计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(1):11−20 [doi: 10.15888/j.cnki.csa.008247]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 ADMM 算法的网络连接数据变量选择①

方佳佳, 李　阳, 郑泽敏

(中国科学技术大学 管理学院统计与金融系, 合肥 230026)
通信作者: 李　阳, E-mail: tjly@mail.ustc.edu.cn

摘　要: 随着科技的发展, 网络连接数据在统计学习、机器学习等领域的应用越来越普遍. 在线性回归模型中, 目前关于网络连接数据的变量选择研究主要针对的是同质性样本, 即样本的个体效应 α 相同, 但在现实中大多数样本的个体效应存在异质性, 在不考虑异质性的情况下会使得模型的估计和预测产生较大偏差. 因此, 当网络数据中个体效应存在组异质性时, 本文提出一种新的变量选择方法 SNC. 利用网络凝聚效应, 我们对变量系数和相连样本个体效应的差异性进行联合惩罚, 同时采用 ADMM 算法进行求解, 并证明了算法的收敛性. 数值模拟和实证分析显示, 我们的方法提高了变量选择的准确性并且降低了预测误差.

关键词: 网络连接数据; 网络凝聚效应; 组异质性; 变量选择; 非凸惩罚

引用格式: 方佳佳,李阳,郑泽敏.基于 ADMM 算法的网络连接数据变量选择.计算机系统应用,2022,31(1):11–20. http://www.c-s-a.org.cn/1003-3254/8247.html

Variable Selection of Network-linked Data Based on ADMM Algorithm

FANG Jia-Jia, LI Yang, ZHENG Ze-Min

(Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China)

Abstract: With the development of science and technology, the application of network-linked data in statistical learning, machine learning and other fields becomes increasingly common. In linear regression models, the current research on the variable selection of network-linked data mainly focuses on the homogeneous samples, namely that the individual effects of the samples are the same. In reality, however, the individual effects of most samples are heterogeneous. As a result, the neglect of the heterogeneity will lead to large deviations in the estimation and prediction of the models. Therefore, this paper proposes a new variable selection method SNC to cope with the situation when there is group heterogeneity in network-linked data. Using the network agglomeration effect, we carry out a joint penalty for the difference between the variable coefficient and the individual effect of the connected samples and solve the problem with ADMM algorithm, with the convergence of the algorithm proved. The results of numerical simulation and example analysis show that this method improves the accuracy of variable selection and reduces the prediction error.

Key words: network-linked data; network agglomeration effect; group heterogeneity; variable selection; nonconvex penalty

随着科学技术的进步, 不同领域的数据都呈现出网络连接的趋势, 许多科学领域都涉及某种形式的网络研究, 例如人际关系研究、学术论文合著和引用、蛋白质相互作用模式等. 20 年前, 关于网络的流行书籍及其研究开始出现[1], 而像 Facebook、MySpace 和 LinkedIn 这样的在线网络社区在近 10 年间也是蓬勃

① 基金项目: 国家自然科学基金 (12101584, 11601501, 11671374, 71731010, 71921001); 中国博士后科学基金 (2021TQ0326, 2021M703100); 2021 年合肥市博士后科研活动项目
收稿时间: 2021-03-21; 修改时间: 2021-04-21; 采用时间: 2021-04-26; csa 在线出版时间: 2021-12-17

兴起, 这更加增强了人们对网络数据的研究兴趣. 网络连接数据由节点和边组成, 社交网络是此类网络模型的一个典型代表. 社交网络中, 每个节点代表一个人, 边代表人与人之间的沟通交流, 此外, 还有商业网络、基因网络等.

目前关于网络连接数据的研究主要分为两个方面. 一方面是关于网络结构的研究. 另一方面主要是将网络连接数据中的结构信息与统计学习中常用的经典模型结合起来研究.

在网络结构方面, 最早被应用于社区检测. 社区检测兴起于物理学和计算机科学领域, 而后开始应用于统计领域. 其中一类社区检测算法是通过在节点的所有可能分区上优化启发式全局准则来检测社区[2,3]. 基于概率模型的方法[4,5] 是另一类社区检测算法. 一些学者从观察到的邻接矩阵中检测社区或潜在结构[6–8], 从其他节点之间的信息估计特定节点之间的边缘概率[9]. 社交网络是此类网络模型的代表, 因此针对社交网络的研究也受到了大量的关注[10,11].

在与经典模型结合方面, 一般是与常用的模型相结合. 例如, 时间序列模型[12], 线性模型[13], 变系数模型[14], 随机效应模型[15], 变化点检测问题[16], 自回归模型[17,18] 等.

线性回归模型是统计学习中的经典模型之一, 应用十分广泛, 关于网络数据的回归模型也开始引起学者的关注. 例如, Asur 等[19] 将网络数据应用于预测模型, 通过研究网络结构来预测现实生活中某一现象的结果. Li 等[13] 将网络连接数据应用于回归预测模型, Zhu 等[17] 和 Tang 等[18] 将网络连接数据与自回归模型相结合, 都表明网络连接数据在回归模型中的研究价值. 随着科技的发展, 数据的采集变得更加容易, 高维数据也越来越受到研究学者的关注, 但是高维数据中存在大量的冗余信息, 如何选出有研究价值的数据? 变量选择领域应运而生. 故将网络连接数据应用到变量选择领域是一个值得研究的课题.

对于线性回归模型, 超高的维度使得传统的普通最小二乘法不再适用. 正则化是稀疏建模和变量选择的有效方法, 通过在目标函数上添加惩罚函数来降低模型的复杂度. 根据惩罚函数的不同, 正则化方法一般可以分为凸正则化和非凸正则化.

凸正则化方法主要包括岭回归、LASSO、弹性网以及 Dantzig Selector 等. 虽然凸正则化的研究已经很成熟, 但由于惩罚函数的凸性, 使得凸正则化估计量都是有偏的. Fan 等提出了一个非凸正则化方法—SCAD (smoothly clipped absolute deviation)[20], 并证明了其 Oracle 性质. 非凸惩罚函数回归的渐进无偏估计, 能进一步降低模型的预测总误差. 此后, 非凸惩罚受到了广泛的关注, 例如 MCP (minimax concave penalty)[21]、限制 Capped-L_1[22]、Hard 阈值惩罚[23] 等.

关于网络连接数据的变量选择问题近年来也有学者做过相关研究[24,25]. 例如 Li 等[24] 和 Kim 等[25] 考虑样本系数之间的网络凝聚效应, 即网络中连接节点表现出相似的行为, 对系数同时施加了 L_1 惩罚和凝聚效应惩罚 $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{\beta}$, 从而能够解决网络连接数据的变量选择问题, 但他们针对的是同质性网络连接数据, 即假设每个样本的个体效应值 $\boldsymbol{\alpha}$ 相同, 并没有考虑到异质性, 异质性是指不同样本的个体效应 $\boldsymbol{\alpha}$ 不同. 在现实生活中, 因为网络凝聚效应的存在而使得网络中的样本存在群组效应, 联系密切的样本组成一个群组, 他们之间的行为会相互影响而慢慢趋同. 针对线性回归模型, 这种群组效应的一个直观体现就是群组内样本的个体效应 $\boldsymbol{\alpha}$ 相同, 不同群组间个体效应 $\boldsymbol{\alpha}$ 不同. 若忽略群组间个体效应的差异性, 将所有样本的个体效应视为相同, 在进行变量选择和预测估计时都会产生较大偏差, 影响模型精度. 故考虑异质性, 能够提高模型精度. 因此, 针对异质性网络连接数据的研究具有重要的价值和实际意义. Li 等[13] 考虑到个体效应之间的异质性, 并惩罚相连样本个体效应的差异性, 提高了回归模型中估计和预测的精度, 但他主要关注的是预测问题, 没有涉及到变量选择.

本文的目标是对因网络凝聚效应而产生个体效应的组异质性的网络连接数据进行变量选择, 我们对组内样本间个体效应的差异性 $\boldsymbol{L}\boldsymbol{\alpha}$ 和变量系数 $\boldsymbol{\beta}$ 进行联合惩罚, 从而保证组内样本的个体效应具有相同的估计值. 本文提出的方法不仅能够处理含有组异质性的网络连接数据的变量选择问题, 而且能够改善变量选择、估计和预测的结果. 在本文中, 我们主要使用 L_1、MCP 和 SCAD 罚函数, 并且运用 ADMM 算法进行求解, 同时证明了算法的收敛性.

1 网络连接数据的变量选择方法

1.1 模型设定

本文中所有的向量都是列向量. 考虑一般的线性回归模型, $\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$ 是 n 维响应变量, $\boldsymbol{X}=(x_1,$

$x_2,\cdots,x_n)^{\mathrm{T}}$是$n\times p$设计矩阵. 假设$\boldsymbol{X}$是固定的且其列已经标准化. 样本$\boldsymbol{X}$的结构网络为$G=(V,E)$, 其中$V=\{1,2,\cdots,n\}$为样本节点集合, $E\subset V\times V$为边的集合. 我们用邻接矩阵$\boldsymbol{A}=(A_{uv})_{n\times n}\in R_{n\times n}$表示该网络以及样本节点和节点之间的连接关系, 若$(u,v)\in E$, 则$A_{uv}=1$, 否则为 0. $A_{uu}=0, A_{uv}=A_{vu}$. 网络G的拉普拉斯矩阵$\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{A}$, $\boldsymbol{D}=diag(d_1,d_2,\cdots,d_n)$为度矩阵, $\boldsymbol{D}$的对角线元素为每个节点的度$d_u=\sum_v A_{uv}$. 建立如下线性回归模型:

$$\boldsymbol{Y}=\boldsymbol{\alpha}+\boldsymbol{X\beta}+\boldsymbol{\varepsilon} \tag{1}$$

其中, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm{T}}$是节点个体效应向量. 假设相连样本的个体效应相等, 不相连样本的个体效应不等, 即样本之间存在组异质性. $\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_p)^{\mathrm{T}}$是模型的回归系数向量. $\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)^{\mathrm{T}}$是$n$维误差向量, $E(\boldsymbol{\varepsilon})=\boldsymbol{0}$, $\boldsymbol{var}(\boldsymbol{\varepsilon})=\sigma^2\boldsymbol{I}_n$.

Li 等[13] 提出了网络连接数据的预测方法 (the regression with network cohesion, RNC), 其主要思想是最小化如下损失函数:

$$\boldsymbol{L}(\boldsymbol{\alpha},\boldsymbol{\beta})=\|\boldsymbol{Y}-\boldsymbol{\alpha}-\boldsymbol{X\beta}\|^2+\mu\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{L\alpha} \tag{2}$$

其中, $\mu>0$是调整参数. RNC 主要是惩罚网络中相连节点个体效应的差异性, 该惩罚可以推导出一个等价的、更直观的形式$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{L\alpha}=\sum_{(u,v)\in E}(\alpha_u-\alpha_v)^2$.

1.2 组异质性网络连接数据的变量选择方法 (SNC)

RNC 中假设各样本的个体效应不相等, 惩罚项$\mu\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{L\alpha}$用来惩罚相连样本个体效应的差异性, 从而能够处理异质性网络连接数据的回归预测问题. 但是由于连接网络中的个体常常存在群组效应, 同一个群组中行为特征存在统一准则而基本相同. 因此, 在本文中我们假设样本之间存在组异质性, 即组内样本 (相连样本) 的个体效应相等, 组间样本 (不相连样本) 的个体效应不相等. 通过对$\boldsymbol{L\alpha}$施加惩罚, 惩罚组内样本个体效应的差异性并压缩至 0, $\boldsymbol{L\alpha}$中的元素是$(\alpha_u-\alpha_v)_{(u,v)\in E}$或其等价形式. 为了产生$\boldsymbol{\beta}$的一个稀疏估计, 我们将同时惩罚$\boldsymbol{\beta}$和$\boldsymbol{L\alpha}$, 这就是我们提出的方法—网络连接数据的变量选择 (variable selection with network cohesion, SNC).

令 $\boldsymbol{\theta}=(\boldsymbol{\beta}^{\mathrm{T}},\boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{H}=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{L}\end{pmatrix}$, 则$\boldsymbol{H\theta}=\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{L\alpha}\end{pmatrix}$.

SNC 的目标函数为:

$$Q(\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{1}{2n}\|\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}\|_2^2+p_\lambda(|\boldsymbol{H\theta}|) \tag{3}$$

在本文中, 对$\boldsymbol{H\theta}$的惩罚主要使用L_1和非凸惩罚, 非凸惩罚包括 MCP 和 SCAD 罚函数. MCP 罚函数为$p_M(t,\lambda)=\lambda\int_0^t\left(1-\frac{x}{a\lambda}\right)_+dx$, $a>0$, SCAD 罚函数为$p_S(t,\lambda)=\lambda\int_0^t\min\{1,\left(a-\frac{x}{\lambda}\right)_+/(a-1)\}dx$, $a>2$.

将 SNC 方法的估计结果与没有对节点个体效应的差异进行惩罚的情况下进行对比, 能够提高估计和预测的精度.

2 算法

直接最小化目标函数 (3) 很难求解出估计量的值, 因为惩罚函数对于每个α_i是不可分的. 因此, 我们通过引入一组新的参数$\boldsymbol{\gamma}=\boldsymbol{H\theta}$来重新参数化准则. 最小化式 (3) 等价于最小化如下约束优化问题:

$$\begin{cases} Q(\boldsymbol{\theta},\boldsymbol{\gamma})=\dfrac{1}{2n}\|\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}\|_2^2+p_\lambda(|\boldsymbol{\gamma}|)\\ \text{s.t.}\quad \boldsymbol{H\theta}-\boldsymbol{\gamma}=\boldsymbol{0}\end{cases}$$

基于文献 [26] 中的思路, 利用增广拉格朗日方法, 通过最小化如下损失函数得到参数的估计:

$$L(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\varphi})=Q(\boldsymbol{\theta},\boldsymbol{\gamma})+\boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{H\theta}-\boldsymbol{\gamma})+\rho/2\|\boldsymbol{H\theta}-\boldsymbol{\gamma}\|_2^2 \tag{4}$$

其中, 对偶变量$\boldsymbol{\varphi}$是拉格朗日乘数, $\rho>0$是惩罚因子. 我们通过交替方向乘子法 (alternating direction multiplier method, ADMM) 来迭代求解$(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\varphi})$的估计. 对于给定的$(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\varphi})$, $L(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\varphi})$关于$\boldsymbol{\gamma}$的最小值是唯一的, 并且在$L_1$惩罚或非凸惩罚下有一个近似的形式. 当给定$(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\varphi})$, 上述最小化问题等价于:

$$\rho/2\|\boldsymbol{\tau}-\boldsymbol{\gamma}\|_2^2+p_\lambda(|\boldsymbol{\gamma}|)$$

其中, $\boldsymbol{\tau}=\boldsymbol{H\theta}+\rho^{-1}\boldsymbol{\varphi}$, 故在$L_1$或非凸惩罚下估计量的近似的形式为:

$$\hat{\boldsymbol{\gamma}}=ST(\boldsymbol{\tau},\lambda/\rho),$$

其中, $ST(t,\lambda)=sign(t)(|t|-\lambda)_+$是 soft 阈值准则, $(x)_+=x, x>0$, 否则$(x)_+=0$.

对于 MCP 罚函数$(a>1/\rho)$, 有:

$$\hat{\gamma}_{ij}=\begin{cases} ST(\tau_{ij},\lambda/\rho), & |\tau_{ij}|\le a\lambda\\ \tau_{ij}, & |\tau_{ij}|>a\lambda\end{cases}$$

对于 SCAD 罚函数$(a>1/\rho+1)$, 有:

$$\hat{\gamma}_{ij}=\begin{cases} ST(\tau_{ij},\lambda/\rho), & |\tau_{ij}|\le\lambda+\lambda/\rho\\ \dfrac{ST(\tau_{ij},a\lambda/(a-1)\rho)}{1-1/((a-1)\rho)}, & \lambda+\lambda/\rho<|\tau_{ij}|\le a\lambda\\ \tau_{ij}, & |\tau_{ij}|>a\lambda\end{cases}$$

算法步骤如算法 1.

算法 1. ADMM 算法

输入: 预测变量$\boldsymbol{X}$, 响应变量$\boldsymbol{Y}$, 邻接矩阵$\boldsymbol{A}$, 惩罚因子ρ, 停止准则η;
输出: $\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\varphi}}$;
目标: 迭代求解获得$\boldsymbol{\theta},\boldsymbol{\gamma}$和$\boldsymbol{\varphi}$.
初始化$\boldsymbol{\theta}^{(0)},\boldsymbol{\gamma}^{(0)}=\boldsymbol{H}\boldsymbol{\theta}^{(0)},\boldsymbol{\varphi}^{(0)}=\boldsymbol{0},m=0,\eta=0.03$.

While $m\geq 0$, **do**
$\boldsymbol{\theta}^{(m+1)}=[n^{-1}(\boldsymbol{X},\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{X},\boldsymbol{I})+\rho\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}]^{-1}$
$*[n^{-1}(\boldsymbol{X},\boldsymbol{I})^{\mathrm{T}}\boldsymbol{Y}+\rho\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\gamma}^{(m)}-\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\varphi}^{(m)}]$;
$\boldsymbol{\gamma}^{(m+1)}=ST(\boldsymbol{\tau}^{(m+1)},\lambda/\rho)$;
$\boldsymbol{\varphi}^{(m+1)}=\boldsymbol{\varphi}^{(m)}+\rho(\boldsymbol{H}\boldsymbol{\theta}^{(m+1)}-\boldsymbol{\gamma}^{(m+1)})$.
If $\boldsymbol{r}^{(m+1)}=\boldsymbol{H}\boldsymbol{\theta}^{(m+1)}-\boldsymbol{\gamma}^{(m+1)}$, $\|\boldsymbol{r}^{(m+1)}\|<\eta$
then
$(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\varphi}})=(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\gamma}^{(m+1)},\boldsymbol{\varphi}^{(m+1)})$;
Break;
Else
$m=m+1$;
End
End

对 ADMM 算法过程中的原始变量进行追踪, $\boldsymbol{r}^{(m+1)}=\boldsymbol{H}\boldsymbol{\theta}^{(m+1)}-\boldsymbol{\gamma}^{(m+1)}$. 停止准则为$\|\boldsymbol{r}^{(m+1)}\|<\eta$, 其中$\eta>0$为一个非常小的常数.

下面考虑 ADMM 算法的收敛性.

命题 1. 对于 MCP 和 SCAD 函数, ADMM 算法的原始残差$\boldsymbol{r}^{(m)}=\boldsymbol{H}\boldsymbol{\theta}^{(m)}-\boldsymbol{\gamma}^{(m)}$和对偶残差$\boldsymbol{s}^{(m+1)}=\rho\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{r}^{(m+1)}-\boldsymbol{r}^{(m)})$满足$\lim\limits_{m\to\infty}\|\boldsymbol{r}^{(m)}\|^2=0$, $\lim\limits_{m\to\infty}\|\boldsymbol{s}^{(m)}\|^2=0$.

命题 1 表明该算法实现了原可行性和对偶可行性, 证明材料见附录. 因此, 它收敛于一个局部最优点. 当采用非凸惩罚函数, 如 MCP 和 SCAD 罚函数时, 此最优点是目标函数的局部最优解. 综上, 算法收敛性和稳定性得到证明. 因为$\boldsymbol{\theta}^{(m)}=((\boldsymbol{\beta}^{(m)})^{\mathrm{T}},(\boldsymbol{\alpha}^{(m)})^{\mathrm{T}})^{\mathrm{T}}$是不稀疏的, 但我们已证明$\boldsymbol{H}\boldsymbol{\theta}^{(m)}=((\boldsymbol{\beta}^{(m)})^{\mathrm{T}},(\boldsymbol{L}\boldsymbol{\alpha}^{(m)})^{\mathrm{T}})^{\mathrm{T}}$是收敛于$\boldsymbol{\gamma}^{(m)}$, 故我们令$\boldsymbol{\gamma}^{(m)}$的前$p$项作为$\boldsymbol{\beta}$的估计值, 即可得到$\boldsymbol{\beta}$的稀疏解.

3 数值模拟

在数值模拟中, 主要比较本文提出的 SNC 方法和没有对个体节点效应的差异性进行惩罚的 LASSO、MCP、SCAD 方法在变量选择和预测方面的效果. 网络凝聚效应下的变量选择方法就是考虑了样本之间的连接关系网络的方法, 即我们的 SNC 方法.无网络凝聚效应下的变量选择方法, 就是不考虑样本之间的连接网络的惩罚方法. 在这里, 我们首先定义几个效果评估指标:

(1) 预测损失 (prediction error, PE): $\boldsymbol{E}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0+\boldsymbol{\alpha}_0-\boldsymbol{X}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\alpha}})^2$;

(2)L_q损失: $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_q, q=1,2,\infty$;

(3) 均方误差 (MSE($\widehat{\boldsymbol{\alpha}}$)): $E(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)^2$;

(4) 假阳性数 (false positives, FP): 真实为反例却被预测为正例的个数;

(5) 假阴性数 (false negatives, FN): 真实为正例却被预测为反例的个数;

(6) 真阳性数 (true positives, TP): 真实为正例预测也为正例的个数;

(7) 真阴性数 (true negatives, TN): 真实为反例预测也为反例的个数;

(8)F_1-score: 2TP/(2TP+FP+FN).

3.1 模拟 1

对于式 (1) 中的线性回归模型, 我们从该模型中随机生成 100 个数据集. 训练样本的大小考虑两种情况$(n,p)=(100,200)$和$(n,p)=(100,500)$, 设计矩阵$\boldsymbol{X}$中的每一行从正态分布$N(\boldsymbol{0},\boldsymbol{\Sigma}),\boldsymbol{\Sigma}=(0.5^{|i-j|})_{1\leq i,j\leq p}$中随机抽样. 真实回归系数$\boldsymbol{\beta}_0=(\underbrace{0.6,\cdots,0.6}_{10},\boldsymbol{0}^{\mathrm{T}})^{\mathrm{T}}$, $\sigma=0.3$为随机误差$\boldsymbol{\varepsilon}$的标准差. 惩罚因子$\rho=1$, λ用交叉验证来选取, 停止条件$\eta=0.03$.

为了生成含有组异质性样本间的邻接矩阵$\boldsymbol{A}$, 我们用 ER 随机图模型生成一个包含$n=100$个节点的样本网络, 样本网络由 4 个不相连的部分G_1,G_2,G_3,G_4组成, 每个部分包含 25 个节点. 每个单独的部分都是一个 ER 随机图, 节点与节点之间以p_b的概率生成边, 即$A_{ij}=1$, 否则为 0, 令$p_b=0.1$. 4 个部分中相连样本的个体节点效应α_i的值分别为 1, −1, 0.5, −0.5, 独立样本的个体节点效应为 0.3.

表 1 展示了两种方法在预测评估指标上的结果对比. 与没有利用相连节点的网络凝聚效应对个体效应进行惩罚的 LASSO、MCP 和 SCAD 结果相比, SNC-LASSO、SNC-MCP 和 SNC-SCAD 都明显改善了估计和预测误差. 这表明将网络凝聚效应加入变量选择模型中, 可以改善模型变量选择、估计和预测的精度.

表 2 展示了两种方法在 100 次模拟实验下变量选择评估指标结果. 我们可以看出各项指标下, SNC 方法的变量选择效果都明显优于没有利用网络凝聚效应进行惩罚的方法. 另外, SNC-MCP 和 SNC-SCAD 都要优于 SNC-LASSO. 尤其对于假阳性数 FP, 100 次模拟中, SNC-LASSO 的 FP 平均为 15.41 (p=200) 和 17.21 (p=500), 而 SNC-MCP 分别为 0.05 (p=200) 和 0.3 (p=500), SNC-SCAD 分别为 1.06 (p=200) 和 0.2 (p=

500), MCP 和 SCAD 变量选择的准确性比 LASSO 显著提高, 主要是由于 LASSO 的有偏性.

3.2 模拟 2

模拟 1 中的结果表明网络凝聚效应惩罚能够改善变量选择、估计和预测效果, 网络凝聚效应主要与邻接矩阵中个体之间产生联系的概率p_b有关, 接下来我们将研究p_b对 SNC 方法的变量选择、估计和预测效果的影响. 模型 2 中的设定与模型 1 类似, 不同的是我们取$p_b = seq(0, 0.02, 0.2)$, R 语言函数 $seq(a, b, c)$ 用于生成一组从 a 到 b, 间隔为 c 的序列.

表 1 不同方法下预测评估指标结果

模型设定	评估指标	无网络凝聚效应下的变量选择方法			网络凝聚效应下的变量选择方法		
		LASSO	MCP	SCAD	SNC-LASSO	SNC-MCP	SNC-SCAD
p=200	PE	0.694 9	0.471 9	0.470 2	0.419 7	0.222 9	0.224 7
	L_1-loss	2.189 9	0.774 0	0.860 6	1.621 4	0.413 8	0.437 8
	L_2-loss	0.519 9	0.246 9	0.244 2	0.443 1	0.158 8	0.162 1
	L_∞-loss	0.250 3	0.134 7	0.129 3	0.221 7	0.095 0	0.096 3
	MSE($\widehat{\alpha}$)	0.398 1	0.397 4	0.397 4	0.168 9	0.156 7	0.157 5
p=500	PE	0.921 9	0.466 4	0.468 5	0.762 8	0.285 5	0.208 1
	L_1-loss	3.135 4	0.639 5	0.741 9	2.467 1	0.569 2	0.388 3
	L_2-loss	0.688 7	0.229 7	0.235 5	0.662 2	0.191 2	0.158 9
	L_∞-loss	0.316 3	0.138 8	0.122 9	0.316 6	0.108 1	0.101 5
	MSE($\widehat{\alpha}$)	0.403 5	0.400 3	0.405 6	0.206 2	0.159 4	0.153 7

表 2 不同方法下变量选择评估指标结果

模型设定	评估指标	无网络凝聚效应下的变量选择方法			网络凝聚效应下的变量选择方法		
		LASSO	MCP	SCAD	SNC-LASSO	SNC-MCP	SNC-SCAD
p=200	FP	25.1	5.63	11.95	15.41	0.05	1.06
	FN	0	0	0	0	0	0
	F_1-score	0.446 7	0.784 1	0.630 9	0.573 9	0.997 6	0.951 8
p=500	FP	35.97	2.78	6.7	0.762 8	0.285 5	0.208 1
	FN	0	0	0	0.03	0	0
	F_1-score	0.359 8	0.882 0	0.753 3	0.561 5	0.944 2	0.990 2

图 1 和图 2 分别展示了p_b对预测和变量选择效果的影响. 从图 1 可以看出, 随着p_b的增大, 即网络的凝聚效应增强, SNC 方法能够明显降低预测损失, 并在$p_b = 0.08$附近趋于稳定. 图 2 表示p_b对F_1分数的影响, F_1分数是查准率和查全率的调和平均数, 当$p_b = 0$即样本之间没有连接关系时, F_1分数值很低. 随着p_b的增大, F_1分数值逐渐增大, 同样地, 在$p_b = 0.08$附近达到最大值, 此时 SNC 方法变量选择的效果较好.

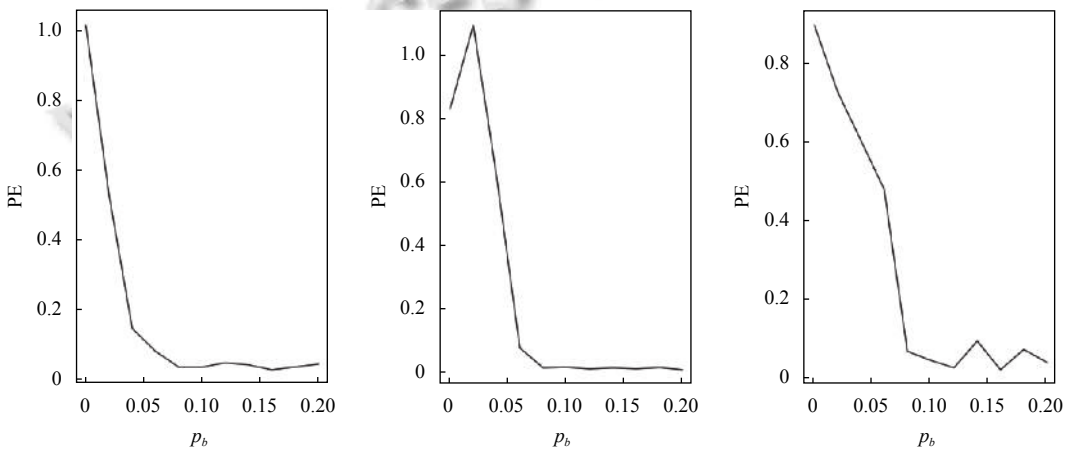

图 1 p_b对预测损失的影响

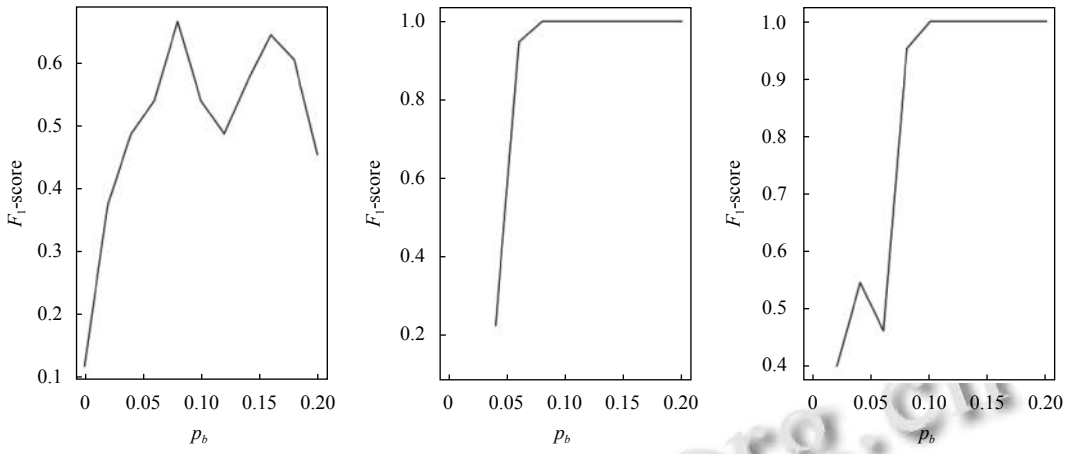

图 2 p_b对 F_1-score 的影响

4 实际数据分析

我们研究的真实数据案例来自于 Teenagers Friends and Lifestyle Study[27]. 这项研究主要是青少年友谊网对他们自身某些行为的影响. 该实际数据与本文中的模型设定保持一致, 因青少年时期学生喜爱团体活动, 故凝聚效应使得网络之间存在组异质性.

Teenagers Friends and Lifestyle Study 旨在确定在青少年早期到中期不良习性的变化过程. 实验记录了 3 个时间点(T_1,T_2,T_3)的数据. 样本来自于 160 名学生, 通过每个学生及学生之间的朋友 (最多 6 个) 关系来建立友谊网络. 研究给出了 3 个时期的友谊网络, 网络中“1”表示“best friend”, “2”表示“just a friend”, “0”表示“no friend”, “10”表示缺失值, 我们根据学生之间的友谊网络来获取邻接矩阵$\boldsymbol{A}$.

本文使用的数据集$\boldsymbol{X}$包含 160 个样本, 40 个特征变量包括青少年的年龄、性别、生活方式、休闲活动以及家庭成员吸烟等情况, 考虑特征之间的交互作用, 最终特征变量为 250 个. 我们的目标是利用友谊网络找出影响青少年不良习性的关键因素, 并预测青少年自身不良行为的活动频率. 我们分别选取 alcohol、tobacco 和 cannabis 作为响应变量$\boldsymbol{Y}$, 对于 tobacco, 元素 1 表示从未抽过烟, 2 表示偶尔吸烟, 3 表示经常吸烟, 故我们将其取对数作为响应变量$\boldsymbol{Y}$的值.

时间点T_1的友谊网络如图 3 所示. 我们只展示了学生之间的“best friendship”(包括“just a friend”和“best friend”). 根据友谊网络建立邻接矩阵$\boldsymbol{A}$时, 当学生i和学生j为“best friend”, 则$A_{ij}=A_{ji}=1$, 否则$A_{ij}=1$.

分别选取 alcohol、tobacco 和 cannabis 作为响应变量来研究影响青少年酗酒、吸烟和吸毒的因素. 将样本随机分成两份: 训练集和测试集, 重复实验 100 次. 由于不知道真实情况下的参数设定, 无法像模拟实验中那样对比假阴性数、假阳性数等指标. 因此, 主要从预测损失和变量选择两个方面来验证 SNC 方法的有效性.

表 3 展示了 SNC 方法 SNC-LASSO、SNC-MCP、SNC-SCAD 与无网络凝聚效应下的变量选择方法 LASSO、MCP 和 SCAD 对青少年不良习性 (酗酒、抽烟以及吸食大麻) 的预测损失, 从结果中可以看出 SNC 方法预测的相对更准确一点. 青少年时期大家都是团体活动, 生活习惯很容易相互影响而慢慢趋同, 而网络凝聚效应正是考虑了这一点, 团体内个体的表现行为更具相似性, 惩罚团体内个体效应的差异性, 提高了个体效应的预测精度, 从而降低了整个模型的预测误差.

为了使挑选出来的变量更具可解释性, 下面我们不考虑特征之间的交互作用, 用 SNC 方法和无网络凝聚效应下的变量选择方法来挑选变量, 并重复实验 100 次, 计算 100 次实验下挑选出来的变量的比例.

表 4 中我们看到, LASSO、MCP 和 SCAD 挑选出更多的冗余变量. 显然, 两种方法下, 特征变量 parent smoking, sibling smoking, “I hang round in the streets”, “I play computer games”和“I go to dance clubs or raves”是最显著的. 青少年时期他们的世界观、人生观和价值观还在形成阶段, 易受他人或团体的影响, 在街上闲逛、经常打电脑游戏、参加俱乐部以及兄弟姐妹抽烟等行为都容易使青少年沾染上不良习性. 通过研究分

析, 我们知道了青少年时期朋友以及家人行为的重要性, 家人、朋友以及整个社会需要给青少年营造一个良好健康的成长环境, 给他们树立积极向上的榜样.

针对各种方法挑选出来变量之后的模型进行回归, 我们得到回归后各变量系数的显著性检验以及调整可决系数R^2和标准误差如表 5 所示.

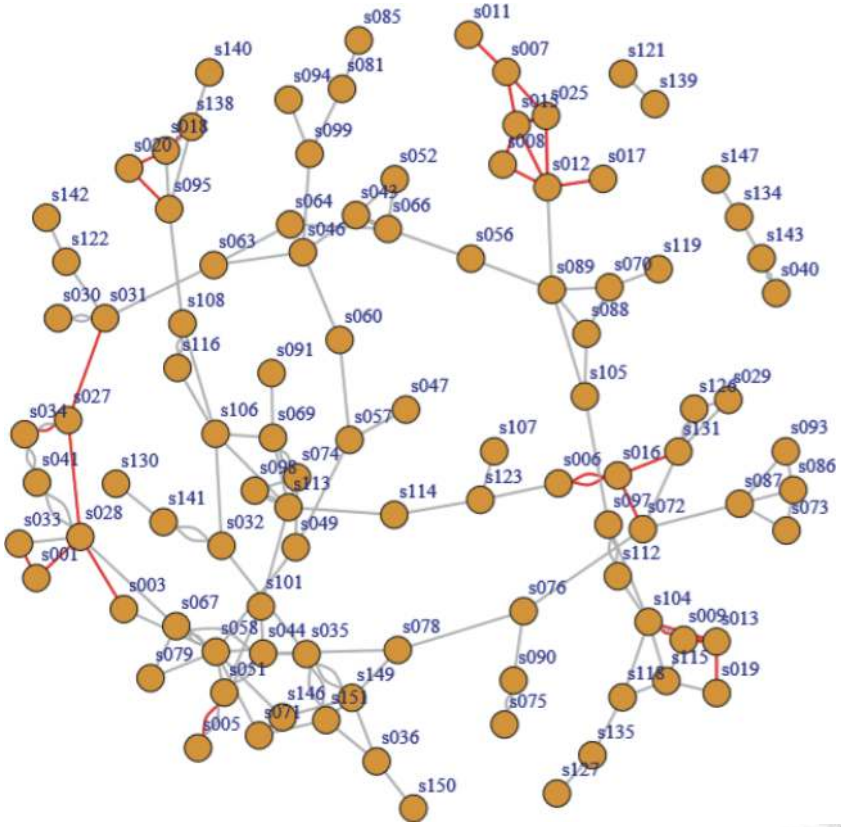

图 3 青少年友谊连接网络

表 3 青少年不良习性的预测损失

时间	响应变量	无网络凝聚效应下的变量选择方法			网络凝聚效应下的变量选择方法		
		LASSO	MCP	SCAD	SNC-LASSO	SNC-MCP	SNC-SCAD
T_1	alcohol	0.732 9	0.670 5	0.621 8	0.548 3	0.445 1	0.425 7
	tobacco	0.748 4	0.694 3	0.620 8	0.603 3	0.535 4	0.307 4
	cannabis	0.503 0	0.428 8	0.460 4	0.408 5	0.309 3	0.336 1
T_2	alcohol	0.497 3	0.424 3	0.454 7	0.432 7	0.356 4	0.358 9
	tobacco	0.610 0	0.540 0	0489 5	0.527 6	0.459 6	0.408 8
	cannabis	0.495 5	0.447 0	0.453 6	0.417 8	0.365 9	0.361 3
T_3	alcohol	0.660 3	0.626 7	0.615 1	0.523 2	0.455 6	0.485 5
	tobacco	0.729 7	0.604 3	0.654 7	0.570 6	0.490 8	0.487 3
	cannabis	0.621 6	0.566 7	0.536 7	0.494 6	0.487 1	0.436 1

由表 5 可知, SNC 方法选取了 sex.F、I hang out in the streets、I play computer games、money、parent.smoking 和 sibling.smoking 6 个变量, 根据值可以看出这些变量都通过了显著性检验. 而 LASSO、MCP 和 SCAD 方法选出了少许的冗余变量. 另外, 从表中的调整可决系数和标准误差来看, SNC 方法的效果也是优于没有网络凝聚效应下的变量选择方法.

5 总结

本文主要对线性回归模型中因网络凝聚效应而产生个体效应的组异质性的网络连接数据进行变量选择, 使用非凸惩罚 MCP 和 SCAD 罚函数同时惩罚变

量系数$\boldsymbol{\beta}$和组内样本的个体效应的差异性$\boldsymbol{L\alpha}$, 使得能够对含有组异质性的网络连接数据筛选出有用变量.

针对本文提出的方法, 我们运用 ADMM 算法进行求解, 并证明了算法的收敛性. 针对 SNC 方法, 本文进行了相关模拟, 从变量选择和预测两个方面来衡量该方法的效果. 从实验结果来看, 无论是预测损失还是变量选择的准确性都有明显改善. 实例分析中, 我们将 SNC 方法应用于青少年友谊网络和生活方式的研究, 分析预测青少年吸烟等不良习性的活动频率以及挑选出影响青少年吸烟等不良习性的特征变量.

表 4 不同方法下挑选出的变量及其比例

变量	无网络凝聚效应下的变量选择方法			网络凝聚效应下的变量选择方法		
	LASSO	MCP	SCAD	SNC-LASSO	SNC-MCP	SNC-SCAD
parent.smoking	1.00	1.00	1.00	1.00	1.00	1.00
sibling.smoking	1.00	1.00	1.00	1.00	1.00	1.00
I hang round in the streets	1.00	1.00	1.00	1.00	1.00	1.00
I play computer games	1.00	1.00	1.00	1.00	1.00	1.00
I go to dance clubs or raves	1.00	1.00	1.00	1.00	1.00	1.00
house	1.00	0.45	0.36	—	—	—
sex.F	1.00	0.94	0.93	0.97	0.92	0.95
Chart	0.98	—	—	—	—	—
Money	1.00	1.00	1.00	1.00	1.00	1.00
I go to something	0.89	0.67	0.70	—	—	—

注: “—”表示变量没有被SNC方法或LASSO、MCP和SCAD方法挑选, 故没有比例值.

表 5 不同方法下挑选出变量的显著性检验

变量	无网络凝聚效应下的变量选择方法			网络凝聚效应下的变量选择方法		
	LASSO	MCP	SCAD	SNC-LASSO	SNC-MCP	SNC-SCAD
sex.F	**	**	**	**	**	**
age	*	0.0678	0.0563	*	—	—
hiphop	0.0766	—	—	—	—	—
I read comics, mags or books	0.0842	—	—	—	—	—
I hang round in the streets	**	**	**	**	**	**
I play computer games	***	***	***	***	***	***
I go to something	*	*	*	*	—	—
I go to dance clubs or raves	0.0896	—	—	—	—	—
money	*	*	*	*	*	*
parent.smoking	**	**	**	**	**	**
sibling.smoking	***	***	***	***	***	***
标准误差	0.3947	0.3269	0.3521	0.305	0.2764	0.2815
R^2	0.6299	0.6378	0.6305	0.6544	0.6620	0.6618

注: *** 表示p值在0~0.001之间, ** 表示p值在0.001~0.01之间, * 表示p值在0.01~0.05之间, “—”表示变量sex.F、age等没有被SNC方法或LASSO、MCP和SCAD方法挑选, 故没有p值.

本文提出的方法, 为含有组异质性网络连接数据的变量选择问题提供了一种解决思路. 我们将变量选择方法进一步拓展了应用领域, 对于基因网络、交通网络、公司网络等网络连接数据, SNC 方法都能适用.

参考文献

1 Barabási AL, Crandall RE. Linked: The new science of networks. American Journal of Physics, 2002, 71(4): 409–410. [doi: 10.1119/1.1538577]

2 Wei YC, Cheng CK. Towards efficient hierarchical designs by ratio cut partitioning. 1989 IEEE International Conference on Computer-Aided Design. Digest of Technical Papers. Santa Clara: IEEE, 1989. 298–301. [doi: 10.1109/ICCAD.1989.76957]

3 Shi JB, Malik JM. Normalized cuts and image segmentation.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888–905. [doi: 10.1109/34.868688]

4 Mariadassou M, Robin S, Vacher C. Uncovering latent structure in valued graphs: A variational approach. The Annals of Applied Statistics, 2010, 4(2): 715–742. [doi: 10.1214/10-AOAS361]

5 Karrer B, Newman MEJ. Stochastic blockmodels and community structure in networks. Physical Review E, 2011, 83(1): 016107. [doi: 10.1103/PhysRevE.83.016107]

6 Zhao YP, Levina E, Zhu J. Consistency of community detection in networks under degree-corrected stochastic block models. Annals of Statistics, 2012, 40(4): 2266–2292. [doi: 10.1214/12-AOS1036]

7 Geng JX, Bhattacharya A, Pati D. Probabilistic community detection with unknown number of communities. Journal of the American Statistical Association, 2019, 114(526): 893–905. [doi: 10.1080/01621459.2018.1458618]

8 Zhang Y, Levina E, Zhu J. Community detection in networks with node features. Electronic Journal of Statistics, 2016, 10(2): 3153–3178. [doi: 10.1214/16-EJS1206]

9 Zhao YP, Wu YJ, Levina E, *et al*. Link prediction for partially observed networks. Journal of Computational and Graphical Statistics, 2017, 26(3): 725–733. [doi: 10.1080/10618600.2017.1286243]

10 郑永广, 岳昆, 尹子都, 等. 大规模社交网络中高效的关键用户选取方法. 计算机应用, 2017, 37(11): 3101–3106. [doi: 10.11772/j.issn.1001-9081.2017.11.3101]

11 张书旋, 康海燕, 闫涵. 基于 Skyline 计算的社交网络关系数据隐私保护. 计算机应用, 2019, 39(5): 1394–1399. [doi: 10.11772/j.issn.1001-9081.2018112556]

12 Krampe J. Time series modeling on dynamic networks. Electronic Journal of Statistics, 2019, 13(2): 4945–4976. [doi: 10.1214/19-EJS1642]

13 Li TX, Levina E, Zhu J. Prediction models for network-linked data. The Annals of Applied Statistics, 2019, 13(1): 132–164. [doi: 10.1214/18-AOAS1205]

14 Lee J, Li G, Wilson JD. Varying-coefficient models for dynamic networks. Computational Statistics & Data Analysis, 2020, 152: 107052. [doi: 10.1016/j.csda.2020.107052]

15 Hoff PD. Additive and multiplicative effects network models. arXiv: 1807.08038, 2018.

16 Cheng JJ, Chen MJ, Zhou MC, *et al*. Overlapping community change-point detection in an evolving network. IEEE Transactions on Big Data, 2018, 6(1): 189–200. [doi: 10.1109/TBDATA.2018.2880780]

17 Zhu XN, Pan R, Li GD, *et al*. Network vector autoregression. The Annals of Statistics, 2017, 45(3): 1096–1123. [doi: 10.1214/16-AOS1476]

18 Tang YM, Bai Y, Huang T. Network vector autoregression with individual effects. Metrika, 2021, 84(6): 875–896. [doi: 10.1007/s00184-020-00805-y]

19 Asur S, Huberman BA. Predicting the future with social media. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Toronto: IEEE, 2010. 492–499. [doi: 10.1109/WI-IAT.2010.63]

20 Fan JQ, Li RZ. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 2001, 96(456): 1348–1360. [doi: 10.1198/016214501753382273]

21 Zhang CH. Nearly unbiased variable selection under minimax concave penalty. The Annals of Statistics, 2010, 38(2): 894–942. [doi: 10.1214/09-AOS729]

22 Zhang T. Analysis of multi-stage convex relaxation for sparse regularization. Journal of Machine Learning Research, 2010, 11: 1081–1107.

23 Zheng ZM, Fan YY, Lv JC. High dimensional thresholded regression and shrinkage effect. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2014, 76(3): 627–649. [doi: 10.1111/rssb.12037]

24 Li CY, Li HZ. Network-constrained regularization and variable selection for analysis of genomic data. Bioinformatics, 2008, 24(9): 1175–1182. [doi: 10.1093/bioinformatics/btn081]

25 Kim S, Pan W, Shen XT. Network-based penalized regression with application to genomic data. Biometrics, 2013, 69(3): 582–593. [doi: 10.1111/biom.12035]

26 Ma SJ, Huang J. A concave pairwise fusion approach to subgroup analysis. Journal of the American Statistical Association, 2017, 112(517): 410–423. [doi: 10.1080/01621459.2016.1148039]

27 Pearson M, Michell L. Smoke rings: Social network analysis of friendship groups, smoking and drug-taking. Drugs: Education, Prevention and Policy, 2000, 7(1): 21–37. [doi: 10.1080/713660095]

28 Teng P. Convergence of a block coordinate descent method for non differentiable minimization. Journal of Optimization Theory and Applications, 2001, 109: 475–494. [doi: 10.1023/A:1017501703105]

附录 A. 命题 1 的证明

命题 1 描述了算法的收敛性, 下面我们开始证明. 由$\boldsymbol{\gamma}^{(m+1)}$的定义可知, 对任意$\boldsymbol{\gamma}$:

$$L(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\gamma}^{(m+1)},\boldsymbol{\varphi}^{(m)}) \le L(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\gamma},\boldsymbol{\varphi}^{(m)})$$

令:

$$f^{(m+1)} = \inf_{\boldsymbol{H\theta}^{(m+1)}-\boldsymbol{\gamma}=0}\left\{\frac{1}{2n}\|\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(m+1)}\|^2 + p_\lambda(|\boldsymbol{\gamma}|)\right\}$$
$$= \inf_{\boldsymbol{H\theta}^{(m+1)}-\boldsymbol{\gamma}=0} L(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\gamma},\boldsymbol{\varphi}^{(m)})$$

故, $L(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\gamma}^{(m+1)},\boldsymbol{\varphi}^{(m)}) \le f^{(m+1)}$.

令t为整数, $\boldsymbol{\varphi}^{(m+t-1)} = \boldsymbol{\varphi}^{(m)} + \rho\sum_{i=1}^{t-1}(\boldsymbol{H\theta}^{(m+i)} - \boldsymbol{\gamma}^{(m+i)})$, 有:

$$L(\boldsymbol{\theta}^{(m+t)},\boldsymbol{\gamma}^{(m+t)},\boldsymbol{\varphi}^{(m+t-1)})$$
$$= \frac{1}{2n}\|\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(m+t)}\|_2^2 + \boldsymbol{\varphi}^{(m+t-1)T}(\boldsymbol{H\theta}^{(m+t)} - \boldsymbol{\gamma}^{(m+t)})$$
$$+ \frac{\rho}{2}\|\boldsymbol{H\theta}^{(m+t)} - \boldsymbol{\gamma}^{(m+t)}\|_2^2 + p_\lambda(|\boldsymbol{\gamma}^{(m+t)}|)$$
$$= \frac{1}{2n}\|\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(m+t)}\|_2^2 + \boldsymbol{\varphi}^{(m)T}(\boldsymbol{H\theta}^{(m+t)} - \boldsymbol{\gamma}^{(m+t)})$$
$$+ \rho\sum_{i=1}^{t-1}(\boldsymbol{H\theta}^{(m+i)} - \boldsymbol{\gamma}^{(m+i)})(\boldsymbol{H\theta}^{(m+t)} - \boldsymbol{\gamma}^{(m+t)})$$
$$+ \frac{\rho}{2}\|\boldsymbol{H\theta}^{(m+t)} - \boldsymbol{\gamma}^{(m+t)}\|_2^2 + p_\lambda(|\boldsymbol{\gamma}^{(m+t)}|) \le f^{(m+t)}$$

由于目标函数$L(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\varphi})$关于$(\boldsymbol{\theta},\boldsymbol{\gamma})$导的, 并且是$\boldsymbol{\varphi}$函数, 基于文献 [28] 的定理 4.1, $(\boldsymbol{\theta}^{(m)},\boldsymbol{\gamma}^{(m)})$有个极值点, 记为$(\boldsymbol{\theta}^*,\boldsymbol{\gamma}^*)$, 故有:

$$f^* = \lim_{m\to\infty} f^{(m+1)} = \lim_{m\to\infty} f^{(m+t)}$$
$$= \inf_{\boldsymbol{H\theta}^*-\boldsymbol{\gamma}=0}\{\frac{1}{2n}\|\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^*\|^2 + p_\lambda(|\boldsymbol{\gamma}|)\}$$

并且对于任意$t \ge 0$, 有:

$$\lim_{m\to\infty} L(\boldsymbol{\theta}^{(m+t)},\boldsymbol{\gamma}^{(m+t)},\boldsymbol{\varphi}^{(m+t-1)})$$
$$= \left\{\frac{1}{2n}\|\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^*\|^2 + p_\lambda(|\boldsymbol{\gamma}|)\right.$$
$$\left. + \lim_{m\to\infty}\boldsymbol{\varphi}^{(m)T}(\boldsymbol{H\theta}^* - \boldsymbol{\gamma}^*) + (t-1/2)\rho\|\boldsymbol{H\theta}^* - \boldsymbol{\gamma}^*\|^{1/2}\right. \le f^*$$

因此, $\lim_{m\to\infty}\|\boldsymbol{r}^{(m)}\|^2 = \boldsymbol{r}^* = \|\boldsymbol{H\theta}^* - \boldsymbol{\gamma}^*\|^2 = 0$.

因为$\boldsymbol{\theta}^{(m+1)}$使得$L(\boldsymbol{\theta},\boldsymbol{\gamma}^{(m)},\boldsymbol{\varphi}^{(m)})$最小化, 故有$\delta L(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\gamma}^{(m)},\boldsymbol{\varphi}^{(m)})/\delta\boldsymbol{\theta} = 0$.并且:

$$\boldsymbol{\delta L}(\boldsymbol{\theta}^{(\boldsymbol{m}+1)},\boldsymbol{\gamma}^{(\boldsymbol{m})},\boldsymbol{\varphi}^{(\boldsymbol{m})})/\boldsymbol{\delta\theta}$$
$$= \frac{1}{\boldsymbol{n}}(\boldsymbol{X},\boldsymbol{I})^T[(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(\boldsymbol{m}+1)} - \boldsymbol{Y}] + \boldsymbol{H}^T\boldsymbol{\varphi}^{(\boldsymbol{m})}$$
$$+ \rho\boldsymbol{H}^T(\boldsymbol{H\theta}^{(\boldsymbol{m}+1)} - \boldsymbol{\gamma}^{(\boldsymbol{m})})$$
$$= \frac{1}{\boldsymbol{n}}(\boldsymbol{X},\boldsymbol{I})^T[(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(\boldsymbol{m}+1)} - \boldsymbol{Y}] + \boldsymbol{H}^T[\boldsymbol{\varphi}^{(\boldsymbol{m})}$$
$$+ \rho(\boldsymbol{H\theta}^{(\boldsymbol{m}+1)} - \boldsymbol{\gamma}^{(\boldsymbol{m})})]$$
$$= \frac{1}{\boldsymbol{n}}(\boldsymbol{X},\boldsymbol{I})^T[(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(\boldsymbol{m}+1)} - \boldsymbol{Y}] + \boldsymbol{H}^T[\boldsymbol{\varphi}^{(\boldsymbol{m}+1)}$$
$$+ \rho(\boldsymbol{\gamma}^{(\boldsymbol{m}+1)} - \boldsymbol{\gamma}^{(\boldsymbol{m})})]$$

因此:

$$\boldsymbol{s}^{(m+1)} = \rho H^T(\boldsymbol{\gamma}^{(m+1)} - \boldsymbol{\gamma}^{(m)})$$
$$= -\frac{1}{n}(\boldsymbol{X},\boldsymbol{I})^T(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(m+1)} + \frac{1}{n}(\boldsymbol{X},\boldsymbol{I})^T\boldsymbol{Y} - \boldsymbol{H\varphi}^{(m+1)}$$

又因为$\|\boldsymbol{H\theta}^* - \boldsymbol{\gamma}^*\|^2 = 0$, 故

$$\lim_{x\to\infty}\partial L(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\gamma}^{(m)},\boldsymbol{\varphi}^{(m)})/\partial\boldsymbol{\theta}$$
$$= \lim_{x\to\infty}\frac{1}{n}(\boldsymbol{X},\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^{(m+1)} - \frac{1}{n}(\boldsymbol{X},\boldsymbol{I})^{\mathrm{T}}\boldsymbol{Y} + \boldsymbol{H\varphi}^{(m+1)}$$
$$= \frac{1}{n}(\boldsymbol{X},\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{X},\boldsymbol{I})\boldsymbol{\theta}^* - \frac{1}{n}(\boldsymbol{X},\boldsymbol{I})^{\mathrm{T}}\boldsymbol{Y} + \boldsymbol{H\varphi}^* = 0$$

因此, $\lim_{x\to\infty}\|\boldsymbol{s}^{(m)}\|^2 = 0$.