



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于实例与目标相关性的多目标稀疏回归算法

何杜博, 孙胜祥

引用本文:

何杜博,孙胜祥. 基于实例与目标相关性的多目标稀疏回归算法[J]. 控制与决策, 2024, 39(5): 1478–1486.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1653>

您可能感兴趣的其他文章

Articles you may be interested in

基于复杂昂贵仿真的体系效能多目标优化

Complex and expensive simulation based multi-objective optimization to system-of-system effectiveness

控制与决策. 2021, 36(3): 589–598 <https://doi.org/10.13195/j.kzyjc.2019.0844>

输入约束不确定系统的点对点迭代学习控制与优化

Point-to-point iterative learning control and optimization for uncertain systems with constrained input

控制与决策. 2021, 36(6): 1435–1441 <https://doi.org/10.13195/j.kzyjc.2019.0908>

基于KRLS的非均匀采样非线性系统辨识

Identification of non-uniformly sampled nonlinear systems based on KRLS

控制与决策. 2021, 36(12): 3049–3055 <https://doi.org/10.13195/j.kzyjc.2020.0962>

基于改进NSGA-II算法求解多目标资源受限项目调度问题

An improved NSGA-II algorithm for multi-objective resource-constrained project scheduling problem

控制与决策. 2021, 36(3): 669–676 <https://doi.org/10.13195/j.kzyjc.2019.0906>

基于R2指标和目标空间分解的高维多目标粒子群优化算法

R2 indicator and objective space partition based many-objective particle swarm optimizer

控制与决策. 2021, 36(9): 2085–2094 <https://doi.org/10.13195/j.kzyjc.2020.0113>

基于实例与目标相关性的多目标稀疏回归算法

何杜博, 孙胜祥[†]

(海军工程大学 管理工程与装备经济系, 武汉 430033)

摘要: 针对传统多目标回归算法无法处理输入与多输出间的非线性关系,且忽视了数据点在输入与输出之间的结构信息,导致算法泛化性能受限、缺乏稳健性等问题,提出一种基于实例与目标相关性的多目标稀疏回归(multi-target sparse regression with instances and targets correlations, MTR-ITC)算法. 首先,通过嵌入潜变量空间来对复杂的输入与输出以及输出间的关联结构解耦,并利用核技巧和稀疏回归学习输入输出间的非线性关系和输出间的相关结构;然后,引入流形正则化项探索不同实例在输入与输出变量间的相关性,确保模型输出与真实结果在局部和全局结构的一致性,以提升模型泛化性能;最后,提出一种交替优化算法来对目标函数进行求解,使其能快速收敛至全局最优. 在基准测试数据集上的实验表明,所提算法在不同MTR数据集上均具有较好的测试性能.

关键词: 多目标回归; 稀疏学习; 流形学习; 交替优化算法; 核方法; 实例相关性

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1653

引用格式: 何杜博, 孙胜祥. 基于实例与目标相关性的多目标稀疏回归算法[J]. 控制与决策, 2024, 39(5): 1478-1486.

Multi-target sparse regression with instance and target correlations

HE Du-bo, SUN Sheng-xiang[†]

(Department of Management Engineering and Equipment Economics, Naval University of Engineering, Wuhan 430033, China)

Abstract: To address the problem that traditional multitarget regression algorithms only focus on the linear correlation between input features and target outputs, but ignore the structural information between different instances, i.e., instance correlation and target correlation, which leads to limited performance of the algorithm, we propose a multi-target sparse regression algorithm based on instance and target correlation (MTR-ITC). First, we construct latent variable space to decouple the complex input-output and output correlation structures, and impose sparse constraints on the corresponding coefficient matrices to explicit encoding and sparse learning of inter-target correlations in the latent variable space. Then, manifold learning is introduced to explore the correlation between different instances in the input and output space to ensure that the model output is consistent with the real results in terms of local and global structure. Finally, an alternating optimization algorithm is proposed to solve the objective function optimally and converge it to the global optimum efficiently. Experiments on the benchmark test dataset show that the MTR-ITC improves the performance of the algorithm compared to existing representative algorithms, and its good convergence makes it possible to iterate and converge to the global optimum rapidly.

Keywords: multi-target regression; sparse learning; manifold learning; alternating optimization algorithm; kernel method; instance correlations

0 引言

近年来,多目标回归(multi-target regression, MTR)已经广泛应用于数据挖掘^[1-3]、计算机视觉^[4]、医疗诊断^[5]等多个具有复杂输出结构的领域研究中. MTR与迁移学习在学习框架上具有一定的相似之处,两者都可以通过协作学习和知识共享等方式来提升算法在目标域上的性能表现,这使得MTR在机器

学习领域中越来越流行^[6].

MTR面临的挑战在于对输入与输出间的关系建模,以及对不同目标间的相关性进行探索. 为了探索目标变量之间的相关性,传统MTR模型主要通过系数矩阵施加不同形式的正则化或低秩约束来挖掘目标间的相关关系,这些方法依赖于线性回归的框架或某些特定的相关性假设^[7]. 然而,在许多真实的

收稿日期: 2022-09-19; 录用日期: 2023-02-11.

基金项目: 国家社会科学基金项目(18BGL287, 18BGL285, 19CGL073).

责任编委: 牛玉刚.

[†]通讯作者. E-mail: ssx13397129808@163.com.

回归问题中,其输入与输出特征间往往呈现出较强的非线性关系,不同数据集中的目标变量间的依赖关系差异较大,这就导致了传统MTR模型难以对输入与输出特征间的关系以及不同目标变量间的关联结构进行有效学习。同时,MTR问题中的数据往往呈现出较高的维度,多个输出目标间具有一定的耦合关联。为了更好地利用其数据点之间的结构关系,通常假设数据点分布在一个低维的非线性子空间中,即流形^[8]。基于流形学习来构建相应的回归模型,可以有效利用特征空间中不同样本点蕴含的结构信息来提升算法性能,同时还可以避免异常样本的干扰和过拟合。因此,本文提出一种基于实例与目标相关性的多目标稀疏回归(multi-target regression with instances and targets correlation, MTR-ITC)算法。通过引入核技巧,使所提算法可以灵活处理非线性关系;利用潜变量空间的嵌入来对目标间的相关结构进行显式编码和稀疏学习,以便准确高效地挖掘不同目标变量间的相关结构;同时,基于流形学习来有效利用实例间的相关信息建模,从而提升所提算法的回归性能。

1 相关工作

1.1 多目标回归

由于MTR探索了不同目标变量间的相关信息,使其在预测多个具有相关性的变量时,可以获得比单目标回归方法更好的预测性能。与多标签分类(multi-label classification, MLC)方法的划分相同^[9],MTR方法一般可以划分为两类:1)问题转换(problem transformation, PT)类方法;2)算法适应(algorithm adaptation, AA)类方法。

PT类方法是将MTR问题拆解成多个单目标回归子问题,并基于传统的单输出回归模型,如 k 近邻回归、决策树、支持向量回归等,对子问题进行独立求解。鉴于MLC与MTR在学习任务上的相似性^[10],受多标签学习方法中的堆叠泛化和分类链的思想启发,文献[11]提出了单目标堆栈(stacked single-target, SST)算法和集成回归链(ensemble of regressor chain, ERC);文献[12]基于乘积量化方法将输出空间随机离散化,进而通过集成多标签学习模型来处理MTR问题;文献[13]利用决策树集成在结构性输出预测问题上的性能优势,通过树嵌入丰富原始特征集,提出了一种深度树集成(deep tree-ensemble, DTE)方法;文献[14]将泛化后的决策树作为基模型,利用不同的集成框架,对多目标输出进行随机选择(random output selection, ROS),通过集成得到一个新的集成拓展MTR模型;文献[15]通过结合子空间投影与负相关学习来训练专家,进而提出一种新的混合专家模

型来处理MTR问题。以上方法虽然较基准模型能有一定的性能提升,但对目标相关性的考虑仍然是局部的,其性能往往取决于选择的传统回归模型,且当数据集规模较大或输入特征维度较高时,其训练产生的时间成本也会迅速增加。

不同于PT类方法,AA类方法侧重于构造一个多输出结构框架,并在该框架中同时处理输入输出关系以及目标间相关性,使模型可以同时考虑以上因素。文献[16]基于 l_2 正则化设计了新的损失函数,使得单输出结构的支持向量回归模型可以拓展至多输出支持向量回归(multi-output support vector regressor, MSVR)模型,并提出了一种迭代加权最小二乘算法来对参数进行求解;文献[17]提出一种多元回归校正(calibrated multivariate regression, CMR)模型,不仅可以挖掘不同回归任务之间的相关性,同时可以依据不同任务上的噪声水平对其进行正则化校准,使其在样本较少时仍然能够实现较好的性能表现;文献[18]开发了一种可以对结构化多任务进行学习和预测的多目标回归树(multi-target regression trees, MTRT)和多输出随机森林(multi-output random forest, MORF),用于对不同类型的结构化输出进行全局和局部预测;文献[19]将原始目标进行线性组合后重构了目标空间,并构建了一种参数自适应最小绝对收缩与选择算子(self-parameterized least absolute shrinkage and selection operator, SPL)模型来对内在的目标关联和输入输出关系进行同时建模;文献[20]利用 $l_{2,1}$ 正则化和弹性网理论,构建了一个对离群点鲁棒的极限学习机(outlier-robust ELM, OR-ELM)模型,并采用乘子交替方向法来求解。

可以看出,现有多数AA类方法大多侧重于在回归矩阵施加不同的正则化约束^[21-23]。这些回归模型或基于线性回归框架,或依赖于具有强先验知识的目标相关结构假设,并通过在回归矩阵施加各种稀疏性约束来探索不同目标间的相关性^[7]。为了更好地处理输入输出间的非线性关联,各种非线性AA类方法也相继被提出。文献[24]提出一种输出内核学习(learning output kernels, OKL)方法,通过核方法来揭示输出空间中的复杂关联结构,进而更好地处理目标间的非线性关联;文献[25]利用核方法同时捕捉输入与输出间的非线性依赖关系以及输出间的复杂关联,以提高模型预测性能;文献[7, 26]通过嵌入潜变量空间形成多层结构的学习框架,将输入输出与目标输出间的相关性进行解耦。上述AA类方法的优势在于通过对输出变量的相关性进行显式编码和学习,使得模型更具解释性,但其往往并未考虑到数据点在输入空

间与输出空间中蕴含的结构信息,这使得其在学习过程中容易受离群点的干扰以及陷入过拟合.

1.2 流形正则化

流形正则化是一种有利于参数训练和性能提升的学习技巧^[27],其已经成功应用于多标签分类^[28-31]、无监督学习^[32-33]和多输出回归^[4,34-36]等多个问题中.通过流形学习挖掘不同数据点中蕴含的结构信息,可以有效提升算法在分类和回归问题上的性能表现.因此,受多标签学习中对实例相关性和标签相关性学习的启发,假设在MTR问题中,若原始输入空间中的两个实例相邻,则其在输出空间中也应当具有一定的相似性.本文拟将输入空间中不同实例的局部相关结构信息以及输出空间中的不同目标间的全局相关结构纳入至模型的学习过程中,进而保证模型预测输出与真实数据之间的结构一致性,削弱噪声和离群点对模型的不良影响并提高模型的鲁棒性,观察其能否带来相应的性能提升.

1.3 符号表示

对于矩阵 $\mathbf{A} = [a_{i,j}]_{n \times m}$, $i = 1, \dots, n$, $j = 1, \dots, m$, \mathbf{A}^T 为矩阵转置; $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{i,i}$ 为矩阵 \mathbf{A}

的迹; $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{i,j}^2}$ 为矩阵 \mathbf{A} 的 Frobenius 范数; 定义矩阵 \mathbf{A} 的 $\ell_{p,q}$ 范数为

$$\|\mathbf{A}\|_{p,q} = \left[\sum_{i=1}^n \left(\sum_{j=1}^m |a_{i,j}|^p \right)^{\frac{q}{p}} \right]^{\frac{1}{q}}, \quad (1)$$

则矩阵 \mathbf{A} 的 $\ell_{2,1}$ 范数为

$$\|\mathbf{A}\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m a_{i,j}^2}. \quad (2)$$

对一个 k 维向量 $\mathbf{a} = [a_i]$, $\|\mathbf{a}\|_2 = \sqrt{\sum_{i=1}^k a_i^2}$ 为其 ℓ_2 范数. \mathbf{I} 为单位矩阵, $\mathbf{1}$ 为元素全为1的向量.

2 方法提出

2.1 模型基础

给定训练集 $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, $\mathbf{x} \in \mathbf{R}^d$, $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_Q]^T \in \mathbf{R}^Q$ 为对应的多元目标输出,则多目标线性回归模型可以表示为 $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$. 其中: $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_Q]^T \in \mathbf{R}^{Q \times d}$ 为权重或回归矩阵, $\mathbf{w}_i \in \mathbf{R}^d$ 为目标输出 y_i 所对应的权重向量, $\mathbf{b} \in \mathbf{R}^Q$ 为偏置向量, d 和 Q 分别对应于回归模型的输入空间维度和输出空间维度. 基于传统 Ridge 线性回归拓展至多维得到如下罚目标函数:

$$\min_{\mathbf{W}} \|\mathbf{Y} - \mathbf{W}\mathbf{X} - \mathbf{B}\|_F^2 + \lambda \|\mathbf{W}\|_F^2. \quad (3)$$

其中: $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ 为模型输入, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ 为对应的输出, $\mathbf{B} = [\dots, \mathbf{b}, \dots] \in \mathbf{R}^{Q \times N}$ 为模型的偏置矩阵, λ 为非负正则化参数. 显然,仅基于权重矩阵 \mathbf{W} 往往难以同时对输入与输出以及输出间的相关结构进行准确表达. 通过嵌入潜变量空间 $\mathbf{Z} \in \mathbf{R}^K$ 将目标函数(3)重写为

$$\begin{aligned} \min_{\mathbf{V}, \mathbf{U}} \|\mathbf{Y} - \mathbf{U}\mathbf{Z}\|_F^2 + \lambda \|\mathbf{V}\|_F^2; \\ \text{s.t. } \mathbf{Z} = \mathbf{V}\mathbf{X} + \mathbf{B}. \end{aligned} \quad (4)$$

其中: $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N] \in \mathbf{R}^{K \times N}$ 为潜变量矩阵, $\mathbf{z}_i = \mathbf{V}\mathbf{x}_i + \mathbf{b} \in \mathbf{R}^K$ 为相应的 \mathbf{x}_i 在潜变量空间中映射. 显然,式(4)可以通过系数矩阵 \mathbf{V} 和 \mathbf{U} 分别表示潜变量与输入变量的关联关系以及潜变量与输出变量的关联结构,进而对输入输出以及输出间的关系解耦. 结构矩阵 \mathbf{U} 显式编码了潜变量 \mathbf{z} 与输出变量 \mathbf{y} 之间的相关结构,使得不同目标变量的内在关联性可以更好地被表达与解释. 同时,为了更好地探索目标间的内在关联性,对结构矩阵 \mathbf{U} 施加一个 $\ell_{2,1}$ 正则化约束^[37],使学习得到的结构矩阵 \mathbf{U} 具有列稀疏性,得到如下稀疏学习模型:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{Y} - \mathbf{U}\mathbf{Z}\|_F^2 + \lambda \|\mathbf{V}\|_F^2 + \beta \|\mathbf{U}^T\|_{2,1}; \\ \text{s.t. } \mathbf{Z} = \mathbf{V}\mathbf{X} + \mathbf{B}. \end{aligned} \quad (5)$$

其中 β 为调节 \mathbf{U} 列稀疏性的正则化参数. 通过对 \mathbf{U} 施加 $\ell_{2,1}$ 正则化,可以鼓励相关的目标之间共享潜变量,进而捕捉到目标之间的关联结构,得到具有列稀疏性的矩阵 \mathbf{U} . 针对结构矩阵 \mathbf{U} 施加稀疏性约束也可以显著提升模型的灵活性和可拓展性^[26].

2.2 基于核技巧的模型扩展

考虑到模型(5)中的目标函数仍然为线性回归形式,无法处理复杂的非线性关系,基于表示定理(representer theorem)将其中的目标函数进行扩展,使其可以在可再生希尔伯特空间(reproducing kernel Hilbert space, RKHS)中处理非线性的输入输出关系. 首先,基于迹形式改写式(5),得到

$$\min_{\mathbf{U}, \mathbf{V}} \text{tr}((\mathbf{Y} - \mathbf{U}(\mathbf{V}\mathbf{X} + \mathbf{B}))^T (\mathbf{Y} - \mathbf{U}(\mathbf{V}\mathbf{X} + \mathbf{B}))) + \lambda \text{tr}(\mathbf{V}^T \mathbf{V}) + \beta \text{tr}(\mathbf{U} \mathbf{D} \mathbf{U}^T). \quad (6)$$

其中 \mathbf{D} 为对角矩阵,有

$$(\mathbf{D})_{ii} = \frac{1}{2\|\mathbf{u}_i\|_2}, \quad (7)$$

\mathbf{u}_i 为矩阵 \mathbf{U} 的第 i 列. 为了避免分母为0,进一步规定

$$(\mathbf{D})_{ii} = \frac{1}{2\sqrt{\mathbf{u}_i^T \mathbf{u}_i + \zeta}}, \quad (8)$$

其中 $\zeta > 0$ 为一个很小的常数. 利用映射函数 $\phi(\mathbf{x}_i)$ 将 \mathbf{x}_i 映射至 RKHS, 则 $\phi(\cdot)$ 为特征映射函数. 依据线性表示定理(linear representer theorem), 矩阵 \mathbf{V} 可以

用特征映射函数的线性组合来表示,即

$$\mathbf{V} = \mathbf{A}\Phi(\mathbf{X})^T. \quad (9)$$

其中: $\Phi(\mathbf{X}) = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)]$, $\mathbf{A} \in \mathbf{R}^{K \times N}$. 相关的核函数 $k(\cdot, \cdot)$ 满足 $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)\phi(\mathbf{x}_j)^T$. 将式(9)代入(6)中可得

$$\min_{\mathbf{U}, \mathbf{A}} \text{tr}((\mathbf{Y} - \mathbf{UAK})^T(\mathbf{Y} - \mathbf{UAK})) + \lambda \text{tr}(\mathbf{AKA}^T) + \beta \text{tr}(\mathbf{UDU}^T), \quad (10)$$

其中 \mathbf{K} 为定义在 RKHS 上的核矩阵. 由于偏置向量 \mathbf{b} 可以通过在输入特征中额外增加一个维度来替代, 为了简便表示, 式(10)中省略偏置矩阵 \mathbf{B} . 一旦通过训练得到参数矩阵 (\mathbf{U}, \mathbf{A}) , 新的模型输入 \mathbf{x}_t 预测得到的多元输出为

$$\hat{\mathbf{f}}_t = \mathbf{UAK}_t. \quad (11)$$

其中: $\mathbf{K}_t = \Phi(\mathbf{X})^T\phi(\mathbf{x}_t)$, $\phi(\mathbf{x}_t)$ 为 \mathbf{x}_t 所对应的特征映射. 利用非线性核将原始输入特征空间中的数据点映射至 RKHS 中, 使模型拥有了学习非线性关系的能力. 通过式(10)学习得到的多层稀疏 MTR 模型结构如图1所示.

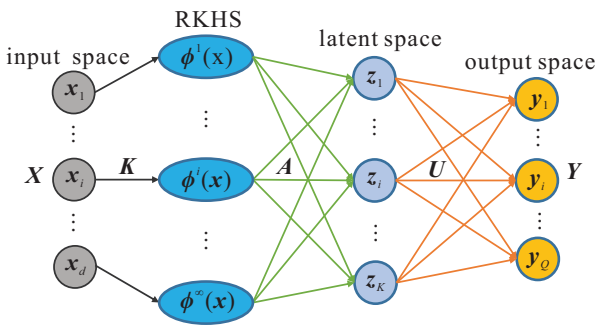


图1 多层多目标稀疏回归模型结构

2.3 实例与目标相关性的流形学习

流形正则化旨在利用不同实例之间的局部或全局相似性, 鼓励具有相似性的实例之间得到相似的预测结果. 即使对于具有复杂相关结构的多元输出回归问题, 其输出 \mathbf{y} 仍可以被合理地假定为输入 \mathbf{x} 的连续和平滑函数^[38]. 基于此假设, 利用不同实例在输入空间中的几何结构信息来保证模型的预测输出也保持相似的几何结构. 考虑训练集 $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ 的近邻关系可以通过一个半正定矩阵 $\mathbf{S}_0 \in \mathbf{R}^{N \times N}$ 表示, 其计算公式如下:

$$[\mathbf{S}_0]_{i,j} = \begin{cases} 1, & \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i); \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

其中 $\mathcal{N}_k(\mathbf{x}_j)$ 表示为 \mathbf{x}_j 的 k 近邻实例集. 基于实例相关性的流形正则化项可以定义为

$$\sum_{i=1}^N \sum_{j=1}^N [\mathbf{S}_0]_{i,j} \|\hat{\mathbf{f}}_i - \hat{\mathbf{f}}_j\|_2^2, \quad (13)$$

其中 $\hat{\mathbf{f}}_i$ 和 $\hat{\mathbf{f}}_j$ 分别为实例 i 和 j 所对应的目标预测向量. 通过最小化式(13), 使得模型充分考虑到不同实例在输入之间的结构关系来优化相应的训练结果. 记模型在训练集上的拟合结果为 \mathbf{F} , 有 $\mathbf{F} = [\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_N] \in \mathbf{R}^{Q \times N}$, 容易得到 $\mathbf{F} = \mathbf{UAK}$. 因此, 流形正则化项(13)可以改写为 $\text{tr}(\mathbf{FL}_0\mathbf{F}^T)$. 其中: $\mathbf{L}_0 = \mathbf{D}_0 - \mathbf{S}_0$ 为所对应的 N 阶图拉普拉斯矩阵, \mathbf{D}_0 为对角矩阵且有 $[\mathbf{D}_0]_{i,i} = \sum_{j=1}^N [\mathbf{S}_0]_{i,j}$.

考虑不同目标输出之间的结构信息也对 MTR 建模同样重要, 针对训练集中不同实例的目标输出设计一个流形正则化项来保证模型输出结果能与真实输出之间保持较为一致的相似结构, 进而帮助模型学习目标的相关性, 减少原始数据中噪声及离群点对模型的干扰. 同样地, 通过 N 阶矩阵 \mathbf{S}_1 来描述训练集上不同实例在输出空间中的相关性, 基于余弦相似度将 \mathbf{S}_1 定义为

$$[\mathbf{S}_1]_{i,j} = \frac{\mathbf{y}_i^T \mathbf{y}_j}{\|\mathbf{y}_i\| \|\mathbf{y}_j\|}, \quad i, j = 1, \dots, N. \quad (14)$$

其中 \mathbf{y}_i 为 \mathbf{x}_i 所对应的多元输出向量. 目标间相关性的正则化项定义为

$$\sum_{i=1}^N \sum_{j=1}^N [\mathbf{S}_1]_{i,j} \|\hat{\mathbf{f}}_i - \hat{\mathbf{f}}_j\|_2^2. \quad (15)$$

同理, 式(15)中的流形正则化项亦可以写为 $\text{tr}(\mathbf{FL}_1\mathbf{F}^T)$, 其中 $\mathbf{L}_1 = \mathbf{D}_1 - \mathbf{S}_1$, \mathbf{D}_1 为对角矩阵且 $[\mathbf{D}_1]_{i,i} = \sum_{j=1}^N [\mathbf{S}_1]_{i,j}$.

2.4 目标函数建立

通过引入式(13)和(15)中的流形正则化项, 使得模型在对输入输出以及输出间的关系建模时可以充分考虑到不同实例在输入与输出之间的相关结构信息. 因此, 最终得到的目标函数如下:

$$\min_{\mathbf{U}, \mathbf{A}} \|\mathbf{Y} - \mathbf{UAK}\|_F^2 + \lambda_1 \text{tr}(\mathbf{AKA}^T) + \lambda_2 \|\mathbf{U}^T\|_{2,1} + \text{tr}(\mathbf{FPF}^T). \quad (16)$$

其中: $\mathbf{P} = \lambda_3 \mathbf{L}_0 + \lambda_4 \mathbf{L}_1$, $\mathbf{F} = \mathbf{UAK}$, $\lambda_1 \sim \lambda_4$ 分别为非负正则化参数. 实例相关性与目标相关性信息分别通过图拉普拉斯矩阵 \mathbf{L}_0 和 \mathbf{L}_1 编码表示, 且通过调节参数 λ_3 与 λ_4 , 可以灵活地权衡实例相关性与目标相关性信息在建模中的重要程度, 进而有效防止过拟合.

3 基于交替优化算法的模型求解

通过交替优化算法可以有效地求解目标问题(16). 记目标函数(16)为 $J(\mathbf{A}, \mathbf{U})$, 通过固定一个矩阵来求另一个矩阵的方法交替求解, 使得模型快速收敛

至全局最优.

3.1 固定矩阵 A 求 U

当矩阵 A 已确定时,问题(16)可以以迹形式写成

$$\min_{U,A} \text{tr}((Y - UAK)^T(Y - UAK)) + \lambda_2 \text{tr}(UDU^T) + \text{tr}(FPF^T). \quad (17)$$

对上式求矩阵 U 的偏导,得到

$$\frac{\partial J}{\partial U} = -(Y - UAK)KA^T + \lambda_2 UD + UAKPKA^T. \quad (18)$$

设置偏导为0,可以得到

$$U(AK(I_N + P)KA^T + \lambda_2 D) = YKA^T, \quad (19)$$

则矩阵 U 可以通过下式迭代得到:

$$U = YKA^T(AK(I_N + P)KA^T + \lambda_2 D)^{-1}. \quad (20)$$

基于式(16),在每次迭代中,可以基于当前计算得到的矩阵 D 计算矩阵 U ,然后可根据新计算得到的 U 对 D 进行更新. 算法1总结了矩阵 U 的计算方法.

算法1 迭代求解结构矩阵.

input: 数据矩阵 X 、 Y , 参数 λ_1 、 λ_2 、 λ_3 , 潜变量维度 K ;

output: 结构矩阵 U .

1) 随机矩阵 $A \in R^{K \times N}$, 计算矩阵 L_0 和 L_1 , 设置 $t = 1$;

2) repeat

3) 基于式(8)计算对角矩阵 $D^{(t+1)}$;

4) 求解式(20)得到矩阵 $U^{(t+1)}$;

5) $t \leftarrow t + 1$;

6) until 收敛;

7) return 结构矩阵 U .

3.2 固定矩阵 U 求 A

矩阵 U 固定时,问题(16)简化为

$$\min_A \text{tr}((Y - UAK)^T(Y - UAK)) + \lambda_1 \text{tr}(AKA^T) + \text{tr}(FPF^T). \quad (21)$$

对上式求矩阵 A 的偏导,得到

$$\frac{\partial J}{\partial A} = -U^T(Y - UAK)K + \lambda_1 AK + U^T UAKPK. \quad (22)$$

同理,设置偏导为0,进一步得到

$$U^T YK = U^T UAKK + \lambda_1 AK + U^T UAKPK. \quad (23)$$

两边同时右乘 K^{-1} ,然后左乘 $(U^T U)^{-1}$,可以得到

$$(U^T U)^{-1} U^T Y = \lambda_1 (U^T U)^{-1} A + AK(I_N + P). \quad (24)$$

显然,式(24)是一个标准的 Sylvester 方程 $A\Theta + \Theta B = C$,其存在闭式解. Θ 对应于未知矩阵 A ,有 $A = \lambda_1 (U^T U)^{-1} B$, $B = K(I_N + P)$, $C = (U^T U)^{-1} U^T Y$. 矩阵 A 可以通过求解 Sylvester 方程得到. 综上所述, MTR-ITC 模型的求解过程如算法2所示.

算法2 基于实例与目标相关性的 MTR 算法.

input: 数据矩阵 X 、 Y , 参数 λ_1 、 λ_3 、 λ_4 , 潜变量维度 K ;

output: 结构矩阵 U , 回归系数矩阵 A .

1) 随机矩阵 $A \in R^{K \times N}$, $U \in R^{Q \times K}$;

2) 计算矩阵 L_0 和 L_1 , 设置 $i = 1$;

3) repeat

4) 基于式(24)计算结构矩阵 $A^{(i+1)}$;

5) 基于算法1求解矩阵 $U^{(i+1)}$;

6) $t \leftarrow t + 1$;

7) until 收敛;

8) return 结构矩阵 U 和系数矩阵 A .

3.3 计算复杂度分析

所提出的交替优化算法的计算复杂度主要来自于算法1和算法2中对矩阵 A 和 U 的计算;算法1中的计算复杂度主要是对 N 阶矩阵求逆,其计算复杂度为 $O(N^3)$;算法2中的计算复杂度主要源自于对 Sylvester 方程的求解,其对应的算法复杂度也为 $O(N^3)$. 假设算法1和算法2所对应的迭代次数分别为 t_1 和 t_2 次,则整个交替优化算法的计算复杂度近似为 $O(t_1 t_2 N^3 + t_2 N^3)$. 由于 $t_1 \ll N$, $t_2 \ll N$, 算法复杂度近似等于 $O(N^3)$. 所提算法的计算复杂度与其他核方法,如 KRR、MMR 和 MSLR 等接近. 实际上,无论是算法1或算法2,其矩阵求逆的算子都是基于对称矩阵的,因此其运算往往可以通过矩阵近似(如 LU 分解等)来提速^[39].

4 实验与分析

4.1 实验设计与评价指标

为了验证本文算法的有效性和可行性,从以下两方面进行验证:1) 通过与目前最常用的数个方法进行比较,观察其是否能在公开数据集上取得较为优势的结果;2) 验证所提出的交替优化算法的收敛性,进而说明算法能在有限次迭代中快速收敛至最优解;3) 通过对实例相关性和目标相关性的正则化参数进行敏感性分析,验证引入实例与目标相关性信息是否可以提升算法性能.

为了对比不同算法在数据集上的性能表现,使用平均相对均方根误差(average relative root mean squared error, aRRMSE)作为衡量指标,计算如下:

$$aRRMSE = \frac{1}{Q} \sum_{i=1}^Q \sqrt{\frac{\sum_{j=1}^{N_{test}} (y_i^{(j)} - \hat{y}_i^{(j)})^2}{\sum_{j=1}^{N_{test}} (y_i^{(j)} - \bar{y}_i)^2}} \quad (25)$$

其中: N_{test} 为测试集的规模, $y_i^{(j)}$ 和 $\hat{y}_i^{(j)}$ 分别为测试集中实例 $\mathbf{x}^{(j)}$ 在目标 i 上的真实值和预测值, \bar{y}_i 为训练集 \mathcal{D}_{train} 在第 i 个目标上的平均值. $aRRMSE$ 衡量了模型在所有输出上的预测结果相对于训练样本均值的均方误差. 所有实验在AMD Ryzen 7 4800 H (2.90 GHz) 处理器, 24 G 内存的计算机中通过 Matlab 2021a 软件平台实现完成.

4.2 对比方法与数据集

在数据实验中, 从 MTR 数据集网站 Mulan^[40] 中选择数个常用的公开数据集作为基准数据集进行对比实验, 验证本文算法的可行性与有效性. 所有数据集信息如表 1 所示. 其中: N 为数据集大小; d 为数据集的输入维度; Q 为输出维度; k 为式(11)中所采取的近邻数, 其大小与数据集的规模成正比; #-fold 为相应数据集采取的交叉验证折数.

表 1 实验所用数据集特征信息

数据集	N	d	Q	k	#-fold
ATP1D	337	411	6	5	10
ATP7D	296	411	6	5	10
OES10	403	298	16	7	10
OES97	334	263	16	7	10
OSALES	639	413	12	9	10
SCM1d	9803	280	16	15	2
SCM20d	8965	61	16	15	2
SCPF	1137	23	3	11	10
SLUMP	103	7	3	3	10
WQ	1059	16	14	11	10

选取现有的主流 AA 类方法, 如 $mKRR$ ^[41]、 $MSVR$ ^[16]、 OKL ^[24]、 $MTFL$ ^[42]、 MMR ^[7] 以及 $MSLR$ ^[26] 作为对比方法. 为了便于比较, 不同模型中的最优正则化参数通过在训练集上基于网格搜索法的交叉验证得到, 并设定网格搜索范围为 $10^{[-3:1:3]}$. 同时, 采用径向基 (RBF) 函数作为非线性回归的特征映射函数, 并设置超参数的取值范围为 $2^{[-6:1:3]}$.

除了选择上述主流算法进行对比, 为验证模型引入实例相关性和目标相关性后是否可能带来性能提升, 通过对参数 λ_3 和 λ_4 进行敏感性分析, 观察调整参数后的模型性能变化.

4.3 实验结果与分析

表 2 总结了所提方法与其他模型在不同数据集上的性能表现. 不难看出, $mKRR$ 作为独立预测多个目标输出的基准算法, 其并未利用不同目标输出间的关联信息, 导致其在不同数据集上的表现明显不如其

他 MTR 方法. $MTFL$ 基于线性结构来进行特征学习, 导致其缺乏处理非线性输入输出关系的能力, 在多数数据集 (如 $OSALES$ 、 $OES97$ 等) 上明显不如其他对比方法. $MSVR$ 算法虽然引入了核技巧来处理非线性关系, 但其仅通过 ℓ_2 范数将不同输出维度的误差进行综合考虑, 并未充分利用到不同输出间内在的相关性, 这也使得其在不同数据集上的表现不如考虑了目标相关性的 MTR 方法 (如 MMR 、 $MSLR$ 和 $MTR-ITC$ 等). OKL 算法同时考虑了目标间和特征与目标之间的非线性关系, 在多数数据集上都明显优于作为基准算法的 $mKRR$ 算法. 相比于其他核方法, 采用多层稀疏结构的 MMR 、 $MSLR$ 以及 $MTR-ITC$ 方法在所有数据集上均能取得较好的测试性能, 这说明引入潜变量空间解耦输入输出以及输出间相关性后, 可以更加准确地对 MTR 中的复杂相关性进行学习. 而且, $MTR-ITC$ 引入了对目标以及样本关系结构的学习, 使其在模型训练过程中更好地探索了样本局部结构以及目标输出之间的相互关系, 进而提升模型的回归性能.

表 2 10 个数据集上 7 种方法的平均相对均方根误差 (aRRMSE) %

数据集	$mKRR$	$MSVR$	OKL	$MTFL$	MMR	$MSLR$	$MTR-ITC$
ATP1D	38.0	38.1	36.4	41.5	33.2	35.2	31.9
ATP7D	48.6	47.8	47.5	55.3	44.3	45.8	43.8
OES10	48.9	44.7	43.2	53.2	40.3	40.7	40.2
OES97	58.7	55.7	53.5	81.8	49.7	48.9	52.8
OSALES	79.9	72.6	71.8	168.2	70.9	74.5	73.3
SCM1D	38.7	53.0	45.6	44.7	30.1	45.9	34.1
SCM20D	49.6	49.4	44.3	64.5	38.8	40.5	35.0
SCPF	85.5	82.8	82.0	89.9	81.2	81.7	78.9
SLUMP	78.9	71.1	69.9	68.1	58.7	56.7	57.9
WQ	92.1	89.9	89.1	96.2	88.9	89.0	88.8
AveRank	5.7	5.2	4.0	6.4	1.9	3.1	1.7

为了验证本文所提算法与对比算法之间在统计上存在显著性差异, 更好地说明引入实例与目标相关性信息的优势, 本文对单一度量下的多个 MTR 方法性能进行排序, 然后分别计算每种方法在所有数据集上的平均排名. 利用显著性水平为 0.05 ($\alpha = 0.05$) 的 Friedman 检验^[43] 来验证所提算法与对比算法之间的性能是否存在显著差异, 其中原假设表示为所提算法与比较算法之间的预测性能相同. 若拒绝该原假设, 则利用 Bonferroni-Dunn 检验作为后验检验来分析比较, 并计算其临界差异 (critical difference, CD) 值, 作为不同算法之间是否具有差异的统计标准. 公式如下:

$$CD = q_\alpha \sqrt{\frac{n(n+1)}{6T}} \quad (26)$$

其中: n 为比较的算法数量, T 为用于实验的数据集数量. 在显著性水平为 $\alpha = 0.05$ 时, 其对应的 $q_\alpha = 2.845$, 因此有 $CD = 2.55 (n = 7, T = 10)$, 其基于

aRRMSE指标所对应的排名如图2所示.

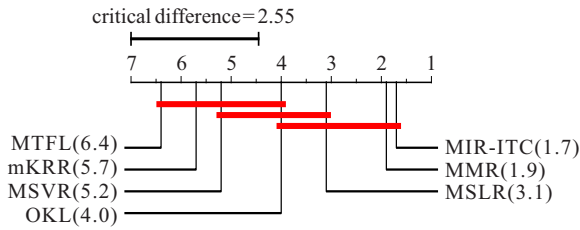


图2 基于aRRMSE的算法平均排名

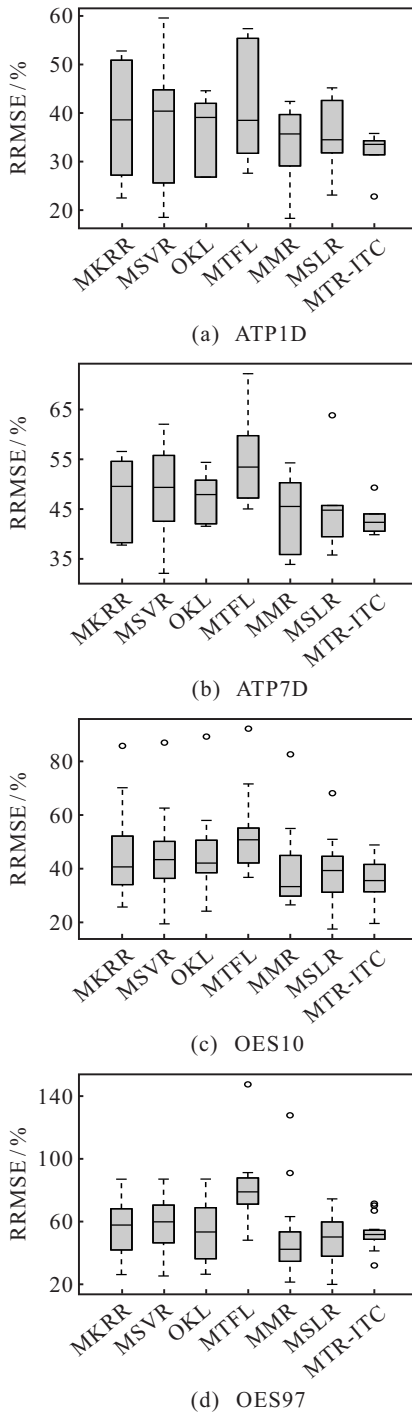


图3 不同算法在多元输出上的RRMSE箱型图

显然,从图2可以看出,3种多层稀疏结构下的3种算法虽然在统计上并未存在明显差别,但MTR-ITC算法在不同性能指标的算法排名上均优于MMR

和MSLR,且与其他算法(如OKL、MSVR等)均存在统计上的差异,这说明本文所提算法在性能上的优越性.此外,比较了多个算法在数据集的不同输出上的RRMSE值差异性,其在部分MTR数据集不同目标上的RRMSE箱型图如图3所示.与其他算法相比,MTR-ITC算法可在所有输出目标上保持较低的误差水平和方差,这说明通过引入实例和目标相关性的正则化项可以确保模型输出的一致性,有助于提升模型的鲁棒性和性能.

4.4 收敛性分析

算法1与算法2的快速收敛是保证交替优化算法能高效求解MTR-ITC算法的关键.因此,在部分具有代表性的数据集(如OSALES)上进行收敛性分析,得到如图4所示的收敛结果.

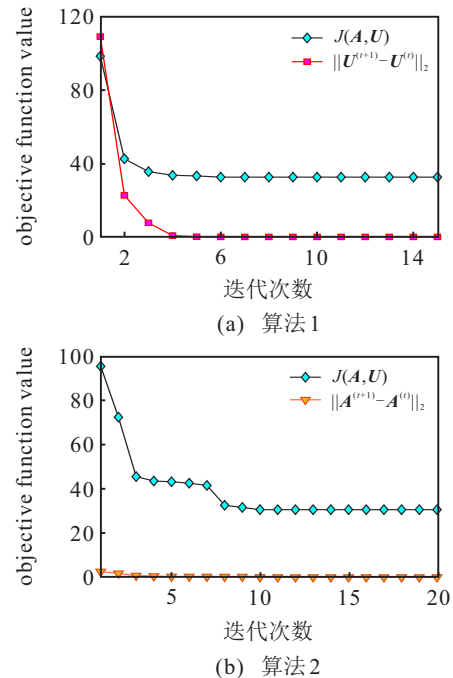


图4 OSALES数据集上算法1和算法2的收敛性

在训练迭代过程中,通过计算算法1中 $\|U^{(t+1)} - U^{(t)}\|_2$ 及其目标函数 $J(A, U)$,进而观察其在有限次迭代中的收敛过程;同理,在基于算法2求解时,计算每次迭代中 $\|A^{(i+1)} - A^{(i)}\|_2$ 以及 $J(A, U)$ 对应的值,最终得到图4所示的收敛曲线.若分别将 $\|U^{(t+1)} - U^{(t)}\|_2 \leq 10^{-5}$ 和 $\|A^{(i+1)} - A^{(i)}\|_2 \leq 10^{-5}$ 作为算法1和算法2的收敛准则,则:算法1在多数数据集上能实现10代内收敛,且算法1的快速收敛也保证了算法2在交替优化时A的更新效率;算法2也可以在20代内实现收敛,这说明MTR-ITC算法具有较好的收敛性.

4.5 敏感性分析

进一步分析正则化参数(λ_3 和 λ_4)对模型性能的影响.通过固定参数 $\lambda_1 = 10^{-2}$ 和 $\lambda_2 = 10^{-2}$,设

置相同的潜变量维度与输出维度,即 $K = Q$. 选择 ATP1D、ATP7D 数据集,通过网格搜索方法调整 λ_3 和 λ_4 的值来观察其对模型性能的影响,并设定参数区间为 $10^{[-3,4]}$,得到结果如图5所示. 无论是实例相关性的正则化参数 λ_3 ,还是目标相关性的正则化参数 λ_4 ,当其对应的值逐渐增大时,算法性能会随之提升然后再下降. 这说明引入实例相关性以及目标相关性的正则化项可以有效提升算法性能,但当算法过分考虑实例相关性以及目标相关性信息时,会导致算法陷入局部最优或过拟合,进而在测试集上表现不佳.

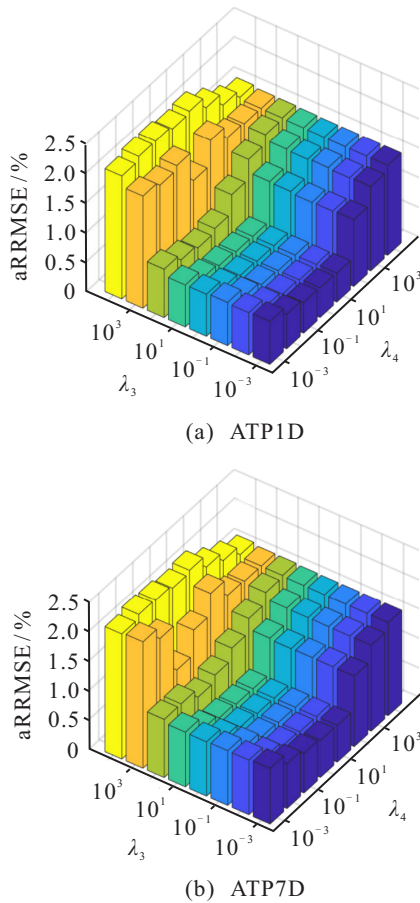


图5 基于正则化参数的模型性能变化

5 结论

本文提出了一种新的MTR模型MTR-ITC,来同时对输入输出以及输出间的依赖关系进行探索和学习. 与以往的MTR方法相比,该方法同时利用输出关联和目标关联信息,引入了流形正则化器,直接从数据中学习拉普拉斯矩阵,而不需要任何目标关联的先验知识,使得算法具有良好的适用性. 通过嵌入潜变量空间来对复杂的非线性输入输出关系以及目标输出间的关联性进行解耦,并在结构矩阵上采用 $\ell_{2,1}$ 范数稀疏正则化,提高了模型的鲁棒性. 实验分析表明,所提方法不仅在不同的数据集上保持了优异性能,其在不同目标输出上的性能表现还具有一致性,这使

得模型结果在处理MTR问题时更加稳定. 在未来的工作中,还可以对潜变量空间的维度进行分析,通过找到最合适的空间维度来对目标内在的关联关系进行解耦和表达,并考虑潜在的结构化变量对输出的影响,进而使得模型更加稳健.

参考文献(References)

- [1] Li H Q, Zhang W, Chen Y, et al. A novel multi-target regression framework for time-series prediction of drug efficacy[J]. *Scientific Reports*, 2017, 7(1): 1-9.
- [2] Kocev D, Džeroski S, White M D, et al. Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition[J]. *Ecological Modelling*, 2009, 220(8): 1159-1168.
- [3] Musicki D. Multi-target tracking using multiple passive bearings-only asynchronous sensors[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2008, 44(3): 1151-1160.
- [4] Zhen X T, Yu M Y, Islam A, et al. Descriptor learning via supervised manifold regularization for multioutput regression[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(9): 2035-2047.
- [5] Wang X Q, Zhen X T, Li Q Z, et al. Cognitive assessment prediction in alzheimer's disease by multi-layer multi-target regression[J]. *Neuroinformatics*, 2018, 16(3): 285-294.
- [6] Borchani H, Varando G, Bielza C, et al. A survey on multi-output regression[J]. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2015, 5(5): 216-233.
- [7] Zhen X T, Yu M Y, He X F, et al. Multi-target regression via robust low-rank learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(2): 497-504.
- [8] Ishibashi H, Higa K, Furukawa T. Multi-task manifold learning for small sample size datasets[J]. *Neurocomputing*, 2022, 473: 138-157.
- [9] 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.
- [10] Spyromitros-Xioufis E, Tsoumakas G, Groves W, et al. Multi-label classification methods for multi-target regression[J/OL]. 2012, arXiv: 1211.6581.
- [11] 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.
- [12] Spyromitros-Xioufis E, Sechidis K, Vlahavas I. Multi-target regression via output space quantization[C]. *2020 International Joint Conference on Neural Networks*. Glasgow, 2020: 1-9.
- [13] Nakano F K, Pliakos K, Vens C. Deep tree-ensembles for multi-output prediction[J]. *Pattern Recognition*, 2022, 121: 108211.
- [14] Breskvar M, Kocev D, Džeroski S. Ensembles for multi-target regression with random output selections[J].

- Machine Learning, 2018, 107(11): 1673-1709.
- [15] Hadavandi E, Shahrabi J, Hayashi Y. SPMoE: A novel subspace-projected mixture of experts model for multi-target regression problems[J]. *Soft Computing*, 2016, 20(5): 2047-2065.
- [16] Sanchez-Fernandez M, de-Prado-Cumplido M, Arenas-Garcia J, et al. SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems[J]. *IEEE Transactions on Signal Processing*, 2004, 52(8): 2298-2307.
- [17] Liu H, Wang L, Zhao T. Multivariate regression with calibration[J]. *Advances in Neural Information Processing Systems*, 2014, 27: 5630.
- [18] Kocev D, Vens C, Struyf J, et al. Tree ensembles for predicting structured outputs[J]. *Pattern Recognition*, 2013, 46(3): 817-833.
- [19] Xiao X S, Xu Y T. Multi-target regression via self-parameterized Lasso and refactored target space[J]. *Applied Intelligence*, 2021, 51(10): 6743-6751.
- [20] da Silva B L S, Inaba F K, Salles E O T, et al. Outlier robust extreme machine learning for multi-target regression[J]. *Expert Systems with Applications*, 2020, 140: 112877.
- [21] Rothman A J, Levina E, Zhu J. Sparse multivariate regression with covariance estimation[J]. *Journal of Computational and Graphical Statistics*, 2010, 19(4): 947-962.
- [22] Lee W, Liu Y F. Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood[J]. *Journal of Multivariate Analysis*, 2012, 111: 241-255.
- [23] 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. Lake Tahoe*, 2012: 3185-3193.
- [24] Dinuzzo F. Learning output kernels for multi-task problems[J]. *Neurocomputing*, 2013, 118: 119-126.
- [25] Rahimzadeh Arashloo S, Kittler J. Multi-target regression via non-linear output structure learning[J]. *Neurocomputing*, 2022, 492: 572-580.
- [26] Zhen X T, Yu M Y, Zheng F, et al. Multitarget sparse latent regression[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(5): 1575-1586.
- [27] Zhu Y, Kwok J T, Zhou Z H. Multi-label learning with global and local label correlation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(6): 1081-1094.
- [28] Fan Y L, Liu J H, Liu P Z, et al. Manifold learning with structured subspace for multi-label feature selection[J]. *Pattern Recognition*, 2021, 120: 108169.
- [29] Huang R, Wu Z J. Multi-label feature selection via manifold regularization and dependence maximization[J]. *Pattern Recognition*, 2021, 120: 108149.
- [30] Cai Z L, Zhu W. Multi-label feature selection via feature manifold learning and sparsity regularization[J]. *International Journal of Machine Learning and Cybernetics*, 2018, 9(8): 1321-1334.
- [31] Fan Y L, Liu J H, Weng W, et al. Multi-label feature selection with constraint regression and adaptive spectral graph[J]. *Knowledge-Based Systems*, 2021, 212: 106621.
- [32] Gao J J, Li F Z, Wang B J, et al. Unsupervised nonlinear adaptive manifold learning for global and local information[J]. *Tsinghua Science and Technology*, 2020, 26(2): 163-171.
- [33] Law M H C, Jain A K. Incremental nonlinear dimensionality reduction by manifold learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(3): 377-391.
- [34] Liu G C, Lin Z C, Yu Y. Multi-output regression on the output manifold[J]. *Pattern Recognition*, 2009, 42(11): 2737-2743.
- [35] Kuleshov A, Bernstein A. Nonlinear multi-output regression on unknown input manifold[J]. *Annals of Mathematics and Artificial Intelligence*, 2017, 81(1): 209-240.
- [36] Zhu X F, Hu R Y, Lei C, et al. Low-rank hypergraph feature selection for multi-output regression[J]. *World Wide Web*, 2019, 22(2): 517-531.
- [37] Liu J, Ji S, Ye J. Multi-task feature learning via efficient $\ell_{2,1}$ -norm minimization[C]. *Arlington: AUA Press*, 2009: 339-348.
- [38] Doquire G, Verleysen M. A graph Laplacian based approach to semi-supervised feature selection for regression problems[J]. *Neurocomputing*, 2013, 121: 5-13.
- [39] Li X M, Wang Y, Zhang Z, et al. RMoR-Aion: Robust multioutput regression by simultaneously alleviating input and output noises[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(3): 1351-1364.
- [40] Tsoumakas 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.
- [41] Shawe-Taylor J, Cristianini N. *Kernel methods for pattern analysis*[M]. Cambridge: Cambridge University Press, 2004: 289-326.
- [42] Gong P H, Ye J P, Zhang C S. Robust multi-task feature learning[J]. *KDD: Proceedings International Conference on Knowledge Discovery & Data Mining*, 2012, 2012: 895-903.
- [43] Demšar J. Statistical comparisons of classifiers over multiple data sets[J]. *Journal of Machine Learning Research*, 2006, 7: 1-30.

作者简介

何杜博(1997—),男,博士生,从事复杂系统建模与优化、装备经济管理研究, E-mail: 1158387625@qq.com;
孙胜祥(1969—),男,教授,博士生导师,从事系统工程、装备经济管理研究, E-mail: sxx13397129808@163.com.