共変量シフトの情報幾何

木村　正成　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 Abstract

- 共変量シフトは学習分布とテスト分布の乖離の一種；
- 本研究では，いくつかの共変量シフト適応の手法群が情報幾何のフレームワークで一般化できることを示す；
- さらにこの一般化によって導入されるパラメータの最適化によって，既存手法を優越する解を達成できる可能性を示唆する．

Definition 1 (共変量シフト仮定)**.** 2つの確率分布$p_{tr}(\boldsymbol{x}, y)$と$p_{te}(\boldsymbol{x}, y)$は，次の条件を満たすとき共変量シフト仮定を満たすという：1) $p_{tr}(\boldsymbol{x}) \neq p_{te}(\boldsymbol{x})$， 2) $\mathrm{supp}(p_{tr}(\boldsymbol{x})) \supset \mathrm{supp}(p_{te}(\boldsymbol{x}))$， 3) $p_{tr}(y|\boldsymbol{x}) = p_{te}(y|\boldsymbol{x})$.

2 α-geodesic

Definition 2 (f-interpolation)**.** ある$a, b, \in \mathbb{R}$, $\lambda \in [0,1]$と$\alpha \in \mathbb{R}$について，f-interpolationを以下で定義する．

$$m_f^{(\lambda,\alpha)}(a,b) = f_\alpha^{-1}\Big((1-\lambda)f_\alpha(a) + \lambda f_\alpha(b)\Big). \tag{1}$$

ここで

$$f_\alpha(x) = \begin{cases} x^{\frac{1-\alpha}{2}} & (\alpha \neq 1) \\ \ln x & (\alpha = 1). \end{cases} \tag{2}$$

Definition 3 (α-表現)**.** ある正測度$m_i^{\frac{1-\alpha}{2}}$について，α-divergenceから誘導される座標系$\boldsymbol{\theta} = (\theta^i)$は$\theta^i = m_i^{\frac{1-\alpha}{2}} = f_\alpha(m_i)$であって，$\theta^i$は正測度$m_i^{\frac{1-\alpha}{2}}$の$\alpha$-表現と呼ばれる．

Definition 4 (α-測地線)**.** 確率密度関数$p(\boldsymbol{x})$と$q(\boldsymbol{x})$を結ぶα-測地線は

$$r_i(t) = c(t) f_\alpha^{-1}\Big\{(1-t)f_\alpha(p(x_i)) + t f_\alpha(q(x_i))\Big\}, \quad t \in [0,1] \tag{3}$$

で定義され，$c(t) = \frac{1}{\sum_{i=1}^n r_i(t)}$とする．

3 Importance weighted ERM and variants

以下のImportance weightingは共変量シフト適応において本質的であることが示されている．

$$\min_{h \in \mathcal{H}} \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} w(\boldsymbol{x}_i^{tr}) \ell(h(\boldsymbol{x}_i^{tr}), y_i^{tr}). \tag{4}$$

ここで，$w : \mathcal{X} \to \mathbb{R}_{\geq 0}$は何らかの重み付け関数.

Proposition 5 (IWERM)**.** 重み付け関数として密度比$p_{te}(\boldsymbol{x})/p_{tr}(\boldsymbol{x})$を用いることで，共変量シフトのバイアスに適応できる：

$$\min_{h \in \mathcal{H}} \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \frac{p_{te}(\boldsymbol{x}_i^{tr})}{p_{tr}(\boldsymbol{x}_i^{tr})} \ell(h(\boldsymbol{x}_i^{tr}), y_i^{tr}). \tag{5}$$

Definition 6 (AIWERM)**.** Let $\lambda \in [0,1]$. $(p_{te}(\boldsymbol{x})/p_{tr}(\boldsymbol{x}))^\lambda$を重み付け関数とすることで，以下のAdaptive importance weighted ERMが定義できる．

$$\min_{h \in \mathcal{H}} \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \left(\frac{p_{te}(\boldsymbol{x}_i^{tr})}{p_{tr}(\boldsymbol{x}_i^{tr})}\right)^\lambda \ell(h(\boldsymbol{x}_i^{tr}), y_i^{tr}). \tag{6}$$

Definition 7. (RIWERM Let $\lambda \in [0,1]$. $p_{te}(\boldsymbol{x})/\lambda p_{tr}(\boldsymbol{x}) + (1-\lambda)p_{te}(\boldsymbol{x})$を重み付け関数とすることで，以下のRelative importance weighted ERMが定義できる．

$$\min_{h \in \mathcal{H}} \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \frac{p_{te}(\boldsymbol{x}_i^{tr})}{\lambda p_{tr}(\boldsymbol{x}_i^{tr}) + (1-\lambda)p_{te}(\boldsymbol{x}_i^{tr})} \ell(h(\boldsymbol{x}_i^{tr}), y_i^{tr}). \tag{7}$$

4 Generalized covariate shift adaptation

Lemma 8 (f-representation of AIWERM)**.** AIWERMによって生成される周辺正測度はf-interpolationを用いて以下で表現できる：

$$p_A^{(\lambda)}(\boldsymbol{x}) = m_f^{(\lambda,1)}(p_{tr}(\boldsymbol{x}), p_{te}(\boldsymbol{x})). \tag{8}$$

Lemma 9 (f-representation of RIWERM)**.** RIWERMによって生成される周辺正測度はf-interpolationを用いて以下で表現できる：

$$p_R^{(\lambda)}(\boldsymbol{x}) = m_f^{(\lambda,3)}(p_{tr}(\boldsymbol{x}), p_{te}(\boldsymbol{x})). \tag{9}$$

上記補題により，以下の一般化が得られる．

Theorem 10 (Geometrically generalized IWERM)**.** ある$\lambda \in [0,1]$と$\alpha \in \mathbb{R}$について，AIWERMとRIWERMは以下のように一般化できる．

$$\hat{h} = \min_{h \in \mathcal{H}} \int_{\mathcal{X} \times \mathcal{Y}} w^{(\lambda,\alpha)}(\boldsymbol{x}) \ell(h(\boldsymbol{x}), y) p_{tr}(\boldsymbol{x}, y) d\boldsymbol{x} dy. \tag{10}$$

ここで

$$w^{(\lambda,\alpha)}(\boldsymbol{x}) = \frac{m_f^{(\lambda,\alpha)}(p_{tr}(\boldsymbol{x}), p_{te}(\boldsymbol{x}))}{p_{tr}(\boldsymbol{x})}. \tag{11}$$

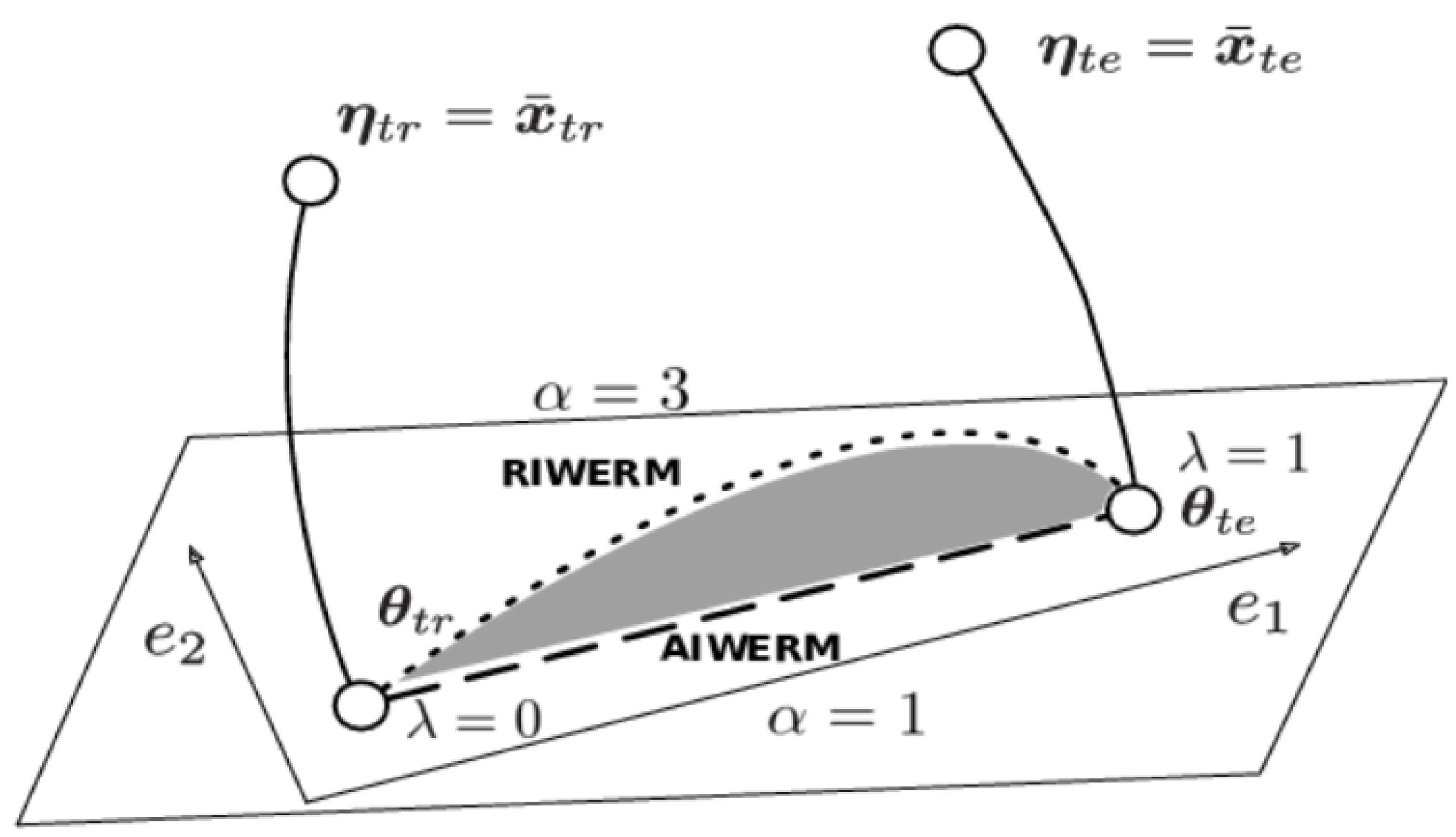

図 1: 共変量シフト適応の幾何．$\boldsymbol{\theta}$-座標系の上で，破線はAIWERMに，点線はRIWERMに対応する．ここで$\lambda = 0$と$\lambda = 1$はそれぞれθ_{tr} (ERM) およびθ_{te} (IWERM)に対応し，$\alpha = 1$と$\alpha = 3$はそれぞれAIWERMとRIWERMの曲線に対応する．

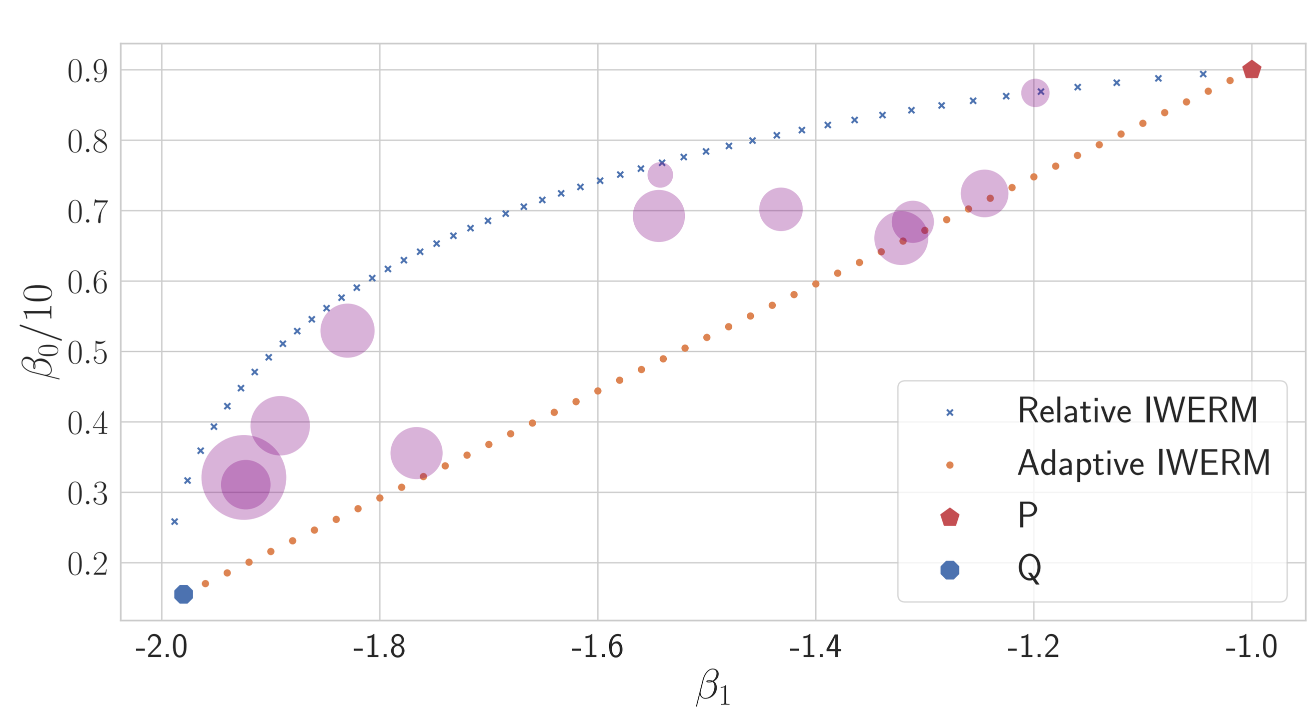

図 2: IGIWERMのベイズ最適化