

引用格式: 刘志刚, 刘森泽. 存在标签噪音的数据中基于低秩矩阵分解的多输出回归[J]. 微电子学与计算机, 2022, 39(3): 50-57. [LIU Z G, LIU S Z. Multi-Target Regression via Low Rank Matrix Factorization for Data with Label Noise[J]. Microelectronics & Computer, 2022, 39(3): 50-57.] DOI: 10. 19304/J. ISSN1000-7180. 2021. 0787

存在标签噪音的数据中基于低秩矩阵分解的多输出回归

刘志刚¹, 刘森泽²

(1 无锡职业技术学院 控制技术学院, 江苏 无锡 214121;

2 南京大学 计算机科学与技术系, 江苏 南京 210023)

摘 要: 多输出回归是指针对一组输入变量来估计其对应的多个连续属性值, 其在数据挖掘领域有着广泛的应用。当前关于多输出回归任务的研究都是基于标签值准确的假设下实现的, 然而在实际情况中, 数据集的部分标签可能并不准确, 即部分标签存在一定的噪声。在这种情况下, 传统多输出回归方法性能较差。为了解决上述情况下的多输出回归问题, 利用大数据中数据样本大的特点来提炼各个标签间的相关性, 从而利用标签间的相关矩阵重构标签。由于多输出问题中的标签个数通常较多, 因此可以一定程度上稀释掉部分标签的噪声干扰。此外, 利用低秩矩阵分解对上述思路建立数学优化问题, 并在此基础上引入核技巧以提升模型的非线性拟合能力。最后, 采用非凸近似的手段求解该优化问题, 从而保证了多输出回归模型的预测性能。实验 18 个数据集上同现有的 6 种多输出回归方法进行了比较, 提出的方法在样本量较大的场景下性能优势较为明显。

关键词: 多输出回归; 标签噪声; 低秩学习; 矩阵分解; 非凸优化

中图分类号: TP391. 4

文献标识码: A

文章编号: 1000-7180(2022)03-0050-08

Multi-Target regression via low rank matrix factorization for data with label noise

LIU Zhigang¹, LIU Senze²

(1 School of Control Technology, Wuxi Institute of Technology, Wuxi 214121, China;

2 Department of Computer Science and Technology, Nanjing University, Nanjing 210023, China)

Abstract: Multiple-target regression refers to a set of input variables to estimate its corresponding multiple continuous attribute values, which has a wide range of applications in the field of data mining. The current research on multi-target regression tasks is based on the assumption that the label value is accurate. However, in actual situations, some labels of the data set may not be accurate, that is, some labels have certain noise. In this case, the traditional multi-target regression methods usually cannot achieve good results. In order to solve the multi-target regression problem in the above situation, the characteristics of large data samples in big data are used to refine the correlation between labels, and then the correlation matrix between labels is used to reconstruct the labels. Since the number of labels in the multi-target problem is usually large, the noise interference of some labels can be diluted to a certain extent. In addition, low-rank matrix decomposition is used to establish a mathematical optimization problem for the above ideas, and on this basis, kernel techniques are introduced to improve the nonlinear fitting ability of the model. Finally, a non-convex approximation method is used to solve the optimization problem, thereby ensuring the prediction performance of the multi-target regression model. The experiment is compared with 6 existing multi-target regression models on 18 datasets. The method proposed in this paper has obvious performance advantages in

收稿日期: 2021-05-28; 修回日期: 2021-07-14

基金项目: 南京大学创新人才支持计划(20201856); 江苏高校“青蓝工程”项目(20193 号, 201812 号)

scenarios with large sample sizes.

Key words: multiple-target regression; label noise; low-rank learning; matrix factorization; non-convex optimization

1 引言

回归分析是一种机器学习方法,它用来建立因变量和自变量的复杂关系.目前,单输出回归是最为常见的一类回归模型.单输出回归关注的是通过创建一个模型来预测未知的某一数据.然而在现实中,通常存在同时预测多个标签的情况.因此,多输出回归模型在传统单输出回归模型的基础上针对多个属性标签的预测需求做了扩展,使得它能对同一输入数据预测多个连续的标签^[1].目前,多输出回归方法已经广泛应用于包括情绪估计^[2]、疾病预测^[3]、计算机视觉^[4]在内的诸多领域.

与传统单输出回归任务相比,多输出回归问题主要面临以下两大挑战:(1)如何对特征数据与多个输出标签之间的映射关系进行建模^[5];(2)如何利用标签之间的相关性^[6]来建立回归模型.为了解决上述多输出回归中的挑战,现有工作主要分为两类方法.其一是为输出空间的每个标签训练一个独立的回归模型,即把多输出回归问题转化成单输出回归问题.该类方法具有简单、直观、运行效率高的特点^[7-8].该类方法主要包括单标签回归法、组合回归链^[1]、随机标签组合^[7]等.单标签方法对输入特征数据和每一个标签值构建多个回归模型.组合回归链通过构建多个回归链模型并且利用投票机制决定最终结果.随机标签组合法是基于各个标签之间的随机线性组合来构建新的标签,但其复杂度略高.

另一类方法则是通过利用多个标签之间的相互关系来建立多输出回归模型.由于各个输出标签之间都有一定的相互关联性,因此在引入相关性后能够提升模型的准确性^[9-10].该类常见的方法主要包括多输出线性回归模型^[9]、多输出核学习^[10]、多输出支持向量回归^[11]、自适应K聚类随机森林^[12]、多任务特征学习^[13].具体地,多输出线性回归模型利用潜在模型参数的协方差矩阵和标签数据的条件协方差矩阵分析标签间的相关性,并将其应用在线性回归中.多输出核模型通过学习多个标签的输出核来确定标签之间的相关性.多输出支持向量回归将传统支持向量回归扩展到多维情况下学习标签间的相关性.多任务特征学习基于新的正则化器的1范数控制所有任务共有的学习特征数量.

但是目前多数关于多输出回归的研究都是基于

假设标签值是准确的前提下实现的.然而在现实中,由于数据采集误差的问题,数据集的部分标签可能存在不准确的情况,即存在一定的噪声.在这种情况下,现有的多输出回归方法通常不能取得较好的效果.

针对这一情况,本文提出一种利用低秩矩阵分解的方法来重构标签矩阵,并以此来建立多输出回归模型.由于多输出问题中的标签个数通常较多,因此可以一定程度上稀释掉部分标签的噪声干扰.此外,我们采用非凸近似的方法求解该矩阵分解问题,从而保证了多输出回归模型的预测性能.最后,实验在18个数据集上同现有的6种多输出回归方法进行了比较,本文提出的方法具有显著性的提升.

2 多输出回归的基本模型

首先回归算法需要从训练集 $\{(x_i, y_i) | 1 \leq i \leq N\}$ 中学习得到函数 $f: X \rightarrow y$.而在多输出回归问题中, $\mathbf{X} \in R^{d \times n}$ 为模型的输入样本,其中 d 表示特征空间的维度, $\mathbf{Y} \in R^{l \times n}$ 则表示 l 维的输出空间,每个样本 $(\mathbf{x}_i, \mathbf{y}_i)$ 由一个 d 维特征向量 $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ 和一个 l 维输出向量 $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{in}]$ 表示,从而需要学习函数 $h: \mathbf{X} \rightarrow \mathbf{Y}$.

具体地,基本的多输出回归模型可以写为:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} + \mathbf{B} \quad (1)$$

式中, $\mathbf{Y} \in R^{l \times n}$ 是被观察到的标签矩阵,其中 l 是标签数, n 是样本数, $\mathbf{W} \in R^{d \times l}$ 为模型权重,其中 d 为每一个特征的维度, $\mathbf{X} \in R^{d \times n}$ 为模型的输入样本, $\mathbf{B}$ 为偏置项.而在给定训练数据的情况下,训练模型权重的优化问题可以表示为:

$$\mathbf{W}^*, \mathbf{B}^* = \underset{\mathbf{W}, \mathbf{B}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{W}\mathbf{X} - \mathbf{B}\|_F^2 + \lambda_0 \|\mathbf{W}\|_F^2 \quad (2)$$

式中, $\|\mathbf{W}\|_F$ 为模型权重 $\mathbf{W}$ 的 Frobenius 范数,用于防止模型过拟合,是一个超参数, λ_0 用于控制正则化程度.为了表示方便,上述优化问题可以写为:

$$\mathbf{W}^* = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{W}\mathbf{X}\|_F^2 + \lambda_0 \|\mathbf{W}\|_F^2 \quad (3)$$

式中,样本数据矩阵 $\mathbf{X}$ 为原矩阵 $\mathbf{X}$ 的最右侧增加一列全为1的列.

3 存在标签噪音的多输出回归问题

3.1 基于矩阵低秩分解的模型优化

由于在多输出学习问题中需要同时预测多个标

签,直接学习从特征空间到标签空间的映射的难度较大.因此,通过借鉴低秩矩阵分解的思想,考虑学习从特征空间到维度较低的隐性标签空间的映射关系.具体地,将标签矩阵分解为 $\mathbf{S}$ 和 $\mathbf{V}$ 两个低秩矩阵,其中矩阵 $\mathbf{V}$ 代表隐性标签矩阵,矩阵 $\mathbf{S}$ 代表隐性标签与原始标签之间的关联矩阵.通过最小化重构误差 $\|\mathbf{Y} - \mathbf{SV}\|_F^2$,可以得到上述矩阵 $\mathbf{S}$ 和 $\mathbf{V}$.因此,模型的权重 $\mathbf{W}$ 也变为对数据样本和隐性标签空间的映射.通过最小化重构误差 $\|\mathbf{V} - \mathbf{WX}\|_F^2$,可以得到新的权重矩阵 $\mathbf{W}$.综上,模型训练的优化问题可以重写为:

$$\begin{aligned} \mathbf{W}^*, \mathbf{S}^*, \mathbf{V}^* = \underset{\mathbf{W}, \mathbf{S}, \mathbf{V}}{\operatorname{argmin}} \quad & \|\mathbf{Y} - \mathbf{SV}\|_F^2 + \\ \lambda_0 \|\mathbf{V} - \mathbf{WX}\|_F^2 + \lambda_1 \|\mathbf{S}\|_F^2 + \lambda_2 \|\mathbf{V}\|_F^2 + \lambda_3 \|\mathbf{W}\|_F^2 \end{aligned} \quad (4)$$

为了简化优化问题,式(4)可以进一步简写为:

$$\begin{aligned} \mathbf{W}^*, \mathbf{S}^* = \underset{\mathbf{W}, \mathbf{S}}{\operatorname{argmin}} \quad & \|\mathbf{Y} - \mathbf{SWX}\|_F^2 + \\ \lambda_0 \|\mathbf{W}\|_F^2 + \lambda_1 \|\mathbf{S}\|_F^2 \end{aligned} \quad (5)$$

3.2 考虑相关性的模型

当前关于多输出回归的研究都是基于标签值是准确的假设下实现的.然而在现实中,数据集的部分标签可能并不准确,即部分标签存在一定的噪声.这种情况下,传统多输出回归问题通常不能取得较好的效果.为解决上述部分标签不准确的问题,利用大数据中数据样本较大的特点来提炼各个标签间的相关性,然后利用标签间的相关性来重构各个标签值,从而达到稀释噪声的作用.这里,用矩阵 $\mathbf{C} \in R^{D \times D}$ 来表征各个标签间的相关性.具体地,重构的标签矩阵可以表示为:

$$\hat{\mathbf{Y}} = \mathbf{CY} \quad (6)$$

并将其带入上述的优化问题中,可得:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{W}, \mathbf{S}} \quad & \|\mathbf{CY} - \mathbf{SZ}\|_F^2 + \lambda_0 \|\mathbf{W}\|_F^2 + \lambda_1 \|\mathbf{CY} - \mathbf{Y}\|_F^2 + \\ \lambda_2 \|\mathbf{C}\|_F^2 + \lambda_3 \|\mathbf{S}\|_F^2 + \lambda_0 \operatorname{rank}(\mathbf{C}) \end{aligned} \quad (7)$$

式中, $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n] \in R^{D \times n}$, $\mathbf{z}_i = \mathbf{W}x_i$ 是潜在空间的潜在变量,其中 $\operatorname{rank}(\mathbf{C})$ 表示矩阵 $\mathbf{C}$ 的秩,其目的是为了通过最小化该矩阵的秩来最大程度地挖掘和利用各个标签之间的相似性以便构建回归模型.但是由于秩函数具有非连续以及非凸的特性,涉及秩函数的最小化问题都需要面对 NP-Hard 的挑战.为此,现有工作主要利用核范数来近似秩函数从而将秩函数转换成一个容易处理的凸函数.但是,通过核范数的定义可知,它只是把矩阵所有非零奇异值加和,因此数值较大的奇异值可能对秩的估计有着较大的影响,从而使其秩近似存在较大的误差,这

为核函数的近似结果造成估计过大的问题.对于此,我们利用伽马范数而非核函数对秩函数直接进行非凸近似,而不需要将其转化为凸函数.不同于现有的基于核范数的方法,伽马范数较好地限制了较大奇异值的作用,从而使得伽马范数能够比核范数更准确地估计矩阵的秩.具体地,矩阵 $\mathbf{C}$ 的伽马范数可以写为:

$$\|\mathbf{C}\|_\gamma \triangleq \sum_i \frac{(1+\gamma)\sigma_i}{\gamma + \sigma_i}, \quad \gamma > 0 \quad (8)$$

式中, σ_i 为矩阵 $\mathbf{C}$ 的奇异值, γ 为伽马范数的参数.其导数表示为:

$$\frac{\partial \|\mathbf{C}\|_\gamma}{\partial \mathbf{C}} = \mathbf{A}_C \operatorname{Diag}([l_1, \dots, l_i, \dots, l_L]) \mathbf{B}_C^T \quad (9)$$

式中, $\mathbf{A}_C$ 和 $\mathbf{B}_C$ 分别表示矩阵 $\mathbf{C}$ 的左奇异值矩阵和右奇异值矩阵, $l_i = \frac{(1+\gamma)\gamma}{(\gamma + \sigma_i)^2}$, $\operatorname{Diag}(q)$ 为将向量 q 中的元素表示为一个对角矩阵.

因此,上述的优化问题的非凸近似可以写为:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{W}, \mathbf{S}} \quad & \|\mathbf{CY} - \mathbf{SZ}\|_F^2 + \lambda_0 \|\mathbf{W}\|_F^2 + \\ & + \lambda_1 \|\mathbf{CY} - \mathbf{Y}\|_F^2 + \lambda_2 \|\mathbf{C}\|_F^2 + \\ & + \lambda_3 \|\mathbf{S}\|_F^2 + \lambda_4 \|\mathbf{C}\|_\gamma \end{aligned} \quad (10)$$

3.3 基于核技巧的非线性模型

通过前述优化问题可知模型的权重 $\mathbf{W}$ 本质上是对输入数据样本和隐性标签空间的线性映射.但对于一个复杂的多输出回归问题,线性映射难以应对复杂的映射关系.因此,借助于核技巧对原线性映射模型进行非线性拓展与改造.具体地,核技巧的原理是首先是通过映射 φ 将数据样本映射到一个特征空间,然后在该特征空间中完成基于权重 w 的线性回归.映射 φ 作用在高维、甚至是无限高维的再生核希尔伯特空间中.具体地,基于核技巧的非线性映射模型可以写为:

$$h(x) = w^T \varphi(x) \quad (11)$$

而根据表示定理可知,该学习模型的最优解可以表示为核函数的线性组合:

$$h^*(x) = \sum_{i=1}^n a_i K(x, x_i) \quad (12)$$

式中, $K(x, x_i) = \varphi^T(x) \varphi(x_i)$. 经过简单的矩阵推导,原权重 w 可以表示为 $w = \sum_{i=1}^n a_i \varphi^T(x_i)$, 因此权重矩阵 $\mathbf{W}$ 可以表示为:

$$\mathbf{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{1i} \varphi^T(x_i) \\ \sum_{i=1}^n a_{2i} \varphi^T(x_i) \\ \vdots \\ \sum_{i=1}^n a_{ni} \varphi^T(x_i) \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} \varphi^T(x_1) \\ \varphi^T(x_2) \\ \vdots \\ \varphi^T(x_i) \end{bmatrix} \\ = \mathbf{A}\Phi^T(X) \quad (13)$$

式中, $\mathbf{A} \in R^{l \times n}$ 为新的学习模型权重矩阵, $\Phi(X) = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)]$. 因此, 前述优化问题可以进一步变为:

$$\min_{C, A, S} tr((CY - SA\Phi^T(X)\Phi(X))^T \\ \times (CY - SA\Phi^T(X)\Phi(X))) \\ + \lambda_0 tr((\mathbf{A}\Phi^T(X))^T (\mathbf{A}\Phi^T(X))) \quad (14) \\ + \lambda_1 \|CY - Y\|_F^2 + \lambda_2 \|C\|_F^2 \\ + \lambda_3 \|S\|_F^2 + \lambda_4 \|C\|_Y$$

根据该矩阵定义 $K \triangleq \Phi(X)^T \Phi(X)$, 可以进一步得到:

$$\min_{C, A, S} tr((CY - SAK)^T (CY - SAK)) \\ + \lambda_0 tr(\mathbf{A}K\mathbf{A}^T) + \lambda_1 \|CY - Y\|_F^2 \quad (15) \\ + \lambda_2 \|C\|_F^2 + \lambda_3 \|S\|_F^2 + \lambda_4 \|C\|_Y$$

3.4 优化问题的求解

该问题可以通过交替迭代更新来解决. 交替迭代更新是一种求解具有可分离变量的优化问题的重要方法, 由于处理速度快, 收敛性能好, 它在统计学习、机器学习等领域有着广泛应用. 在每一次的迭代中, 使用梯度下降法更新变量 A, S, C 中的一个, 与此同时固定其他两个变量. 如此, 整个优化问题将转换为多个简单的子问题.

3.4.1 固定 A 和 S , 更新 C

优化目标函数为:

$$\min_C tr((CY - SAK)^T (CY - SAK)) \\ + \lambda_1 \|CY - Y\|_F^2 + \lambda_2 \|C\|_F^2 + \lambda_4 \|C\|_Y \quad (16)$$

对该目标函数求关于 C 的梯度得:

$$\nabla_C = 2(CY - SAK)Y^T + 2\lambda_1 (CY - Y)Y^T \\ + 2\lambda_2 C + \lambda_4 A_C \text{Diag}([l_1, l_2, \dots, l_L])B_C^T \quad (17)$$

然后, 更新 $C^{+1} = C - \eta_C \nabla_C$ 直到收敛.

式中, η 是步长, 也称学习速率, 可以通过线搜索算法自适应地选择.

3.4.2 固定 S 和 C , 更新 A

优化目标函数为:

$$\min_A tr((CY - SAK)^T (CY - SAK)) \\ + \lambda_1 \|CY - Y\|_F^2 + \lambda_2 \|C\|_F^2 + \lambda_4 \|C\|_Y \quad (18)$$

对该目标函数求关于 A 的梯度得:

$$\nabla_A = -2S^T(CY - SAK)K^T \\ + \lambda_0 (KA^T + K^T A^T) \quad (19)$$

然后更新 $A^{+1} = A - \eta_A \nabla_A$ 直到收敛.

3.4.3 固定 S 和 C , 更新 A

优化目标函数为:

$$\min_S tr((CY - SAK)^T (CY - SAK)) \\ + \lambda_3 \|S\|_F^2 \quad (20)$$

对该目标函数求关于 S 的梯度得:

$$\nabla_S = -2(CY - SAK)(AK)^T + 2\lambda_3 S \quad (21)$$

然后更新 $S^{+1} = S - \eta_S \nabla_S$ 直到收敛.

如此迭代以上步骤直到收敛即可得到优化问题的解.

算法 1 基于低秩矩阵分解的多输出回归模型的训练

Input: 数据矩阵 X , 标签矩阵 Y

Output: A, S, C

初始化 C, W

while (A, S, C 均未收敛) do

固定 A 和 S , 通过(17)更新 C 直到收敛;

固定 S 和 C , 通过(19)更新 A 直到收敛;

固定 C 和 A , 通过(21)更新 S 直到收敛;

end

输出 A, S, C ;

对于算法 1 中的 C 和 W 可以使用如下方式的热启动来得到算法 1 中的初始值:

$$\min_{C, W, S} \|CY - SZ\|_F^2 + \lambda_0 \|W\|_F^2 \\ + \lambda_1 \|CY - Y\|_F^2 + \lambda_2 \|C\|_F^2 + \lambda_3 \|S\|_F^2 \quad (22)$$

在得到 C, S 和 A 的最优解后, 便可以通过以下公式预测输入数据的多个标签:

$$\hat{Y} = SAK \quad (23)$$

4 实验

在这一小节, 将开展充分的实验用以评估本文所提方法的有效性并和现有算法进行性能比较. 实验数据集来自于 Mulan 多目标学习库^[14]中的 18 个数据集, 其中包含了图像类和文本类数据. 各数据集

的具体参数请见表 1.

表 1 数据集的统计
Tab. 1 Statistics of the dataset

数据集	样本数	输入样本维度	标签数	交叉验证层数
EDM	154	16	2	10
SF1	323	10	3	10
SF2	1 066	10	3	10
JURA	359	15	3	10
WQ	1 066	16	14	10
ENB	768	8	2	10
SLUMP	103	7	3	10
ANDRO	49	30	6	10
OSCALES	639	413	12	10
SCPF	1 137	23	3	10
ATP1d	337	411	6	10
ATP7d	296	411	6	10
OES97	334	263	16	10
OES10	403	298	16	10
RF1	9 125	64	8	5
RF2	9 125	576	8	5
SCM1d	9 803	280	16	2
SCM20d	8 966	61	16	2

另外,与现有具有代表性多输出回归模型进行了比较,其中单任务学习(STL)^[15]包括多维支持向量回归(mSVR)^[11],自适应 K 聚类随机森林(AKRF)^[12],而多任务学习包括多特征学习(MT-FL)^[13]、多输出内核岭回归(mKRR)^[16]以及基于旋转森林的多目标学习(RFMT)^[17].此外,为验证本文所提的基于伽马范数的秩函数近似方法对于多输出回归模型的性能提升效果,将所提方法与基于传统核范数的多输出回归模型的性能进行了对比.为区分这两种方法,将所提算法称为 LRMF,而将基于传统核范数的方法称为 LRMFN.

4.1 实验设置

在训练集的标签矩阵中,均匀随机地选择 5% 的元素添加高斯白噪声.该噪声是一个高斯随机变量,他的均值为 0,方差为真实标签值方差的 p 倍.为了直接使用最新的算法进行基准测试,使用相对均方根误差(Relative Root Mean Squared Error, RRMSE)^[17]来衡量模型的性能.

$$\text{RRMSE} = \sqrt{\frac{\sum_{\langle x_i, y_i \rangle \in D_{\text{test}}} (\hat{y}_i - y_i)^2}{\sum_{\langle x_i, y_i \rangle \in D_{\text{test}}} (\bar{Y} - y_i)^2}} \quad (24)$$

式中, (x_i, y_i) 是第 i 个样本 x_i ,而其标签的真实值是 y_i , $\hat{y}_i$ 是 y_i 的预测值, $\bar{Y}$ 是整个项目的标签的平均值.将测试集 D_{test} 中的所有标签变量的 RRMSE 视

作一个单一的度量标准.较低的 RRMSE 值意味着模型有着更好的性能.通过交叉验证从训练集 D_{test} 上的 $10^{[-5:1:3]}$ 的搜索网中选择参数 $\lambda_0, \lambda_1, \lambda_2$ 和 λ_3 , 该方法是固定三个参数的同时调整另一个参数,除此之外也可以通过自适应技术进行参数选择.我们使用径向基函数核为所提方法的核范数.

4.2 实验结果

表 2 和表 3 分别给出了所提算法与现有算法在 $p=0.1$ 和 $p=0.5$ 时的对比结果.通过对比表 2 和表 3 可以发现,随着标签噪声的增大,各算法的 RRMSE 性能均出现下降,这说明标签噪声对多输出回归问题造成了较大的挑战.此外,还可以发现,所提算法在大部分数据集上的表现都优于现有算法,这说明所提算法能够较现有算法更好地在标签存在噪声的情况下完成多输出回归的任务.具体地,在数据集 EDM、SLUMP 和 ANDRO 中,所提算法的 RRMSE 性能有限.这是因为这三个数据集的样本数过少,所提算法无法准确地挖掘提炼各标签之间的关联关系.此外,还可以发现,所提算法在样本数较大的数据集上的表现明显好于在样本数较小的数据集.比如,在数据集 RF1 和 RF2 中,所提算法的性能优势明显高于在 SCPF 和 WQ 数据集上的.这是因为在样本数较大的数据集中,所提算法能够更准确地通过标签间的相关性重构标签,进而稀释掉噪声对标签的影响.另外,还可以发现,基于伽马范数的所提方法比基于核范数的方法性能更好.这说明,与核范数相比,基于非凸近似思想的伽马范数能够更准确地估计一个矩阵的秩,从而使得所提方法能够最大程度地挖掘和利用各个标签之间的相似性以便构建回归模型.

本文所提出的 MMR 算法在绝大部分数据集上的性能表现优于最新的几种算法,其已经被证实可以用于多种多输出回归任务.相对于单任务学习(STL)^[15]和多输出内核岭回归(mKRR)^[16],本文提出的 MMR 算法有了很大的改进,在所有 18 个数据集上均有显著优势,这证明了它在揭示模型中样本间关联的有效性. STL 和 mKRR 被视为独立预测多个标签而无需探索样本间相关性的基本方法,而 MMR (包括 MTSC, ERC, RLC, mSVR, AKRF, MROTS 和 OKL)对上述独立预测的方法进行了较大的改进,这些 MMR 方法都以某种方式探索了标签间的相关性.上述的比较结果证明了本文提出的 LRMF 在通过引入结构矩阵通过健壮性低秩学习对标签间相关性进行建模方面的优势.

表 2 $p=0.1$ 时的性能对比

Tab. 2 Performance comparison when $p=0.1$

	STL	mSVR	AKRF	MTFL	mKKR	RFMT	LRMFN	LRMFG
EDM	75.7	74.6	75.9	86.3	84.5	74.5	76.5	75.3
SF1	114.8	103.7	113.1	112.9	111.9	105.1	104.8	103.2
SF2	115.7	105.5	114.4	113.8	117.5	104.6	105.8	103.5
JURA	59.7	62.1	62.6	61.5	64.2	59.2	59.9	58.8
WQ	91.5	90.6	92.9	97.4	93.4	91.1	90.8	89.6
ENB	12.5	23.3	25.1	33.6	27.1	12.3	13.6	12.1
SLUMP	69.5	71.9	73.5	68.8	79.5	70.2	72.8	71.5
ANDRO	61.8	63.6	63.8	81.4	64.8	62.3	63.7	62.2
OSCALES	75.5	78.6	78.9	169.5	80.5	74.2	75.1	73.5
SCPF	84.9	83.6	83.8	90.6	86.3	84.1	83.8	82.6
ATP1d	38.2	38.8	41.7	42.3	38.8	38.2	37.9	36.7
ATP7d	53.3	48.4	54.2	55.7	49.5	47.2	48.1	46.9
OES97	53.7	56.4	59.4	82.6	59.8	53.6	53.5	52.6
OES10	43.5	48.5	45.4	53.9	49.6	42.5	42.9	41.7
RF1	12.2	12.9	12.4	98.3	18.3	11.2	10.7	9.5
RF2	13.4	16.6	16.3	111.6	14.5	11.9	11.8	10.6
SCM1d	37.8	38.7	37.7	45.3	39.5	35.9	36.2	34.1
SCM20d	49.9	51.8	67.9	66.8	51.6	46.7	46.9	45.5

表 3 $p=0.5$ 时的性能对比

Tab. 3 Performance comparison when $p=0.5$

	STL	mSVR	AKRF	MTFL	mKKR	RFMT	LRMFN	LRMFG
EDM	76.5	75.8	77.3	88.4	86.7	77.8	77.3	76.3
SF1	115.7	104.9	115.3	114.7	115.6	106.1	105.7	104.2
SF2	116.8	107.3	116.8	115.6	118.9	105.2	106.4	104.7
JURA	60.6	63.2	63.5	62.7	65.3	61.2	60.8	59.6
WQ	92.7	91.8	93.7	98.5	94.3	91.5	91.7	90.5
ENB	13.7	24.2	26.4	34.8	28.3	14.3	14.9	13.3
SLUMP	71.3	73.7	74.7	69.7	80.9	70.6	73.9	72.7
ANDRO	62.7	64.4	64.7	82.6	65.7	65.2	64.5	63.3
OSCALES	76.7	79.8	79.7	170.7	81.4	75.6	76.3	74.6
SCPF	86.2	84.8	84.9	91.7	87.5	85.3	84.7	83.6
ATP1d	39.4	39.9	42.6	43.5	40.1	38.1	39.7	37.5
ATP7d	54.5	49.6	55.4	56.9	50.7	48.9	49.0	47.7
OES97	55.8	57.9	61.1	84.8	62.4	55.2	54.7	53.8
OES10	44.7	49.7	47.7	56.5	52.3	45.3	44.8	43.5
RF1	13.8	14.6	13.7	99.8	19.6	11.3	11.8	10.6
RF2	14.8	18.2	17.9	113.8	15.8	12.2	12.7	11.5
SCM1d	39.4	39.8	39.6	47.6	42.3	36.3	37.8	35.6
SCM20d	51.6	52.7	68.6	68.5	53.4	48.9	48.4	47.3

此外,为从统计推断的角度证明各算法在 RRMSE 性能上的差异,我们首先提出原假设 H_0 : 各算法的 RRMSE 性能不存显著性差异,再采用 Friedman 检验对该假设进行验证,其相关结果如表 4 所示.从表 4 中可以看出 Friedman 统计特征值为 62.23,这远大于显著性水平在 $\alpha=0.05$ 时的临界值 2.52,因此原假设 H_0 被拒绝.由此可推断各算法在 RRMSE 性能上存在显著性差异.

表 4 Friedman 统计检验结果
Tab. 4 The results of Friedman statistical test

Friedman 检验统计特征	特征值
Friedman 统计特征值	62.23
显著性水平 $\alpha=0.05$ 时的临界值	2.52
统计推断结论	拒绝

由于各算法在 RRMSE 性能上存在显著性差异,则需要利用 Bonferroni-Dunn 检验来进一步探究各算法孰优孰劣. 在 Bonferroni-Dunn 检验中,算法间 RRMSE 性能的差异可由平均排序和显著性差异值(Critical Difference, CD)来衡量. 当某两个算法的平均排序差值大于 CD 值时,说明这两个算法中平均排序较高的算法的 RRMSE 性能显著地优于平均排序较低的算法. 这里,显著性差异值 CD 由公式(25)所示:

$$CD = q_{\alpha} \sqrt{\frac{J(J+1)}{6K}}$$

(25)

式中, J 表示参与排序的算法个数, K 表示实验数据集的个数, q_{α} 表示显著性水平为 α 时的临界值. 这里,选取 $\alpha = 0.05$, 而对应的临界值 $q_{\alpha} = 2.498$, 则 $CD = 1.32$. 而各算法之间 RRMSE 性能差异的排序图在图 1 中进行了充分的展示. 平均排序最高的算法位于 CD 图的最右边,相反平均排序最低的算法位于 CD 图的最左边. 从图 1 中可以看出,本文算法的 RRMSE 性能明显高于 STL、mSVR、AKRF、MTFL 和 mKKR. 另外,所提算法与 LRMFN 算法的预测性能虽然没有显著的差异,但是却比 LRMFN 略高. 由此验证了所提基于伽马范数的低秩矩阵分解方法比现有的基于核范数的方法性能更好.

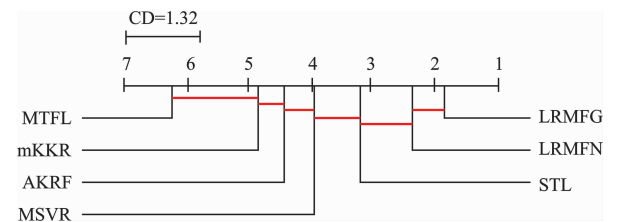


图 1 算法排序对比

Fig. 1 The comparison of algorithm rank

5 结束语

本文针对多输出回归中的标签噪声问题,提出了一种基于低秩矩阵分解的多输出回归方法. 该方法通过重构观测到的标签矩阵从而稀释掉部分标签噪声. 然后,利用低秩矩阵分解对该思路建立数学优化问题. 为保证该优化问题的求解准确性,采用伽马

范数对该优化问题进行了非凸转换. 在 18 个数据集上的实验结果表明,本文算法相对于现有算法具有显著的性能优势. 另外,本文所提出的基于伽马范数的非凸近似方法还可以应用于其他低秩矩阵分解问题.

参考文献:

[1] SPYROMITROS-XIOUFIS E, TSOUMAKAS G, GROVES W, et al. Multi-target regression via input space expansion; treating targets as inputs[J]. Machine Learning, 2016, 104(1): 55-98. DOI:10.1007/s10994-016-5546-z.

[2] XIA R, LIU Y. A multi-task learning framework for emotion recognition using 2D continuous space[J]. IEEE Transactions on Affective Computing, 2017, 8(1): 3-14. DOI:10.1109/TAFFC.2015.2512598.

[3] WANG X, WANG F, HU J Y. A multi-task learning framework for joint disease risk prediction and comorbidity discovery[C]//2014 22nd International Conference on Pattern Recognition. Stockholm: IEEE, 2014: 220-225. DOI: 10.1109/ICPR.2014.47.

[4] RANJAN R, PATEL V M, CHELLAPPA R. HyperFace: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 41(1): 121-135. DOI: 10.1109/TPAMI.2017.2781233.

[5] ZHANG M L, ZHOU Z H. A review on multi-label learning algorithms [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837. DOI:10.1109/TKDE.2013.39.

[6] BREIMAN L, FRIEDMAN J H. Predicting multivariate responses in multiple linear regression[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1997, 59(1): 3-54. DOI:10.1111/1467-9868.00054.

[7] TSOUMAKAS G, SPYROMITROS-XIOUFIS E, VREKOU A, et al. Multi-target regression via random linear target combinations[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Nancy: Springer, 2014: 225-240. DOI:10.1007/978-3-662-44845-8_15.

[8] MELKI G, CANO A, KECMAN V, et al. Multi-target support vector regression via correlation regressor chains[J]. Information Sciences, 2017, 415-416: 53-69. DOI:10.1016/j.ins.2017.06.017.

- [9] RAI P, KUMAR A, DAUME H. Simultaneously leveraging output and task structures for multiple-output regression[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc. , 2012: 3185-3193.
- [10] ÁLVAREZ M A, ROSASCO L, LAWRENCE N D. Kernels for vector-valued functions: a review[J]. Foundations and Trends[®] in Machine Learning, 2012, 4(3): 195-266.
- [11] SÁNCHEZ-FERNÁNDEZ M, DE-PRADO-CUMPLIDO M, ARENAS-GARCÍA J, et al. SVM multi-regression for nonlinear channel estimation in multiple-input multiple-output systems[J]. IEEE Transactions on Signal Processing, 2004, 52(8): 2298-2307. DOI:10.1109/TSP.2004.831028.
- [12] HARA K, CHELLAPPA R. Growing regression forests by classification: applications to object pose estimation[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014: 552-567. DOI: 10.1007/978-3-319-10605-2_36.
- [13] EVGENIOU A, PONTIL M. Multi-task feature learning[J]. Advances in Neural Information Processing Systems, 2007(19): 41.
- [14] TSOU MAKAS G, SPYROMITROS-XIOUFIS E, VILCEK J, et al. MULAN: A Java library for multi-label learning[J]. Journal of Machine Learning Research, 2011(12): 2411-2414.
- [15] EVGENIOU T, MICCHELLI C A, PONTIL M. Learning multiple tasks with kernel methods[J]. Journal of Machine Learning Research, 2005(6): 615-637.
- [16] SHAWE-TAYLOR J, CRISTIANINI N. Kernel methods for pattern analysis[M]. Cambridge: Cambridge University Press, 2004.
- [17] RODRÍGUEZ J J, JUEZ-GIL M, LÓPEZ-NOZAL C, et al. Rotation Forest for multi-target regression[J]. International Journal of Machine Learning and Cybernetics, 2021:1-26. DOI:10.1007/s13042-021-01329-1.

作者简介:

刘志刚 男,(1982-),硕士,副教授.研究方向为机器学习.

刘森泽(通讯作者) 男,(1987-),博士,助理研究员.研究方向为机器学习. E-mail: yinnuo0423@gmail.com