

# 変数選択付きカーネル密度比推定に基づく多施設の予後予測解析

理化学研究所 革新知能統合研究センター	松井孝太
理化学研究所 革新知能統合研究センター	熊谷亘
名古屋工業大学 情報工学専攻	金森研太
名古屋大学 医学系研究科	錦見満暁
東京工業大学 情報理工学院	金森敬文

## 1 はじめに

疾患の予後予測解析は、適切な治療方針の選択や疾患の重症化リスクの予測などに役立つ可能性があり、多くは多施設研究として様々な疾患に対して実施されている。多施設研究は、短期間で比較的多くの患者を集積できるなどのメリットがある一方で、施設ごとに患者の臨床あるいは遺伝的情報に大きな変動が生じる可能性がある。このような問題は、共変量シフトとして適切にモデリングを行うことで変動による予測精度の低下を抑止することができる。提案法である変数選択付きカーネル密度比推定法 [1] は、上述のモデリングを実現すると同時に、モデリングに必要な変数を削減し、実際の運用コストを下げられ得ることを示す。

## 2 問題設定

予後予測問題は、ある疾患の患者の予後が良好 ( $Y = 0$ ) か不良 ( $Y = 1$ ) かを予測する 2 値判別問題と解釈できる。共変量シフトは、患者の背景情報を表す共変量  $X$  とアウトカム  $Y$  の確率分布に対して  $P_{tr}(Y | X) = P_{te}(Y | X)$  を仮定する。ここで、 $P_{tr}, P_{te}$  はそれぞれ予測モデルを学習するデータの分布とそれを運用するテストデータの分布を表す。直感的には、施設毎に患者の背景情報に変動はあるが、同様な背景情報を持つ患者は予後も同様であることを期待する仮定である。共変量シフトを加味した予測モデルの学習は、以下の密度比による重み付き経験損失最小化によって実現される：

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^n w(\mathbf{x}_i^{tr}) \ell(h_{\theta}(\mathbf{x}_i^{tr}), y_i^{tr}). \quad (1)$$

ここで、 $w(\mathbf{x}) = \frac{P_{tr}(\mathbf{x})}{P_{te}(\mathbf{x})}$  であり、 $\ell$  は損失関数、 $h_{\theta}$  はパラメータ  $\theta$  を持つ仮説である。(1) の学習を正しく行うためには、学習データとテストデータの密度比  $w(\mathbf{x})$  を精度良く推定することが必要である。一方、多施設研究においては共変量の中に測定コストの高いものが含まれている場合があり、運用上ボトルネックとなる。そこで本研究では、変数選択付きカーネル密度比推定によって密度比の推定と、推定に寄与する変数の選択を同時に実現する。

## 3 方法

提案法は、以下のアルゴリズムによって実行される：

**Step 1** カーネル密度比推定 [2] によって  $w$  の推定量  $\hat{w}$  を得る。

**Step 2**  $\hat{w}_{\xi}(\mathbf{x}) = \hat{w}(\xi \circ \mathbf{x})$  とおく。ここで、 $\circ$  は成分毎の積を表す。以下の問題を解いて  $\xi$  の推定値  $\hat{\xi}$  を得る：

$$\hat{\xi} = \arg \min_{\xi \in [0,1]^d} L_n(\hat{w}_{\xi}) + \eta \|\xi\|_1 \quad (2)$$

ここで、 $L_n$  は経験損失関数を表す。 $\eta$  は正則化パラメータである。

出力  $\hat{w}_{\hat{\xi}}(\mathbf{x})$ 。

$\hat{w}_{\hat{\xi}}(\mathbf{x})$  が推定されれば、(1) の  $w(\cdot)$  に plug-in してモデルの学習を行うことができる。提案法の理論的性質や適用例については当日報告する。

## 参考文献

- [1] K. Matsui et al., “Variable Selection for Nonparametric Learning with Power Series Kernels”, arXiv:1806.00569.
- [2] T. Kanamori et al., “Statistical analysis of kernel-based least-squares density-ratio estimation.” Machine Learning 86.3 (2012): 335-367.