



部分空間の整合による 回帰モデルのテスト時適応

(Test-time Adaptation for Regression by Subspace Alignment, ICLR2025)

足立 一樹[†], 山口 真弥^{*}, 熊谷 充敏^{*}, 濱上 知樹[†]

^{*}NTT株式会社 [†]横浜国立大学

2026年9月4日

- 学習時と運用時の環境の違いにより入力分布が変化するとモデル運用時に精度が低下する

学習環境



学習データ



深層学習モデル

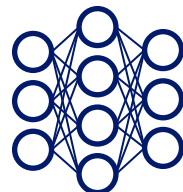


猫

運用環境



運用環境データ



深層学習モデル

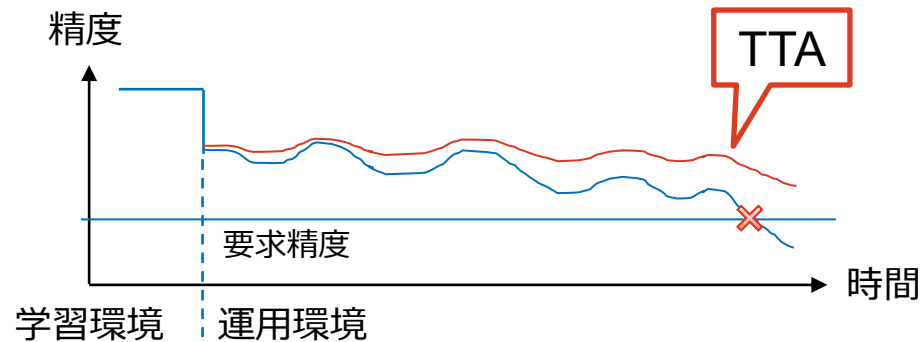
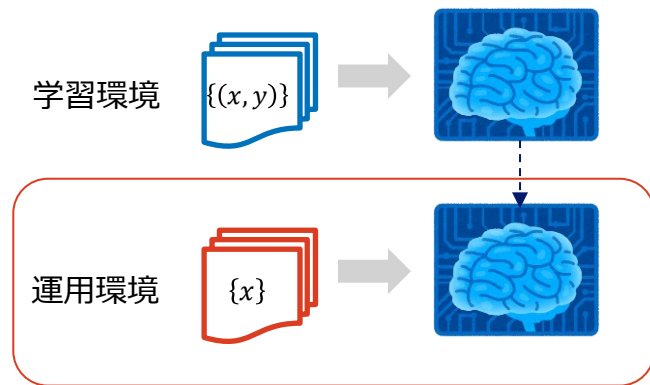


???

- 運用環境に合わせて人手でデータ前処理やモデル調整を行うのは高コスト
→ モデル適応

Test-time Adaptation (TTA)

- 学習済みモデルを運用環境に持ち込み、運用環境データだけを用いて適応
- モデルを自律的に適応させることが可能



- 既存のTTA手法はほぼ全てが分類モデルが対象
- 連続値を予測する回帰タスクも基本的なタスクであり重要
- 回帰モデルに対してもTTAを適用したい

- Test-time entropy minimization (Tent) [Wang et al., ICLR2021]
 - 分布外データに対して予測の確信度（エントロピー）と精度が強く相関することを利用
 - 予測出力のエントロピーを最小化するようにモデルを更新
- Bayesian adaptation to covariate shift (BACS) [Zhou et al., NeurIPS2021]
 - エントロピー最小化 + モデルパラメータの事後分布の正則化
- Efficient anti-forgetting test-time adaptation (EATA) [Niu et al., ICML2022]
 - エントロピー最小化 + データ選択 + 継続学習
- 予測出力の確信度（エントロピー）を利用する手法が多い

- 既存のTTA手法は回帰モデルに直接適用できない
 - 予測出力の確信度（エントロピー）を計算するためには、予測分布を知る必要がある

$$H = \sum_{y=1}^C -\hat{P}(y|x) \log \hat{P}(y|x)$$

- 分類モデルは予測分布（= 各クラスの確率） $\hat{P}(y|x)$ が直接出力されるため、エントロピー計算が容易
- 通常の回帰モデルは予測値（出力分布の平均）のみが出力されるため確信度の計算が困難
- 確信度に依存しないアプローチが必要

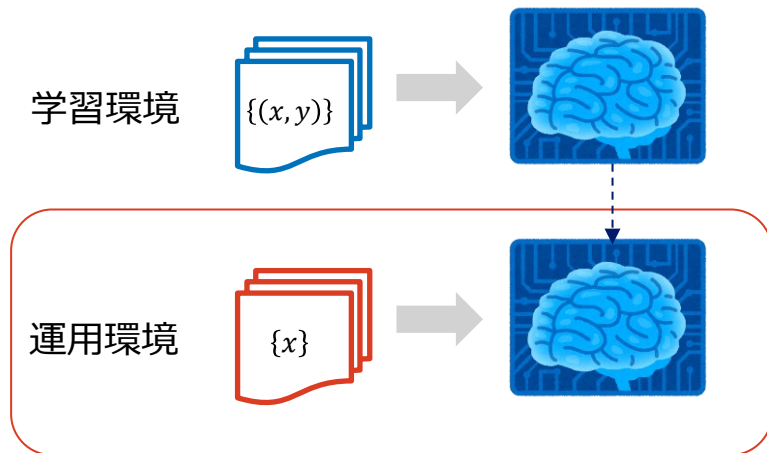
問題設定

- Input

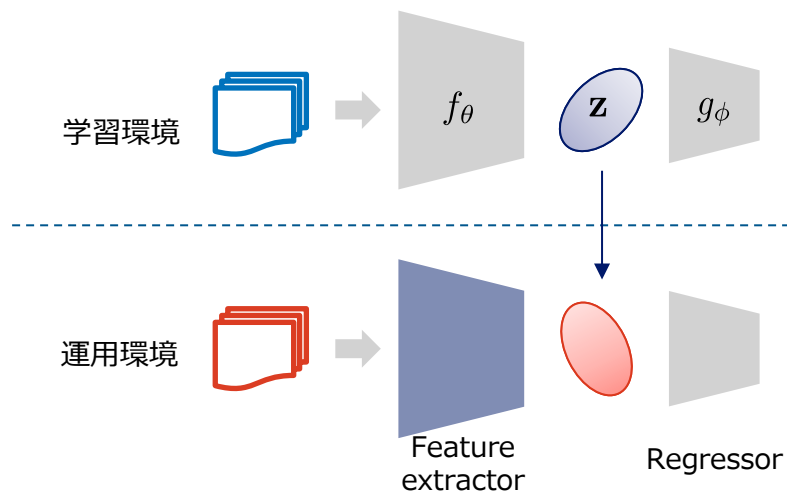
- 学習済み回帰モデル $f_{\theta}: \mathcal{X} \rightarrow \mathbb{R}$
- 運用環境データセット (教師なし) $\{\mathbf{x}_i^t\}$

- Output

- 運用環境データセットに対する予測
- 適応済みモデル $f_{\theta'}$



- 出力の確信度は使えない
 - 運用環境データの特徴分布を学習データと合わせる
 - モデル学習後に、学習データで特徴ベクトルの統計量を計算し運用環境に持ち出す
 - 運用環境データの統計量を学習データと一致させるように特徴抽出器を更新



- ナイーブな実現方法

- 特徴ベクトルの各次元の平均・分散を一致させる

学習データの平均・分散：

$$\mu^s = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_i^s, \quad \sigma^{s2} = \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i^s - \mu^s)^2$$

運用環境データの平均・分散（ミニバッチ）： $\hat{\mu}^t = \frac{1}{B} \sum_{i=1}^B \mathbf{z}_i^t, \quad \hat{\sigma}^{t2} = \frac{1}{B} \sum_{i=1}^B (\mathbf{z}_i^t - \hat{\mu}^t)^2$

- ガウス分布同士のKLを最小化

$$\mathcal{L}_{\text{TTA}} = \sum_{d=1}^D D_{\text{KL}} \left(\mathcal{N}(\mu_d^s, \sigma_d^{s2}) \| \mathcal{N}(\hat{\mu}_d^t, \sigma_d^{t2}) \right) + D_{\text{KL}} \left(\mathcal{N}(\hat{\mu}_d^t, \hat{\sigma}_d^{t2}) \| \mathcal{N}(\mu_d^s, \sigma_d^{s2}) \right)$$

- 問題点

- 回帰は分類に比べて特徴ベクトルの多様性が低くなりやすい [1]
- 実際、特徴空間の限られた部分空間しか使われていないことを実験的に確認
→ オリジナルの特徴空間で分布を合わせても部分空間への影響が小さい

Dataset	#Valid dims.	#Subspace dims.
SVHN	353	14
UTKFace	2041	76
Biwi Kinect*	713	34.5
California Housing (100 dims.)	45	40

*The average over a model trained on each gender/target is reported.

- 実際に特徴ベクトルが分布している部分空間上で分布を合わせる
 - 学習データの特徴ベクトルでPCAを行い、寄与度top- k の主成分が張る部分空間を使う

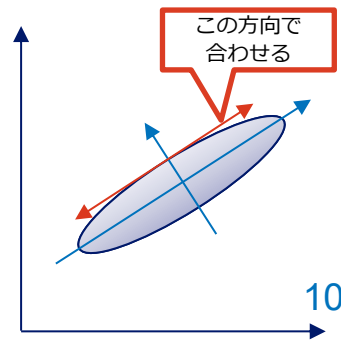
学習データ特徴ベクトルの分散・共分散行列: $\Sigma_s = \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i^s - \boldsymbol{\mu}^s)(\mathbf{z}_i^s - \boldsymbol{\mu}^s)^\top$

- Σ_s の固有ベクトル $\mathbf{v}_i$ が主成分、固有値 λ_i が対応する主成分軸上の分散: $\Sigma_s \mathbf{v}_i = \lambda_i \mathbf{v}_i$
- $\mathbf{v}_1, \dots, \mathbf{v}_k$ ($\lambda_1 > \dots > \lambda_k$) が張る部分空間にターゲットデータを射影

$$\mathbf{z}_i^{t'} = (\mathbf{z}_i^t - \boldsymbol{\mu}^s) \mathbf{V}_k, \quad \mathbf{V}_k = [\mathbf{v}_1, \dots, \mathbf{v}_k]$$

- 部分空間でのKL:

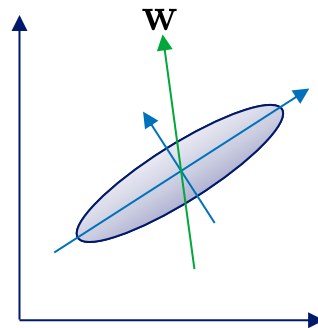
$$\mathcal{L}_{\text{TTA}} = \sum_{d=1}^k D_{\text{KL}} (\mathcal{N}(0, \lambda_d) \| \mathcal{N}(\hat{\mu}_d^{t'}, \hat{\sigma}_d^{t'^2})) + D_{\text{KL}} (\mathcal{N}(\hat{\mu}_d^{t'}, \hat{\sigma}_d^{t'^2}) \| \mathcal{N}(0, \lambda_d))$$



- 部分空間の重み付け

- 回帰の予測は特徴ベクトルを線形変換して行うとする

$$\hat{y} = \mathbf{w}^\top \mathbf{z} + b$$



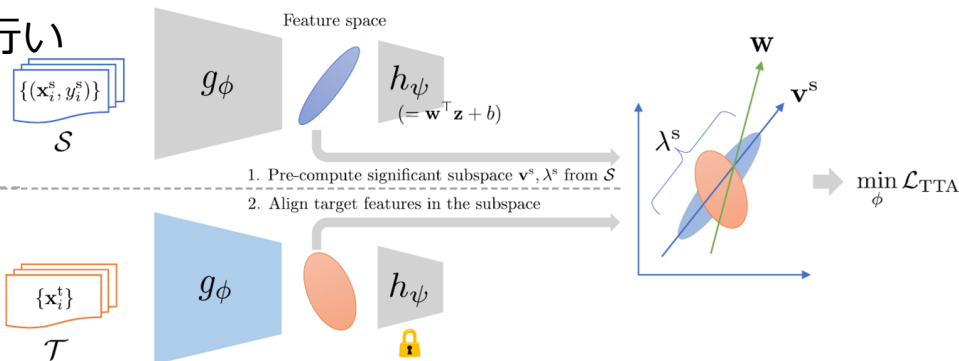
- 予測に対して重要なのは特徴空間の $\mathbf{w}$ の方向なので、その方向は重点的に合わせたい
→ 各次元のKLに重みづけする
 - › $\mathbf{w}$ と同じ方向を向いている基底ほど重点的にKLを小さくしたい

$$\alpha_d = 1 + |\mathbf{w}^\top \mathbf{v}_d|$$

$$\mathcal{L}_{\text{TTA}} = \sum_{d=1}^k \alpha_d \left\{ D_{\text{KL}} \left(\mathcal{N}(0, \lambda_d) \parallel \mathcal{N}(\hat{\mu}_d^{t'}, \hat{\sigma}_d^{t'2}) \right) + D_{\text{KL}} \left(\mathcal{N}(\hat{\mu}_d^{t'}, \hat{\sigma}_d^{t'2}) \parallel \mathcal{N}(0, \lambda_d) \right) \right\}$$

• 手法まとめ

1. 学習データで回帰モデルを訓練
2. 学習データの特徴ベクトルでPCAを行い $\{(\mathbf{v}_1, \lambda_1), \dots, (\mathbf{v}_k, \lambda_k)\}$ を計算



3. 運用環境で $\mathcal{L}_{\text{TTA}}$ を小さくする方向に特徴抽出器を更新
 - › 正規化層のアフィンパラメータのみ更新

実験（一部のみ紹介）

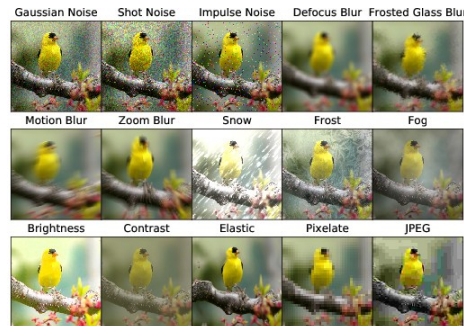
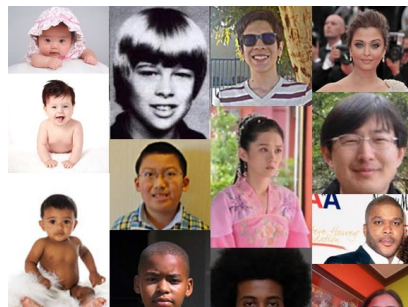
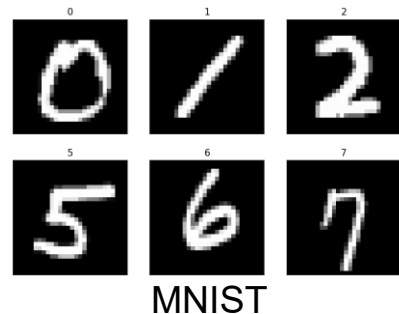
- 回帰タスクでTTA精度を検証

- タスク

- › SVHN-MNIST: 数字の予測を回帰として解く
 - » 学習時：SVHN
 - » テスト時：MNIST
- › UTKFace: 顔写真からの年齢予測
 - » 学習時：オリジナル画像
 - » テスト時：ImageNet-Cと同様の画像破損を付加

- モデル

- › ResNet-26/50



- SVHN → MNIST

Method	$R^2(\uparrow)$	RMSE ($\downarrow$)	MAE ($\downarrow$)
Source	0.406	2.232	1.608
DANN	$0.307_{\pm 0.09}$	$2.406_{\pm 0.16}$	$1.489_{\pm 0.09}$
TTT	$0.288_{\pm 0.02}$	$2.443_{\pm 0.03}$	$1.597_{\pm 0.03}$
BN-adapt	$0.396_{\pm 0.00}$	$2.251_{\pm 0.01}$	$1.458_{\pm 0.00}$
Prototype	$0.491_{\pm 0.00}$	$2.065_{\pm 0.01}$	$1.479_{\pm 0.01}$
FR	$0.369_{\pm 0.01}$	$2.300_{\pm 0.02}$	$1.631_{\pm 0.02}$
VM	$-685.1_{\pm 27.63}$	$75.83_{\pm 1.52}$	$75.78_{\pm 1.52}$
RSD	$0.252_{\pm 0.12}$	$2.497_{\pm 0.20}$	$1.703_{\pm 0.20}$
SSA (ours)	$0.511_{\pm 0.03}$	$2.024_{\pm 0.06}$	$1.209_{\pm 0.04}$
Oracle	$0.874_{\pm 0.00}$	$1.028_{\pm 0.00}$	$0.575_{\pm 0.00}$

- UTKFace (Corruption別)

Method	Defocus blur	Motion blur	Zoom blur	Contrast	Elastic transform	Jpeg comp.	Pixelate	Gaussian noise	Impulse noise	Shot noise	Brightness	Fog	Snow	Mean
Source	0.410	0.159	0.658	-3.906	0.711	0.069	0.595	-2.536	-2.539	-2.522	0.661	-0.029	-0.544	-0.678
DANN	0.512	0.586	0.637	-0.720	0.729	0.698	0.807	-4.341	-3.114	-3.744	0.590	-0.131	-0.425	-0.609
TTT	0.748	0.761	0.773	0.778	0.826	0.772	0.861	0.525	0.532	0.477	0.775	0.397	0.493	0.671
BN-Adapt	0.727	0.759	0.763	0.702	0.826	0.778	0.850	0.510	0.510	0.446	0.790	0.392	0.452	0.654
Prototype	-1.003	-1.020	-1.016	-0.719	-0.967	-0.908	-0.974	-0.514	-0.512	-0.512	-1.004	-0.823	-0.822	-0.830
FR	0.794	0.839	0.849	0.756	0.899	0.825	0.946	0.509	0.522	0.458	0.861	0.408	0.428	0.700
VM	-2.009	-1.991	-2.037	-1.889	-1.918	-1.918	-1.751	-2.181	-2.207	-2.176	-1.927	-2.250	-2.197	-2.035
RSD	0.789	0.833	0.851	0.749	0.897	0.825	0.941	0.502	0.503	0.445	0.862	0.419	0.500	0.701
SSA (ours)	0.803	0.839	0.851	0.792	0.899	0.829	0.943	0.580	0.592	0.560	0.863	0.440	0.517	0.731
Oracle	0.856	0.890	0.889	0.862	0.917	0.873	0.960	0.635	0.652	0.635	0.895	0.519	0.671	0.789

- テスト時適応（TTA）：運用環境で教師なしデータのみでモデルを適応
- 回帰モデルのためのテスト時適応手法を提案
 - 学習データで事前計算した特徴統計量を用い、運用環境データの特徴分布を一致させる
- 回帰タスクで性能向上を確認

- Kazuki Adachi, Shin'ya Yamaguchi, Atsutoshi Kumagai, Tomoki Hamagami
“Test-time Adaptation for Regression by Subspace Alignment”
International Conference on Learning Representations (ICLR), 2025.



arXiv