

# 基于样本特征核矩阵的稀疏双线性回归



邵政毅<sup>1</sup> 陈秀宏<sup>1,2</sup>

1 江南大学人工智能与计算机学院 江苏 无锡 214122

2 江苏省媒体设计与软件技术重点实验室 江苏 无锡 214122

**摘要** 在许多实际应用中出现了大量的冗余数据,这些数据可能是高维的,这时进行回归预测将会出现过拟合的现象,并且还会出现预测精度偏低等问题。另外,大多数回归方法都是基于向量的,忽略了矩阵数据原始位置之间的关系。为此,文中提出了一种基于样本特征核矩阵的稀疏双线性回归(Kernel Matrix-based Sparse Bilinear Regression, KMSBR)方法。该方法直接将数据矩阵作为输入,其是通过左右回归系数矩阵而建立的,利用样本的特征核矩阵和  $L_{2,1}$  范数,能够同时实现对样本及样本特征的选择,且考虑了数据的原始位置,提高了算法的性能。在若干数据集上的实验结果表明,KMSBR 能有效地选择相对重要的样本和特征,从而提高算法的运行效率,且其预测精度优于已有的几种回归模型。

**关键词**: 特征核矩阵; 线性回归; 样本与特征提取; 稀疏性; 左右回归矩阵

**中图分类号** TP391.4

## Sample Feature Kernel Matrix-based Sparse Bilinear Regression

SHAO Zheng-yi<sup>1</sup> and CHEN Xiu-hong<sup>1,2</sup>

1 School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China

2 Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi, Jiangsu 214122, China

**Abstract** There are a large number of redundant data in many real applications, which may be high dimensional. In this case, there will be many problems in regression prediction, such as overfitting and low prediction accuracy. In addition, most regression methods are based on vectors, ignoring the relationship between the original positions of matrix data. To this end, a sample kernel matrix-based sparse bilinear regression (KMSBR) method is proposed. The KMSBR model which use the sample feature kernel matrix and  $L_{2,1}$ -norm is established through the left and right regression coefficient matrix. Thus, the KMSBR can implement the selection of samples and its features simultaneously. Experimental results on several data sets show that KMSBR can effectively select samples and its features, thus improve the efficiency of the algorithm, and the prediction accuracy is better than the existing regression models.

**Keywords** Feature kernel matrix, Linear regression, Sample and feature extraction, Sparsity, Left and right regression matrix

## 1 引言

在许多实际应用中,构建学习模型并通过相应的算法来处理数据以解决不同的问题是常用的方法。在这些方法中,回归分析或曲线拟合在实际应用中较为常见。回归模型通过估计一个或多个自变量(输入)与一个因变量(输出)之间的关系来预测给定输入的连续输出值,使模型拟合误差之和最小<sup>[1]</sup>。回归模型可以被广泛应用于金融、经济、电力能源、医疗卫生等领域。但是,如果数据集中存在大量次要的样本与特征,则回归模型容易出现过拟合现象,从而降低预测的精确性。因此,除了预测功能外,选择重要的样本和特征对于回归分析也是至关重要的。其可以提高模型的泛化能力和可解释性,降低冗余样本与特征对预测结果的影响,提高预测精度。

常见的回归方法分为线性回归<sup>[2]</sup>与非线性回归<sup>[3]</sup>。传统的线性回归方法包括岭回归、LASSO (Least Absolute Shrinkage and Selection Operator) 回归等,但这类方法无法

解决非线性问题。而典型的非线性回归方法有支持向量回归(Support Vector Regression, SVR)、最小二乘支持向量回归(Least Squares Support Vector Regression, LSSVR)等。支持向量回归(SVR)利用正则化参数在模型复杂度和总误差之间进行权衡,从而预测输出的输出值<sup>[4-6]</sup>。在面对非线性回归问题时,SVR引入核函数,将有限维的原始空间映射到新的空间,然后在特征空间中进行回归操作。与SVR不同的是,最小二乘支持向量回归(LSSVR)通过最小化平方误差来降低模型的计算复杂度<sup>[7-8]</sup>。与SVR相比,LSSVR模型虽然具有较强的统计鲁棒性和较低的计算复杂度,但缺乏模型解的稀疏性,因此它提取重要样本的能力相对较差。Tian等<sup>[9]</sup>提出了一种线性规划支持向量回归方法(Linear Programming Support Vector Regression, LPSVR),该方法引入  $L_1$  范数为正则化项。但是,上述方法仍然不具有特征选择的能力,而且计算量也很大。Zhang等提出了一种基于双稀疏优化的最小二乘回归方法(Bi-Sparse Optimization-based Least Squares

Regression, BSOLSR)<sup>[10]</sup>和基于双稀疏优化的可解释回归方法(Bi-Sparse Optimization Regression, BSOR)<sup>[11]</sup>,这些方法通过重构行核矩阵和列核矩阵来提高数值预测的准确性,但它们都需要迭代地求解两个线性方程组,因此其复杂度较高。BSOLSR模型仅使用一对回归系数向量,这可能会导致此模型存在较大的回归误差。

现有的特征选择算法大多是基于向量的,如基于线性判别分析的特征选择算法 Fisher Score<sup>[12]</sup>。Liu等<sup>[13]</sup>在特征选择的基础上加入多任务学习,提出了多任务特征选择方法。Nie等<sup>[14]</sup>在鲁棒线性回归中加入稀疏约束,提出了鲁棒特征选择算法(Robust Feature Selection, RFS)。Cai等<sup>[15]</sup>提出了一种特征选择方法,该方法将 $L_{2,1}$ 范数作为损失函数。Liu等<sup>[16]</sup>提出了一种全局和局部结构保存特征选择方法(Global and Local Structure Preservation for Feature Selection, GL-SPFS),该方法在RFS算法全局结构的基础上考虑了数据的局部结构。这些特征选择方法在学习前都需要先将原始的矩阵数据扫描成向量,然后再进行其他的处理。将矩阵数据向量化会引起一些问题,首先,向量数据通常维度都较高,这会导致许多方法出现小样本问题。其次,向量化忽略了原始矩阵数据的位置信息,破坏了数据之间的几何关系。基于此类问题,Hou等<sup>[17]</sup>提出了一种稀疏矩阵回归算法(Sparse Matrix Regression, SMR),用于对矩阵数据进行直接特征选择,有效地改善了特征选择的分类性能。

在处理非线性问题时,为了更为有效地利用原始数据的拓扑信息,本文在BSOLSR方法的基础上,结合稀疏矩阵回归SMR方法,构建了一种新的基于样本特征核的稀疏双线性回归(KMSBR)模型,以在双稀疏优化的基础上提高数值预测的准确性。与同样对样本与特征进行选择的BSOLSR模型相比,KMSBR模型使用了多对回归系数向量,提高了预测精度。此外,KMSBR还使用 $L_{2,1}$ 范数作为稀疏正则化项,可得到稀疏的多对回归系数向量。相比传统的非线性回归模型,KMSBR通过多对回归系数向量同时实现对样本及特征的选择,从而选择重要的样本与特征。因此,KMSBR在提高预测准确性的同时,也具有一定的可解释性。

## 2 相关研究

给定一个训练样本集 $\mathbf{T} = \{(x_i, y_i)\}_{i=1}^n$ ,其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \in R^d$  ( $i = 1, 2, \dots, n$ )为第 $i$ 个样本, $\mathbf{F} = \{f_m\}_{m=1}^d$  ( $f_m \in R^n$ )为特征集, $y_i$  ( $y_i \in R$ )为其对应的连续输出值, $d$ 为特征维数, $n$ 为样本个数。对于回归问题,目标是寻找一个回归函数 $f: x_i \rightarrow y_i$ ,使训练集 $\mathbf{T}$ 的预测误差值最小。对于测试样本 $x$ ,通过回归函数给出其预测值。

下文介绍几种典型的基于回归的特征选择方法。

### 2.1 鲁棒特征选择 (RFS)

RFS<sup>[15]</sup>是一种基于向量的特征选择方法,该方法通过对回归系数 $W$ 施加 $L_{2,1}$ 范数,使得 $W$ 稀疏,并通过稀疏的回归系数选择数据特征。RFS的目标函数为:

$$L(W, b) = \sum_{i=1}^n \|W^T x_i + b - y_i\|_F + \alpha \|W\|_{2,1} \quad (1)$$

其中, $W \in R^{d \times c}$ ,  $b \in R^c$  ( $c$ 为类别数)为偏置项, $\alpha$  ( $\alpha > 0$ )为拉格朗日参数。这里使用 $L_{2,1}$ 范数来实现样本特征的选择,即选

择样本数据中相对重要的特征。虽然RFS在许多应用中都取得了显著的效果,但它只接受向量作为输入。当数据集为矩阵数据集时,矢量化会忽略原始矩阵的位置信息。

### 2.2 最小二乘支持向量回归 (LSSVR)

LSSVR<sup>[5-6]</sup>是一种基于核函数的特征选择方法。与支持向量回归模型(SVR)相比,LSSVR使用平方误差和(SSE)来代替SVR中的总误差,故LSSVR的模型如下:

$$\min_{w, b, \zeta} \frac{1}{2} (\|w\|_2^2 + \alpha \sum_{i=1}^n \zeta_i^2) \quad (2)$$

$$\text{s. t. } (w^T \phi(x_i) + b) + \zeta_i = y_i, \zeta_i \in R, i = 1, \dots, n$$

其中, $w$ 为权重向量,截距 $b$ 为无约束变量, $\zeta_i$ 为第 $i$ 个样本的误差, $\alpha$  ( $\alpha > 0$ )为拉格朗日参数,函数 $\phi(x)$ 为核函数。LSSVR仅是对样本进行冗余度分析,但不包含对特征的稀疏分析。

### 2.3 基于双稀疏优化的最小二乘回归方法 (BSOLSR)

BSOLSR<sup>[10]</sup>是一种基于核函数的特征选择方法,它可以选择重要的样本与特征。BSOLSR可表示为以下两个优化问题:

$$\min_{\bar{\lambda}, \eta} \frac{1}{2} (\bar{\lambda}^T \text{diag}(\theta^{(\eta)}) \bar{\lambda} + \alpha \sum_{i=1}^n \eta_i^2) \quad (3)$$

$$\text{s. t. } \bar{\lambda}^T A_i + \eta_i = y_i, \bar{\lambda} \in R^{n+1}, \eta_i \in R, i = 1, \dots, n$$

$$\min_{\bar{\mu}, \zeta} \frac{1}{2} (\bar{\mu}^T \text{diag}(\pi^{(\zeta)}) \bar{\mu} + \beta \sum_{i=1}^m \zeta_i^2) \quad (4)$$

$$\text{s. t. } \bar{\mu}^T B_i + \zeta_i = y_i, \bar{\mu} \in R^{d+1}, \zeta_i \in R, i = 1, \dots, n$$

其中, $\eta = (\eta_1, \dots, \eta_n)^T$ 和 $\zeta = (\zeta_1, \dots, \zeta_n)^T$ 是两个松弛变量, $\alpha > 0$ 和 $\beta > 0$ 是两个正参数, $\bar{\lambda}$ 和 $\bar{\mu}$ 是两个回归系数向量。 $A$ 和 $B$ 为所有训练样本的行核矩阵和列核矩阵,即:

$$A = \begin{bmatrix} \mu^T u_{11} & \dots & \mu^T u_{n1} \\ \vdots & \ddots & \vdots \\ \mu^T u_{1n} & \dots & \mu^T u_{nn} \\ 1 & \dots & 1 \end{bmatrix}, B = \begin{bmatrix} \lambda^T v_{11} & \dots & \lambda^T v_{n1} \\ \vdots & \ddots & \vdots \\ \lambda^T v_{1d} & \dots & \lambda^T v_{nd} \\ 1 & \dots & 1 \end{bmatrix} \quad (5)$$

其中, $\mu_{ij} = (k(x_{i1}, x_{j1}), \dots, k(x_{id}, x_{jd}))^T$ ,  $v_{im} = (k(x_{i1}, x_{1m}), \dots, k(x_{im}, x_{mm}))^T$ 。交替求解式(3)和式(4)可得到稀疏的回归系数向量 $\bar{\lambda}$ 和 $\bar{\mu}$ 。

## 3 基于样本特征核矩阵的稀疏双线性回归 (KMSBR)

BSOLSR算法使用两个稀疏回归模型来分别实现对样本与特征的选择。而提出的KMSBR算法则可以同时实现样本选择和特征选择的统一模型,并利用RFS中的 $L_{2,1}$ 范数来表示损失函数和稀疏约束。

对于任意两个训练样本 $x_i$ 和 $x_j$  ( $i, j = 1, \dots, n$ ),假设基函数 $\phi(x_m)$ 将 $x_i$ 的第 $m$ 个特征值 $x_{im}$ 从原始特征空间映射到一个新的特征空间 ( $m = 1, \dots, d$ ),则称函数 $k(x_{im}, x_{jm}) = \phi(x_{im})^T \phi(x_{jm})$ 为 $x_i$ 和 $x_j$ 的第 $m$ 个特征的核函数。本文使用如下线性核函数:

$$k(x_{im}, x_{jm}) = \phi(x_{im})^T \phi(x_{jm}) = x_{im}^T x_{jm} \quad (6)$$

则定义如下矩阵:

$$K_i = \begin{bmatrix} k(x_{i1}, x_{11}) & k(x_{i2}, x_{12}) & \dots & k(x_{id}, x_{1d}) \\ k(x_{i1}, x_{21}) & k(x_{i2}, x_{22}) & \dots & k(x_{id}, x_{2d}) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_{i1}, x_{n1}) & k(x_{i2}, x_{n2}) & \dots & k(x_{id}, x_{nd}) \end{bmatrix} \quad (7)$$

该矩阵为样本 $\mathbf{x}_i$ 的特征核矩阵,它的第 $j$ 行为 $\mathbf{x}_i$ 与 $\mathbf{x}_j$ 的特征核向量,称之为样本 $\mathbf{x}_i$ 与 $\mathbf{x}_j$ 的行核向量;第 $m$ 列为 $\mathbf{x}_i$ 和所有样本的第 $m$ 个特征的核向量,称之为 $\mathbf{x}_i$ 与所有样本的第 $m$ 个特征间的列核向量。

广义双线性回归模型(GBR)<sup>[18]</sup>使用一组左回归系数向量和右回归系数向量来实现对样本与特征的选择,但是该模型的拟合误差较大。为了提升模型的回归预测能力,使用 $s < \min(n, d)$ 对左回归向量 $\{\mathbf{u}_l\}_{l=1}^s$ 和右回归向量 $\{\mathbf{v}_l\}_{l=1}^s$ 建立以下基于样本特征核矩阵的双线性回归(KMSBR)模型。

$$\min_{\mathbf{U}, \mathbf{V}, b} \sum_{i=1}^n (\text{Tr}(\mathbf{U}^T \mathbf{K}_i \mathbf{V}) + b - y_i)^2 + \alpha \|\mathbf{U}\|_{2,1} + \beta \|\mathbf{V}\|_{2,1} \quad (8)$$

其中,  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_s] \in R^{n \times s}$  和  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s] \in R^{d \times s}$  分别为左、右回归系数矩阵,回归向量对数 $s$ 用来平衡回归模型的学习与泛化能力。 $\alpha$ 和 $\beta$ 为非负参数, $b \in R$ 为模型偏差。

模型(8)中行稀疏的左回归矩阵 $\mathbf{U}$ 可以从 $\mathbf{K}_i$ 中选择行,即实现样本的选择,而行稀疏的右回归矩阵 $\mathbf{V}$ 可以从 $\mathbf{K}_i$ 中选择列,即实现特征的选择。因此,该模型能同时实现对样本及特征的选择。模型(8)含有3个变量 $\mathbf{U}, \mathbf{V}$ 和 $b$ ,同时求解会比较困难,因此可以利用交替迭代法进行求解。求解过程如下。

(1) 固定 $\mathbf{V}$ 和 $b$ ,更新 $\mathbf{U}$

当 $\mathbf{V}$ 和 $b$ 固定时,目标函数(8)可表示为:

$$\sum_{i=1}^n \left( \sum_{l=1}^s (\mathbf{u}_l^T \mathbf{K}_i \mathbf{v}_l + b - y_i)^2 + \alpha \text{Tr}(\mathbf{U}^T \mathbf{D}^{(u)} \mathbf{U}) \right) \quad (9)$$

其中,对角矩阵 $\mathbf{D}^{(u)}$ 的第 $l$ 个主对角元素定义为:

$$d_{ll}^{(u)} = \begin{cases} \frac{1}{2 \|\mathbf{u}^l\|_2^2}, & \mathbf{u}^l \neq 0 \\ \delta, & \text{otherwise} \end{cases} \quad (10)$$

其中, $\mathbf{u}^l$ 为 $\mathbf{U}$ 的第 $l$ 行, $\delta$ 为充分小的正常数。

对式(9)关于 $\mathbf{u}_j$ 求导并令其等于零,可以得到:

$$\mathbf{u}_j = -[\alpha \mathbf{D}^{(u)} + \sum_{i=1}^n \mathbf{K}_i \mathbf{v}_j (\mathbf{v}_j)^T (\mathbf{K}_i)^T]^{-1} \times \sum_{i=1}^n \left[ \sum_{l \neq j} (\mathbf{u}_l)^T \mathbf{K}_i \mathbf{v}_l + b - y_i \right] \mathbf{K}_i \mathbf{v}_j \quad (11)$$

(2) 固定 $\mathbf{U}$ 和 $b$ ,更新 $\mathbf{V}$

当固定 $\mathbf{U}$ 和 $b$ 时,目标函数(8)可表示为:

$$\sum_{i=1}^n \left( \sum_{l=1}^s (\mathbf{u}_l^T \mathbf{K}_i \mathbf{v}_l + b - y_i)^2 + \beta \text{Tr}(\mathbf{V}^T \mathbf{D}^{(v)} \mathbf{V}) \right) \quad (12)$$

其中,对角矩阵 $\mathbf{D}^{(v)}$ 的第 $r$ 个主对角元素定义为:

$$d_{rr}^{(v)} = \begin{cases} \frac{1}{2 \|\mathbf{v}^r\|_2^2}, & \mathbf{v}^r \neq 0 \\ \delta, & \text{otherwise} \end{cases} \quad (13)$$

其中, $\mathbf{v}^r$ 为 $\mathbf{V}$ 的第 $r$ 行。

对式(12)关于 $\mathbf{v}_j$ 求导并令其等于零,可以得到:

$$\mathbf{v}_j = -[\beta \mathbf{D}^{(v)} + \sum_{i=1}^n (\mathbf{K}_i)^T \mathbf{u}_j (\mathbf{u}_j)^T \mathbf{K}_i]^{-1} \times \sum_{i=1}^n \left[ \sum_{l \neq j} (\mathbf{v}_l)^T (\mathbf{K}_i)^T \mathbf{u}_l + b - y_i \right] (\mathbf{K}_i)^T \mathbf{u}_j \quad (14)$$

(3) 固定 $\mathbf{U}$ 和 $\mathbf{V}$ ,更新 $b$

当固定 $\mathbf{U}$ 和 $\mathbf{V}$ 时,目标函数(8)等价于:

$$\min_b \sum_{i=1}^n (\text{Tr}(\mathbf{U}^T \mathbf{K}_i \mathbf{V}) + b - y_i)^2 \quad (15)$$

其最优解为:

$$b = \frac{1}{n} \sum_{i=1}^n (y_i - \text{Tr}(\mathbf{U}^T \mathbf{K}_i \mathbf{V})) \quad (16)$$

以上过程可迭代地进行,直到达到最大迭代次数或满足如下终止迭代条件:

$$|L(t+1) - L(t)| < \tau \quad (17)$$

其中,  $L(t) = \sum_{i=1}^n (\text{Tr}(\mathbf{U}^{(t)})^T \mathbf{K}_i \mathbf{V}^{(t)}) + b^{(t)} - y_i)^2 + \alpha \|\mathbf{U}^{(t)}\|_{2,1} + \beta \|\mathbf{V}^{(t)}\|_{2,1}$ ,  $\tau > 0$  为一个充分小的正实数。整个过程如算法1所示。

#### 算法1 KMSBR算法

输入: 训练样本 $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ ; 正则化参数 $\alpha$ 和 $\beta$ ; 左右回归向量的对数 $s \in \{1, 2, \dots\}$ ; 核函数 $k(\cdot, \cdot, \cdot)$ ; 最大迭代次数 $T_{\max}$

输出:  $\mathbf{U}^{(T_{\max})}$  和  $\mathbf{V}^{(T_{\max})}$

1. 初始化:  $\mathbf{U}^{(0)} = [\mathbf{I}_{s \times s}, \mathbf{0}_{s \times (n-s)}]^T$ ,  $\mathbf{V}^{(0)} = [\mathbf{I}_{s \times s}, \mathbf{0}_{s \times (d-s)}]^T$ ,  $t=0$
2. while no convergence do
3. 由式(6)计算各样本 $\mathbf{x}_i$ 的特征核矩阵 $\mathbf{K}_i$
4. 由式(16)更新 $b$ 的值
5. 由式(10)计算对角矩阵 $\mathbf{D}^{(u)}$
6. 由式(11)更新左回归矩阵 $\mathbf{U}$
7. 由式(13)计算对角矩阵 $\mathbf{D}^{(v)}$
8. 由式(14)更新右回归矩阵 $\mathbf{V}$
9. 更新迭代次数  $t=t+1$
10. end while

## 4 实验和结果分析

本节通过实验说明KMSBR的有效性,并将它和其他7种算法进行比较,即SMR<sup>[17]</sup>,SVR<sup>[2-4]</sup>,BSOLSR<sup>[10]</sup>,BSOR<sup>[11]</sup>,LASSO<sup>[19]</sup>,LSSVR<sup>[5-6]</sup>和RFS<sup>[14]</sup>。这里使用的11个不同规模的数据集均取自UCI数据库<sup>[20]</sup>,各数据集的样本数和特征数如表1所列。本文评价指标分别为均方误差(MSE)、平均绝对误差(MAE)、定义样本缩减率(IRR)和特征缩减率(FRR)等预测统计量。实验平台为MATLAB R2016a。

表1 数据集统计

Table 1 Data set statistics

data set	number of samples	number of features	number of training sample	number of testing sample
bodyfat	252	14	176	76
kidney	76	6	53	23
house	20640	6	4368	16272
PM10	500	7	350	150
ml-prove	4589	56	3212	1377
SCAD1	70	205	49	21
NO2	500	7	350	150
airfoil_self_noise	1503	5	1052	451
abalone	4177	8	2505	1672
winequality-red	1599	11	1117	482
dermatology	366	33	254	112

### 4.1 数据集

实验中将每个数据集分别随机地划分为训练集和测试集,各数据集的训练样本数以及测试样本数如表1所列。

从表1中各数据集的统计信息可以看出,house,ml-prove,airfoil self\_noise,abalone及winequality-red等数据集相对较大,而ml-prove和SCAD1则是高维的。其中ml-prove是高维的大数据集,其他的都是中小规模的数据集。

### 4.2 性能指标

实验中,对于11个数据集集中的每个数据,采用最小-最大准则将不同特征值归一化到区间 $[0, 1]$ 。

评价各回归模型预测性能的指标通常有均方误差(MSE)和平均绝对误差(MAE),它们的定义分别为:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (18)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (19)$$

其中,  $y_i$  是真实的标签,  $\hat{y}_i$  是样本  $x_i$  ( $i=1, 2, \dots, n$ ) 的预测值。MSE 和 MAE 值越接近于零, 回归模型的预测精度就越高。

此外, 还使用重要样本数 #  $II_s$  和重要特征数 #  $IF_s$  来分别分析比较不同回归模型的样本选择能力和特征选择能力。对于某种回归模型, 为了度量其识别重要样本和特征的能力, 分别定义样本缩减率 (IRR) 和特征缩减率 (FRR) 为:

$$IRR = \left(1 - \frac{\#II_s}{|F|}\right) \times 100\% \quad (20)$$

$$FRR = \left(1 - \frac{\#IF_s}{|T|}\right) \times 100\% \quad (21)$$

其中,  $F$  为训练集中的样本数, 而  $T$  为训练集中的特征数, #  $II_s$  为重要样本数, #  $IF_s$  为重要特征数。这里, 样本的重要性和特征的重要性可以分别用以下指标来描述:

$$II(x_i) = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \times 100\% \quad (22)$$

$$FI(f_m) = \frac{\mu_m}{\sum_{m=1}^d \mu_m} \times 100\% \quad (23)$$

其中,  $x_i$  为第  $i$  个样本,  $f_m$  是特征集中的第  $m$  个特征;  $\lambda_i = \|u^i\|_1$  为左回归矩阵  $U$  中第  $i$  行元素的绝对值之和, 它表示样本  $x_i$  的相对重要性;  $\mu_m = \|v^m\|_1$  为右回归矩阵  $V$  中第  $m$  行所有元素绝对值之和, 它表示所有的样本第  $m$  个特征  $f_m$  的重要性 ( $m=1, \dots, d$ )。设置充分小的正实数  $\sigma_1$  和  $\sigma_2$ , 如果  $\lambda_i < \sigma_1$ , 那么对应的样本  $x_i$  为次要的样本, 其  $II$  值取为零; 否则, 样本  $x_i$  为相对重要的样本, 而且  $II$  值越大, 对应的样本就越重要, 因此它为更具代表性的训练样本。同理, 如果  $\mu_m < \sigma_2$ , 那么对应的特征  $f_m$  为次要的特征, 其  $FI$  值可取为零; 否则, 特征  $f_m$  为相对重要的特征。这样, 可以从训练集和特征集中选择一些相对重要的样本和特征, 从而提高样本分类的效率。

### 4.3 参数设置

由模型(8)可知, 非负参数  $\alpha$  和  $\beta$  以及回归向量个数  $s$  的大小决定了算法的预测精确度。假设非负参数  $\alpha$  和  $\beta$  的取值范围为  $[10^{-4}, 10^{-3}, \dots, 10^1]$ ,  $s$  为  $[1, \min(n, d)]$  内的正整数。采用网格搜索的方法, 在以上参数范围内找到在测试集上使得 MSE 最优的参数组合。各回归模型在 11 个数据集上的最优参数  $\alpha, \beta$  以及 KMSBR 中的  $\alpha, \beta$  和  $s$  参数值如表 2 所列。

表 2 在 11 个测试集上各回归模型的参数值

Table 2 Parameter values of each regression model on 11 test sets

data set	KMSBR			SMR	SVR	BSOLSR		LASSO	LSSVR	RFS	BSOR	
	$\alpha$	$\beta$	$s$	$\alpha$	$\alpha$	$\alpha$	$\beta$	$\alpha$	$\alpha$	$\alpha$	$\alpha$	$\beta$
bodyfat	$10^{-2}$	$10^{-2}$	1	$10^{-4}$	$10^3$	$10^1$	$10^2$	$10^0$	$10^3$	$10^{-3}$	$10^1$	$10^1$
kidney	$10^{-1}$	$10^{-1}$	1	$10^{-4}$	$10^2$	$10^2$	$10^3$	$10^0$	$10^0$	$10^{-2}$	$10^2$	$10^3$
house	$10^{-2}$	$10^{-2}$	2	$10^{-2}$	$10^3$	$10^1$	$10^1$	$10^0$	$10^4$	$10^{-1}$	$10^1$	$10^1$
PM10	$10^{-2}$	$10^{-2}$	2	$10^{-4}$	$10^4$	$10^2$	$10^4$	$10^0$	$10^4$	$10^{-3}$	$10^1$	$10^1$
ml-prove	$10^{-2}$	$10^{-2}$	1	$10^{-2}$	$10^2$	$10^1$	$10^1$	$10^0$	$10^2$	$10^{-3}$	$10^2$	$10^2$
SCAD1	$10^0$	$10^0$	1	$10^{-3}$	$10^1$	$10^2$	$10^3$	$10^0$	$10^2$	$10^{-2}$	$10^2$	$10^2$
NO2	$10^{-4}$	$10^{-4}$	1	$10^{-2}$	$10^5$	$10^2$	$10^2$	$10^0$	$10^4$	$10^{-3}$	$10^2$	$10^2$
airfoil_self_noise	$10^{-2}$	$10^{-2}$	1	$10^{-3}$	$10^4$	$10^2$	$10^2$	$10^0$	$10^1$	$10^{-3}$	$10^1$	$10^1$
abalone	$10^{-4}$	$10^{-4}$	1	$10^{-2}$	$10^4$	$10^2$	$10^3$	$10^0$	$10^3$	$10^{-3}$	$10^1$	$10^3$
winequality-red	$10^{-4}$	$10^{-4}$	1	$10^{-2}$	$10^3$	$10^2$	$10^2$	$10^0$	$10^2$	$10^{-3}$	$10^1$	$10^1$
dermatology	$10^{-4}$	$10^{-4}$	1	$10^{-3}$	$10^1$	$10^1$	$10^3$	$10^0$	$10^2$	$10^{-3}$	$10^0$	$10^3$

### 4.4 计算复杂性

由算法 1 可以看出, KMSBR 的计算步骤最多的是用函数(11)更新左回归矩阵  $U$ , 用函数(14)更新右回归矩阵  $V$ 。函数(11)是函数(9)正则化最小二乘问题的解, 其计算复杂度为  $O(n \times s^2)$ 。相应地, 函数(14)中的计算复杂度为  $O(n \times s^2)$ 。假设总迭代次数为  $T$ , KMSBR 的计算复杂度为  $O(n^2 \times s^2 \times T)$ 。一般  $T$  小于 30,  $s$  远小于  $n$ 。

### 4.5 实验结果与分析

KMSBR 算法和对比算法在 11 个训练集上进行训练, 并分别在训练集上找到参数最优的回归模型, 然后将这些回归模型应用于独立测试集, 得到预测输出值。

对于 11 个测试集中的每个测试集, 使用式(18)一式(21)分别计算回归模型的 MSE, MAE, FRR 和 IRR 值, 如表 3—表 6 所列。

表 3 在 11 个测试集上各回归模型预测的 MSE 值

Table 3 MSE values predicted by each regression model on 11 test sets

data set	KMSBR	SMR	SVR	BSOLSR	LASSO	LSSVR	RFS	BSOR
bodyfat	0.0004	0.0004	0.0717	0.0320	0.0024	<b>0.0003</b>	0.0006	0.0480
kidney	<b>0.0685</b>	0.0891	0.2105	0.0748	0.2613	0.0774	0.1058	0.1038
house	<b>0.0340</b>	0.0559	0.0943	0.0562	0.1032	0.0461	0.1154	0.0549
PM10	<b>0.0263</b>	0.0291	0.0921	0.2024	0.0276	0.0264	0.0431	0.0366
ml-prove	<b>0.0001</b>	0.2430	0.0519	0.2489	0.1319	0.0008	0.0025	0.2398
SCAD1	<b>0.0350</b>	0.0507	0.2712	0.0500	0.0699	0.0473	0.0368	0.1237
NO2	0.0166	0.0255	0.0696	0.1092	0.0202	<b>0.0129</b>	0.0176	0.0500
airfoil_self_noise	<b>0.0303</b>	0.0374	0.2157	0.0329	0.0419	0.0328	0.0984	0.0342
abalone	<b>0.1678</b>	0.1690	0.2316	0.1789	0.1741	0.1680	0.1760	0.1726
winequality-red	<b>0.0178</b>	0.0263	0.0874	0.0252	0.0252	0.0202	0.0330	0.0257
dermatology	0.0217	0.0197	0.0663	0.0955	0.0305	<b>0.0147</b>	0.0281	0.0946

表4 在11个测试集上各回归模型预测的MAE值

Table 4 MAE values predicted by each regression model on 11 test sets

data set	KMSBR	SMR	SVR	BSOLSR	LASSO	LSSVR	RFS	BSOR
bodyfat	0.0130	0.0121	0.2667	0.1470	0.0387	<b>0.0096</b>	0.0167	0.1771
kidney	<b>0.2052</b>	0.2407	0.3986	0.2092	0.4381	0.2151	0.2676	0.2798
house	<b>0.1416</b>	0.1780	0.2323	0.1822	0.2407	0.1656	0.2551	0.1802
PM10	0.1309	0.1360	0.2647	0.4152	0.1341	<b>0.1300</b>	0.1660	0.1497
ml-prove	<b>0.0024</b>	0.4861	0.2105	0.4916	0.3283	0.0257	0.0124	0.4781
SCAD1	<b>0.1381</b>	0.1966	0.4867	0.1905	0.2173	0.1656	0.1436	0.2746
NO2	0.0969	0.1223	0.2397	0.3007	0.1096	<b>0.0815</b>	0.0955	0.1713
airfoil_self_noise	<b>0.1458</b>	0.1604	0.4118	0.1500	0.1648	0.1495	0.2686	0.1495
abalone	<b>0.3461</b>	0.3465	0.4160	0.3695	0.3600	0.3490	0.3690	0.3601
winequality-red	<b>0.1042</b>	0.1371	0.2693	0.1324	0.1247	0.1121	0.1397	0.1351
dermatology	0.1126	0.1124	0.2156	0.2633	0.1357	<b>0.0964</b>	0.1328	0.2641

表5 在11个测试集上6种回归模型的FRR值

Table 5 FRR values of the five regression models on 11 test sets

data set	KMSBR	SMR	BSOLSR	LASSO	RFS	BSOR
bodyfat	50.00	64.29	42.86	<b>71.43</b>	21.43	64.29
kidney	<b>83.33</b>	<b>83.33</b>	50.00	16.67	16.67	33.33
house	62.50	12.50	<b>75.00</b>	62.50	12.50	37.50
PM10	42.86	42.86	<b>57.14</b>	28.57	28.57	28.57
ml-prove	<b>75.00</b>	66.07	73.21	66.07	73.21	73.21
SCAD1	82.93	<b>95.12</b>	81.95	87.32	82.44	82.44
NO2	42.86	28.57	<b>57.14</b>	28.57	28.57	42.86
airfoil_self_noise	40.00	<b>60.00</b>	<b>60.00</b>	<b>60.00</b>	<b>60.00</b>	<b>60.00</b>
abalone	<b>50.00</b>	37.50	12.50	37.50	12.50	12.50
winequality-red	<b>54.55</b>	<b>54.55</b>	45.45	27.27	27.27	45.45
dermatology	48.48	33.33	<b>90.91</b>	48.48	27.27	30.30

表6 在11个测试集上6种回归模型的IRR值

Table 6 IRR values of the five regression models on 11 test sets

data set	KMSBR	SMR	SVR	BSOLSR	LSSVR	BSOR
bodyfat	36.93	36.93	60.23	43.18	<b>66.48</b>	46.02
kidney	54.72	86.79	28.30	<b>98.11</b>	41.51	56.60
house	99.91	0.04	29.40	<b>99.95</b>	30.72	<b>99.95</b>
PM10	98.29	0.29	43.14	<b>99.71</b>	43.14	94.86
ml-prove	<b>61.99</b>	52.18	53.80	40.41	54.98	40.41
SCAD1	55.10	73.47	26.53	<b>97.96</b>	44.90	91.84
NO2	91.71	95.14	22.86	<b>99.71</b>	17.71	84.86
airfoil_self_noise	59.32	96.10	58.56	<b>99.81</b>	60.65	96.10
abalone	99.24	74.69	21.52	<b>99.88</b>	30.30	43.16
winequality-red	98.48	92.66	65.98	<b>99.91</b>	68.13	88.99
dermatology	94.88	90.55	94.08	<b>97.64</b>	95.28	95.28

由表3和表4可知,在大多数测试集上,KMSBR一般比其他回归模型具有更好的MSE和MAE值。具体来说,LSSVR在bodyfat,NO2和dermatology数据集上获得了最佳的MSE,而KMSBR在剩余的8个数据集上获得了最佳的MSE。关于MAE,LSSVR在bodyfat,PM10,NO2和dermatology数据集上分别获得了最佳的MAE值,但它们与KMSBR的MAE相差不是太明显,而且KMSBR在剩余的数据集上明显地获得了最佳的MAE值。

由表3和表4可知,在大样本以及高维的数据集中,KMSBR算法具有更好的预测性能,而在小样本或是低维的数据集中,KMSBR算法的预测性能则与其他几种回归方法相差不多。这主要是因为在大样本或是高维的数据集中,具有判别性的特征或样本会更多,而KMSBR算法结合了稀疏矩阵回归方法,利用了 $L_{2,1}$ 范数稀疏正则化项,具有可以同时选择出更多具有判别性的特征与样本的优势,因此在这类数据集中,KMSBR算法的性能会比其他对比的回归算法好。

KMSBR算法在大部分数据集上的MSE和MAE值优于其他回归模型的主要原因之一可能是在其模型中使用了成对的左回归系数和右回归系数。由于左回归系数矩阵和右回归

系数矩阵具有稀疏性,因此该模型可以分别对样本和特征进行选择,从而提高预测的准确性。另外,与BSOLSR不同的是,KMSBR模型中使用了多对左右回归系数向量,这样可以进一步减少拟合误差,有利于预测精度的提高。总之,由实验结果可知,KMSBR模型的预测精度优于其他现有的回归算法。

为进一步说明算法的有效性,接下来比较相关算法对特征和样本的选择能力,即算法的特征缩减率和样本缩减率,如表5和表6所列。由于SVR和LSSVR仅仅是对样本进行稀疏化,因此在FRR值的统计结果中(见表5)没有列出它们的相关特征缩减率。同理,在IRR值的统计结果中(见表6)也没有列出LASSO和RFS的样本缩减率。

由表5和表6可知,KMSBR回归模型的FRR值和IRR值相对于其他模型虽然在多数测试集上不是最优的,但它至少不是最差的,因此其具有一定的特征选择和样本选择能力。在KMSBR回归模型中,由于利用了左右回归矩阵对样本的特征核矩阵进行了回归,并根据左回归矩阵的行稀疏性实现了样本的选择,根据右回归矩阵的行稀疏性实现了特征的选择,因此其预测精度高于其他方法。注意到在左右回归矩阵中的非稀疏行中可能包含绝对值较大的元素,这些行的元素绝对值之和将会影响特征和样本的选择,从而在某些测试集上其特征和样本的缩减率小于其他回归算法。

#### 4.6 收敛性

为说明算法的收敛性,考虑KMSBR的目标函数值随迭代次数的变化趋势,这里仅给出该算法目标函数在bodyfat,kidney数据集上的变化情况,而在其他数据集上的情况与之类似。其中,各数据集的训练样本数为176,53,特征数为14,6,最大迭代次数为50。结果如图1所示,本文算法的目标函数随着迭代次数的增加逐渐减小,且当迭代次数增加到一定值时,目标函数值趋于稳定,这说明本文提出的KMSBR算法是具有收敛性的。

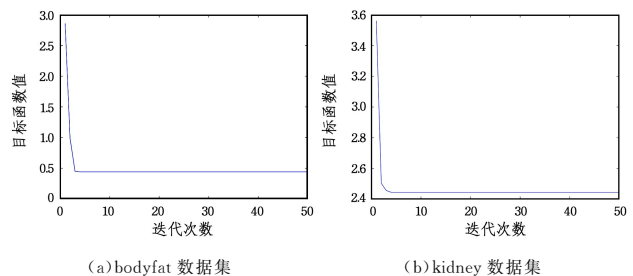


图1 部分数据集的目标函数收敛图

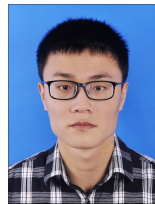
Fig. 1 Convergence figure of objective function of partial data sets

**结束语** 本文提出了一种基于样本特征核矩阵的稀疏双线性回归模型(KMSBR)。该模型在BSOLSR的基础上,结合稀疏矩阵回归SMR方法,使用了多对回归系数向量,还使用 $L_{2,1}$ 范数作为稀疏正则化项,得到了稀疏的多对回归系数向量。利用该模型及得到的算法可以同时选择相对重要的样本和特征。在11个真实的数据集上的实验结果表明,KMSBR的预测性能普遍优于现有的SMR, SVR, LSSVR, BSOLSR, LASSO和RFS方法。

虽然KMSBR在高维大样本数据集中具有良好的预测性能,但如何进一步提高其在低维小样本数据集中的预测性能仍需继续研究。KMSBR在对相对重要的样本和特征的选择能力方面仍然有所不足,还需要进一步对特征和样本的缩减率进行研究,从而改善算法的运行效率。

### 参 考 文 献

- [1] CHATTERJEE S, HADI A S. Regression analysis by example (5th edition)[M]. John Wiley & Sons Inc, 2012; 31-34.
- [2] DAVID J O. Linear regression [M]. Springer, 2017; 1-15.
- [3] RHINEHART R R. Nonlinear regression modeling for engineering applications: modeling, model validation, and enabling design of experiments [M]. John Wiley & Sons, 2016; 12-30.
- [4] BASAK D, SRIMANTA P, PATRANBIS D C. Support vector regression[J]. Neural Information Processing Letters & Reviews, 2007, 11(10): 203-224.
- [5] CHU W, KEERTHI S S, CHONG J O. Bayesian support vector regression using a unified loss function[J]. IEEE Transactions on Neural Networks, 2004, 15(1): 29-44.
- [6] ZHAO Y, SUN J. Robust support vector regression in the primal[J]. Neural Networks, 2008, 21(10): 1548-1555.
- [7] SUYKENS J A K, GESTEL V T, BRABANTER J D, et al. Least squares support vector machines [M]. World Scientific, 2002; 10-12.
- [8] CHEN C F. A robust weighted least squares support vector regression based on least trimmed squares[J]. Neurocomputing, 2015, 168: 941-946.
- [9] TIAN Y, GAO G, YUE J, et al. Two-phase multi-kernel LP-SVR for feature sparsification and forecasting[J]. Neurocomputing, 2016, 214: 594-606.
- [10] ZHANG Z W, HE J, GAO G X, et al. Bi-sparse optimization-based least squares regression [J]. Applied Soft Computing, 2019, 77: 300-315.
- [11] ZHANG Z, GAO G, YAO T, et al. An interpretable regression approach based on bi-sparse optimization [J]. Applied Intelligence, 2020, 50(11): 4117-4112.
- [12] STORK D G, HART P E, PATTERN R O. Classification [M]. Hoboken, New York, USA, 2001; 307-321.
- [13] LIU J, JI S W, YE J P. Multi-Task Feature Learning Via Efficient  $l_{2,1}$ -Norm Minimization [C]// Conference on Uncertainty in Artificial Intelligence. 2009; 339-348.
- [14] NIE F, HUANG H, CAI X, et al. Efficient and robust feature selection via joint  $l_{2,1}$ -norms minimization [C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2010; 1813-1821.
- [15] CAI X, NIE F, HUANG H. Exact Top-k Feature Selection via  $l_{2,0}$ -Norm Constraint [C]// International Joint Conference on Artificial Intelligence. AAAI Press, 2013; 1240-1246.
- [16] LIU X, WANG L, ZHANG J, et al. Global and Local Structure Preservation for Feature Selection [J]. IEEE Transactions on Neural Networks & Learning Systems, 2014, 25(6): 1083-1095.
- [17] HOU C, JIAO Y, NIE F, et al. 2D Feature Selection by Sparse Matrix Regression [J]. IEEE Transactions on Image Processing, 2017, 26(9): 4255-4268.
- [18] GABRIEL K R. Generalised bilinear regression [J]. Biometrika, 1998, 85: 689-700.
- [19] TIBSHIRANI R. Regression shrinkage and selection via the Lasso: a retrospective [J]. Journal of the Royal Statistical Society; Series B, 1994, 73(3): 267-288.
- [20] Репозиторий UCI. UCI Machine Learning Repository; Data Sets [EB/OL]. [2018-09-24]. <http://archive.ics.uci.edu/ml>.



**SHAO Zheng-yi**, born in 1996, M.S candidate. His main research interests include pattern recognition, linear regression, etc.