KYOTO UNIVERSITY

介入効果推定の方法

鹿島 久嗣
京都大学

DEPARTMENT OF INTELLIGENCE SCIENCE AND TECHNOLOGY

目次：
介入効果推定の方法

1. 介入効果推定問題
2. 介入効果推定法
3. 反事実の欠損への対処
4. 傾向スコアによる偏り補正
5. 偏り補正な表現学習
6. グラフ上での介入効果推定

近頃、機械学習界隈でも
話題の因果推論が…

深層学習の技術と合体して…

同じく話題のグラフ学習と合体

介入効果推定問題

機械学習の目的：
予測 と 発見 による意思決定の自動化あるいは補助

予測型の機械学習

- 過去のデータをもとに、将来のデータについて予測する
- 「これから何が起こるのか？」がわかれば、
それに基づく意思決定ができる

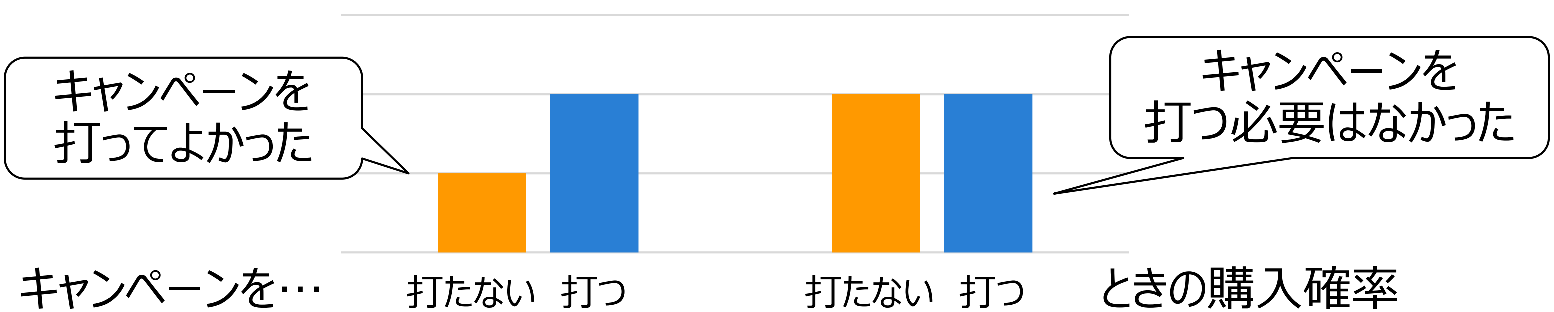
より意思決定に
つながりやすい

発見型の機械学習

- 過去のデータから、何らかの知見を得る
- 「いま何が起きているのか？」がわかれば、
正しい認識に基づく意思決定ができる

一般的な予測モデリングの問題点：意思決定（介入）のもたらす効果を考慮していない

- 一般的な予測モデリングにもとづく意思決定：
  1. 各ユーザ$\mathbf{x}^{(i)}$に対して、購入確率$f\left(\mathbf{x}^{(i)}\right)$ を推定する
  2. $f\left(\mathbf{x}^{(i)}\right)$が大きいユーザ$\mathbf{x}^{(i)}$にキャンペーンを打つ（クーポン発行など）
- 問題点：別にキャンペーンを打たなくても買う人はいるのでは？
  - キャンペーンによる効果（＝購買可能性の増分）をみていない



両方ともキャンペーンを打った結果は良いが…

介入効果推定の目的：
意思決定（介入）の効果を考慮した予測モデリング

- 対象（人）に何らかの働きかけを行い、
  特定の結果（行動など）を促したい場面：
  - ユーザに対して、キャンペーンを打つことで、購買行動を期待
  - 患者に対して、ある治療を行うことで、治癒を期待
  - 顧客に対して、サービスの更新を打診することで、更新を期待
  - 有権者に対して、投票を促すことで、投票行動を期待
- それぞれの対象に対して、適切な働きかけ（介入）を行いたい：
  - どの対象に介入すべきか？ ある対象に介入すべきか否か？
- 働きかけの「効果」を考えた予測モデル化が必要

意思決定のあるべき姿：
介入効果のある対象に介入すべき

- 対象は介入／非介入に対する反応によって4タイプに分けられる

介入した場合の結果	購入する	確実（Sure Thing）	説得可能（Persuadable）
	買わない	あまのじゃく（Do-Not-Disturb）	見込みナシ（Lost Cause）
		購入する	買わない
		介入しなかった場合の結果	

- 「説得可能」カテゴリに介入するべき
- 「あまのじゃく」に介入してはいけない
- その他は介入しても意味がない（介入するだけ無駄）：
  従来の予測モデリングだと「確実」カテゴリにも介入している可能性

介入効果推定と通常の予測モデリングの違い：過去の介入結果を含むデータから介入の効果を予測

- 通常の予測モデリング：
  - $\mathbf{x}$：対象の表現（性別、年齢など）
  - y：結果

  が訓練データとして与えられ、対象と結果の関係を予測

- 介入効果推定：
  - $\mathbf{x}$：対象の表現（性別、年齢など）
  - z：介入の有無
  - y：結果

  購入の有無 or 購入金額

  が訓練データとして与えられ、介入の効果を予測

$$\{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{N}$$

訓練データ

$$\{(\mathbf{x}^{(i)}, z^{(i)}, y^{(i)})\}_{i=1}^{N}$$

介入効果推定問題：
介入の有無を伴うデータから介入効果の予測モデルを得る

- データ： $\{(\mathbf{x}^{(i)}, z^{(i)}, y^{(i)})\}_{i=1}^{N}$

  i番目の対象$\mathbf{x}^{(i)}$に対して介入をした／しなかった ($z^{(i)} \in \{0,1\}$) ところ、結果が$y^{(i)}$（購入の有無 or 購入金額）だった

- 目的：対象$\mathbf{x}$への介入効果 $\tau = y_1 - y_0$ を予測したい
  - 介入した場合の結果をy_1、しなかった場合の結果をy_0とする

- ただし： $\mathbf{x}^{(i)}$に対して$\tau^{(i)} = y_1^{(i)} - y_0^{(i)}$は直接は観測されない
  - 観測されるのは$y_0^{(i)}$か$y_1^{(i)}$のいずれか一方のみ
  - 本質的なデータ欠損：原理的に両方ともは観測できない

「反事実」

介入効果推定法

介入効果推定問題における技術的課題：反事実の欠損と観測の偏り

- 前提： $\mathbf{x}^{(i)}$に対して、介入結果$y_1^{(i)}$か非介入結果$y_0^{(i)}$のいずれか一方のみが観測される
  - （我々の知りたい）介入効果$\tau^{(i)} = y_1^{(i)} - y_0^{(i)}$は観測不可
- 介入効果推定に立ちはだかる２つの技術的課題：

1. 反事実の欠損：
   介入結果と非介入結果のいずれか一方しか観測できない
2. 観測の偏り：
   介入結果と非介入結果のいずれか一方に偏って観測される

介入効果推定法：
反事実の欠損と観測の偏りへの対処

- 介入効果推定に立ちはだかる２つの技術的課題：

1. 反事実の欠損：
   介入結果と非介入結果のいずれか一方しか観測できない

   ⇨ なんとか推定したらよいのでは？

2. 観測の偏り：
   介入結果と非介入結果のいずれか一方に偏って観測される

   ⇨ なんとか偏りを補正したらよいのでは？

反事実の欠損への対処

反事実の欠損への対処：
2モデルアプローチと目的変数の変換アプローチ

- 介入効果を測るには、実際の結果と反事実を比較する必要がある
- 反事実の欠損に対処するための方法：

1. 2モデルアプローチ：反事実を補完する
   - 介入・非介入それぞれに予測モデルをつくる
2. 目的変数を変換するアプローチ：介入効果を直接推定する
   - 単一の予測モデルで介入効果を直接推定する

2モデルアプローチ：
介入・非介入それぞれの結果をモデル化して、差をとる

- 2モデルのアプローチ：介入・非介入それぞれの予測モデルをつくる
  1. 介入ありの場合のモデル：$f_1(\mathbf{x})$
     - $z^{(i)} = 1$の場合の訓練データ$\left\{\left(\mathbf{x}^{(i)}, z^{(i)}, y^{(i)}\right)\right\}_{i:z^{(i)}=1}$ から推定
     - $z^{(i)} = 0$の場合の介入効果の推定は $\tau^{(i)} = f_1(\mathbf{x}^{(i)}) - y^{(i)}$
  2. 介入なしの場合のモデル：$f_0(\mathbf{x})$
     - $z^{(i)} = 0$の場合の訓練データ$\left\{\left(\mathbf{x}^{(i)}, z^{(i)}, y^{(i)}\right)\right\}_{i:z^{(i)}=0}$から推定
     - $z^{(i)} = 1$の場合の介入効果の推定は $\tau^{(i)} = y^{(i)} - f_0(\mathbf{x}^{(i)})$
- 将来のデータ$\mathbf{x}$の介入効果の推定は $\tau(\mathbf{x}) = f_1(\mathbf{x}) - f_0(\mathbf{x})$

※ $f_0(\mathbf{x})$と$f_1(\mathbf{x})$をまとめて一つのモデル$f(\mathbf{x}, z)$としてもよい

目的変数の変換アプローチ：介入効果を直接推定するモデル

- 結果変数の変換：$\eta^{(i)} = 2\left\{z^{(i)}y_1^{(i)} - (1 - z^{(i)})y_0^{(i)}\right\}$

ランダム化試験
A/Bテスト

- お気持ち：介入$z^{(i)}$が「ランダムに決められていたとしたら」
$z^{(i)}$の期待値は1/2なので、$\eta^{(i)}$の期待値は $\mathrm{E}\left[\eta^{(i)}\right] = \tau^{(i)}$

- 訓練データの変換：

$$\left\{\left(\mathbf{x}^{(i)}, z^{(i)}, y^{(i)}\right)\right\}_{i=1}^{N} \Longrightarrow \left\{\left(\mathbf{x}^{(i)}, \eta^{(i)}\right)\right\}_{i=1}^{N}$$

- $\eta^{(i)}$を予測するモデル $\eta^{(i)} \approx f(\mathbf{x}^{(i)})$ をつくる
  - 入力から介入効果を直接推定するモデル

介入効果に基づく決定：予測介入効果の高い順に介入する

- 新たな対象に対して、介入すべきかどうかを決めたい
- 介入効果推定に基づく介入決定：

1. すべての対象 $\mathbf{x}$ に対して介入効果 $\tau(\mathbf{x})$ を推定する：
2. $\tau(\mathbf{x})$ が大きい順に介入する
   - $\tau(\mathbf{x}) > 0$ であるものは介入効果がプラスなので基本的には介入すればよい
   - コストとの兼ね合いでどこまで介入するかを決定する

傾向スコアによる観測の偏りの補正

観測の偏り：
観測データの偏りは推定の偏りを生む

- $\eta^{(i)} = 2\left\{z^{(i)}y_1^{(i)} - (1 - z^{(i)})y_0^{(i)}\right\}$ は介入効果$\tau^{(i)}$のよい推定
  - 介入 $z^{(i)}$ が「ランダムに決められていたとしたら」$\mathrm{E}\left[\eta^{(i)}\right] = \tau^{(i)}$
- もし、$z^{(i)}$がランダム（ $\Pr\left[z^{(i)} = 1\right] = 1/2$ ）でなかったら？
  - 営業担当は買いそうな客にキャンペーンを打つ傾向がある
  - 医者は効きそうな患者に薬を与える傾向がある

$\Pr\left[z^{(i)} = 1\right]$を傾向スコアと呼ぶ

- 例えば $\Pr\left[z^{(i)} = 1\right] = 2/3$のとき $\eta^{(i)} = 2\left\{\frac{2}{3}y_1^{(i)} - \frac{1}{3}y_0^{(i)}\right\} > \tau^{(i)}$

⟹ 介入効果を過大評価（ $\eta^{(i)} > \tau^{(i)}$ ）している

観測の偏りを取り除く：逆確率重みづけによる偏りの除去

- 介入$z^{(i)}$がランダムならOK、そうでなければ過大評価・過小評価
- もし、$z^{(i)}$がランダムでなかったら？

⟹ 「あたかもランダム」に介入したと見えるよう補正する

- 逆確率重みづけ法：傾向スコア$\Pr[z^{(i)} = 1]$の逆数をかけて補正
- 例： $\Pr[z^{(i)} = 1] = 2/3$のとき $\eta^{(i)} = 2\left\{\frac{2}{3}y_1^{(i)} - \frac{1}{3}y_0^{(i)}\right\} > \tau^{(i)}$

「あたかもランダム」補正　$\times 1/2\Pr[z^{(i)} = 1]$ ⇓　⇓ $\times 1/2\Pr[z^{(i)} = 0]$

$$\eta^{(i)} = 2\left\{\frac{1}{2}y_1^{(i)} - \frac{1}{2}y_0^{(i)}\right\} > \tau^{(i)}$$

逆重みづけによる介入効果推定の手続き：傾向スコア推定と介入効果モデル推定の2段階

1. 傾向スコアの推定：
   訓練データ $\left\{\left(\mathbf{x}^{(i)}, z^{(i)}\right)\right\}_{i=1}^{N}$ から、傾向スコアのモデル
   $g(\mathbf{x}) = \Pr[z = 1|\mathbf{x}]$ を推定する

- 適当な 確率的二値分類器（ロジスティック回帰、NN、...）を利用可能

2. 介入効果モデルの推定：
   $\tilde{\eta}^{(i)} = \frac{z^{(i)} y_1^{(i)}}{g\left(\mathbf{x}^{(i)}\right)} - \frac{\left(1 - z^{(i)}\right) y_0^{(i)}}{1 - g\left(\mathbf{x}^{(i)}\right)}$ として、訓練データ $\left\{\left(\mathbf{x}^{(i)}, \tilde{\eta}^{(i)}\right)\right\}_{i=1}^{N}$
   から介入効果モデル $\tilde{\eta}^{(i)} \approx f(\mathbf{x}^{(i)})$を推定する

偏り補正な表現学習

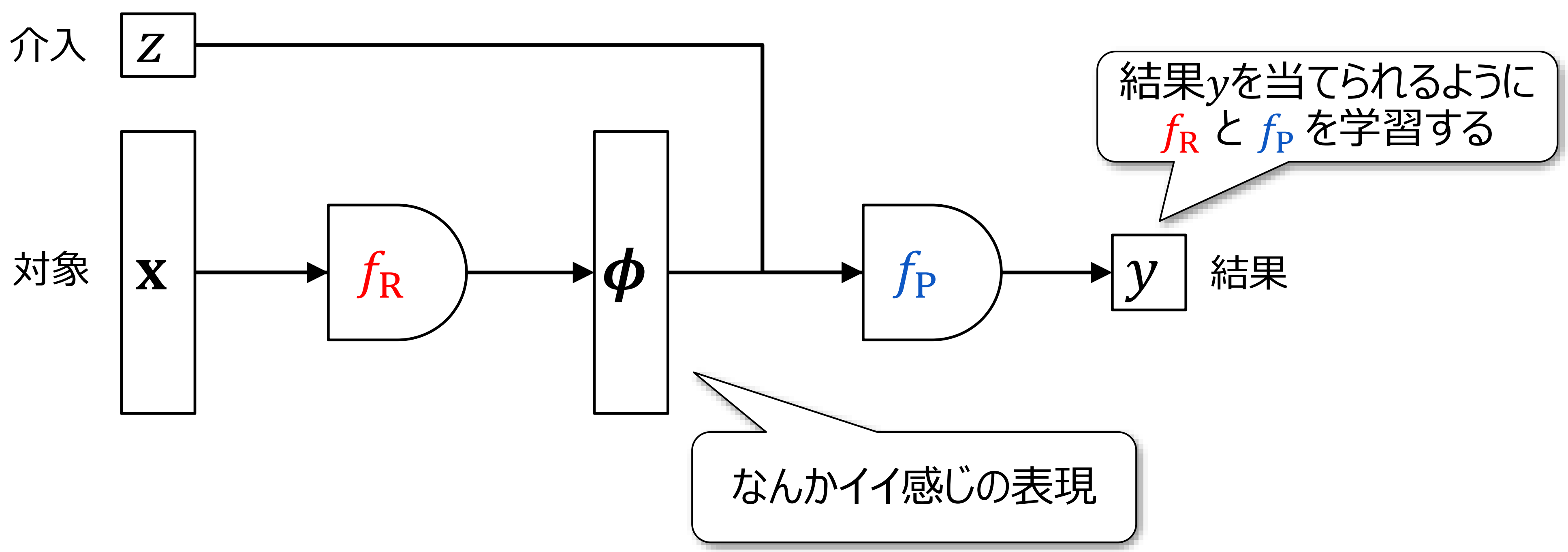
Learning Representations for Counterfactual Inference
Johansson *et al*. (ICML2016)

Estimating Individual Treatment Effect: Generalization Bounds and Algorithms
Shalit *et al*. (ICML2017)

深層学習による介入効果推定：
介入効果推定に適した表現学習は可能か？

- 期待：深層学習なら、「表現学習」で、反実仮想をなんかいい感じで補完してくれるはずだ（?）

想定されるネットワーク構造



介入効果測定のための表現学習：「あたかもランダム」な介入を実現する表現の獲得

- 逆重みづけにおける「あたかもランダム」補正：
  介入・非介入が「あたかもランダム」（＝半々）で決定されたかのような状況を作り上げた

言い換えると ⟺ $\mathbf{x}$ を見ても $z = 1$ なのか $z = 0$ なのか分からない

- 「あたかもランダム」な介入を実現する$\mathbf{x}$の表現$\boldsymbol{\phi}$とは：

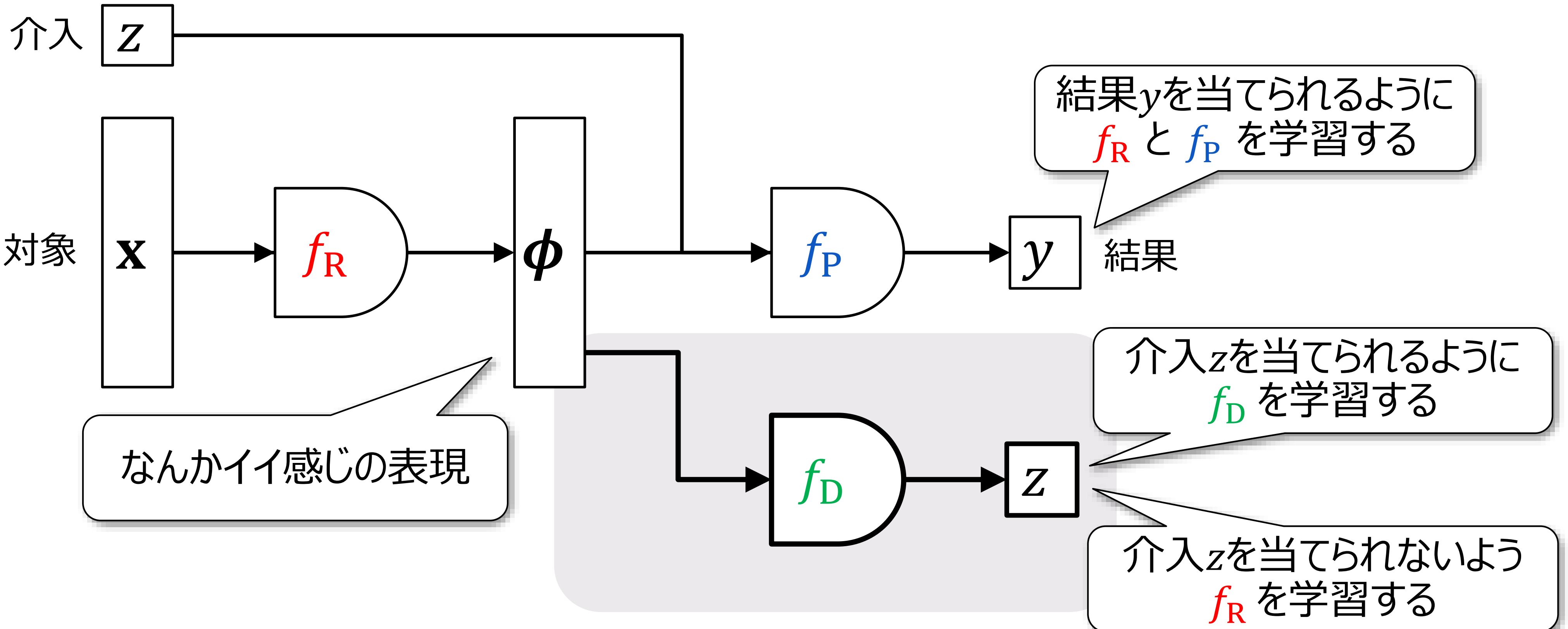
  $\boldsymbol{\phi}$ を見ても $z = 1$ なのか $z = 0$ なのか分からないこと

  （あるいは、$\boldsymbol{\phi} \mid z = 1$ の分布と $\boldsymbol{\phi} \mid z = 0$ の分布が同じ）

あたかもランダム介入な表現学習：
介入が予測できないような表現を獲得するよう学習

- $\boldsymbol{\phi}$ を見ても $z=1$ なのか $z=0$ なのか分からないように、表現抽出器 f_{R} を学習する（識別器 f_{D} は介入 z を当てようと頑張る）

$\Longleftrightarrow$ $\boldsymbol{\phi} \mid z=1$ の分布と $\boldsymbol{\phi} \mid z=0$ の分布間距離を最小化する



グラフ上での介入効果推定

Learning Individual Causal Effects from Networked Observational Data
Guo *et al*. (WSDM2020)

グラフ上で介入効果推定問題：
従来の介入効果推定問題 ＋ （ソーシャル）ネットワーク

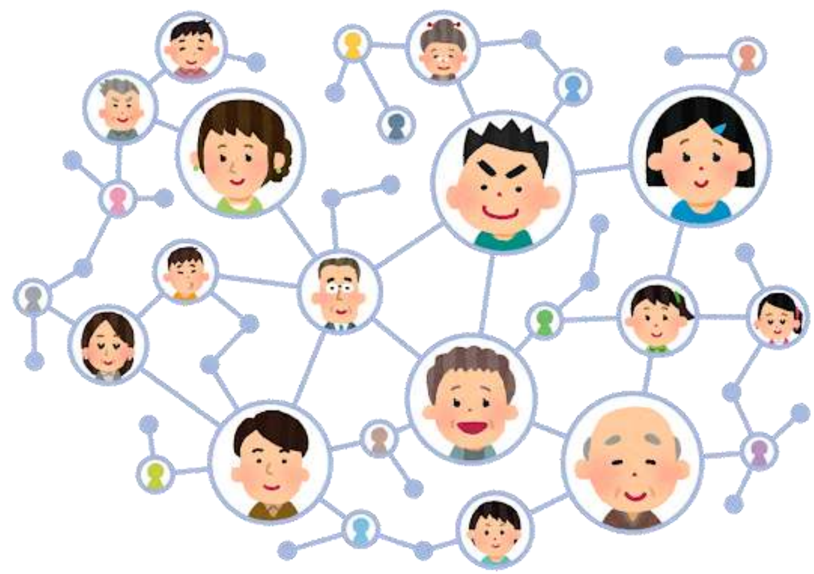
- 介入効果推定問題：
  - $\mathbf{x}$：対象の表現（性別、年齢など）
  - z：介入の有無
  - y：結果

  が訓練データとして与えられ、
  介入の効果を予測

$$\left\{\left(\mathbf{x}^{(i)}, z^{(i)}, y^{(i)}\right)\right\}_{i=1}^{N}$$

＋

- 対象間のつながり

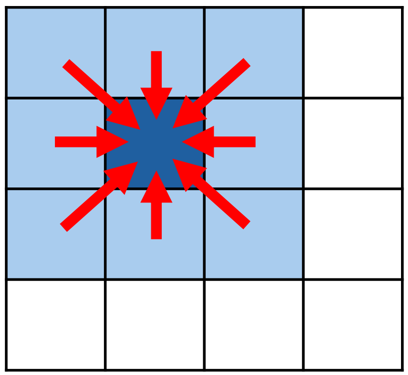


$$G = (V, E)$$

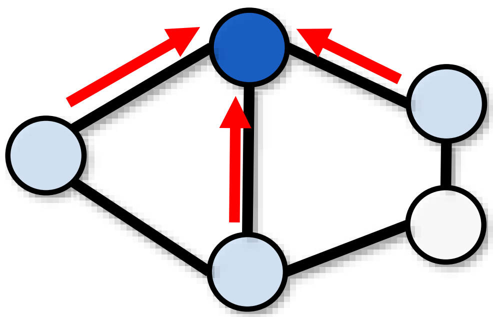
$$V = \{1,2, \dots, N\}$$

グラフ深層学習：
近年のグラフ機械学習の発展

- グラフ構造からの特徴抽出にニューラルネットワークを利用
- グラフ畳み込みニューラルネットワーク（GCN／GNN）
  - 画像畳み込みニューラルネットワーク（CNN）：
    各ピクセルがその近傍ピクセルの情報を取り込む
  - グラフ畳み込み：各ノードが周辺ノードの情報を取り込む



画像畳み込み



グラフ畳み込み

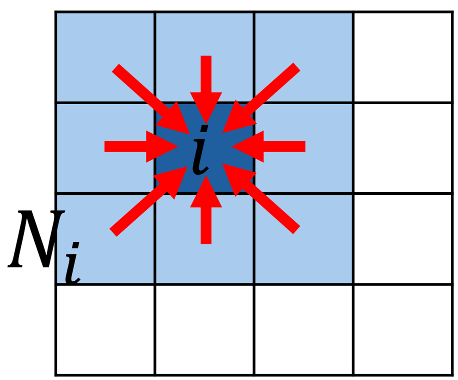
グラフ畳み込みニューラルネットワーク：周辺構造を取り込んだノード表現の獲得

- 周辺の情報を取り込む畳み込み操作：

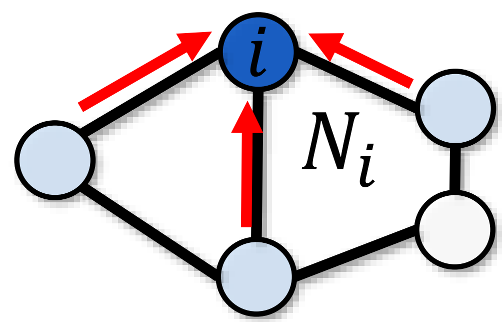
i 番目のピクセル（ノード）の ℓ 層目の表現

$$\mathbf{h}_i^{(\ell)} = \sigma\left(\boldsymbol{V}^{(\ell-1)} \mathbf{h}_i^{(\ell-1)} + \sum_{j \in N_i} \boldsymbol{W}^{(\ell-1)} \mathbf{h}_j^{(\ell-1)} \right)$$

i 番のピクセル（ノード）の 近傍ピクセル（ノード）集合



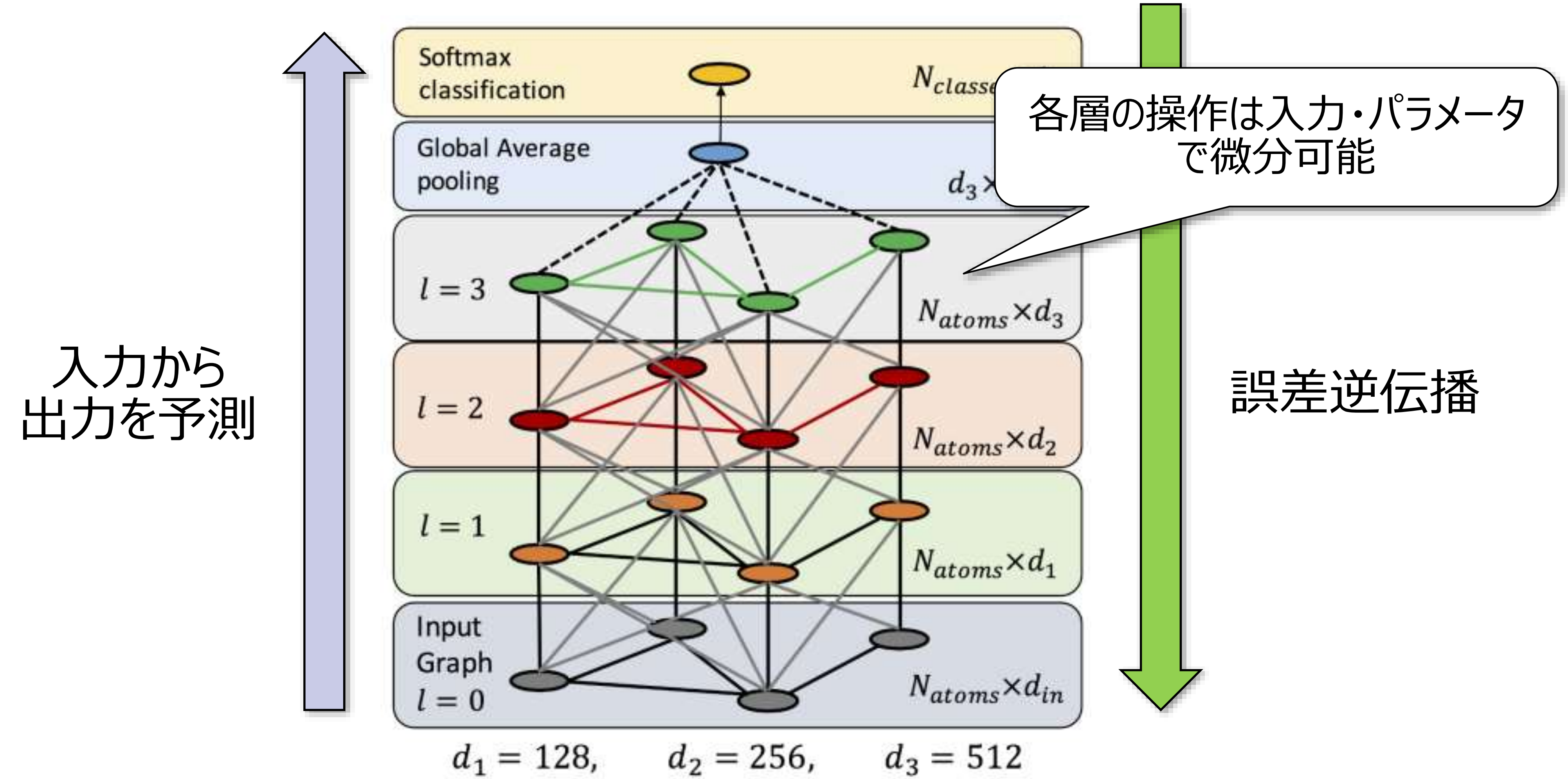
画像畳み込み



グラフ畳み込み

グラフニューラルネットワークの学習：特徴抽出と予測モデルの一気通貫学習

- グラフNNは誤差逆伝播（自動微分）によって学習可能

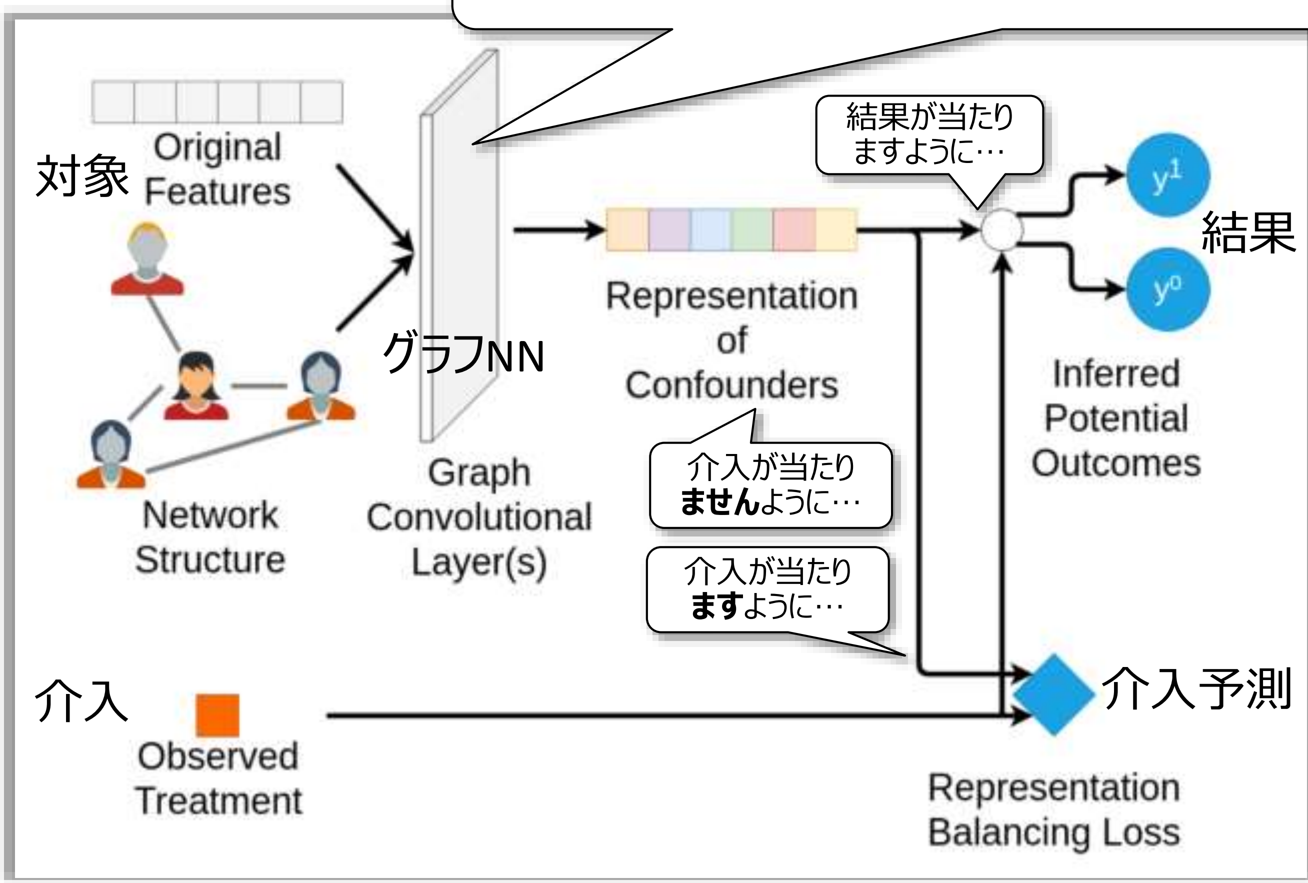


Pope *et al.* (2019)

グラフ深層学習によるグラフ上での介入効果推定問題：「あたかもランダム」表現学習＋グラフNN

- 枠組みは「あたかもランダム」表現学習と同じ

グラフからの表現抽出にグラフNNを用いる（だけ）



まとめ：
介入効果推定の方法

1. 介入効果推定問題
2. 介入効果推定法
3. 反事実の欠損への対処
4. 傾向スコアによる偏り補正
5. 偏り補正な表現学習
6. グラフ上での介入効果推定

近頃、機械学習界隈でも
話題の因果推論が…

深層学習の技術と合体して…

同じく話題のグラフ学習と合体