北京大学

硕士研究生学位论文

题目：通过术前指标对经皮冠状动脉介入治疗术后无复流现象的预测

姓　　名：	吕渊
学　　号：	1101210028
院　　系：	数学科学学院
专　　业：	概率论与数理统计
研究方向：	机器学习
导师姓名：	姚远 教授

二零一五年四月

版权声明

摘　要

本文研究通过病人的各项检测指标预测经皮冠状动脉介入治疗(PCI)术后是否出现无复流现象(no-reflow phenomenon)的问题。首先，由于数据集中有大量缺失，所以我们对其进行预处理，删除完全缺失的自变量，对其他自变量进行填补。其次，根据相关医学研究和统计学方法，筛选出显著的变量。最后，建立模型对无复流现象是否发生进行预测，并将变量范围缩小到术前变量，用不同算法的交叉检验提高模型预测准确度。

关键词：PCI，无复流现象，缺失数据填补，交叉检验

目 录

表格

插图

第一章 绪 论

1.1 背景介绍

经皮冠状动脉介入治疗(percutaneous coronary intervention, PCI)，是指经心导管技术疏通狭窄甚至闭塞的冠状动脉管腔，从而改善心肌的血流灌注的治疗方法。无复流现象(no-reflow phenomenon)是指冠状动脉闭塞，血流中断后重新恢复血流，却无心肌组织的有效灌注的现象。无复流现象可发生于经皮冠状动脉介入治疗(PCI)术后，并且是造成不良预后的重要因素。因此，通过病人的各项数据预测是否发生无复流是一项有价值的工作。

检测无复流现象对冠心病患者进行介入治疗的预后有重要价值。无复流者易发生心包积液及早期左室重塑和充血性心力衰竭，且充血性心力衰竭时间长；无复流者左室舒张末期容积在愈合过程中进行性增大，而有复流者下降；但对心律失常及冠脉时间的发生无明显影响。而PCI术后是否发生无复流可根据临床特点、冠状动脉造影及冠状动脉内超声结果进行初步判断。但是，预测手术成败效果，是决定是否手术的一个依据，所以实际上不能用全部指标，而是用术前指标比较合理。因此，仅用术前指标进行预测，并提高预测准确性，能为医生的判断提供更大的帮助。

1.2 本文的目标和思路框架

在这次研究中，我们获得了北京安贞医院和301医院提供的数据，包含1214条病例的记录，每条记录包含37个术前指标、36个术后指标和一个二值的因变量无复流，数据集中含有大量数据缺失。

1、缺失数据的处理

由于数据集中有大量缺失，所以需要对其进行预处理，删除完全缺失的自变量，对其他自变量进行填补。

2、变量选择和模型建立

根据相关医学研究和统计学方法，筛选出显著的变量。建立模型对无复流现象是否发生进行预测。

3、预测结果

通过交叉检验预测结果，分为用全部指标预测和仅通过术前指标预测两部分。

第二章　数据处理

2.1 数据介绍

安贞医院、朝阳医院、301医院三家医院整合的数据共包含2581 个病人的73 项指标。无复流是二值变量，目标是根据病人的术前术后指标对是否出现无复流现象进行分类。

73项指标分为两类：

一、入院即刻或PCI术前采集的指标（统称术前指标）：

性别，年龄，身高，体重，吸烟史，糖尿病史，高血压史，PCI 史，脑梗塞史，既往调脂药，既往阿司匹林，既往ADP 拮抗剂，既往ACEI，既往利尿剂，既往β受体阻滞剂，既往CA 拮抗剂，梗死前心绞痛，收缩压，舒张压，心率，入院诊断，killip 分级，梗死部位，中性粒细胞，血红蛋白，白蛋白，肌酐，总胆固醇，甘油三酯，LDLC，HDLC，随机血糖，apoa1，apob，LPa，高敏C 反应蛋白，BNP，TNI，PCI前CK，PCI前CKMB，内皮素，PCI前阿司匹林，PCI前低分子肝素，PCI前ADP拮抗剂，PCI 前2b3a拮抗剂，PCI前ACEI，PCI前ARB，PCI前硝酸酯，PCI 前β阻滞剂，PCI前钙拮抗剂，PCI前溶栓，PCI 前他汀，症状到PCI 时间等54项；

二、PCI术中采集的指标：

病变支数，罪犯血管，病变位置，罪犯血管狭窄程度，术前TIMI血流，罪犯血管血栓数量，侧支循环分级，PCI 术中2b3a，PCI术中硝酸酯，PCI术中钙拮抗剂，术中腺苷，PCI 术中血栓抽吸，预扩张，后扩张，支架直径1，支架长度1，最大扩张压力，支架数量，球囊扩张次数等19 项。

这些指标的取值范围有些是连续取值，如心率、血压等，有些是分段取值，如性别、梗死部位等。

值得注意的是，数据集中包含着相当数量的缺失数据，各指标的缺失比例如下表所示：

缺失比例	0-10%	10%-20%	20%-30%	30%-40%	40%-100%
指标数量	38	6	4	14	12

表 2.1　数据缺失比例

2.2 初步处理

2.2.1 缺失响应值的数据

二值变量无复流是我们需要进行预测的目标，因此在预处理中首先剔除7个无观测值的记录。

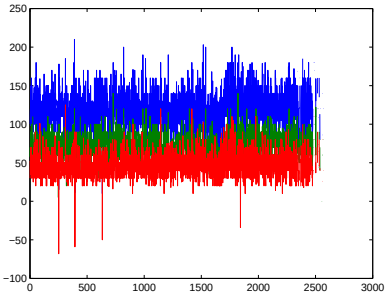
2.2.2 明显有错误的数据

将每个指标的数据画出散点图，找出异常点，进行处理。下面举三个例子说明对不同类型的异常指标的处理方法。

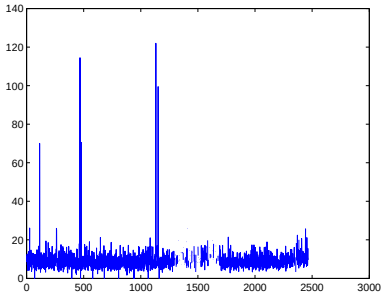
指标收缩压、舒张压，散点图如下图a所示，红线表示收缩压减去舒张压的值，可见有些病人的收缩压低于舒张压，这是明显错误的，这类观测我们将其剔除。

指标中性粒细胞，散点图如下图b所示，个别观测超出正常范围很多，将异常点记为缺失值，保留观测。

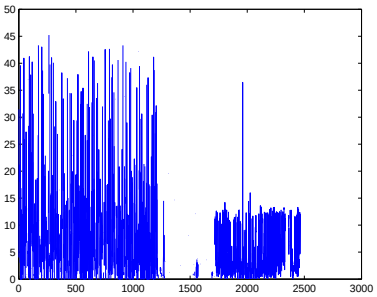
指标高敏C反应蛋白，散点图如下图c所示，不同医院的观测明显有差异，但尚不明确差异原因，故而暂时不做处理。



(a) 收缩压和舒张压



(b) 中性粒细胞



(c) 高敏c反应蛋白

图 2.1　例：明显有错误的数据

2.2.3 其他预处理

对取值范围很大的变量取对数$log(x + 0.1))$，如TNI、PCI前CK、PCI前CKMB等。

对取值大部分为0，其他取值离散程度很大的变量取为二值变量，如LPa、BNP等。

2.3 处理缺失数据

2.3.1 使用完全数据

删除所有有缺失值的观测，只用全部变量都观测到的个体分析。完全观测的病人共450位，约为所有病人的1/6。

2.3.2 随机填补

对于有缺失的变量，从其观测到的个体中等概率抽取，填补缺失的观测。

2.3.3 KNN填补

对数据矩阵标准化，按照欧式距离选出k个邻居。对于连续变量，用邻居的均值填补缺失观测；对于离散变量，从邻居中等概率抽取，填补缺失的观测。

传统的k近邻（KNN）方法，要求所有数据中，至少有一个变量是完全无缺失的。但是这在我们的数据中无法得到保证。有一种处理方法是先用均值或随机填补所有的缺失数据，再在其上用KNN填补，但是这种方法会减小对于有相同变量缺失的观测之间的距离。

在此，我们采用另一种方式从某种程度上解决这个问题。

定义

$$D(x,y)=\sqrt{(\sum_{i=1}^{p}d(x_i-y_i))}$$

其中，$d(x,y)$定义为：

若x和y都没有缺失，则$d(x,y)=(x-y)^2$

若x和y都缺失，则$d(x,y)=E(X-Y)^2$

若只有y缺失，则$d(x,y)=E(x-Y)^2$。

这个方法的问题在于，数据中的某些变量是离散取值的，这样填补无法区分离散变量和连续变量，但由于数据中的离散变量都是二值变量或是有序的离散取值，故而这样定义的距离是有意义的。

另外，为了保证距离的可加性，首先依然要对变量进行标准化处理。

KNN填补的具体过程如下：

Algorithm 1 KNN填补

for all $i \in$ 观测 **do**
  for all $j \in$ 变量 **do**
    if x_{ij} 缺失 **then**
      找到最近的$x_{i1}, x_{i2}, \cdots, x_{ik}$
      $x_{ij} = \frac{\sum_{l=1}^{k} x_{il}}{k}$
    end if
  end for
end for

填补后即可得到无缺失值的数据集。

第三章　变量选择和模型建立

3.1 变量选择

3.1.1 逻辑斯蒂回归模型

模型1：（Logistic回归模型）

$$\log(\frac{P[Y=1|X]}{1-P[Y=1|X]}) = \beta_0 + X\beta_1$$

对每个指标分别做逻辑斯蒂回归，选出相对显著的20个变量，选择标准是赤池信息量AIC 尽量小。

$$AIC = 2k - 2\ln(L)$$

其中k为模型的独立参数个数，L为模型的极大似然函数。

当误差为独立同分布的正态分布时，AIC有如下表达：

$$AIC = 2k + n\ln(\frac{RSS}{n})$$

其中RSS为残差平方和。

3.1.2 Lasso方法

模型2：（Lasso方法）

$$\min \|Y - X\beta\|_2^2,$$

s.t.

$$\|\beta\|_1 \leq t, \text{ for some } t > 0$$

它是如下组合优化的convex relaxation：

$$\min \|Y - X\beta\|_2^2,$$

s.t.

$$\|\beta\|_0 \leq s, \text{ for some } s > 0$$

Lasso还可以表示为：

$$\min \|Y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

Lasso方法也可用作变量选择。

3.2 模型建立

3.2.1 支持向量机（SVM）

SVM是解决如下优化问题的方法：

$$\min_{\omega,b,\xi} \frac{1}{2}\langle\omega,\omega\rangle + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to } y_i(\langle\omega,\phi(x_i)\rangle + b) \geq 1 - \xi_i, \forall i = 1,2,\ldots,n$$

分类函数为：

$$F(x) = sgn(\langle\omega,\phi(x)\rangle + b)$$

这个优化问题满足KKT条件，故其对偶问题的最优解也是原问题的最优解，且有：

$$\omega^* = \sum_{i=1}^{n}\omega_i\phi(x_i)$$

若存在

$$K(x,y) = \langle\phi(x),\phi(y)\rangle,$$

则其对偶问题为：

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$$

$$\text{subject to } \begin{cases} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases}$$

分类函数为：

$$F(x) = \sum_{i=1}^{n} \alpha_i^* y_i K(x, x_i) + b^*$$

常见的核函数有：

Linear Kernel: $K(x, y) = x^T y$

Radial Basis Kernel: $K(x, y) = \exp(-\gamma \|x - y\|^2)$

Polynomial Kernel: $K(x, y) = (\gamma x^T y + constant)^d$

Sigmoid Kernel: $K(x, y) = \tanh(\gamma x^T y + constant)$

SVM主要针对线性可分情况进行分析，对于线性不可分的情况，SVM通过非线性映射将低维空间线性不可分的样本映射到高维特征空间，从而使其线性可分。SVM的核心部分核函数的价值在于，虽然它是将特征进行从低维到高维的转化，但仅在低维上进行计算，而将实质上的分类效果表现在了高维上，避免了直接在高维空间中的复杂计算。

Logistic回归和SVM都是常见的分类算法。两者的区别在于，从目标函数来看，前者采用logistical loss，而后者采用hinge loss。这两个损失函数的目的都是增加对分类影响较大的数据点的权重，减少与分类关系较小的数据点的权重。其做法是，Logistic回归通过非线性映射，减小离分类平面较远的点的权重，而SVM的处理方法是只考虑支持向量，也就是和分类最相关的少数点。另外，SVM在转化为对偶问题后，只需要计算与少数几个支持向量的距离，在进行复杂核函数计算时有很大优势，能够大大简化模型和计算量。

3.2.2 随机森林（Random Forest）

随机森林的基本思想如下所述：

首先，用Bootstrap抽样从原始训练集（含N个观测）中抽取k个样本，每个样本都含有N个观测；

其次，对每个样本分别建立一个决策树模型，总共得到k种分类结果；

最后，根据k种分类结果对每条观测进行计分，决定其最终分类。

随机森林的分类准确率高，但解释性较差，而且很容易出现过拟合的现象。因此，在筛选变量时需要更加小心。

第四章　预测结果

4.1 Lasso结合SVM的预测结果

4.1.1 Logistic回归选择变量

在原数据集上，删除所有含缺失值的数据，得到450条完整数据。在其上判断变量的显著性，选择出20个指标如下。

显著程度	指标名称
1	PCI术中钙拮抗剂
2	PCI术中2b3a
3	罪犯血管血栓数量
4	IABP
5	PCI前CKMB
6	随机血糖
7	PCI前CK
8	侧枝循环分级
9	PCI术中血栓抽吸
10	killip分级
11	LDLC
12	舒张压
13	TNI
14	脑梗塞史
15	PCI前硝酸酯
16	PCI前溶栓
17	高血压史
18	症状到PCI时间
19	球囊扩张次数
20	预扩张

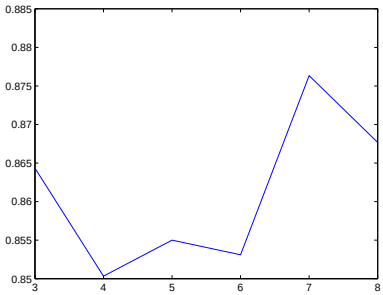
表 4.1　Logistic回归选择出前20位指标

4.1.2 用全部指标进行预测

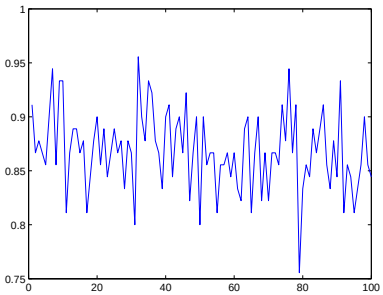
4.1.2.1 用完整数据进行预测

删除所有含缺失值的观测，直接用Lasso方法选择变量，用SVM进行分类。

通过5-fold交叉检验循环100次获得模型判断准确率。



(a) Lasso选择的变量个数对判断准确率的影响



(b) Lasso选择7个变量时的判断准确率

图 4.1　用完整观测进行预测

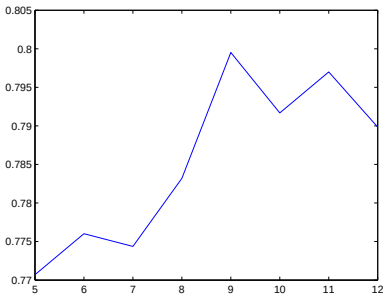
从图中可以看出，用Lasso选择7个变量时预测准确率达到最大。

选择7个变量时，这100次循环的准确率曲线如图。取平均值得到模型准确率，约为87%。

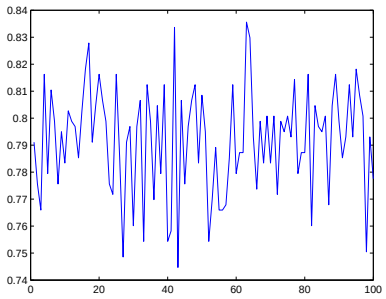
但是实际上，有完整观测的数据仅有450条，占全部数据的1/6。虽然这种模型的预测准确率很高，但并不实用。

4.1.2.2 随机填补的判断准确率

对原数据集进行预处理后，通过Logistic回归选出显著的变量，仅用这些变量进行预测。随机分出训练集合测试集后，对两个数据集都采取随机填补的方法填补缺失值。同样通过5-fold 交叉检验循环100 次获得模型判断准确率。



(a) Lasso选择的变量个数对判断准确率的影响



(b) Lasso选择9个变量时的判断准确率

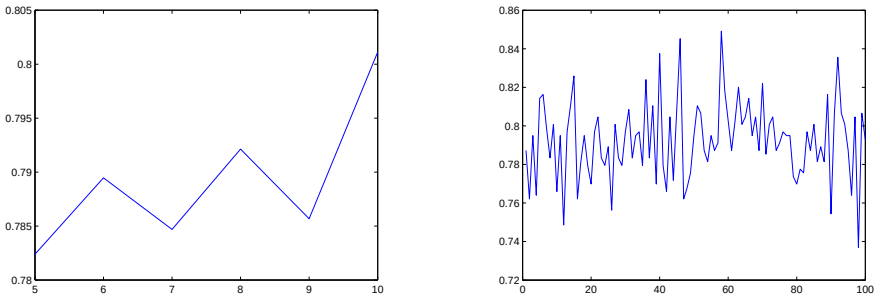
图 4.2　随机填补缺失值进行预测

从图中可以看出，用Lasso选择9个变量时预测准确率达到最大。

选择9个变量时，这100次循环的准确率曲线如图。取平均值得到模型准确率，约为79%。

4.1.2.3 KNN填补的判断准确率

对原数据集进行预处理后，通过Logistic回归选出显著的变量，仅用这些变量进行预测。随机分出训练集合测试集后，对两个数据集都采取KNN填补的方法填补缺失值。同样通过5-fold 交叉检验循环100 次获得模型判断准确率。



(a) Lasso选择的变量个数对判断准确率的影响

(b) Lasso选择10个变量时的判断准确率

图 4.3 KNN填补缺失值进行预测

从图中可以看出，用Lasso选择10个变量时预测准确率达到最大。

选择10个变量时，这100次循环的准确率曲线如图。取平均值得到模型准确率，约为79%。

4.1.3 仅用术前指标的预测

预测手术成败效果，是决定是否手术的一个依据，所以实际上不能用全部指标，而是用术前指标比较合理。

4.1.3.1 初步预测

训练集和测试集都仅保留术前指标，重新进行变量选择和模型建立。

Logistic回归选出如下相对显著的术前指标：

显著程度	指标名称
1	PCI前CKMB
2	随机血糖
3	PCI前CK
4	killip分级
5	LDLC
6	舒张压
7	TNI
8	脑梗塞史
9	PCI前硝酸酯
10	PCI前溶栓
11	高血压史
12	症状到PCI时间

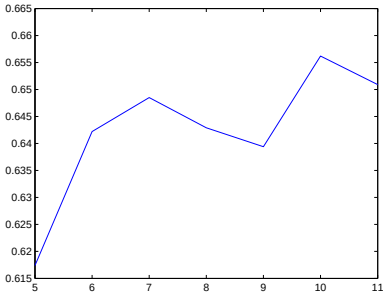
表 4.2　Logistic回归选择术前指标

删除所有含缺失值的观测，直接用Lasso方法选择变量，用SVM进行分类。通过5-fold交叉检验循环100次获得模型判断准确率。

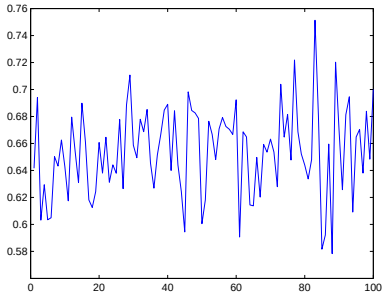
采用完整术前观测数据时，从图中可以看出，用Lasso选择10个变量时预测准确率达到最大。选择10 个变量时，这100次循环的准确率曲线如图。取平均值得到模型准确率，约为65.4%。

采用随机填补时，从图中可以看出，用Lasso选择6个变量时预测准确率达到最大。选择6个变量时，这100次循环的准确率曲线如图。取平均值得到模型准确率，约为72%。

采用KNN填补时，从图中可以看出，用Lasso选择7个变量时预测准确率达到最大。选择7个变量时，这100次循环的准确率曲线如图。取平均值得到模型准确率，约为76%。

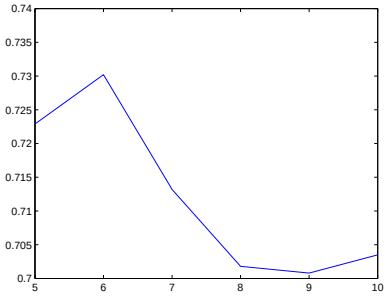


(a) Lasso选择的变量个数对判断准确率的影响

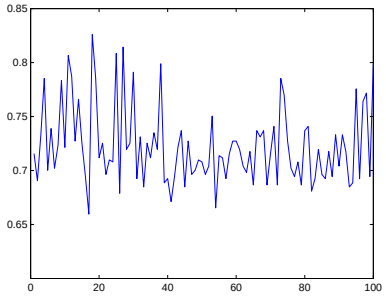


(b) Lasso选择10个变量时的判断准确率

图 4.4 用术前完整观测进行预测



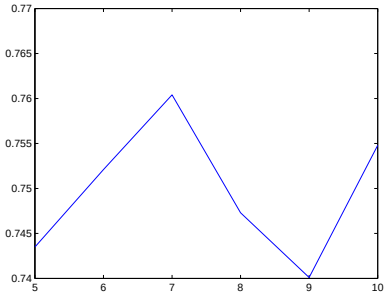
(a) Lasso选择的变量个数对判断准确率的影响



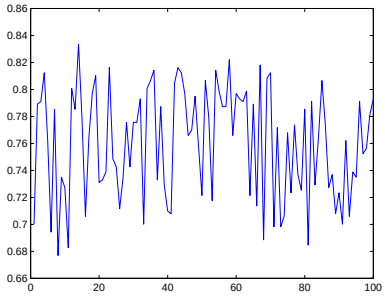
(b) Lasso选择6个变量时的判断准确率

图 4.5 用术前观测随机填补进行预测

显然，由于重要的术中变量的缺失，观测准确率相比用全部指标进行判断要低很多。



(a) Lasso选择的变量个数对判断准确率的影响



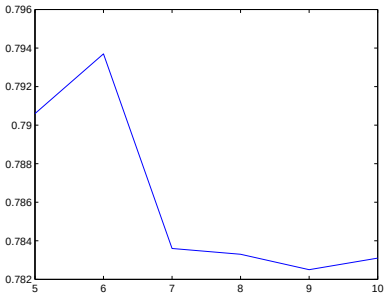
(b) Lasso选择6个变量时的判断准确率

图 4.6　用术前观测KNN填补进行预测

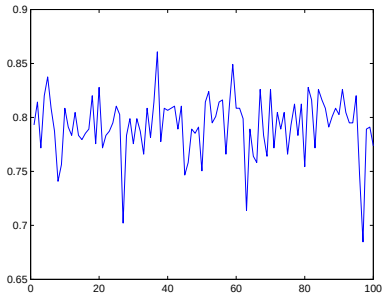
4.1.3.2 模型改进

以上的方法仅考虑了术前指标，而实际上，对于训练集，我们是可以同时获取术前和术后指标的，并且可以从中学习到术前指标和术后指标的相关性。我们可以利用这种相关性在测试集中填补术后指标，从而结合起来预测是否发生无复流现象。

我们将数据集分为训练集和测试集，首先将测试集的所有术后指标置为空，然后与训练集放在一起用KNN方法进行填补。通过测试集进行SVM学习建立模型，应用在生成的训练集上进行预测。



(a) Lasso选择的变量个数对判断准确率的影响



(b) Lasso选择6个变量时的判断准确率

图 4.7　用KNN填补测试集的术后指标进行预测

从图中可以看出，用Lasso选择6个变量时预测准确率达到最大。选择6 个变量时，这100次循环的准确率曲线如图。取平均值得到模型准确率，约为79%，预测准确率已经达到了比较好的程度。

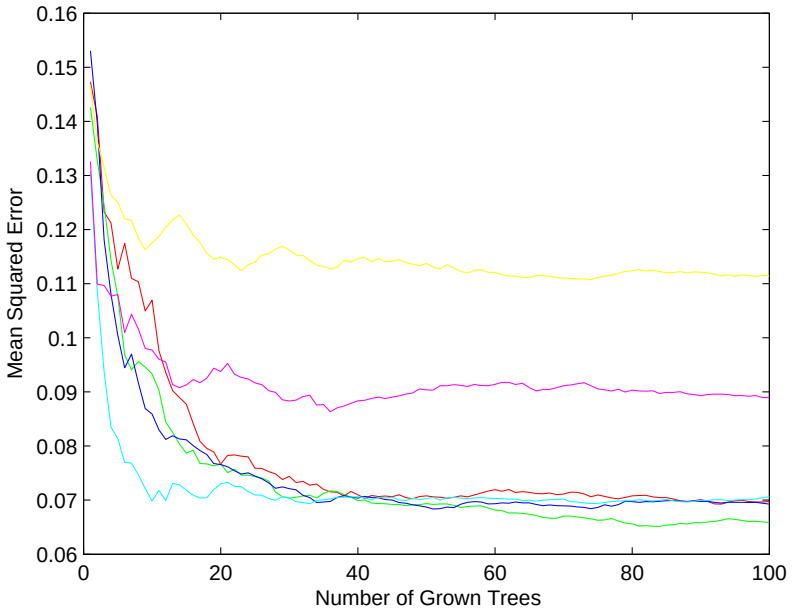


图 4.8　最优Leaf Size

4.2 随机森林的预测结果

4.2.1 变量选择

首先确定最优的Leaf Size。这里首先对具有完整观测值的450条观测形成的数据集进行测试，用红、绿、蓝、亮蓝、粉红、黄六种颜色分别表示Leaf Size取为1,5,10,20,50,100时的均方误差曲线，如下图：

由图中可见，在Leaf Size取为5时，均方误差最小。

接下来，用随机森林对变量的重要性做出比较。

由图中可见，有三个变量的重要性非常大，分别是：罪犯血管血栓数量、PCI术中2b3a和PCI术中钙拮抗剂。遗憾的是，这三个指标都不是术前指标。

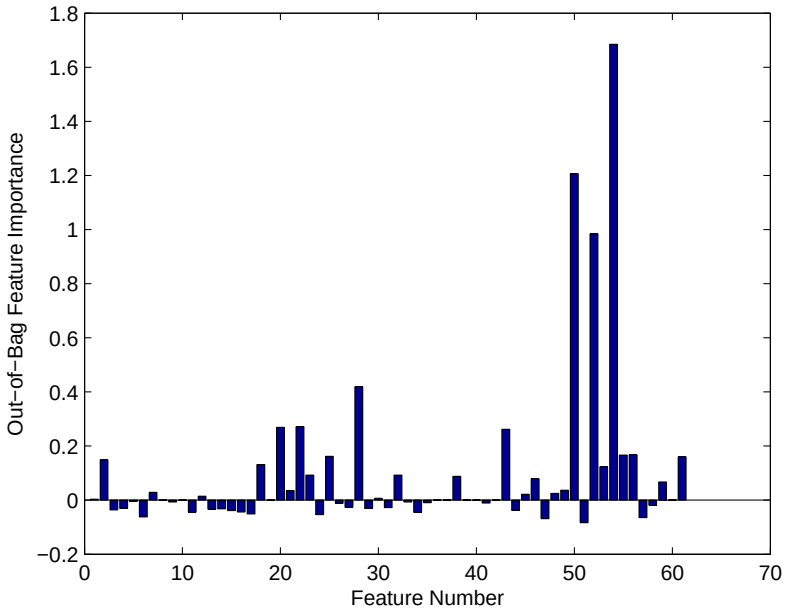
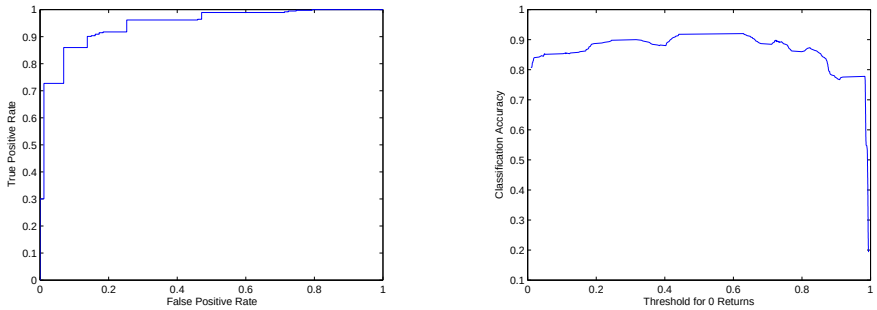


图 4.9　变量重要性

4.2.2 用全部指标进行预测

4.2.2.1 用完整数据进行预测

删除所有含缺失值的观测，对剩下的450条完整观测应用随机森林，用如上挑选出的三个重要变量进行预测，结果如下：



(a) ROC　　(b) 准确率

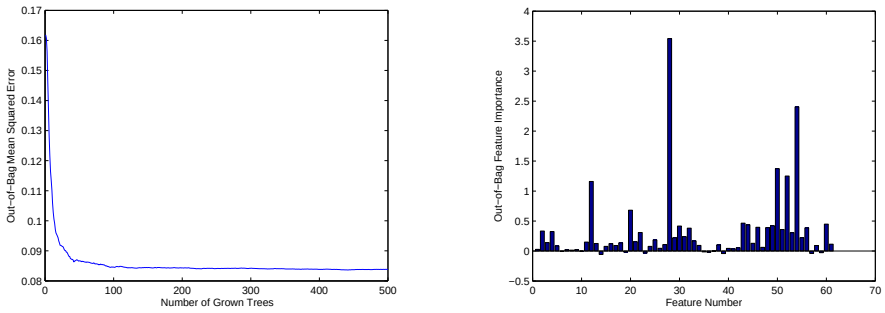
图 4.10　用完整观测进行预测

从图中可见，此时的预测准确率达到90%以上（由于因变量是二元取值，即0和1，此处自然的以0.5为分界点，小于0.5则判定为0，大于等于0.5则判定

为1）。

4.2.2.2 随机填补的判断准确率

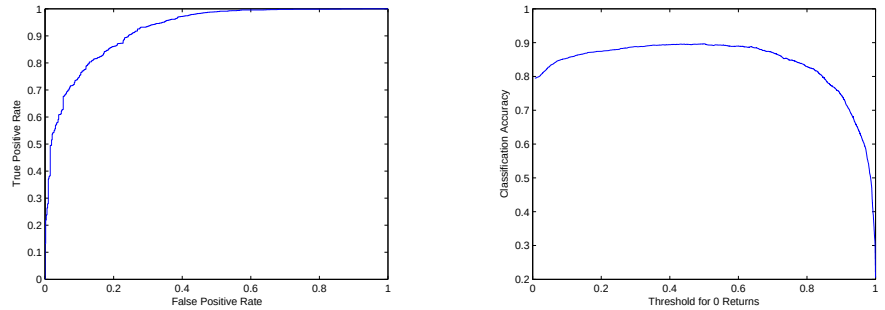
对原始数据进行第二章说明的初步处理后，剩余2569条观测，其中含有大量缺失数据。用随机填补的方法处理数据后用随机森林进行建模，结果如下：



(a) MSE　　(b) 变量重要性

图 4.11　随机填补后用随机森林建模

筛选出重要变量6个，分别是：既往利尿剂、Killip分级、随机血糖、罪犯血管血栓数量、PCI术中2b3a和PCI术中钙拮抗剂。其中，前三个是术前指标。



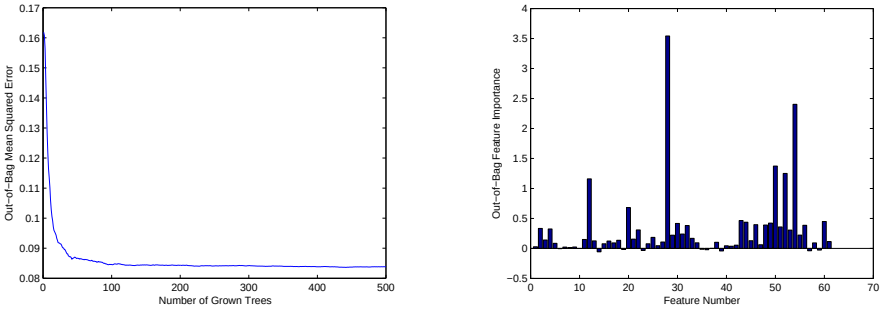
(a) ROC　　(b) 预测准确率

图 4.12　随机填补后用随机森林预测

由图中可见，此时预测准确率达到89.65%。

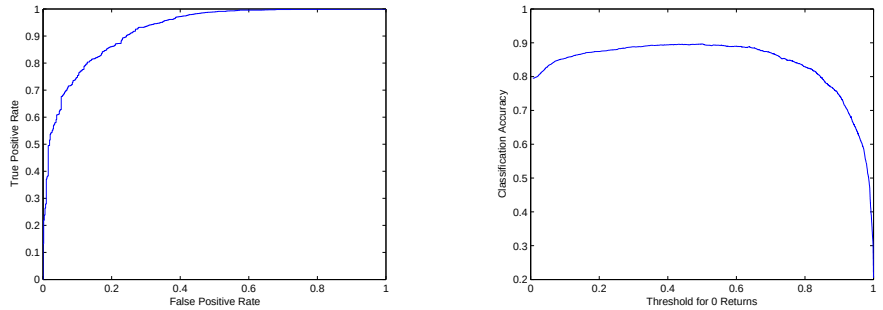
4.2.2.3 KNN填补的判断准确率

用K近邻填补的方法处理数据（取$k = 5$）后用随机森林进行建模，结果如下： 筛选出重要变量8个，分别是：既往利尿剂、Killip分级、随机血糖、BNP、IABP、罪犯血管血栓数量、PCI术中2b3a和PCI术中钙拮抗剂。其中，前5个是术前指标。 由图中可见，此时预测准确率达到90.15%。



(a) MSE　　(b) 变量重要性

图 4.13　k近邻填补后用随机森林建模



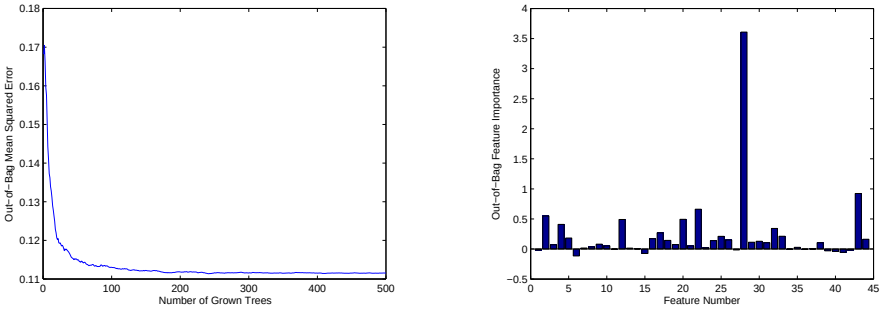
(a) ROC　　(b) 预测准确率

图 4.14　k近邻填补后用随机森林预测

4.2.3 仅用术前指标的预测

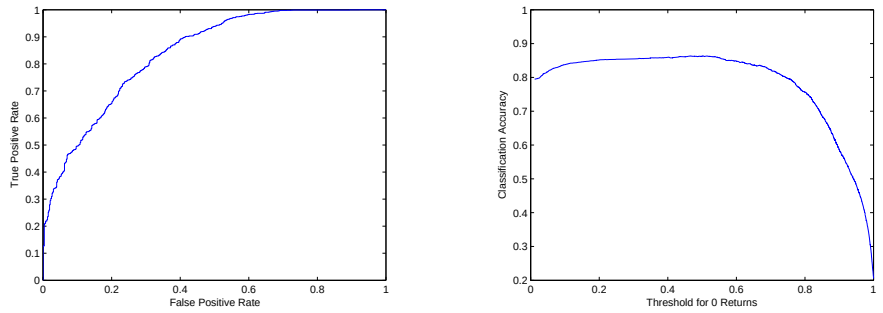
4.2.3.1 初步预测

训练集和测试集都仅保留术前指标，重新进行变量选择和模型建立。随即填补后用随机森林建模，筛选出重要变量6个，分别是：年龄、既往利尿剂、Killip分级、中性粒细胞、随机血糖和IABP。由图中可见，此时预测准确率达到86.30%。



(a) MSE　　(b) 变量重要性

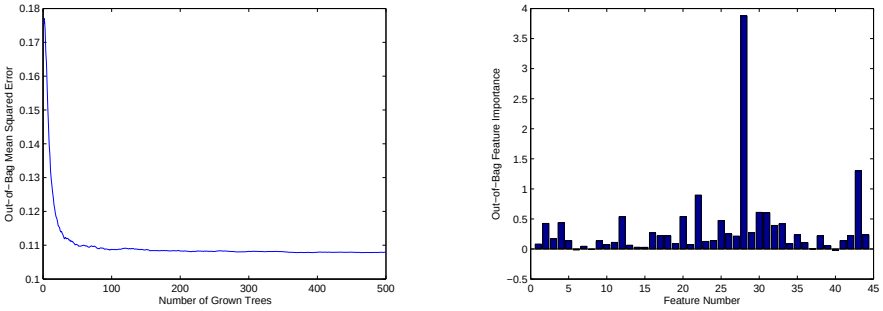
图 4.15　随机填补后用随机森林建模



(a) ROC　　(b) 预测准确率

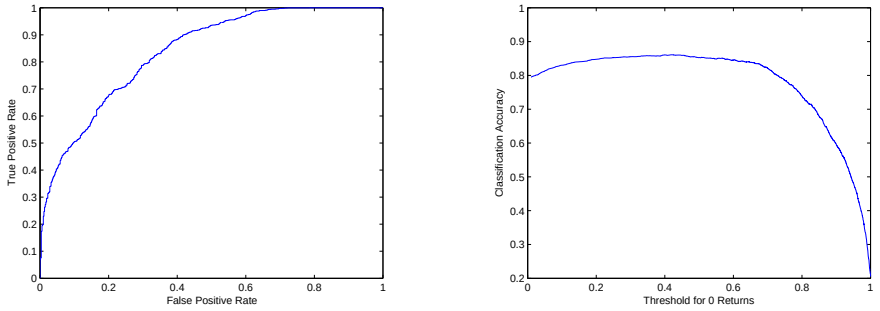
图 4.16　随机填补后用随机森林预测

k近邻填补（取$k=5$）后用随机森林建模，筛选出重要变量7个，分别是：既往利尿剂、Killip分级、中性粒细胞、随机血糖、BNP、PCI前CK和IABP。由图中可见，此时预测准确率达到85.29%。



(a) MSE　　(b) 变量重要性

图 4.17　k近邻填补后用随机森林建模



(a) ROC　　(b) 预测准确率

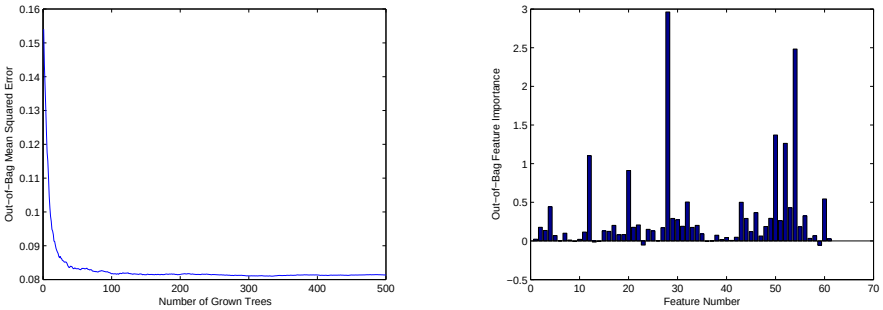
图 4.18　k近邻填补后用随机森林预测

4.2.3.2 模型改进

随机选取2000个观测作为训练集，剩下的569个观测做测试集。先将测试集的非术前指标置为空，再进行随机填补。

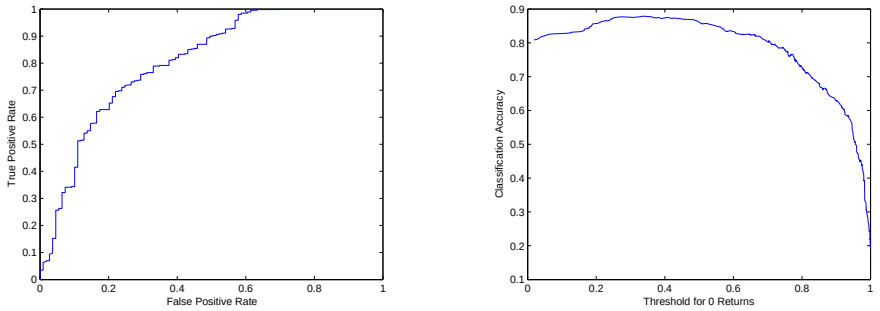
用随机森林建模，结果如下：筛选出重要变量6个，分别是：既往利尿剂、Killip分级、随机血糖、罪犯血管血栓数量、PCI术中2b3a和PCI术中钙拮抗剂。其中，前三个是术前指标。

将测试集的非术前指标置为空，再根据训练集的数据进行填补，用训练集生成的随机森林进行预测。由图中可见，此时预测准确率达到86.99%。预测准确率曲线在分界点取为0.3375时最大，达到87.87%。



(a) MSE (b) 变量重要性

图 4.19 训练集随机填补后用随机森林建模



(a) ROC (b) 预测准确率

图 4.20 测试集随机填补后用随机森林预测

0/0	0/1	1/0	1/1	准确率
444	58	16	51	86.99%

表 4.3 预测准确率

参考文献

[1]Jin-Zhu Jia. Statistical Computing[M]. 2012.

[2]J Friedman, T Hastie, R Tibshirani. Elements of Statistical Learning. 2001.

[3]Hao Zhang, Berg, A.C., Maire, M., Malik, J.SVM-KNN:Discriminative Nearest Neighbor Classification for Visual Category Recognition. 2006.

[4]K-nearest neighbor. http://scholarpedia.org/article/K-nearest_neighbor.

[5]http://www.math.pku.edu.cn/teachers/yaoy/reference/LuYu_201303_BigHeart.pdf

致　谢

在论文即将完成之际，我的心情无法平静，三年的研究生生活如白驹过隙，在此我以一颗感恩的心对所有关心、爱护我的老师和同学表示最真诚的感谢。

感谢陈大岳老师对我的悉心指导和关怀，在我面临选择和人生的十字路口时给我指点迷津，热诚鼓励。

感谢姚远老师在在论文课题和平时学习生活中对我无微不至的悉心指导.毕业论文的选题、框架、思路和每一个证明的细节都离不开您的指导。

感谢张鹏老师、贾金柱老师、蒋达全老师等对我的教育培养。你们在课堂上教会我很多知识是我论文的基础和前提，在此，我要向诸位老师深深地鞠上一躬。

感谢我身边的同学,我的同门,我的室友,我特别要感谢闫博巍同学对我在论文上的帮助.感谢我的室友秦莉同学、吴贵超同学、杨雪芹同学，研究生三年里我们结下了深厚的友谊。在平时生活中大家相互帮助、共同成长。虽然毕业之后我们各奔东西,但相信我们是一生的好朋友。

最后感谢我的爸爸妈妈，焉得谖草，言树之背，养育之恩，无以回报，你们永远健康快乐是我最大的心愿。

北京大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名：　　日期：　　年　月　日

学位论文使用授权说明

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

按照学校要求提交学位论文的印刷本和电子版本；

学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务；

学校可以采用影印、缩印、数字化或其它复制手段保存论文；

在不以赢利为目的的前提下，学校可以公布论文的部分或全部内容。

（保密的论文在解密后应遵守此规定）

论文作者签名：　　导师签名：

日期：　　年　月　日

学位论文出版授权声明

本人已经认真阅读《“中国精品学位论文全文数据库”建设章程》，同意将本人的学位论文提交给“中国精品学位论文全文数据库”项目的产品开发与运作方——北京北大方正电子有限公司全文发表，并可按“中国精品学位论文全文数据库稿酬支付说明”享受相关权益。同意论文提交后滞后：□半年；□一年；□二年发布。

作者签名：_____　　导师签名：_____

_____年__月__日　　_____年__月__日

“中国精品学位论文全文数据库”稿酬支付说明

作者信息：姓名：_____　　学号：_____　　所在院系：_____

提交论文类型：□硕士论文，　　□博士论文

授权作者可以选择报酬方式：

□1、 唯一授权：本人论文电子版**独家**授权给北大方正，并可选择下列报酬方式(三选一)

□（1） 读书卡：硕士论文作者将获得有效期5年价值449元的方正Apabi数字图书馆读书卡1张，博士论文作者将获得有效期10年价值669元的方正Apabi数字图书馆读书卡1张；

□（2） 现金＋读书卡：硕士论文作者将获得30元现金和有效期3年价值269元的方正Apabi数字图书馆读书卡1 张，博士论文作者将获得80元现金和有效期5年价值449元的方正Apabi数字图书馆读书卡1张。

□（3） 销售分成：作者每年获得作者本人提交论文销售收入的10 %作为著作权使用费。本人同意提供并填写完整、正确的个人信息,并在下列任一个人信息发生改变时及时通知北大方正（发邮件到ApabiCEDD@founder.com，或上网http://www.apabi.com在线提交），若因下列个人信息填写不完整、不正确或未将变化及时通知北大方正而发生的著作权使用费无法按期支付等问题由本人负责。北大方正及项目其它参与方负责对作者的下列个人信息保密：

作者姓名：　　开户行：　　银行帐号：

Email：　　手机：

□2、 非唯一授权：本人论文电子版以**非独家**授权方式授权给北大方正，并获得下列报酬:

读书卡: 硕士论文作者将获得有效期3年价值269元的方正Apabi数字图书馆读书卡1张，博士论文作者将获得有效期5年价值449元的方正Apabi数字图书馆读书卡1张。

同时，本人导师将获得方正Apabi数字图书馆读书卡1张（硕士论文作者的导师的读书卡有效期5年，价值449元人民币；博士论文作者的导师的读书卡有效期10年，价值669元人民币）。

北大方正将按照作者选择的论文授权方式和报酬方式，及时支付作者应得的报酬。北大方正将提供数字版权保护技术（DRM）确保合法使用作者论文。同时，北大方正负责为作者提供网上查询论文销售情况及作者个人信息的服务，服务主页：http://www.apabi.com。

注：1、本“稿酬支付说明”一式两份，一份由作者本人保存，另一份做为领取稿酬凭证，交送给图书馆。

2、“学位论文授权使用声明”请直接装订在论文影印本的后面。

3、授权作者需提交“学位论文授权使用声明”的复印件、稿酬支付说明原件到图书馆论文采收处，领取稿酬。