基于联合 Trans-Lasso 算法的高维线性回归模型参数估计

刘　毅

(华北水利水电大学数学与统计学院,郑州 450046)

摘要：针对高维线性回归模型参数估计问题,基于迁移学习提出一种新算法,即联合 Trans-Lasso 算法。该算法使用高维统计技术有效结合大量与目标样本相似的辅助样本及少量的目标样本,在充分考虑辅助样本信息性的前提下对目标模型进行参数估计,有效降低了估计误差。通过数值模拟将迁移弹性网算法、OrcaleTrans-Lasso 算法、联合 Trans-Lasso 算法与传统的 Lasso 算法估计性能进行比较,结果表明,联合 Trans-Lasso 算法的估计误差最小,提高率最高。

关键词：高维;Lasso;迁移学习;参数估计

中图分类号：O211.9　　**文献标志码**：A　　**文章编号**：1674 - 8646(2023)20 - 0022 - 03

Parameter Estimation of High-dimensional Linear Regression Model Based on Joint Trans-Lasso Algorithm

Liu Yi

(School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou 450046, China)

Abstract: Aiming at the parameter estimation problem of high-dimensional linear regression model, the study proposes a new algorithm based on transfer learning-Joint Trans-Lasso algorithm. This algorithm uses high-dimensional statistical techniques to effectively combine a large number of auxiliary samples similar to the target sample, and a small number of target samples. Under the premise of fully considering the information of auxiliary samples, the parameters of the target model are estimated, which effectively reduces the estimation error. At the same time, the estimated performance of the migration elastic net algorithm, Orcale Trans-Lasso algorithm, the Joint Trans-Lasso algorithm and the traditional Lasso algorithm are compared through numerical simulation. The results show that the Joint Trans-Lasso algorithm has the smallest estimation error and the highest improvement rate.

Key words: High-dimensional; Lasso; Transfer learning; Parameter estimation

0　引言

回归分析是广泛使用的统计方法之一,可了解结果与一组协变量的关联。但现代统计分析以高维统计为主,即统计模型中有较多参数,在高维回归分析中表现为自变量的个数远大于样本数。经典的处理方法是充分利用先验信息,如稀疏性来提取最相关的某些变量参数(如 Lasso 估计[1]、弹性网络、岭回归等)。高维问题的特点是变量较多,但用于研究的目标数据量较少,达不到研究需要的样本量,导致建立的模型或算法在实际应用中难以表现出较好的性能。解决此类问题的有效方法是迁移学习[2],它将一些有用的信息从相似的任务迁移到原始任务,以达到较好的学习及预测效果,即将一些与目标模型相关且样本量足够的数据作为辅助样本进行研究,可有效解决高维回归问题。迁移学习得到了广泛应用,例如在某些生物或医学研究,由于伦理或成本问题难以获得生物学或临床结果,可利用迁移学习从不同但相关的生物学结果中收集信息,提高结果的预测性及估计性。还可用于商品推荐[3],许多网络平台都希望通过预测客户购买可能性来向其推荐个性化商品,但每个客户的历史采购数据有限,可将客户点击数据作为辅助数据,通过迁移学习来对购买任务进行预测。学者对其具体应用进行了深入研究,Pan[4]等研究了其在客户评论分类中的应用,Hajiramezanali[5]等研究了其在医疗诊断中的应用,Wang[6]等研究了拼车平台中的乘车调度问题。Ma[7]等对辅助样本及目标样本的高维问题进行探讨,分析了多源高维线性回归问题。还有人提出了几种L_1惩罚或约束的最小化方法,将其用于高维线性回归的预测及估计[8-10]。Bastani[3]等利用高维统计技术提出了一种结合大量辅助数据及少量目标数据的新型两步估计器。Li Sai[11]等考虑在迁移学习的基础上使用一些

收稿日期:2023 - 08 - 04

作者简介:刘　毅(1998 -),男,硕士研究生。研究方向:高维统计。

来自不同但可能相关的回归模型辅助样本及目标模型样本对目标模型进行参数估计及预测分析。Tian[12]等研究了高维广义线性模型(GLM)下的迁移学习问题。本研究分析了处理高维线性回归模型参数估计问题的几种迁移学习算法,对其性能进行评估及比较。

1 几种迁移学习算法

考虑高维线性回归模型中的迁移学习,目标模型可写成:

$$y_i^{(0)} = [x_i^{(0)}]^T\beta + \varepsilon_i^{(0)}, i = 1, \cdots, n_0 \tag{1}$$

其中,$[(x_i^{(0)})^T, y_i^{(0)}], i = 1, \cdots, n_0$,是独立样本,$\beta \in \mathbb{R}^p$是回归系数,$\varepsilon_i^{(0)}$是独立分布的随机噪声,使得$\mathbb{E}[\varepsilon_i^{(0)} \mid x_i^{(0)}] = 0$。在高维情况下,$p$可以大于$n_0$且大部分情况下比$n_0$大得多,假设$\beta$是稀疏的,用$s$表示$\beta$的非零元素数量,且$s$比$p$小得多。从辅助模型观察$K$个辅助研究样本$[(x_i^{(k)})^T, y_i^{(k)}]$。

$$y_i^{(k)} = (x_i^{(k)})^T w^{(k)} + \varepsilon_i^{(k)}, i = 1, \cdots, n_k, k = 1, \cdots, K \tag{2}$$

其中,$w^{(k)} \in \mathbb{R}^p$是第k次研究的真实系数向量,$\varepsilon_i^{(k)}$是随机噪声,使得$\mathbb{E}[\varepsilon_i^{(k)} \mid x_i^{(k)}] = 0$。回归系数$w^{(k)}$未知,且与目标$\beta$是不同的。利用目标数据$[(x_i^{(0)})^T, y_i^{(0)}], i = 1, \cdots, n_0$及第$k$个辅助数据$[(x_i^{(k)})^T, y_i^{(k)}], i = 1, \cdots, n_k, k = 1, \cdots, K$来研究模型(1)。

辅助样本是在对目标模型进行参数估计时提供一些有用信息的样本,因此用于辅助研究的辅助模型与目标模型之间具有一定的相似性。辅助样本具有信息性的前提是该辅助模型与目标模型相似。使用$w^{(k)}$与β之间的差异稀疏性来表示第k个辅助研究的信息水平。设$\delta^{(k)} = \beta - w^{(k)}$表示$w^{(k)}$与$\beta$之间的差异性。信息辅助样本是差异性足够稀疏的样本,即$w^{(k)}$与β之间的差大部分为零。用集合A_0来表示信息辅助样本集:

$$A_0 = \left\{1 \leqslant k \leqslant K: \| \delta^{(k)} \|_0 \leqslant h\right\} \tag{3}$$

对于一个向量$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_p)^T \in \mathbb{R}^p$,定义几种范数如下:$\|\alpha\|_0$为$\alpha$中非零元素的个数,$\|\alpha\|_1 = \sum_{j=1}^{p} |\alpha_j|$,$\|\alpha\|_2 = \sqrt{\sum_{j=1}^{p} \alpha_j^2}$。

1.1 迁移弹性网算法

弹性网算法是一种综合 Lasso 回归与岭回归的回归算法。在 Lasso 回归进行变量选择时,有时会筛掉某些对研究有利的变量,而利用岭回归研究问题则不能保证稀疏假设。故考虑利用迁移弹性网算法来研究高维线性回归问题。该算法利用辅助数据对辅助模型的回归参数进行估计,利用L_1与L_2惩罚项,利用目标数据及估计出的辅助模型回归参数对目标模型参数进行估计。

算法 1:迁移弹性网算法

输入目标数据$(X^{(0)}, Y^{(0)})$及信息辅助样本数据$\left\{X^{(k)}, Y^{(k)}\right\}_{k \in A}$

输出$\hat{\beta}$

计算

$$\hat{\omega}^A = \arg\min_{\omega \in \mathbb{R}^p} \left\{ \frac{1}{2n_A} \sum_{k \in A} \|y^{(k)} - X^{(k)}\omega\|_2^2 + \lambda_\omega (\|\omega\|_1 + \|\omega\|_2) \right\} \tag{4}$$

其中,$\lambda_\omega = \sqrt{\log p \backslash n_A}$

令

$$\hat{\beta} = \hat{\omega}^A + \hat{\delta}^A \tag{5}$$

其中,

$$\hat{\delta}^A = \arg\min_{\delta \in \mathbb{R}^p} \left\{ \frac{1}{2n_0} \|y^{(0)} - X^{(0)}(\hat{\omega}^A + \delta)\|_2^2 + \lambda_\delta (\|\delta\|_1 + \|\delta\|_2) \right\} \tag{6}$$

其中,$\lambda_\delta = \sqrt{\log p \backslash n_0}$

1.2 Orcale Trans-Lasso 算法

Orcale Trans-Lasso 算法是由 LiSai 等提出的一种处理高维线性回归问题的迁移学习算法,使用所有目标数据及信息辅助样本计算$\hat{\omega}^A$,使用$\hat{\omega}^A$对$\delta^{(k)}$进行估计。

算法 2:Orcale Trans-Lasso 算法

输入目标数据$(X^{(0)}, Y^{(0)})$及信息辅助样本数据$\left\{X^{(k)}, Y^{(k)}\right\}_{k \in A}$

输出$\hat{\beta}$

计算

$$\hat{\omega}^A = \arg\min_{\omega \in \mathbb{R}^p} \left\{ \frac{1}{2n_A} \sum_{k \in A} \|y^{(k)} - X^{(k)}\omega\|_2^2 + \lambda_\omega \|\omega\|_1 \right\} \tag{7}$$

其中,$\lambda_\omega = \sqrt{\log p \backslash n_A}$

令

$$\hat{\beta} = \hat{\omega}^A + \hat{\delta}^A \tag{8}$$

其中,

$$\hat{\delta}^A = \arg\min_{\delta \in \mathbb{R}^p} \left\{ \frac{1}{2n_0} \|y^{(0)} - X^{(0)}(\hat{\omega}^A + \delta)\|_2^2 + \lambda_\delta \|\delta\|_1 \right\} \tag{9}$$

其中,$\lambda_\delta = \sqrt{\log p \backslash n_0}$

1.3 联合 Trans-Lasso 算法

算法 2 通过对辅助模型的回归系数$w^{(k)}$及其与目标模型的回归系数β之间的差距$\delta^{(k)}$的估计得到结果,但估计量与真实值之间总是存在一定的差距,为了缩

小这个差距，引入一个新的量$\gamma^{(k)}=\beta-w^{(k)}-\delta^{(k)}$，表示$w^{(k)}+\delta^{(k)}$与$\beta$之间的差距，将辅助数据与真实数据回归系数之间的差距分为更详细的两部分进行估计，得到更精确的结果。在联合 Lasso 算法中，信息辅助样本集合更新为：

$$A_0=\left\{1\leqslant k\leqslant K:\|\delta^{(k)}\|_0\leqslant h,\|\gamma^{(k)}\|_0\leqslant h\right\} \tag{10}$$

算法 3：联合 Trans-Lasso 算法

输入目标数据$(X^{(0)},Y^{(0)})$及信息辅助样本数据$\left\{X^{(k)},Y^{(k)}\right\}_{k\in A}$

输出$\hat{\beta}$

计算

$$\hat{\omega}^A=\arg\min_{\omega\in\mathbb{R}^p}\left\{\frac{1}{2n_A}\sum_{k\in A}\|y^{(k)}-X^{(k)}\omega\|_2^2+\lambda_\omega\|\omega\|_1\right\} \tag{11}$$

其中，$\lambda_\omega=\sqrt{\log p\backslash n_A}$

计算

$$\hat{\delta}^A=\arg\min_{\delta\in\mathbb{R}^p}\left\{\frac{1}{2n_0}\|y^{(0)}-X^{(0)}(\hat{\omega}^A+\delta)\|_2^2+\lambda_\delta\|\delta\|_1\right\} \tag{12}$$

对一些常数C_2，$\lambda_\delta=\sqrt{\log p\backslash n_0}$

令

$$\hat{\beta}=\hat{\omega}^A+\hat{\delta}^A+\hat{\gamma}^A \tag{13}$$

其中，$\hat{\gamma}^A=\arg\min_{\gamma\in\mathbb{R}^p}\left\{\frac{1}{2n_0}\|y^{(0)}-X^{(0)}(\hat{\omega}^A+\hat{\delta}^A+\gamma)\|_2^2+\lambda_\gamma\|\gamma\|_1\right\}$ (14)

其中，$\lambda_\gamma=\sqrt{\log p\backslash n_0}$

2 数值模拟

通过数值模拟评估 Lasso、迁移弹性网、Orcale Trans-Lasso 及联合 Trans-Lasso 4 种方法的性能，令 $p=500$，$n_0=150$，$n_1,\cdots,n_K=100$，$K=20$。$x_i^{(k)}$和$\varepsilon_i^{(k)}$都是独立同分布的，对于目标参数β，令$\beta_j=0.3$，$s=16$，$j\in\left\{1,\cdots,s\right\}$，其他参数都设为零。如图 1、图 2 所示：

图 1 与图 2 的横坐标代表信息辅助样本集A_0的不同取值，纵坐标表示各种算法在对模型参数进行估计时产生的均方误差。

由图 1、图 2 可知，与传统的 Lasso 算法相比，迁移弹性网、Orcale Trans-Lasso、联合 Trans-Lasso 在对高维回归模型参数进行估计时误差较小，表明这三种算法在处理此类问题时能够表现出较好的性能。且 Lasso 的估计性能并不随着信息辅助样本集合的改变而变化，三种迁移学习算法的估计误差随着信息辅助样本集合的增大而减小。

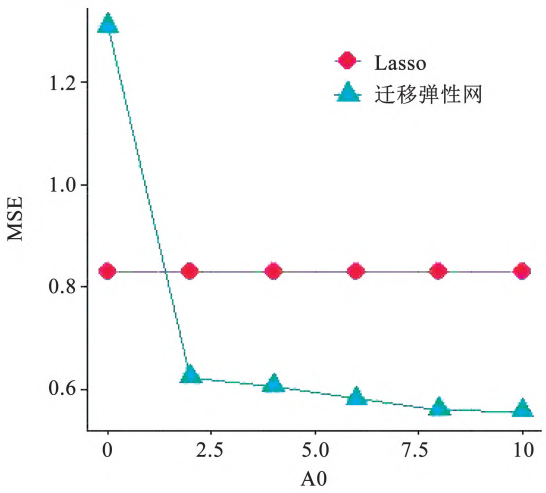

图 1　迁移弹性网与 Lasso 的估计误差

Fig. 1　Estimation error of transfer elastic network and Lasso

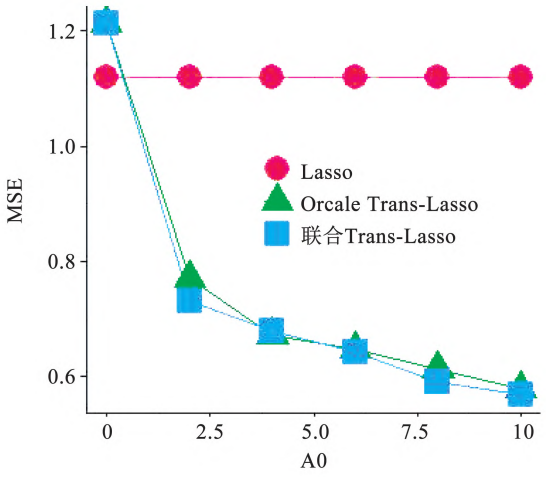

图 2　联合 Trans-Lasso、Orcale Trans-Lasso 及 Lasso 的估计误差

Fig. 2　Estimation error of combined Trans-Lasso, Orcale Trans-Lasso and Lasso

表 1　Lasso、迁移弹性网、Orcale Trans-Lasso 及联合 Lasso 均方误差对比

Tab. 1　Comparison of mean square error of Lasso, migration elastic network, Orcale Trans-Lasso and Joint Lasso

	迁移弹性网	Orcale Trans- Lasso	联合 Trans-Lasso
MSE	0.134031	0.116598	0.102461
LassoMSE	0.829622	1.119774	1.119774
绝对误差	-0.69559	-1.00318	-1.01731
提升率	83.84%	89.58%	90.85%

其中 MSE 等于参数β的估计值与真实值的差的平方的平均值，绝对误差等于 MSE 与 LassoMSE 之间的差，提升率等于绝对误差与 LassoMSE 的比值。

从表 1 可以看出，三种算法与传统的 Lasso 算法相比都有较高的提升率，其中迁移弹性网算法、Trans-Lasso 算法、联合 Lasso 算法的提升率分别为 83.84%、89.58%、90.85%。联合 Trans-Lasso 算法的提升率最高，说明联合 Lasso 算法处理高维回归模型的参数估计问题会表现出更好的性能。

（下转第 28 页）

[11] Gao Fangfang, Tian Wei, Wang Zhen, et al. Effect of diameter of multi-walled carbon nanotubes on mechanical properties and microstructure of the cement-based materials[J]. Construction and Building Materials,2020,260:120452.
[12] 高芳芳. 高温冷却后多壁碳纳米管混凝土力学性能研究[D]. 西安:长安大学,2022.
[13] 陆富龙. 碳纳米管对水泥基材料的作用机理研究[D]. 广州:广州大学,2019.
[14] 董怀斌,李长青,邹霞辉. 电场诱导碳纳米管在聚合物中定向有序排列的研究进展[J]. 材料导报 A,2018,32(04):427-433.
[15] 杜润红,闫伟,杜春良,等. 外加电场对碳纳米管在正丁醇中的定向分散[J]. 天津工业大学学报,2020,39(04):15-19.
[16] 彭刚,蔡晓兰,周蕾. 碳纳米管的分散性研究进展[J]. 化工新型材料,2015(09):39-41.
[17] 郑小青. 碳纳米材料对水泥基材料耐久性的影响研究[J]. 东莞理工学院学报,2021,28(03):94-99.
[18] Feng Tao, Liu Neng, Wang Shunjie, et al. Research on the dispersion of carbon nanotubes and their application in solution-processed polymeric matrix composites: a review [J]. 2021,10(06): 559-576.
[19] 张芳芳,李雷,张轲. 碳纳米管功能化表面修饰研究进展[J]. 广州化工, 2013,41(17):31-33.
[20] 杨景红,梅亚平,潘成,等. 多壁碳纳米管表面修饰改性研究[J]. 安徽化工,2023,49(03):91-96.
[21] Le Thi Mai Hoa. A Study on the effects of potassium permanganate on the functionalization of multi-walled carbon nanotubes[J]. ECS Journal of Solid State Science and Technology,2022,11(11):1004.
[22] 范杰,李庚英,王中坤. 表面处理碳纳米管对水泥砂浆性能影响的研究[J]. 新型建筑材料,2019,49(08):43-47.
[23] 徐涛,杨静晖,傅强,等. 等离子体处理碳纳米管的研究现状及展望[J]. 物理学进展,2008(01):78-82.
[24] 赵毅毅. 等离子体改性碳纳米管及碳纳米管气凝胶吸附染料分子的性能及机理研究[D]. 西安:西北大学,2022.
[25] 蒋威. 海水腐蚀对改性碳纳米管水泥基材料性能的影响[D]. 苏州:苏州科技大学,2019.
[26] 张作钦,潘贵翔,李杰,等. CO_2等离子体改性多壁碳纳米管对沥青基炭材料微结构及性能影响[J]. 铁道标准设计,2022,66(09):142-147.
[27] Zhu Yuanheng, Sun Min, Li Zhendong, et al. Influence of plasma modified carbon nanotubes on the resistance sensitiveness of cement [J]. Journal of Wuhan University of Technology-Mater Sci Ed,2023, 38(01):136-140.
[28] 许苗,李彩兰,蔡永杰,等. 碳纳米管非共价功能化的应用进展[J]. 化学与生物工程,2023,40(01):1-7.
[29] 肖昱强. 碳纳米管水泥砂浆复合材料导电性和压阻性研究[D]. 西安:长安大学,2021.
[30] 南茜. 非共价功能化法修饰碳纳米管在溶剂及高分子中的分散研究[D]. 太原:太原理工大学,2017.
[31] 陈泽宇,刘 静,蒲春生,等. 表面活性剂分散多壁碳纳米管机理及性能评价[J]. 精细化工,2022,39(02):269-275,410.
[32] 秦煜,唐元鑫,阮鹏臻,等. 碳纳米管悬浮液分散质量影响因素试验研究[J]. 应用化工,2022,51(05):1287-1290.
[33] 卜路霞,李京京,高琳琳,等. SDBS 对多壁碳纳米管悬浮液分散性的影响[J]. 电镀与精饰,2019,41(07):10-13.
[34] 杜健民,刘政,庄文娟,等. 分散剂 PVP 和 PC 对 MWCNTs 的分散效果及机理研究[J]. 混凝土,2023(04):64-68.
[35] 黎恒杆,王玉林,罗昊,等. 多壁碳纳米管分散性及其对复合材料层间剪切力学性能影响影响[J]. 炭素技术,2021,40(06):26-32.
[36] 刘婉玥,杨文刚,姜欣荣,等. 碳纳米管预分散及其对复合材料层间剪切力学性能影响[J]. 炭素技术,2021,40(06):26-32.
[37] 朱鼎. 碳纳米管超高性能混凝土力学性能及抗硫酸盐侵蚀试验研究[D]. 常州:常州大学,2021.

(上接第 24 页)

3 结论

研究了在信息辅助样本已知的情况下几种处理高维线性回归问题算法的性能。结果表明,与传统的 Lasso 估计相比,迁移弹性网算法、Orcale Trans-Lasso 算法、联合 Trans-Lasso 算法的估计误差都远远小于 Lasso 估计,其中联合 Lasso 算法的估计误差最小,说明这几种迁移学习算法都能较好地解决此类高维回归问题。但迁移学习在统计学中的应用较少,可考虑在信息辅助样本未知的情况下联合 Trans-Lasso 算法及其他迁移学习算法,探讨其是否能表现出较好的性能。

参考文献:

[1] Caruana R. Multitask learning[J]. Machine Learning,1997,28(01): 41-75.
[2] Torrey L,Shavlik J,Walker T,et al. Transfer learning via advice taking [J]. Springer Berlin Heidelberg,2010(08):425-436.
[3] Bastani H. Predicting with proxies: transfer learning in high dimension [J]. Management Science,2021,67(05):2964-2984.
[4] Pan SJ,Yang Q. A survey on transfer learning[J]. IEEE transactions on Knowledge and Data Engineering,2009,22(10):1345-1359.
[5] Hajiramezanali E, Zamani Dadaneh S, Karbalayghareh A, et al. Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data[J]. Advances in Neural Information Processing Systems,2018,31:111-117.
[6] Wang Z, Qin Z, Tang X, et al. Deep reinforcement learning with knowledge transfer for online rides order dispatching[C]//2018 IEEE International Conference on Data Mining (ICDM). IEEE,2018.
[7] Ma R,Tony Cai T,Li H. Global and simultaneous hypothesis testing for high-dimensional logistic regression models [J]. Journal of the American Statistical Association,2021,116:984-998.
[8] Candès E,Tao T. Rejoinder: the Dantzig selector: statistical estimation when p is much larger than n[J]. The Annals of Statistics,2007,35 (06):2392-2404
[9] Zhang CH. Nearly unbiased variable selection under minimax concave penalty[J]. The Annals of Statistics,2010,38(02):894-942.
[10] Mullainathan S, Obermeyer Z. Does machine learning automate moral hazard and error? [J]. American Economic Review, 2017, 107 (05):476-480.
[11] Li Sai, Cai TT, Li H. Transfer learning for high-dimensional linear regression: prediction, estimation and minimax optimality [J]. Journal of the Royal Statistical Society: Series B Statistical Methodology,2019,84:149-173.
[12] Tian Y,Feng Y. Transfer learning under high-dimensional generalized linear models[J]. Journal of the American Statistical Association, 2022(04):1-14.