第 37 卷第 7 期　　西南大学学报(自然科学版)　　2015 年 7 月
Vol. 37　No. 7　　Journal of Southwest University (Natural Science Edition)　　Jul.　2015

DOI: 10.13718/j.cnki.xdzk.2015.07.014

Elastic Net 方法在 Cox 模型变量选择中的研究①

李春红[1]，　韦新星[2]

1. 广西大学 数学与信息科学学院，南宁 530004；2. 河池学院 数学与统计学院，广西 宜州 546300

摘要： 将 Elastic Net 方法运用于 Cox 模型的变量选择中，建立 EN-Cox 模型，证明了 Cox 模型的 Elastic Net 估计具有组效应性质. 数值模拟验证了 Elastic Net 方法能将强相关变量全部选入 Cox 模型，而 Lasso 方法无此功效. 数值实例验证了 Elastic Net 方法运用于 Cox 模型的可行性，表明 EN-Cox 在拟合效果和预测能力方面均优于传统 Cox 模型.

关　键　词： Elastic Net 方法；Cox 模型；变量选择；强相关；组效应性质

中图分类号：O212.1　　**文献标志码：A**　　**文章编号：** 1673-9868(2015)07-0095-07

Cox 模型[1]是对生存数据进行多因素分析的最常用方法. 它不仅能分析出生存时间与各个协变量的变化规律，而且对基准风险函数的分布没有严格要求. 因此，自 1972 年英国统计学家 Cox 提出该模型后，它便被广泛运用到生物学、医学、保险学、经济学等众多领域.

但是，Cox 模型要求自变量之间相互独立，至少不允许存在很强的相关性，同时还要求所研究的样本量 n 大于预测变量个数 p[2]. 为此，在处理高维度且变量间有强相关性的生存数据时，传统 Cox 模型就不再适用了.

虽然文献[3－5]能通过在系数的绝对值之和上增加一个约束条件来降低高维生存资料的维数，进而改善 Cox 模型的拟合效果，但是当 $p \gg n$ 时，它却最多只能选择 n 个自变量，且缺少解释分组效应信息的能力. 为了克服这些缺陷，文献[6]提出了 Elastic Net 方法. 本文将 Elastic Net 用于 Cox 模型中，建立 EN-Cox 模型，对高维度、强相关变量的变量选择问题[7－9]进行了有益的探索.

1　相关原理介绍

1.1　Cox 模型

常用的 Cox 模型为：

$$h(t \mid \boldsymbol{X}) = h_0(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}) = h_0(t)\exp\left(\sum_{k=1}^{p}\beta_k X_k\right) \tag{1}$$

其中：$\boldsymbol{X} = (X_1, X_2, \cdots, X_p)^{\mathrm{T}}$ 为协变量，$h_0(t)$ 为基准风险函数，$\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^{\mathrm{T}}$ 为回归系数.

Cox 模型中的参数往往采用偏似然函数来估计[10－11].

记观测数据为(Z_i, δ_i, X_i)，$i = 1, 2, \cdots, n$，Z_i 为第 i 个个体的研究时间，δ_i 为示性函数，事件删失时

---

① 收稿日期：2014-05-28
基金项目：国家自然科学基金(11361007)；广西自然科学基金(2014GXNSFCA118001，2012GXNSFBA053010).
作者简介：李春红(1971-)，女，广西南宁人，副教授，主要从事应用统计研究.

$\delta_i=0$，事件发生时 $\delta_i=1$，R_i 为 t_i 时刻个体的风险集，则 Cox 模型的偏似然函数为：

$$L(\boldsymbol{\beta})=\prod_{i=1}^{n}\left\{\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i)}{\sum_{j\in R_i}\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_j)}\right\}^{\delta_i} \tag{2}$$

其中 $\boldsymbol{X}_i=(x_{i1},x_{i2},\cdots,x_{ip})$ 为第 i 个观测样本的 p 个协变量，且(2)式对有结的情形成立. 通常情况下，把 $L(\boldsymbol{\beta})$ 看作一般的似然函数，则通过求其极大值点，可得到 β 的极大似然估计.

在 Cox 模型中，时间 Z_i 处的偏残差[12-14] 定义为 $\hat{\boldsymbol{r}}_i=(\hat{r}_{i1},\hat{r}_{i2},\cdots,\hat{r}_{ip})$，具体表达式为：

$$\begin{aligned}\hat{r}_{ir}&=x_{ir}-E(x_{ir}\mid R_i)=\\&x_{ir}-\frac{\sum\limits_{l\in R_i}x_{lr}\exp(\boldsymbol{\beta}^{T}x_l)}{\sum\limits_{l\in R_i}\exp(\boldsymbol{\beta}^{T}x_l)}\end{aligned} \tag{3}$$

它反映了观察值 x_{ir} 与给定风险集 R_i 时 x_{ir} 的条件期望之间的差异.

1.2 Elastic Net 方法

Elastic Net，即弹性网技术，是由 Lasso 衍生的一种更能有效处理高维小样本数据的方法. 它是在 Lasso 的基础上，通过引入系数的二次惩罚而得到的[6].

假设数据的样本量为 n，预测变量个数为 p，响应变量 $\boldsymbol{y}=(y_1,\cdots,y_n)^{\mathrm{T}}$，预测变量 $\boldsymbol{x}_j=(x_{1j},\cdots,x_{nj})^{\mathrm{T}}$，$j=1,\cdots,p$，$\boldsymbol{X}$ 是由 p 个预测变量组成的矩阵. 事先对响应变量进行中心化处理，对预测变量进行标准化处理，于是，对普通线性模型，Lasso 方法[3] 可定义为：

$$\hat{\beta}(\text{Lasso})=\arg\min\{\|\boldsymbol{y}-\boldsymbol{X\beta}\|^2+\lambda\|\boldsymbol{\beta}\|_1\} \tag{4}$$

其中：$\|\boldsymbol{\beta}\|_1=\sum\limits_{j=1}^{p}|\beta_j|$，$\lambda$ 为非负的调整参数.

对(4)式引入系数的二次惩罚，得到 Elastic Net 方法[6] 的定义为：

$$\hat{\beta}(\text{EN})=\arg\min\{\|\boldsymbol{y}-\boldsymbol{X\beta}\|^2+\lambda_2\|\boldsymbol{\beta}\|^2+\lambda_1\|\boldsymbol{\beta}\|_1\} \tag{5}$$

其中：λ_1 和 λ_2 为非负的调整参数，$\|\boldsymbol{\beta}\|^2=\sum\limits_{j=1}^{p}\beta_j^2$，$\|\boldsymbol{\beta}\|_1=\sum\limits_{j=1}^{p}|\beta_j|$. $\lambda_1=0$ 时得到的是岭估计，$\lambda_2=0$ 时得到的是 Lasso 估计.

Elastic Net 本质上是一个严格的凸优化问题，故有以下引理.

引理 1[6,15] 假设 $x_i=x_j$，$i,j\in\{1,2,\cdots,p\}$，如果优化问题是严格凸的，则对 $\forall\lambda>0$，有 $\hat{\beta}_i=\hat{\beta}_j$ 成立.

引理 2[15] 给定数据($\boldsymbol{X}$，$\boldsymbol{y}$)以及参数(λ_1，λ_2)，$\boldsymbol{y}$ 已经中心化且 $\boldsymbol{X}$ 已经标准化. 令 $\hat{\beta}(\lambda_1,\lambda_2)$ 表示 Elastic Net 估计. 假设 $\hat{\beta}_i(\lambda_1,\lambda_2)\hat{\beta}_j(\lambda_1,\lambda_2)>0$.

定义

$$D_{\lambda_1,\lambda_2}(i,j)=\frac{1}{\|y\|_1}|\hat{\beta}_i(\lambda_1,\lambda_2)-\hat{\beta}_j(\lambda_1,\lambda_2)|$$

则

$$D_{\lambda_1,\lambda_2}(i,j)\leqslant\frac{1}{\lambda_2}\sqrt{2(1-\rho)}$$

其中：$\rho=\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$ 为样本相关系数，且 $\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2=2(1-\rho)$. $D_{\lambda_1,\lambda_2}(i,j)$ 刻划了两个变量系数估计之间的差距. 相关系数越大，它们系数估计的差距就越小. 如果 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 强相关，则相关系数 $\rho\to1$. 由此可知，$\hat{\beta}_i$ 与 $\hat{\beta}_j$ 之间的差异几乎为零. 于是，该引理表明，Elastic Net 具有组效应性质，即强相关变量得到的系数估计大致相同.

2　Elastic Net 方法在 Cox 模型中的应用

2.1　Cox 模型 Elastic Net 估计的定义

传统的Cox模型只适用于 $p < n$，且自变量之间相互独立，至少不允许存在强相关性的情况. 当 $p \gg n$ 时，在 Cox 模型下，第 i 个个体的风险函数为

$$h(t_i) = h_0(t_i)\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i)$$

其中 $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)^{\mathrm{T}}$ 为协变量矩阵. 令 $h_0(t)$ 恒定，则此时 Cox 模型的似然函数为：

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\exp(\sum_{k=1}^{p}\beta_k x_{ik})}{\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}$$

进一步，Cox 模型部分对数似然函数[19] 为：

$$l(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i - \ln[\sum_{j\in R_i}\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_j)]\} =$$
$$\frac{1}{n}\sum_{i=1}^{n}\{\sum_{k=1}^{p}\beta_k x_{ik} - \ln[\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})]\} \tag{6}$$

回归参数 β 的估计就是通过极大化(6) 式求得.

于是，可通过极小化(6) 式的相反数并添加上适当的惩罚项来定义 Cox 模型的 Elastic Net 估计. 结合(5) 式，具体过程如下：

$$\hat{\beta}(\text{EN-Cox}) = \arg\min\left\{\frac{1}{n}\sum_{i=1}^{n}\{-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i + \ln[\sum_{j\in R_i}\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_j)]\} + \lambda_2\|\boldsymbol{\beta}\|^2 + \lambda_1\|\boldsymbol{\beta}\|_1\right\} \tag{7}$$

2.2　Cox 模型 Elastic Net 估计的性质

由引理 1,2 可知，Elastic Net 具有组效应，强相关变量得到的系数估计大致相同. 现研究 Cox 模型 Elastic Net 估计的组效应性质.

定理 1　对Cox模型，给定数据$(Z_i, \delta_i, \boldsymbol{X}_i)$以及参数$(\lambda_1, \lambda_2)$，响应变量已经中心化且自变量已经标准化. 令$\hat{\beta}(\lambda_1, \lambda_2)$ 表示 Elastic Net 估计. 其中，$\hat{\beta}_a(\lambda_1, \lambda_2)$ 和$\hat{\beta}_b(\lambda_1, \lambda_2)$ 是任意一组强相关变量 x_a 和 x_b 的系数. 假设$\hat{\beta}_a(\lambda_1, \lambda_2)\hat{\beta}_b(\lambda_1, \lambda_2) > 0$. 定义

$$D_{\lambda_1,\lambda_2}(a, b) = |\hat{\beta}_a(\lambda_1, \lambda_2) - \hat{\beta}_b(\lambda_1, \lambda_2)|$$

则

$$D_{\lambda_1,\lambda_2}(a, b) \to 0$$

证明：

由于$\hat{\beta}_a(\lambda_1, \lambda_2)\hat{\beta}_b(\lambda_1, \lambda_2) > 0$，故符号函数 $\operatorname{sgn}\{\hat{\beta}_a(\lambda_1, \lambda_2)\} = \operatorname{sgn}\{\hat{\beta}_b(\lambda_1, \lambda_2)\}$，且$\hat{\beta}_a(\lambda_1, \lambda_2) \neq 0$，$\hat{\beta}_b(\lambda_1, \lambda_2) \neq 0$.

现令$\hat{\beta}_m(\lambda_1, \lambda_2) \neq 0$，$\hat{\beta}(\lambda_1, \lambda_2)$ 满足

$$\left.\frac{\partial L(\lambda_1, \lambda_2, \beta)}{\partial \beta_m}\right|_{\beta = \hat{\beta}(\lambda_1, \lambda_2)} = 0$$

其中

$$L(\lambda_1, \lambda_2, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i + \ln[\sum_{j\in R_i}\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_j)]\} + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|^2 =$$
$$\frac{1}{n}\sum_{i=1}^{n}\{-\sum_{k=1}^{p}\beta_k x_{ik} + \ln[\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})]\} + \lambda_1\sum_{k=1}^{p}|\beta_k| + \lambda_2\sum_{k=1}^{p}\beta_k^2$$

则

$$\frac{1}{n}\left(-\sum_{i=1}^{n}x_{ia}+\sum_{i=1}^{n}\frac{\sum_{j\in R_i}x_{ja}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}{\sum_{j\in R_i}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}\right)+\lambda_1\operatorname{sgn}\{\hat{\beta}_a(\lambda_1,\lambda_2)\}+2\lambda_2\hat{\beta}_a(\lambda_1,\lambda_2)=0$$

于是

$$\hat{\beta}_a(\lambda_1,\lambda_2)=\frac{1}{2\lambda_2}\left\{\frac{1}{n}\sum_{i=1}^{n}x_{ia}-\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\in R_i}x_{ja}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}{\sum_{j\in R_i}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}-\lambda_1\operatorname{sgn}\{\hat{\beta}_a(\lambda_1,\lambda_2)\}\right\} \tag{8}$$

同理

$$\hat{\beta}_b(\lambda_1,\lambda_2)=\frac{1}{2\lambda_2}\left\{\frac{1}{n}\sum_{i=1}^{n}x_{ib}-\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\in R_i}x_{jb}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}{\sum_{j\in R_i}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}-\lambda_1\operatorname{sgn}\{\hat{\beta}_b(\lambda_1,\lambda_2)\}\right\} \tag{9}$$

由(8),(9)式得到

$$\hat{\beta}_a(\lambda_1,\lambda_2)-\hat{\beta}_b(\lambda_1,\lambda_2)=\frac{1}{2\lambda_2 n}\sum_{i=1}^{n}\left[x_{ia}-\frac{\sum_{j\in R_i}x_{ja}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}{\sum_{j\in R_i}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}-x_{ib}+\frac{\sum_{j\in R_i}x_{jb}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}{\sum_{j\in R_i}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}\right] \tag{10}$$

由(3),(10)式可变形为

$$\begin{aligned}\hat{\beta}_a(\lambda_1,\lambda_2)-\hat{\beta}_b(\lambda_1,\lambda_2)&=\frac{1}{2\lambda_2 n}\sum_{i=1}^{n}[x_{ia}-(x_{ia}-\hat{r}_{ia})-x_{ib}+(x_{ib}-\hat{r}_{ib})]=\\&\frac{1}{2\lambda_2 n}\sum_{i=1}^{n}(\hat{r}_{ia}-\hat{r}_{ib})=\frac{1}{2\lambda_2 n}\left(\sum_{i=1}^{n}\hat{r}_{ia}-\sum_{i=1}^{n}\hat{r}_{ib}\right)=\\&\frac{1}{2\lambda_2}[E(\hat{r}_{ia})-E(\hat{r}_{ib})]\end{aligned} \tag{11}$$

由于 x_a 和 x_b 强相关，即相关系数 $\rho\to 1$，故

$$|x_{ia}-x_{ib}|\to 0 \qquad |E(x_{ia})-E(x_{ib})|\to 0$$

从而

$$\begin{aligned}&|[x_{ia}-E(x_{ia}\mid R_i)]-[x_{ib}-E(x_{ib}\mid R_i)]|\leqslant\\&|x_{ia}-x_{ib}|+|E(x_{ia}\mid R_i)-E(x_{ib}\mid R_i)|\to 0\end{aligned}$$

由(3)式有

$$|\hat{r}_{ia}-\hat{r}_{ib}|\to 0 \qquad |E(\hat{r}_{ia})-E(\hat{r}_{ib})|\to 0$$

故在(11)式中

$$\begin{aligned}&|\hat{\beta}_a(\lambda_1,\lambda_2)-\hat{\beta}_b(\lambda_1,\lambda_2)|=\\&\frac{1}{2\lambda_2}|[E(\hat{r}_{ia})-E(\hat{r}_{ib})]|\to 0\end{aligned}$$

即 $D_{\lambda_1,\lambda_2}(a,b)\to 0$. 这表明Cox模型的Elastic Net估计具有组效应性质，相关系数越大，它们系数估计的差距就越小.

2.3 数值模拟

定理 1 从理论上揭示了 Cox 模型的 Elastic Net 估计具有组效应性质. 对于具有强相关性的变量，EN-Cox 能将强相关变量全部选入模型，而并非只选择其中的一个，且相关系数越大，它们系数估计的差距就越小. 现通过数值模拟来加以验证.

设 $x_i\sim N(0,1)$，$i=1,2,\cdots,10$，其中 $x_3=x_2$，$x_{10}=x_9$，$x_4=x_1+\frac{1}{2}x_2+\frac{1}{2}x_3$. 考虑 Cox 模

型 $h(t)=h_0(t)\exp(\sum_{i=1}^{10}\beta_i x_i)$，$t\sim U[0,1]$，且真实参数为$(-1,1,1,0,\frac{2}{3},0,0,0,2,2)^{\mathrm{T}}$. 由假设可知，在真实模型中有两组强相关变量，分别为 x_3 与 x_2，x_{10} 与 x_9，且 x_4,x_1,x_2 与 x_3 之间存在共线性.

将上述模型模拟 1 000 次，得到 $n=1\ 000$，$p=10$ 的样本. 为了验证 Elastic Net 方法能将强相关变量全部选入 Cox 模型而 Lasso 方法无此功效，分别运用 Lasso 方法和 Elastic Net 方法进行变量筛选. 由于 Elastic Net 方法的解经转换后可表达为 Lasso 方法的解的形式，故可利用最小角回归算法解决其算法问题. 结合 Lars 算法[16]，在 R 软件[17-18]中进行变量筛选，得到的系数估计见表 1.

表 1　数值模拟得到的系数估计值

方法	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Lasso	−0.996 1	1.996 4	0	0	0.668 2	0.001 8	−0.003 2	0.001 1	3.998 9	0
EN	−0.995 9	0.995 8	0.995 8	0	0.665 3	0.001 6	−0.002 2	0.000 6	1.998 4	1.998 4

分析表 1 可得如下比较结果：

1）对于与其它变量存在共线性的 x_4，Lasso 方法和 Elastic Net 方法均没有将其选入模型. 这表明，这两种方法在处理一般共线性问题上均是有效的.

2）对于强相关变量 x_3 与 x_2，Lasso 方法只筛选出了 x_2，而 Elastic Net 方法则把二者同时选入模型；同样，对于强相关变量 x_{10}与 x_9，Lasso 方法也只选出了 x_9，而 Elastic Net 方法则把 x_{10}与 x_9 同时选入模型. 这表明，对于存在强相关性的数据，运用 Elastic Net 方法筛选出的变量更接近于真实模型.

2.4　实例分析

本实例来自于对某高校在校大学生手机卡使用情况的问卷调查. 调查采用问卷填写方式，共收集了从 2007 年 1 月至 2014 年 1 月 7 年内的 389 份问卷，其中 9 份为无效问卷，将其剔除后最终得到 380 份有效问卷. 随机抽取其中的 300 份作为训练集，剩余的 80 份作为检验集. 现把手机卡在研究中的持续使用时间记为 t. 若在研究结束前手机卡停止使用，则将退出原因标注为流失；若在研究结束前手机卡没有停止使用或无法判断其是否停止使用，则标注为删失. 部分数据见表 2.

表 2　部分数据样本

编号	t/a	退出原因	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	6	删失	2	6	1	1	2	1	1	3	2	2
2	1	流失	2	3	1	1	1	1	1	2	2	1
3	1	流失	1	3	2	2	1	1	2	2	2	1
4	1	流失	1	6	1	1	1	1	2	2	1	1
5	4	删失	2	4	2	2	1	1	1	5	2	4
…	…	…	…	…	…	…	…	…	…	…	…	…
380	2	流失	2	2	2	2	2	1	1	2	3	1

表 2 中：$x_1=1,2$ 分别表示男性和女性；$x_2=1,2,3,4,5,6,7$ 分别对应大一至研三 7 个年级；$x_3=1,2$ 表示是否学生干部；$x_4=1,2$ 表示是否少数民族；$x_5=1,2$ 表示是否农业户口；$x_6=1,2$ 表示是否生源地就读；$x_7=1,2$ 表示是否移动；$x_8=1,2,3,4,5$ 分别表示月均电话费用为 0～20 元、20～40 元、40～60 元、60～80 元及 80 元以上；$x_9=1,2,3$ 分别表示售后服务质量为好、中、差；$x_{10}=1,2,3,4$ 分别表示月均生活费用为 0～400 元、400～800 元、800～1 200 元及 1 200 元以上.

经分析，上述生存资料的变量并非相互独立，而是存在一定的相关性，甚至某些变量间还存在着很强的相关性，如 x_2 与 x_8 的相关系数为 0.111，而 x_8 与 x_{10}的相关系数高达 0.837，不符合传统 Cox 模型的条件.

一方面，运用 Cox 模型向后逐步法，对整理好的数据进行变量筛选. 筛选准则为 $\alpha=0.05$，剔除准则为 $\alpha=0.10$. 另一方面，运用 EN-Cox，对整理好的数据进行变量筛选. 考虑到 Elastic Net 方法的解经转换后可表示为 Lasso 方法的解的形式，故可利用最小角回归算法解决其算法问题. 结合 Lars 算法[16]，在 R 软件[17-18]中进行变量筛选.

两种方法的筛选结果见表 3.

表 3 逐步法及 EN-Cox 得到的系数估计值

方法	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
逐步法	0	−0.194	0	0	0	−0.310	0	0.190	0.734	0
EN-Cox	0	−0.112	0.004	−0.023	0.124	−0.284	0	0.019	0.313	0.026

表 3 显示，由逐步法筛选出的变量有 4 个，而由 EN-Cox 筛选出的变量有 8 个.

进一步分析表 3，可得到如下比较结果：

1）两种方法均没有将 x_1 和 x_7 选入模型，表明手机卡的使用情况与被调查者的性别及所属的通讯运营商无关；此外，x_2 和 x_6 均为负值，表明那些属于高年级、不在生源地就读的学生，其流失概率往往较小，而 x_8 和 x_9 均为正值，表明学生的月均电话费用越高或运营商的售后服务质量越差，客户的流失概率越大. 这些均与实际情况相符，说明将 Elastic Net 方法运用于 Cox 模型是可行的.

2）对于具有较强相关性的变量 x_8 与 x_{10}，逐步法只选择了 x_8，而 EN-Cox 能将两者同时选入模型，这表明 EN-Cox 能将强相关变量全部选入模型；同时注意到，在所有系数的估计值中，二者系数的差距最小，这表明 EN-Cox 能体现变量间的相关性，且相关系数越大，它们系数估计的差距就越小，即体现了 EN-Cox 的组效应性质.

3）在拟合效果方面，根据 Nagelkerke[19-20] 给出的在删失生存数据条件下的模型评价标准 $R^2=1-\exp\left\{-\dfrac{2}{n}\left[l(\hat{\beta})-l(0)\right]\right\}$，可以得到逐步法的 R^2 为 0.193，而 EN-Cox 的 R^2 为 0.364，R^2(EN-Cox)$>R^2$(逐步法)，表明 EN-Cox 的模型拟合效果优于逐步法.

4）在预测能力方面，利用检验集中的数据，对逐步法和 EN-Cox 进行预测能力比较. 结果见表 4.

表 4 预测能力比较

方 法	预测错误率	方 法	预测错误率
逐步法	0.412 5	EN-Cox	0.212 5

表 4 显示 EN-Cox 的预测出错率较低，表明 EN-Cox 的预测能力优于逐步法.

3 结 论

Cox 模型是一种处理生存数据的经典方法，但美中不足的是，其往往只适用于处理低维度、不相关或不存在强相关的生存数据，故在处理高维度、变量间有强相关性的生存数据时，传统 Cox 模型就不再适用了.

为克服这一缺陷，本文鉴于 Elastic Net 能有效处理高维小样本、强相关数据的事实，将 Elastic Net 方法运用于 Cox 模型的变量选择中，证明 Cox 模型的 Elastic Net 估计具有组效应性质，并通过数值模拟验证了 Elastic Net 方法能将强相关变量全部选入 Cox 模型而 Lasso 方法无此功效. 此外，结合某高校在校大学生手机卡的使用实例，通过比较逐步法和 EN-Cox 的输出结果，进一步肯定了 Elastic Net 方法运用于 Cox 模型的可行性，再次验证了 EN-Cox 模型能将强相关变量全部选入模型，且在预测能力和拟合效果方面均优于传统 Cox 模型. 值得一提的是，调查问卷只代表了调查时间段内被调查者的消费习惯和消费结构，故在进行预测时，所得结果是具有特定性和时效性的，但当样本增大时，所得预测结果还是有一定的合理性的.

综上，Elastic Net 方法能有效克服传统 Cox 模型的不足，使 Cox 模型的拟合效果和预测能力得到改善.

参考文献：

[1] COX D R. Regression Models and Life Tables [J]. Journal of Royal Statistical Society，1972，34：187−220.

[2] 闫丽娜，覃 婷，王 彤. Lasso 方法在 Cox 模型中的应用 [J]. 中国卫生统计，2012，29(1)：58−64.

[3] TIBSHITANI R. Regression Shrinkage and Selection via the Lasso [J]. Journal of the Royal Statistical Society，Series B，1996，58：267−288.

[4] MEINSHAUSEN N. Relaxed Lasso [J]. Computational Statistics and Analysis，2007，52(1)：374−393.

[5] 曹 芳，朱永忠. 基于多重共线性的 Lasso 方法 [J]. 江南大学学报：自然科学版，2012，11(1)：87−90.

[6] ZOU H，HASTIE T. Regularization and Variable Selection via the Elastic Net [J]. Journal of the Royal Statistical Society，Series B，2005，67(1)：301－320.
[7] 毕伯竹. 高维多重共线性数据的变量选择问题 [D]. 济南：山东大学，2011.
[8] TIBSHITANI R. The Lasso Method for Variable Selection in the Cox Model [J]. Statistics in Medicine，1997，16(4)：385－395.
[9] YUAN M，LIN Y. Model Selection and Estimation in Regression with Grouped Variables [J]. Journal of the Royal Statistical Society，Series B，2006，68(1)：49－67.
[10] 王启华. 生存数据统计分析 [M]. 北京：科学出版社，2006：232－237.
[11] 田　俊. Cox 回归模型及其参数估计 [J]. 福建医学院学报，1987，21(2)：115－117.
[12] 郜艳晖，何大卫. Cox 模型的残差分析和影响诊断 [J]. 现代预防医学，2000，27(1)：48－50.
[13] SCHOENFELD D. Partial Residuals for the Proportional Hazards Regression Model [J]. Biometrika，1982，69(1)：239－241.
[14] KENNETH R H. Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression [J]. Statistics in Medicine，1995，14：1707－1723.
[15] 卢　颖. 广义线性模型基于 Elastic Net 的变量选择方法研究 [D]. 北京：北京交通大学，2011.
[16] Efron B，Hastie T，Johnstone I，et al. Least Angle Regression [J]. The Annals of statistics，2004，32(2)：407－499.
[17] 吴喜之. 复杂数据统计方法——基于 R 的应用 [M]. 北京：中国人民大学出版社，2012.
[18] 王斌会. 多元统计分析及 R 语言建模 [M]. 广州：暨南大学出版社，2010.
[19] 闫丽娜. 惩罚 Cox 模型和弹性网技术在高维数据生存分析中的应用 [D]. 太原：山西医科大学，2011.
[20] NAGELKERKE N. A Note on a General Definition of the Coefficient of Determination [J]. Biometrika，1991，78(3)：691－692.

A Study of the Elastic Net Method in Variable Selection for the Cox Model

LI Chun-hong[1]，　WEI Xin-xing[2]

1. College of Mathematics and Information Science，Guangxi University，Nanning 530004，China；
2. College of Mathematics and Statistics，Hechi University，Yizhou Guangxi 546300，China

Abstract：In this study，the elastic net method was used for variable selection of the Cox model，and an EN-Cox model was established. The estimations of the Cox model obtained proved to have the group effect property. A numerical simulation showed that the elastic net method could select all the strong correlation variables into the Cox model while the Lasso method could not. An empirical analysis further confirmed the feasibility of applying the elastic net method to the Cox model and verified that the EN-Cox model was superior to the traditional Cox model in fitting effect and forecasting ability.

Key words：elastic net method；Cox model；variable selection；strong correlation；group effect property

责任编辑　张　构