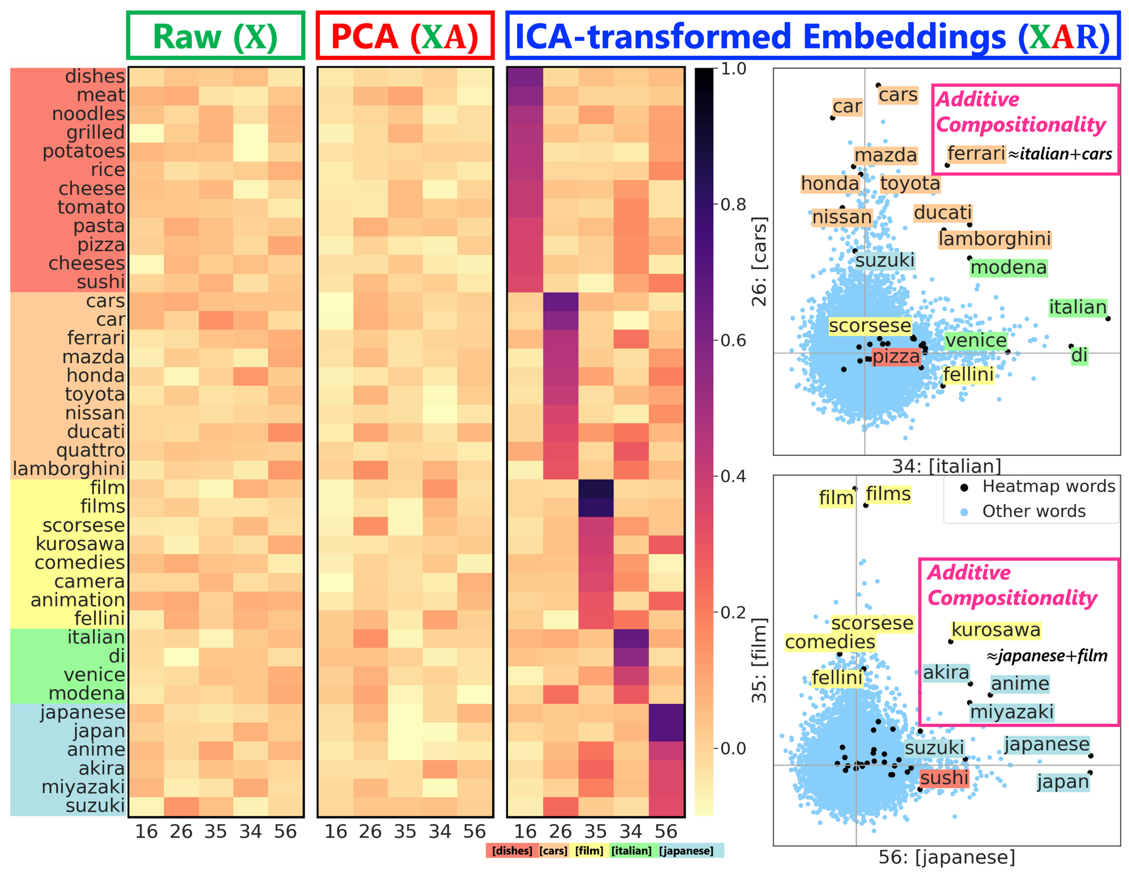
$\Pi(\mu,\nu)$: set of transportation plans between μ and ν, distributions on $\mathbb{R}^d$

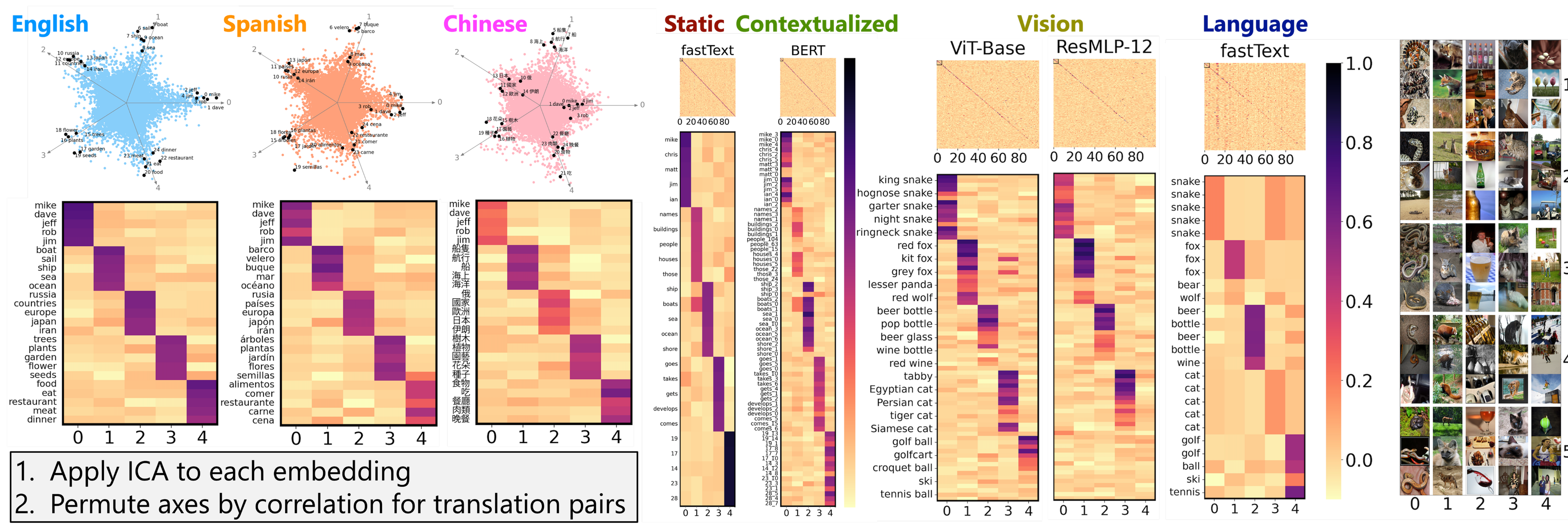


Natural Language Processing
自然言語処理

独立成分に基づく埋め込み表現の解釈と普遍的形状の解明
Discovering Universal Geometry in Embeddings with ICA (EMNLP2023)

- 単語や画像の埋め込みに独立成分分析(ICA)を適用して得られる成分が解釈可能であることがわかった.
- 解釈可能な成分の言語・モデル・ドメインの違いを超えた普遍性が明らかになった.





1. Apply ICA to each embedding
2. Permute axes by correlation for translation pairs

依存関係の大きさは意味の関連性を表す
Norm of Word Embedding Encodes Information Gain (EMNLP2023)

$$\mathrm{KL}(p(\cdot|w) \parallel p(\cdot)) = \sum_{w' \in V} p(w'|w) \log \frac{p(w'|w)}{p(w')}$$

$p(\cdot|w)$: Co-occurrence Distribution of Word w

$p(\cdot)$: Unigram Distribution of the Corpus

- 共起分布とユニグラム分布の乖離をKL情報量で測定したものが「単語が持つ情報の大きさ」を表すという定式化をした.
- このKL情報量は単語ベクトルの長さの2乗に相当することを理論と実験で示した.

$$\|\tilde{u}_w\|^2 \simeq 2\mathrm{KL}(w)$$

- $\tilde{u}_w$: Whitening-like Transformed u_w

$$\tilde{u}_w := \mathrm{Cov}(v_w)^{\frac{1}{2}}(u_w - \bar{u})$$

- $\mathrm{KL}(w) = \mathrm{KL}(p(\cdot|w) \parallel p(\cdot))$