

因果推論チーム： 因果探索法の理論と応用

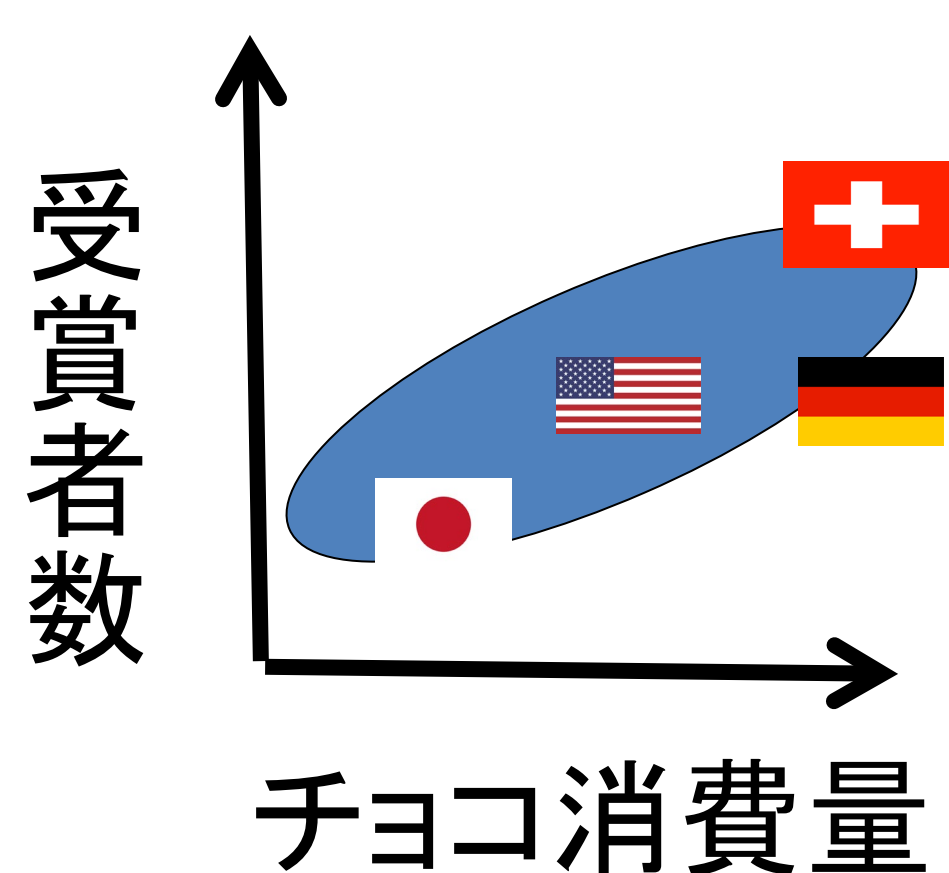
観察データによる因果仮説の探索

実験を行うには仮説が必要

実験が倫理的に難しい場合もある

実験の前に観察データで仮説を発見+絞込み

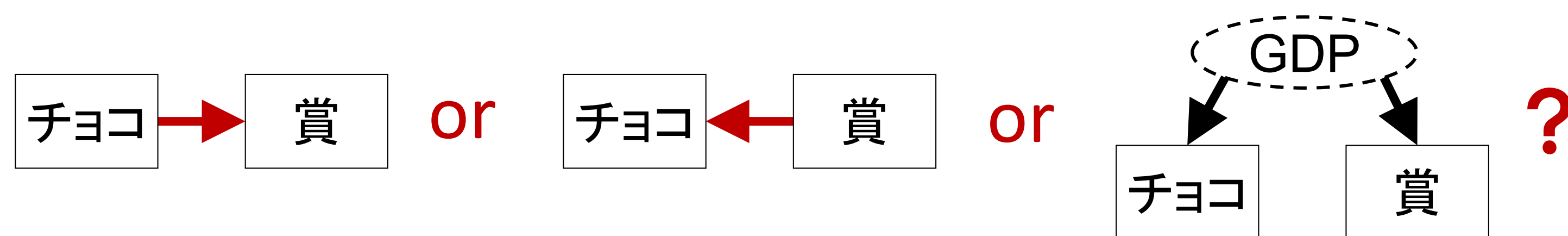
観察データの難しさ



ギャップを
埋めるには？

潜在共通原因
に対処!

GDP



(Maurage et al., 2013, Journal of Nutrition)

複数の因果関係がそのようなデータを与え得る

1. 因果関係が決定論的な場合の因果方向推定

P. Blöbaum, D. Janzing, T. Washio, S. Shimizu, B. Schölkopf
(AISTATS2018; PeerJ, 2019)

モデル:

$$E_{\alpha} := \phi(C) + \alpha N$$

Independence postulate:

$$\text{Cov}[\phi', \text{Var}[N|C]p_C] = 0.$$

主結果:

Theorem 1 (Error Asymmetry) Let the assumptions 1-5 in Section 2.3 hold. Then the following limit always holds

$$\lim_{\alpha \rightarrow 0} \frac{\mathbb{E}[\text{Var}[C|\tilde{E}_{\alpha}]]}{\mathbb{E}[\text{Var}[\tilde{E}_{\alpha}|C]]} \geq 1,$$

with equality only if the functional relation in Assumption 3 is linear.

確率的グラフィカルモデルにおける高速 推論手法の提案 (特別研究員 Li Chao):

確率的グラフィカルモデルとは

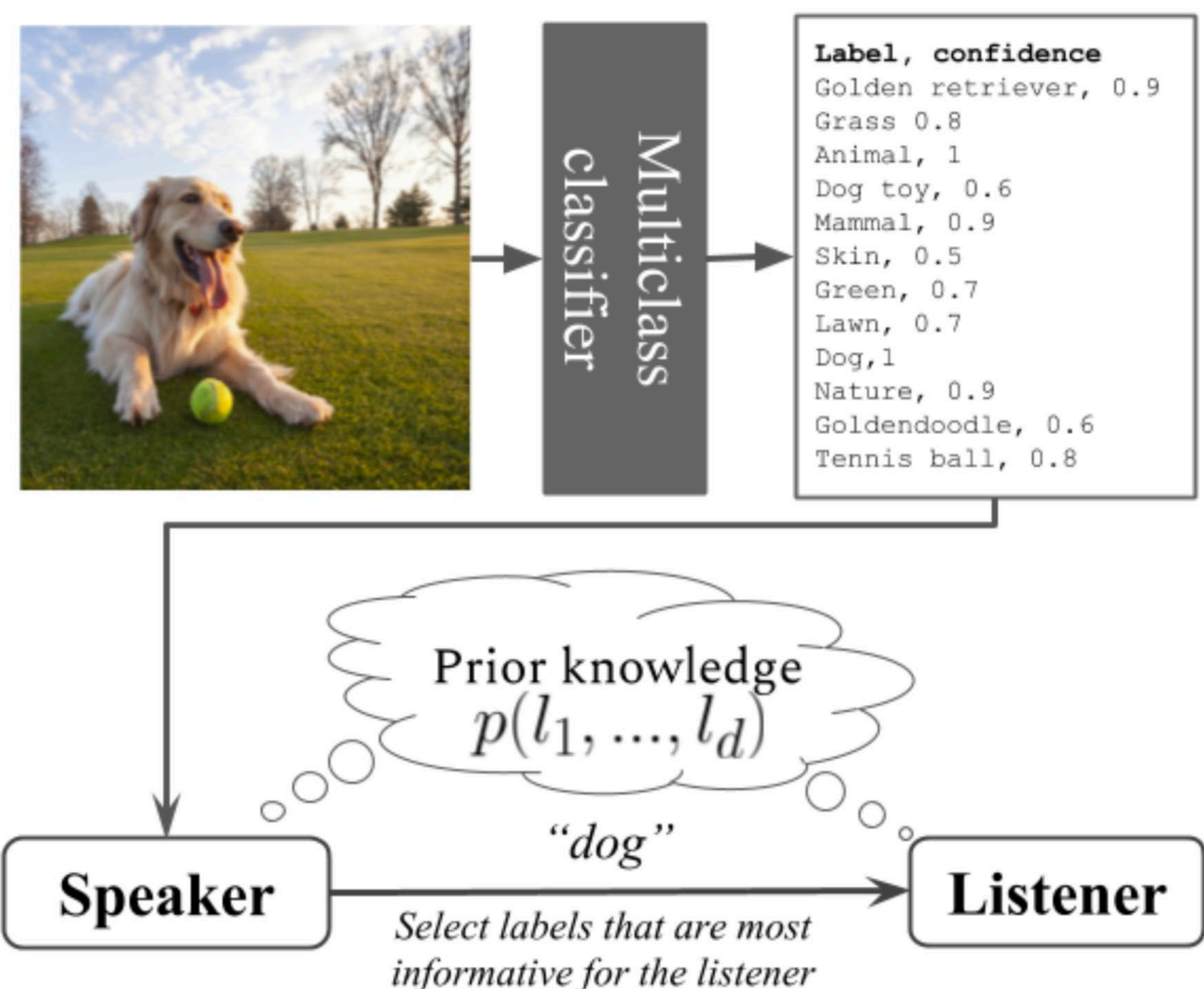
因果モデル解析、未知データの予測などによく用いられる手法の一つである

確率的グラフィカルモデルでは、変数間の条件付き独立性の仮定を取り入れることで、複数の変数の同時確率分布をコンパクトな形で表現することができる

応用1、情報量の多い画像ラベルの選択

目標: d 個ラベルの集合から情報量が多い k ($k < d$) 個ラベルを選択する
(Ilor & Gal, 2018, arxiv)

エントロピーの変化量の近似計算のため、ラベル変数間のグラフ構造をChow-liu木と仮定



応用2、コンピュータ適応型テスト

目標: 限られた問題数で学生たちの知識の現状の情報を得る
(Martin & Vomlel, 2015, UAI)

情報ゲインの近似計算のため、テスト項目を表す変数は学生の
能力値が与えられると独立になると仮定

情報ゲインの推論アルゴリズム

モデル情報ゲイン:

$$\Delta H(i) = H[p(l_1, \dots, l_d)] - H[p(l_1, \dots, l_d | l_i)],$$

$H(\cdot)$ はエントロピー

定理:

変数 X と Z は Y が所与とすると独立になるなら、そのとき、

$$\begin{aligned} I(X; Y, Z) &= H(X) - H(X|Y, Z) \\ &= H(X) - H(X|Y) \\ &= I(X; Y) \end{aligned}$$

ローカル計算:

従来研究でおいだ仮定しない場合、ある変数 i の独立markov blanketが分かれば、変数 i の情報ゲイン $\Delta H(i)$ の計算は d の指数オーダーからその変数のmarkov blanketのサイズ k の指数オーダーに減少