アフィンカップリング型モデル変換による転移学習

南　俊匠　　総合研究大学院大学　統計科学専攻　博士課程(5年一貫制)5年

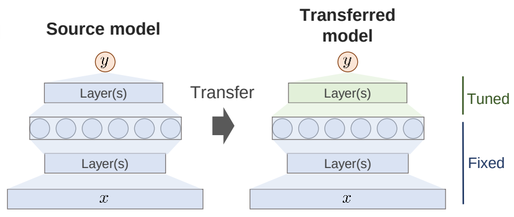
1．はじめに

転移学習 (Transfer learning)

- 目標ドメイン（サンプル空間と確率分布の組）での学習性能の向上のために，関連するドメインのタスクで得られた情報を**目標ドメインの学習に再利用**するための方法論.
- 十分なデータで訓練された元ドメインのモデルが与えられた下で，モデルやその特徴表現を目標ドメインの学習に適切に転移することで，**目標ドメインのデータが限られていても高い予測性能**を持つモデルを構築できる.

ニューラルネットワークによる転移学習

- **特徴抽出型**：中間層の特徴表現を目標ドメインのモデルの入力変数に用いる.
- **ファインチューニング**：訓練済みモデルのパラメータ初期値として再学習する.



本研究の目的

特徴表現の変換を用いた転移学習において，転移モデルの最適な構造を導き，その理論的性質を明らかにする.

2．変数変換を用いた転移学習

- 回帰問題において，元ドメインの特徴表現 $f_s(x)$ と目標ドメインの出力 y を使って新しい変数 z を作り，それに対してモデルの訓練を行うアプローチを考える.

例えば...　元ドメインと目標ドメインの差分を学習 [Kuzborskij+ (2013)]

1. 差分のデータ $z_i = y_i - f_s(x_i)$ を計算する.
2. z に対する回帰モデル $\hat{g}(x)$ を訓練する.
3. 元ドメインの情報を足し合わせる．$\hat{f}_t(x) = \hat{g}(x) + f_s(x)$.

⬇ この手順を一般化すると...

学習プロセス [Du+ (2017)]

1. 変換	関数 $\phi: \mathcal{Y} \times \mathcal{F}_s \mapsto \mathbb{R}$ を用いて，作業変数 $z = \phi(y, f_s(x))$ を作る.
2. 学習	観測データ $\{(x_i, z_i)\}_{i=1}^n$ を用いて，モデル $\hat{g}(x)$ を訓練する.
3. 再変換	関数 $\psi: \mathbb{R} \times \mathcal{F}_s \mapsto \mathcal{Y}$ を用いて，予測モデル $\hat{f}_t(x) = \psi(\hat{g}(x), f_s(x))$ を得る.

具体例

- 元タスクとの差分を学習

作業変数 $z = y - f_s(x)$　　予測モデル $\hat{f}_t(x) = \hat{g}(x) + f_s(x)$

⇒ $\phi = y - f_s(x)$, $\psi = \hat{g}(x) + f_s(x)$

- 予測値が元タスクと近くなるように学習　λ：正則化パラメータ

目的関数 $\frac{1}{n}\sum_{i=1}^n \left[\{y_i - g(x_i)\}^2 + \lambda\{g(x_i) - f_s(x_i)\}^2\right]$

$\left(= (1+\lambda)\frac{1}{n}\sum_{i=1}^n \left\{\frac{y_i + \lambda f_s(x_i)}{1+\lambda} - g(x_i)\right\}^2 + \text{const.}\right)$.

⇒ $\phi = \frac{y + \lambda f_s(x)}{1+\lambda}$, $\psi = \hat{g}(x)$

3．リスク上界とモデリング

- いくつかの仮定の下，この学習プロセスの期待誤差について次の不等式が得られる.

$$\mathbb{E}_{x,y}\left[\left|f_t(x) - \hat{f}_t(x)\right|^2\right] \le 3\,\mathbb{E}_{x,y}\left[\mu_\psi(f_s(x))^2\left|\psi^{-1}(f_t(x), f_s(x)) - \phi(f_t(x), f_s(x))\right|^2\right] \quad \text{変換関数の乖離度}$$
$$+ 3\sigma^2\,\mathbb{E}_{x,y}\left[\mu_\psi(f_s(x))^2\mu_\phi(f_s(x))^2\right] \quad \text{ノイズの分散}$$
$$+ 3\,\mathbb{E}_{x,y}\left[\mu_\psi(f_s(x))^2\left|z - \hat{g}(x)\right|^2\right] \quad \text{変換後の予測誤差}$$

※ $\mu_\phi(f_s(x))$, $\mu_\psi(f_s(x))$ ：$f_s(x)$ を固定したときの ϕ, ψ のリプシッツ定数

- この不等式から次のことがわかる.
  - $\phi = \psi^{-1}$ のときに第一項はゼロになる.
  - ψ が一つ目の引数について線形で表されるとき，第二項は最小値 1 をとる.

$$\left(\mu_\psi = \max|\psi'|,\ \mu_{\psi^{-1}} = \max\left|\frac{1}{\psi'}\right| = \frac{1}{\min|\psi'|},\ \mu_\psi^2\mu_\phi^2 = \mu_\psi^2\mu_{\psi^{-1}}^2 = \left(\frac{\max|\psi'|}{\min|\psi'|}\right)^2\right)$$

⬇

  - 適当な関数 $g_1, g_2 : \mathcal{F}_s \mapsto \mathbb{R}$ を用いて $\phi = \frac{y - g_1(f_s(x))}{g_2(f_s(x))}$, $\psi = g_1(f_s(x)) + g_2(f_s(x))g(x)$ と書くことができ，このとき第三項の最小化問題は次のようになる.

$$\min_{g_1,g_2,g}\ \mathbb{E}_{x,y}\left[\mu_\psi(f_s(x))\left|\frac{y - g_1(f_s(x))}{g_2(f_s(x))} - g(x)\right|^2\right] = \min_{g_1,g_2,g}\ \mathbb{E}_{x,y}\left|y - g_1(f_s(x)) - g_2(f_s(x))g(x)\right|^2$$

- 結果として，次のような転移モデルのクラス $\mathcal{H}$ が導かれる.

提案モデル　適当な関数クラス $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ に対し，

$$\mathcal{H} = \left\{x \mapsto \underline{g_1(f_s(x)) + g_2(f_s(x)) \cdot g_3(x)} \mid g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2, g_3 \in \mathcal{G}_3\right\}$$

4．パラメータ推定

提案モデル　$x \mapsto g_1(f_s(x)) + g_2(f_s(x)) \cdot g_3(x)$

カーネル法

- 適当な正定値カーネル k_1, k_2, k_3 に対し，対応する再生核ヒルベルト空間をそれぞれ $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$，特徴写像をそれぞれ Φ_1, Φ_2, Φ_3 と表す.

目的関数 $\frac{1}{n}\sum_{i=1}^n \{y_i - \overbrace{\langle\alpha, \Phi_1(f_s(x_i))\rangle}^{g_1(f_s(x))} - \overbrace{\langle\beta, \Phi_2(f_s(x_i))\rangle\langle\gamma, \Phi_3(x_i)\rangle}^{g_2(f_s(x))\ g_3(x)}\}^2 + \lambda_1\|\alpha\|_2^2 + \lambda_2\|\beta\|_2^2 + \lambda_3\|\gamma\|_2^2$

⬇ リプレゼンター定理　$\alpha = \sum_{i=1}^n a_i\Phi_1(f_s(x_i))$, $\beta = \sum_{i=1}^n b_i\Phi_2(f_s(x_i))$, $\gamma = \sum_{i=1}^n c_i\Phi_3(x_i)$.

$$= \frac{1}{n}\|y - K_1a - (K_2b) \circ (K_3c)\|_2^2 + \lambda_1 a^\top K_1 a + \lambda_2 b^\top K_2 b + \lambda_3 c^\top K_3 c$$

⇒ $$F(a,b,c) = \frac{1}{n}\sum_{i=1}^n \left(y_i - k_1^{(i)\top}a - b^\top M^{(i)}c\right)^2 + \lambda_1 a^\top K_1 a + \lambda_2 b^\top K_2 b + \lambda_3 c^\top K_3 c$$

$(K_I)_{i,j} = k_I(x_i, x_j)$,　$k_I^{(i)} = [k_I(x_i, x_1)\ \cdots\ k_I(x_i, x_n)]^\top$ for $I \in \{1,2,3\}$,　$M^{(i)} = k_2^{(i)} \otimes k_3^{(i)\top}$,　$\circ$: Hadamard product.

- 結果として，低ランクテンソル回帰の問題に帰着し，block relaxation algorithm [Zhou+ (2013)] のようなアルゴリズムを用いてパラメータを推定することができる.

[Zhou+ (2013)]

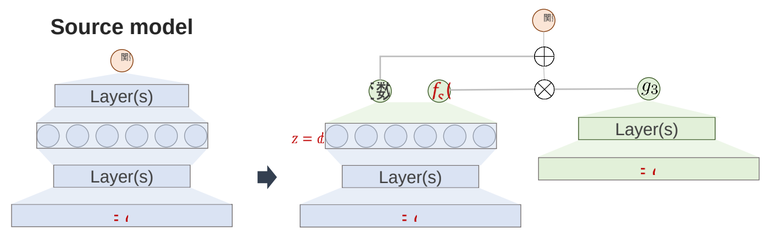
Algorithm 1 Block relaxation algorithm

Initialize : $a_0, b_0 \neq 0, c_0 \neq 0$
repeat
　$b_{t+1} = \arg\min_b F(a_t, b, c_t)$
　$c_{t+1} = \arg\min_c F(a_t, b_{t+1}, c)$
　$a_{t+1} = \arg\min_a F(a, b_{t+1}, c_{t+1})$
until Convergence

※それぞれの最小化問題は陽に解くことができる.

ニューラルネットワーク

- ニューラルネットワークの場合，ネットワーク構造の中に組み込むことができる.



- これは，アフィンカップリング則を使用した可逆ニューラルネットワーク [Dinh+ (2015)] の一種に相当する.

5．推定誤差の理論解析

- カーネル法を使ってモデリングした場合について，汎化誤差の収束レートや excess risk の上界を次のように導くことができる.

定理 1. いくつかの仮定の下，n に依存しない定数 c_1, c_2, c_3 が存在し，任意の関数 $h \in \mathcal{H}$ と実数 $\eta > 0$ に対し，少なくとも確率 $1 - e^{-\eta}$ で次の不等式が成り立つ.

$$\mathop{\mathbb{E}}_{x,y}\ell(y, h(x)) \le \frac{1}{n}\sum_{i=1}^n \ell(y_i, h(x_i)) + \tilde{O}\left(\left(\sqrt{\frac{R_s}{n}} + \frac{\mu_\ell^2 c^2 + \sqrt{\eta}}{n}\right)\left(\sqrt{L}c + \sqrt{L\eta}\right) + \frac{c^2 L + L\eta}{n}\right).$$

ただし，$R_s = \inf_{\alpha:\|\alpha\|_2^2 \le \lambda_\alpha^{-1}\hat{R}_s} \mathop{\mathbb{E}}_{x,y}\ell(y, \langle\alpha, \Phi_1\rangle)$ とする.

⇒ ドメイン同士が関連しているほど早く収束する.

定理 2. いくつかの仮定の下，実数 $\eta > 0$ に対し，少なくとも確率 $1 - (2(M+1)^2 + 1)e^{-\eta}$ で次の不等式が成り立つ.

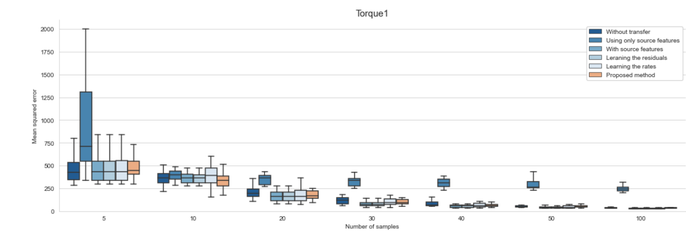
$$\mathop{\mathbb{E}}_{x,y}(y - \hat{h}(x))^2 - \mathop{\mathbb{E}}_{x,y}(y - h^*(x))^2 \le O\left(2^{\frac{s}{1+s}} \cdot n^{-\frac{s}{1+s}}\right).$$

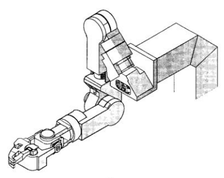
ただし，$\hat{h}(x) = \arg\inf_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^n (y - h(x))^2$, $h^*(x) = \arg\inf_{h\in\mathcal{H}} \mathbb{E}_{x,y}(y - h(x))^2$ とし，s はカーネルによるグラム行列の固有値の減衰率に依存する実数である.

6．実験

- ロボットアームにおける各関節の位置，測度，加速度から関節にかかるトルクを予測する.

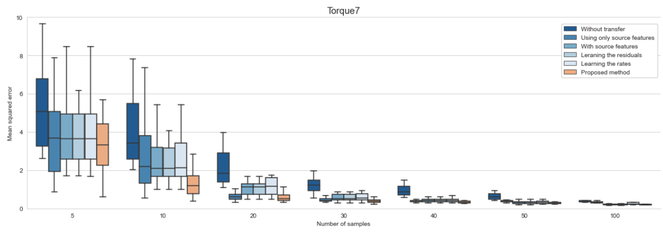
Torque2-7 to Torque1





[Vijayakumar+ (2010)]

Torque1-6 to Torque7



⇒
- 小サンプル下で高精度
- タスク間の関連性が高いほど他手法と比較して優位

7．まとめ

- 本研究では，元ドメインで得られた特徴表現と変換関数を用いた教師あり転移学習において，アフィンカップリング則に基づく転移モデルの最適性とその理論的性質を明らかにした.
- 変換関数をデータから推定することで，ドメイン間の関係性の理解につながる可能性がある.
- 本研究では回帰問題に焦点を絞った．一般の問題への展開は今後の課題の一つである.