



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



基于自适应图学习的多目标特征选择算法

何杜博, 孙胜祥, 梁新, 谢力, 张侃

引用本文:

何杜博,孙胜祥,梁新,谢力,张侃. 基于自适应图学习的多目标特征选择算法[J]. *控制与决策*, 2024, 39(7): 2295–2304.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0127>

您可能感兴趣的其他文章

Articles you may be interested in

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>

基于向量角分解的高维多目标进化算法

Many-objective evolutionary algorithm based on vector angle decomposition

控制与决策. 2021, 36(3): 761–768 <https://doi.org/10.13195/j.kzyjc.2019.0925>

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>

基于弱关联的自适应高维多目标进化算法

A weak association-based adaptive evolutionary algorithm for many-objective optimization

控制与决策. 2021, 36(8): 1804–1814 <https://doi.org/10.13195/j.kzyjc.2019.1723>

基于低秩矩阵恢复的视觉显著性目标检测与细化

Saliency object detection and refinement based on low rank matrix recovery

控制与决策. 2021, 36(7): 1707–1713 <https://doi.org/10.13195/j.kzyjc.2019.1795>

基于自适应图学习的多目标特征选择算法

何杜博, 孙胜祥, 梁新, 谢力[†], 张侃

(海军工程大学 管理工程与装备经济系, 武汉 430033)

摘要: 针对多目标回归中的特征选择问题, 提出一种基于自适应图学习的多目标特征选择算法, 在单个框架中同时考虑 3 种关系结构: 输入特征与目标输出、不同目标输出以及样本间的相关结构, 并基于上述结构信息进行特征选择. 首先, 在传统稀疏回归模型中对系数矩阵施加低秩约束, 利用低秩学习对特征间相关性以及目标间的依赖关系进行解耦学习; 然后, 构建基于样本局部结构信息的自适应图学习项, 充分利用样本间的相似结构进行特征选择; 进一步地, 引入基于输出相关性的结构矩阵优化项, 使模型能够更加充分地考虑目标间的相关性; 最后, 提出一种交替优化算法求解目标函数, 并从理论上证明算法的收敛性. 在公开数据集上的实验表明, 所提方法相较于现有主流的多目标特征选择方法具有更好的性能和适用性.

关键词: 特征选择; 稀疏回归; 多目标回归; 交替优化算法; 自适应图学习

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0127

引用格式: 何杜博, 孙胜祥, 梁新, 等. 基于自适应图学习的多目标特征选择算法[J]. 控制与决策, 2024, 39(7): 2295-2304.

Multi-target feature selection algorithm based on adaptive graph learning

HE Du-bo, SUN Sheng-xiang, LIANG Xin, XIE Li[†], ZHANG Kan

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

Abstract: Feature selection not only enhances the efficiency of regression modelling but also reduces the detrimental effects of feature redundancy and noises. This paper proposes a multi-target feature selection algorithm based on adaptive graph learning. Specifically, the method imposes a low-rank constraint on the regression matrix, enabling simultaneous modelling of inter-target, input-output and inter-sample relationships within a general framework. The similarity-induced graph matrix is learned to adaptively preserve samples' similarity structure to alleviate the influence of noises and outliers. Furthermore, we introduce a manifold regularizer to preserve the global target correlations to ensure the global target correlations structure of data in the subsequent learning process. An alternative optimization algorithm is presented to solve the final objective function. Extensive experiments conducted on real-world data sets demonstrate that the proposed method is superior to state-of-the-art multi-target feature selection methods.

Keywords: feature selection; sparse regression; multi-target regression; alternating optimization algorithm; adaptive graph learning

0 引言

多目标回归 (multi-target regression, MTR), 又称为多输出或多变量回归, 其目的在于利用一组共同的输入变量, 同时对多个连续型变量进行预测. MTR 同时考虑了特征和目标以及目标之间的相关关系, 具备良好的可解释性和计算效率^[1], 已经广泛应用于数据挖掘^[2-4]、计算机视觉^[5]、医疗诊断^[6]、股市预测^[7]、负荷预测^[8] 等多个具有复杂输出关系的领域研究中.

由于需要同时对多个目标变量进行预测, MTR 通常需要收集大量可能相关的特征来形成一组共同的特征集对目标进行建模表示, 导致其原始数据中往往具有较高的特征维度. 然而, 高维特征不仅容易陷入“维度诅咒”, 导致模型在训练学习过程中产生较大的内需求和计算成本, 其不相关和冗余特征也会显著降低模型的学习效率和泛化性能. 特征选择是一种有效的降维技术, 其通过特定方法或准则从原始的

收稿日期: 2023-02-06; 录用日期: 2023-04-15.

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

责任编辑: 巩敦卫.

[†]通讯作者. E-mail: xieli_hg@163.com.

特征集中选择最为相关和冗余性较低的特征子集进行训练和学习,使得学习得到的模型更具解释性的同时减少了模型的训练成本。

近年来,针对回归问题的特征选择方法大致可以分为以下3类:过滤式、包裹式和嵌入式。过滤式方法采用特定的评分准则如互信息、拉普拉斯得分^[9]、Fisher得分^[10]对特征进行排序,选择得分较高的特征构成特征子集。过滤式方法以较低的计算成本高效剔除大量的无关特征,但易将高度冗余的特征纳入选择的特征子集中,且过滤式方法忽略了所选特征与模型的关系,其在学习模型上可能无法获得令人满意的性能。包裹式方法通过特定的策略来选择特征子集,再将其输入至模型中进行训练迭代和更新优化,直至达到相应的性能要求。由于每次特征选择都需要通过模型检验性能,这也使其计算成本大幅增加。另外,当所选学习模型不同时,包裹式方法得到的特征子集也可能不一致,其特征选择的稳定性和适应性还有待完善。嵌入式方法作为过滤式和包裹式方法的折中,不仅可以解决过滤式方法中特征集冗余度较高的问题,且相比于包裹式方法有效降低了计算成本。因此,嵌入式方法凭借其优越的性能表现受到越来越多的关注。

MTR面临的挑战在于同时对输入输出以及目标间的相关性进行建模^[11]。准确利用目标间的相关性信息,可以使MTR模型获得比单目标回归模型更加优越的性能表现。因此,现有的MTR方法往往注重目标间相关性的探索挖掘,其常用策略为在学习过程中对权重矩阵施加各种低秩或稀疏约束^[12],使得学习得到的权重矩阵具有良好的稀疏结构,以便更好地对目标间相关性进行表示。虽然稀疏低秩等约束可有效挑选出信息量较大的特征,但其在特征选择过程中并未考虑到特征和样本的局部和全局结构。大量文献证明^[13-14],特征与样本的局部和全局结构信息可以提升不同问题下的特征选择性能。

综上所述,本文提出一种自适应图学习的特征选择算法。首先,通过对权重矩阵施加低秩约束对输入特征之间以及输出目标之间的相关性进行解耦,以便更为准确地学习到原始数据集中不同的相关性信息。其次,引入对样本的全局和局部结构的自适应图学习,减少噪声和离群点对特征选择模型性能的影响,进一步增加所选择特征集的稳定性和抗噪性。再次,考虑不同目标输出间的统计相关性信息,设计基于输出相关性的流形正则化约束,使得模型在特征选择中能充分考虑不同目标间的全局相关性,进一步

提升特征选择模型在MTR中的性能效果。最后,提出一种交替迭代优化算法对不同的参数矩阵依次迭代求解,并从理论上证明其收敛性。在8个不同领域的MTR公开数据集上的实验验证了所提出特征选择算法在MTR问题下的可行性和有效性。

1 相关工作

多目标特征选择是近年来研究的一个新领域。以嵌入式方法为例,Yuan等^[15]基于多层多输出结构提出了一种嵌入式的稀疏结构特征选择(sparse structural feature selection, SSFS)模型,对目标函数、回归矩阵以及结构矩阵联合施加稀疏性约束,更好地进行特征选择。Sechidis等^[16]提出了一种用于多标签分类和MTR的特征选择方法。该方法计算高阶目标交互,并考虑特征之间的相关性、冗余性和互补性,保证得到的特征子集可以具有较少的冗余和较高的相关性。Zhu等^[17]提出一种结合稀疏特征选择和低秩线性回归的嵌入式方法来选择原始特征的最佳子集进行回归建模。Zhang等^[18]利用低秩约束寻找输出变量之间的相关性,并通过在系数矩阵上施加 $\ell_{2,1}$ 范数正则化来进行特征选择。近年来,不同过滤式多目标特征选择方法也相继被提出。Hashemi等^[19]提出了一种融合多准则决策的多目标特征选择方法,基于VIKOR算法来衡量输入特征与输出目标集之间的相关程度,并以此为依据对特征进行排序。Petković等^[20]基于预测聚类树集成以及RRelief方法拓展得到了两组多目标特征的得分,通过融合两组方法的特征得分得到最佳的特征排名。

实际上,当多标签分类特征选择方法可以输出连续型变量时,其同样可用于MTR任务中。Fan等^[21]提出一种基于标签相关性以及特征冗余的有监督特征选择方法,其利用低维嵌入来挖掘不同标签的相关性,并通过保持原始标签空间中的全局和局部结构来提升特征选择的性能。Xu等^[22]通过最大化特征方差和特征与标签之间的依赖性进行特征提取,进而在多标签分类问题中取得更好的性能表现。Zhang等^[23]提出一种流形正则化的嵌入式多标签特征选择方法,其基于原始特征空间构造低维嵌入,以便更好地适应标签分布和捕捉局部的标签相关性。Zhu等^[24]引入流形正则化对标签的全局及局部相关性进行挖掘来提升多标签学习的性能。从上述研究中可以看出,在特征选择中,维持样本在全局和局部的相似结构不仅可以为特征选择提供相应的补充信息,还有利于模型对各种关系的学习,如输入特征之间的相关结构,输出特征之间的依赖关系以及输入输出之间的关联

关系^[25]. 现有针对MTR的特征选择模型鲜有同时考虑上述信息, 因此本文试图在同一框架中利用上述信息进行多目标特征选择, 并观察其特征选择的性能表现.

2 算法描述

2.1 基于低秩学习的多目标回归

给定MTR训练集 $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$, 记 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T \in \mathbb{R}^{n \times d}$ 代表特征矩阵, 则 $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]^T \in \mathbb{R}^{n \times q}$ 为对应的目标矩阵, 其中 $\mathbf{y}_i = [y_{i,1}, \dots, y_{i,q}]^T$ 为 \mathbf{x}_i 对应的输出向量, 则多输出Ridge回归的目标函数如下:

$$\min_{\mathbf{W}, \mathbf{b}} \|\mathbf{X}\mathbf{W} + \mathbf{1}_n \mathbf{b}^T - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{W}\|_F^2. \quad (1)$$

其中: $\mathbf{W} \in \mathbb{R}^{d \times q}$ 为回归系数矩阵, $\mathbf{b} \in \mathbb{R}^q$ 为偏置向量, α 为正则化参数. 为了对输入变量进行特征选择, 通过对 \mathbf{W} 施加 $\ell_{2,1}$ 正则化约束, 使学习得到的 \mathbf{W} 具有行稀疏性, 进而实现对输入特征的选择. 可以得到如下目标函数:

$$\min_{\mathbf{W}, \mathbf{b}} \|\mathbf{X}\mathbf{W} + \mathbf{1}_n \mathbf{b}^T - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{W}\|_{2,1}. \quad (2)$$

显然, 式(2)并未考虑到不同输出之间存在一定的耦合关系, 其仅仅是单输出结构的简单拓展, 且仅基于系数矩阵 \mathbf{W} 难以同时对MTR数据集中输入与输出以及不同输出之间的相关关系进行准确的学习和描述. 因此, 通过对 \mathbf{W} 施加一个低秩假设, 即 $\mathbf{W} = \mathbf{A}\mathbf{B}$, $\mathbf{A} \in \mathbb{R}^{d \times r}$, $\mathbf{B} \in \mathbb{R}^{r \times q}$, 且 $r \leq \min(d, q)$. 式(2)变成

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{b}} \|\mathbf{X}\mathbf{A}\mathbf{B} + \mathbf{1}_n \mathbf{b}^T - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{A}\mathbf{B}\|_{2,1}. \quad (3)$$

矩阵 \mathbf{A} 可以看成是将原始的 d 维特征线性转换至一个 r 维的潜变量空间, 然后利用矩阵 \mathbf{B} 实现对潜变量空间与输出空间之间的关系学习. 利用潜变量空间嵌入, 矩阵 \mathbf{A} 和 \mathbf{B} 分别表示输入特征之间的冗余关系以及不同输出间的耦合关系, 进而对原始的输入输出以及输出间的相关性进行解耦. 进一步地, 利用低秩假设可以得到更为鲁棒的特征选择结果, 进而减少冗余特征以及异常变量对特征选择结果的干扰.

2.2 基于样本局部结构信息的自适应图学习

在MTR问题的特征选择中, 希望特征选择后样本能够保持与特征选择之前一样的局部结构, 这样不仅保证了原始数据中的结构信息不丢失, 且更有利于后续的建模学习. 现有研究大多通过学习样本间的近邻图矩阵 \mathbf{S} 来获取样本的局部分布结构和信息, 给定特征矩阵 \mathbf{X} 和其对应的权重矩阵 \mathbf{W} , 依据文献[26]可以得到如下目标函数:

$$\min_{\mathbf{W}} \sum_{i,j=1}^n \|\mathbf{x}_i^T \mathbf{W} - \mathbf{x}_j^T \mathbf{W}\|_2^2 s_{i,j}. \quad (4)$$

其中: $\mathbf{W} \in \mathbb{R}^{d \times d}$, $\mathbf{S} = [s_{i,j}] \in \mathbb{R}^{n \times n}$, 元素 $s_{i,j}$ 表示样本 \mathbf{x}_i 与 \mathbf{x}_j 的相似度. 传统方法往往基于高斯核函数来计算不同样本之间的相似度, 即当样本 \mathbf{x}_i 为样本 \mathbf{x}_j 的 k 近邻时, 样本间相似度为

$$s_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right), \quad (5)$$

其中 σ 为核参数. 虽然式(5)被广泛使用, 但是其仍然存在一定局限性: 1) 学习到的相似关系是基于原始数据得到的, 当原始数据中存在一定水平的噪声和冗余特征时, 难以学习到正确的样本近邻关系; 2) 样本间相似结构的学习与特征选择模型之间是相互独立的, 即使两者在学习中都达到了最优, 其最终仍然可能只是次优的结果; 3) 相似度值对近邻参数 k 和核参数 σ 十分敏感, 当算法复杂度较高时, 调参仍然会造成较大的时间成本.

为了更好地学习到原始数据中蕴含的真实近邻关系, 同时减少需要调整的参数数量, 将系数矩阵与相似关系矩阵同时进行学习, 并通过迭代优化的手段逐步求解不同参数, 使模型在特征选择过程中可以同时考虑到真实样本中的特征关系以及局部的结构信息, 进而自适应地学习到正确的样本近邻关系, 而不是仅依据高斯核函数计算一个固定的图矩阵. 因此, 设计如下的目标函数:

$$\begin{aligned} \min_{\mathbf{S}, \mathbf{W}} \sum_{i,j=1}^n (\|\mathbf{x}_i^T \mathbf{W} - \mathbf{x}_j^T \mathbf{W}\|_2^2 s_{i,j} + \gamma \|\mathbf{s}_i\|_2^2). \\ \text{s.t. } \forall i, \mathbf{1}^T \mathbf{s}_i = 1, s_{i,i} = 0. \\ \begin{cases} s_{i,j} \geq 0, j \in \mathcal{N}(i); \\ s_{i,j} = 0, j \notin \mathcal{N}(i). \end{cases} \end{aligned} \quad (6)$$

其中: γ 为非负正则化参数; \mathbf{s}_i 为相似矩阵 \mathbf{S} 的第 i 列, 通过引入 $\|\mathbf{s}_i\|_2^2$ 来避免平凡解; $\mathcal{N}(i)$ 表示第 i 个样本的近邻集合; $\mathbf{1}^T \mathbf{s}_i = 1$ 可以提升模型对冗余和噪声的鲁棒性. 结合式(3)和(6)得到如下目标函数:

$$\begin{aligned} \min_{\mathbf{S}, \mathbf{A}, \mathbf{B}} \sum_{i,j=1}^n (\|\mathbf{x}_i^T \mathbf{A}\mathbf{B} - \mathbf{x}_j^T \mathbf{A}\mathbf{B}\|_2^2 s_{i,j} + \gamma \|\mathbf{s}_i\|_2^2). \\ \text{s.t. } \forall i, \mathbf{1}^T \mathbf{s}_i = 1, s_{i,i} = 0. \\ \begin{cases} s_{i,j} \geq 0, j \in \mathcal{N}(i); \\ s_{i,j} = 0, j \notin \mathcal{N}(i). \end{cases} \end{aligned} \quad (7)$$

基于式(7)可以保证学习得到最终的模型输出中的近邻关系与原始数据中保持一致, 其有利于后续对不同输出相关结构的学习. 同时, 保留原始样本间的近邻关系, 更有利于减少冗余或不相关特征的影响.

2.3 基于输出相关性的结构矩阵优化

在MTR问题中,由于不同输出目标间的相关关系未知,将输出间的相关性引入至学习过程中,可以更加充分地考虑到不同输出间的相关性对特征选择的影响.因此,考虑样本在输入空间中的几何分布结构的同时,还将目标空间中的结构信息纳入模型中,如不同输出目标之间的相似度信息.采用常用的余弦相似度量来衡量不同目标向量间的相似性,其计算公式如下:

$$\tilde{s}_{i,j} = \frac{\langle \mathbf{y}^i, \mathbf{y}^j \rangle}{\|\mathbf{y}^i\| \|\mathbf{y}^j\|}, \quad (8)$$

其中 \mathbf{y}^i 和 \mathbf{y}^j 分别为第 i 个和第 j 个目标向量.式(8)衡量了第 i 个目标向量与第 j 个目标向量在输出空间中的相似度.对系数矩阵 \mathbf{B} 而言,当目标输出向量 \mathbf{y}^i 与 \mathbf{y}^j 之间越相似时,其对应的权重向量 \mathbf{b}_i 和 \mathbf{b}_j 也应具有一定的相似性.基于以上假设可以得到如下的目标函数:

$$\min_{\mathbf{B}} \sum_{i,j=1}^q \|\mathbf{b}_i - \mathbf{b}_j\|_2^2 \tilde{s}_{i,j}. \quad (9)$$

式(9)可以鼓励相似的目标输出在其对应的权重向量也保持相似性,其优势在于可以使特征选择模型在监督学习中更加充分地利用不同目标输出间的相似性信息,进而提升MTR任务的特征选择性能.

2.4 目标函数建立

结合式(3)、(7)和(9),得到基于自适应图学习的多目标特征选择算法的目标函数如下:

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{b}} \|\mathbf{XAB} + \mathbf{1}_n \mathbf{b}^T - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{AB}\|_{2,1} + \\ & \beta \left(\sum_{i,j=1}^n \|\mathbf{x}_i^T \mathbf{AB} - \mathbf{x}_j^T \mathbf{AB}\|_2^2 s_{i,j} + \gamma \|\mathbf{s}_i\|_2^2 \right) + \\ & \lambda \sum_{i,j=1}^q \|\mathbf{b}_i - \mathbf{b}_j\|_2^2 \tilde{s}_{i,j}. \\ & \text{s.t. } \forall i, \mathbf{1}^T \mathbf{s}_i = 1, s_{i,i} = 0. \\ & \begin{cases} s_{i,j} \geq 0, j \in \mathcal{N}(i); \\ s_{i,j} = 0, j \notin \mathcal{N}(i). \end{cases} \end{aligned} \quad (10)$$

其中 α 、 β 、 γ 和 λ 为非负正则化参数.目标函数(10)可以分为3个部分:1)针对权重矩阵的低秩稀疏学习;2)样本输出局部相关结构学习;3)基于输出相关性的结构矩阵优化.相较于传统低秩稀疏回归模型,式(10)更加充分地考虑了回归模型在拟合多元输出时的局部结构相似性,同时利用原始数据中不同目标间的相似结构对结构矩阵 \mathbf{B} 进行了优化,使得模型在有监督学习过程中可以充分利用原始数据中的目标相关性,保证基于所选特征子集进行回归建模时保

存原始数据中蕴含的相关性信息,有利于后续MTR建模的性能提升.

由于式(10)并未对 \mathbf{b} 约束,依据KKT条件,对式(10)求 \mathbf{b} 的偏导并令其为0,得到 \mathbf{b} 的最优值为

$$\mathbf{b}^T = \frac{1}{n} (\mathbf{1}_n^T \mathbf{Y} - \mathbf{1}_n^T \mathbf{XAB}). \quad (11)$$

将式(11)的结果代入(10),则目标函数可以简化为

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}, \mathbf{S}} \|\mathbf{H}(\mathbf{XAB} - \mathbf{Y})\|_F^2 + \alpha \|\mathbf{AB}\|_{2,1} + \\ & \beta \left(\sum_{i,j=1}^n \|\mathbf{x}_i^T \mathbf{AB} - \mathbf{x}_j^T \mathbf{AB}\|_2^2 s_{i,j} + \gamma \|\mathbf{s}_i\|_2^2 \right) + \\ & \lambda \sum_{i,j=1}^q \|\mathbf{b}_i - \mathbf{b}_j\|_2^2 \tilde{s}_{i,j}. \\ & \text{s.t. } \forall i, \mathbf{1}^T \mathbf{s}_i = 1, s_{i,i} = 0. \\ & \begin{cases} s_{i,j} \geq 0, j \in \mathcal{N}(i); \\ s_{i,j} = 0, j \notin \mathcal{N}(i). \end{cases} \end{aligned} \quad (12)$$

当学习到最优的参数矩阵 \mathbf{A} 、 \mathbf{B} 和 \mathbf{S} 时,通过计算 \mathbf{AB} 中行向量 $(\mathbf{AB})_i$ 的 ℓ_2 范数,对其进行降序排列可以得到相应特征的重要度排名,进而实现对MTR任务的特征选择.

3 基于交替优化策略的求解优化

虽然目标函数(12)对于所有参数矩阵并不是联合凸的(如 \mathbf{A} 、 \mathbf{B} 和 \mathbf{S}),但是不同矩阵在其他参数矩阵固定时其优化是一个凸问题,可通过交替优化的方式迭代求解交替优化的方式迭代求解.

3.1 固定 \mathbf{S} 求 \mathbf{A} 和 \mathbf{B}

当固定矩阵 \mathbf{S} 时,目标函数退化为

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}} \|\mathbf{H}(\mathbf{XAB} - \mathbf{Y})\|_F^2 + \alpha \|\mathbf{AB}\|_{2,1} + \\ & \beta \sum_{i,j=1}^n \|\mathbf{x}_i^T \mathbf{AB} - \mathbf{x}_j^T \mathbf{AB}\|_2^2 s_{i,j} + \lambda \sum_{i,j=1}^q \|\mathbf{b}_i - \mathbf{b}_j\|_2^2 \tilde{s}_{i,j}. \end{aligned} \quad (13)$$

根据KKT定理,将 $\|\mathbf{AB}\|_{2,1}$ 改写为 $\text{tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{DAB})$,则式(13)可以简化为

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}} \|\mathbf{H}(\mathbf{XAB} - \mathbf{Y})\|_F^2 + \alpha \text{tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{DAB}) + \\ & \beta \text{tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{X}^T \mathbf{LXAB}) + \lambda \text{tr}(\mathbf{B} \tilde{\mathbf{L}} \mathbf{B}^T). \end{aligned} \quad (14)$$

其中: \mathbf{L} 和 $\tilde{\mathbf{L}}$ 分别为 $s_{i,j}$ 和 $\tilde{s}_{i,j}$ 所对应的拉普拉斯矩阵; $\mathbf{D} \in \mathbb{R}^{d \times d}$ 为对角矩阵,其第 i 个对角元素为

$$D_{i,i} = \frac{1}{2 \|\mathbf{AB}\|_{2,1}^2}, i = 1, 2, \dots, d. \quad (15)$$

同样地,固定矩阵 \mathbf{B} ,并对式(14)求 \mathbf{A} 的偏导,令偏导为0后有

$$\mathbf{A}^* = \mathbf{P}^{-1} \mathbf{X}^T \mathbf{HYB}^T (\mathbf{B} \mathbf{B}^T)^{-1}, \quad (16)$$

其中 $\mathbf{P} = \mathbf{X}^T \mathbf{H} \mathbf{X} + \alpha \mathbf{D} + \beta \mathbf{X}^T \mathbf{L} \mathbf{X}$. 同理, 固定矩阵 \mathbf{A} 可以得到如下优化问题:

$$\min_{\mathbf{A}} \text{tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{B} - 2\mathbf{B}^T \mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{Y}) + \lambda \text{tr}(\mathbf{B} \tilde{\mathbf{L}} \mathbf{B}^T). \quad (17)$$

令其对 \mathbf{B} 的偏导为0, 可以得到如下等式:

$$\mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{B} + \lambda \mathbf{B} \tilde{\mathbf{L}} = \mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{Y}. \quad (18)$$

显然, 式(18)的形式是标准Sylvester方程 $\mathcal{A}\Theta + \Theta\mathcal{B} = \mathcal{C}$, 其存在闭式解. Θ 对应于未知矩阵 \mathbf{B} , 有 $\mathcal{A} = \mathbf{A}^T \mathbf{P} \mathbf{A}$, $\mathcal{B} = \lambda \tilde{\mathbf{L}}$, $\mathcal{C} = \mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{Y}$, \mathbf{B} 可以通过求解Sylvester方程得到. 综上所述, 矩阵 \mathbf{A} 和 \mathbf{B} 的求解过程如算法1所示.

算法1 求解参数矩阵 \mathbf{A} 和 \mathbf{B} .

input: 输入矩阵 $\mathbf{X} \in \mathbb{R}^{n \times d}$, 输出矩阵 $\mathbf{Y} \in \mathbb{R}^{n \times q}$, 正则化参数 α, β 和 λ , 近邻数 k 和秩值 r ;

output: $\mathbf{A} \in \mathbb{R}^{d \times r}$, $\mathbf{B} \in \mathbb{R}^{r \times q}$.

step 1: 初始化 $\mathbf{D} = \mathbf{I} \in \mathbb{R}^{d \times d}$;

step 2: repeat

step 2.1: 依据式(16)计算矩阵 \mathbf{A} ;

step 2.2: 依据式(18)计算矩阵 \mathbf{B} ;

step 2.3: 依据式(15)计算矩阵 \mathbf{D} 和 \mathbf{P} ;

until 收敛.

3.2 固定 \mathbf{A} 和 \mathbf{B} 求 \mathbf{S}

固定矩阵 \mathbf{A} 求 \mathbf{B} , 则式(12)退化为

$$\min_{\mathbf{S}} \sum_{i,j=1}^n (\|\mathbf{x}_i^T \mathbf{A} \mathbf{B} - \mathbf{x}_j^T \mathbf{A} \mathbf{B}\|_2^2 s_{i,j} + \gamma \|\mathbf{s}_i\|_2^2). \quad (19)$$

s.t. $\forall i, \mathbf{1}^T \mathbf{s}_i = 1, s_{i,i} = 0.$

$$\begin{cases} s_{i,j} \geq 0, j \in \mathcal{N}(i); \\ s_{i,j} = 0, j \notin \mathcal{N}(i). \end{cases}$$

首先, 利用欧氏距离计算所有训练样本的 k 近邻样本集合, 若样本 j 并不属于样本 i 的近邻集合, 则设置其对应的 $s_{i,j} = 0$. 由于不同的 \mathbf{s}_i 相互独立, 可以并行求解. 因此将式(19)改写为

$$\min_{\substack{\mathbf{1}^T \mathbf{s}_i = 1, \\ s_{i,i} = 0, s_{i,j} \geq 0}} \sum_{j=1}^n (\|\mathbf{x}_i^T \mathbf{A} \mathbf{B} - \mathbf{x}_j^T \mathbf{A} \mathbf{B}\|_2^2 s_{i,j} + \gamma s_{i,j}^2). \quad (20)$$

记矩阵 $\mathbf{G} = [g_1, \dots, g_n] \in \mathbb{R}^{n \times n}$ 中的元素 $g_{i,j} = \|\mathbf{x}_i^T \mathbf{A} \mathbf{B} - \mathbf{x}_j^T \mathbf{A} \mathbf{B}\|_2^2$, 则可以将式(22)改写为

$$\min_{\mathbf{1}^T \mathbf{s}_i = 1, s_{i,i} = 0, s_{i,j} \geq 0} \frac{1}{2} \left\| \mathbf{s}_i + \frac{1}{2\gamma} \mathbf{g}_i \right\|_2^2. \quad (21)$$

记标量 ψ 和向量 $\boldsymbol{\eta} \geq 0$ 分别为约束条件 $\mathbf{1}^T \mathbf{s}_i = 1$ 和 $s_{i,j} \geq 0$ 对应的拉格朗日乘子, 则问题(21)对应的拉格朗日函数为

$$\begin{aligned} \mathcal{L}(\mathbf{s}_i, \psi, \boldsymbol{\eta}) = & \frac{1}{2} \left\| \mathbf{s}_i + \frac{\mathbf{g}_i}{2\gamma} \right\|_2^2 - \psi (\mathbf{1}^T \mathbf{s}_i - 1) - \boldsymbol{\eta}^T \mathbf{s}_i = \\ & \frac{1}{2} \sum_{j=1}^n \left(s_{i,j} + \frac{g_{i,j}}{2\gamma} \right)^2 - \psi \left(\sum_{j=1}^n s_{i,j} - 1 \right) - \sum_{j=1}^n \eta_j s_{i,j}. \end{aligned} \quad (22)$$

通过KKT条件, 可以得到

$$\begin{cases} s_{i,j} + \frac{g_{i,j}}{2\gamma} - \psi - \eta_j = 0, \\ \sum_{j=1}^n s_{i,j} = 1, \\ s_{i,j} \eta_j = 0, \\ \eta_j \geq 0. \end{cases} \quad (23)$$

根据式(23)计算可以得到

$$\begin{aligned} s_{i,j} &= \left(-\frac{g_{i,j}}{2\beta} + \psi \right)_+, \\ \psi &= \frac{1}{k} + \frac{1}{2k\beta} \sum_{j=1}^k g_{i,j}, \end{aligned}$$

其中 $g_{i,1}, g_{i,2}, \dots, g_{i,n}$ 为升序排列的数列. 为了使 \mathbf{s}_i 中有 k 个非零的元素, 可以使 $s_{i,k} > 0$ 和 $s_{i,k+1} = 0$, 进而得到 $-\frac{g_{i,k}}{2\beta} + \psi > 0$ 和 $-\frac{g_{i,k+1}}{2\beta} + \psi \leq 0$, 则有

$$\begin{cases} \gamma > \left(kg_{i,k} - \sum_{j=1}^k g_{i,j} \right) / 2, \\ \gamma \leq \left(kg_{i,k+1} - \sum_{j=1}^k g_{i,j} \right) / 2. \end{cases} \quad (24)$$

令 $\gamma = \frac{kg_{i,k+1} - \sum_{j=1}^k g_{i,j}}{2}$, 则可以得到 $s_{i,j}$ 最终的解析解

$$s_{i,j} = \begin{cases} \frac{g_{i,k+1} - g_{i,j}}{k}, j \leq k; \\ kg_{i,k+1} - \sum_{j=1}^k g_{i,j}, \\ 0, j > k. \end{cases} \quad (25)$$

最终提出的特征选择算法的求解步骤如下所示.

算法2 基于交替优化的多目标回归特征选择.

input: 矩阵 $\mathbf{X} \in \mathbb{R}^{n \times d}$, $\mathbf{Y} \in \mathbb{R}^{n \times q}$, 正则化参数 α, β 和 λ , 近邻参数 k 和低秩参数 r ;

output: $\mathbf{A} \in \mathbb{R}^{d \times r}$, $\mathbf{B} \in \mathbb{R}^{r \times q}$, $\mathbf{S} \in \mathbb{R}^{n \times n}$.

step 1: 依据式(5)得到所有样本的 k 近邻样本集.

step 2: 令 \mathbf{W} 为单位矩阵, 依据式(6)计算近邻样本的相似度, 得到初始化的相似度矩阵 \mathbf{S} .

step 3: repeat

step 3.1: 依据算法1更新矩阵 \mathbf{A} 和 \mathbf{B} ;

step 3.2: 依据式(25)更新相似度矩阵 \mathbf{S} 中的元素 $s_{i,j}$;

step 3.3: 更新相似度矩阵 S 所对应的拉普拉斯矩阵 L ;

until 收敛.

step 4: 计算矩阵 AB 的行向量 $(AB)_i$ 的 ℓ_2 范数, 对其进行降序排列, 即得到输入特征的重要度排序.

3.3 算法复杂度分析

在每次迭代中, 算法 1 的主要计算复杂度来源于对 $P^{-1}X^T H Y B^T (B B^T)^{-1}$ 的计算以及对 Sylvester 方程的求解, 其所对应的计算复杂度分别为 $\max\{\mathcal{O}(r^3), \mathcal{O}(d^3), \mathcal{O}(ndq), \mathcal{O}(dqr)\}$ 和 $\mathcal{O}(q^3)$. 算法 2 的主要计算复杂度来源于对矩阵 G 的计算, 其计算复杂度为 $\max\{\mathcal{O}(n^2 d), \mathcal{O}(n^2 q)\}$. 由于 $r \leq \min(d, q)$, $n, d \gg r, q$, 且实验发现算法 2 在不同数据集上均可以在 30 次迭代内收敛, 所提出的多目标特征选择算法的计算复杂度为 $\max\{\mathcal{O}(d^3), \mathcal{O}(nd^2)\}$.

4 实验与结果分析

4.1 数据集

参考相关文献^[15-16], 在 8 个高维实际数据集上测试所提出的算法, 所有数据集均来自公开数据库网站 Mulan^[27]. 不同数据集的详细信息如表 1 所示, N 表示数据集的样本数量, d 表示样本的输入特征数, q 表示样本的输出目标数, #-fold 表示在此数据集上采用的交叉验证折数. 考虑数据集 Rf1 和 Rf2 中存在一定的数据缺失, 本文采取特征均值进行相应的缺失填充.

表 1 实验所用数据集特征信息

数据集	N	d	q	#-fold
Atp1d	337	411	6	10
Atp7d	296	411	6	10
Oes10	403	298	16	10
Oes97	334	263	16	10
Rf1	9 125	64	8	2
Rf2	9 125	576	8	2
Scm1d	9 803	280	16	2
Scm20d	8 966	61	16	2

4.2 对比方法

本文选择如下所示的不同嵌入式多目标特征选择方法与所提方法进行性能对比, 以验证所提算法的适用性和有效性.

1) MTFs^[28]: 通过 $\ell_{2,1}$ 范数正则化来对权重矩阵施加行稀疏约束.

2) RFS^[29]: 通过对损失函数和权重矩阵施加 $\ell_{2,1}$ 范数正则化来提升多目标特征选择模型的鲁棒性.

3) SSFS^[15]: 通过低维嵌入构建多层回归结构, 并对损失函数、权重和结构矩阵采取联合 $\ell_{2,1}$ 范数正则

化.

4) HLMR-FS^[17]: 在传统稀疏线性回归中引入了超图拉普拉斯正则化项来保持样本间的关联结构, 通过低秩约束寻找不同目标变量间关联结构.

5) LFR-FS^[18]: 通过低秩约束来捕捉不同目标间的相关性, 并通过在损失函数以及回归矩阵上设计 $\ell_{2,p}$ 范数正则化, 利用正交子空间的学习使得多个输出之间共享相同的低秩数据结构.

为检验本文方法的有效性, 首先考虑将未经过特征选择的原始数据集进行 MTR 建模, 并将其作为评价基准, 对比上述主流多目标特征选择方法以及本文所提出的方法在测试数据集上的性能提升. 为保证不同方法在实验中的一致性, 首先利用特征选择方法得到原始数据集中特征的重要度排序, 依据排序来选择相应的特征子集. 然后, 采用多目标核岭回归 (multi-target kernel ridge regression, mKRR)^[30] 方法来衡量特征选择方法在不同目标输出上的性能表现. 对于 mKRR 方法, 采用高斯核函数作为相应的核函数, 且核参数与正则化参数的设置范围为 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$. 对不同的数据集, 选择 70% 的样本作为训练集, 剩下的则作为测试集. 在训练集中通过交叉验证来确定相应的模型最优参数. 考虑部分样本集 (如 Rf1、Rf2、Scm1d 和 Scm20d) 的样本规模较大, 若交叉验证的折数较多则会造成较大的时间成本, 因此其采用两折交叉验证, 其余样本集则采取 10 折交叉验证. 为了保证实验结果的可靠性, 所有数据集上的性能结果均通过 10 次实验后取均值得到.

4.3 评价指标

为了衡量特征选择后 MTR 模型 mKRR 在不同目标上的综合性能表现, 采用如下指标: 平均相对均方根误差 (average relative root mean squared error, aRRMSE)^[1]. 给出 aRRMSE 的计算公式为

$$aRRMSE = \frac{1}{q} \sum_{i=1}^q \sqrt{\frac{\sum_{j=1}^{N_{\text{test}}} (y_i^{(j)} - \hat{y}_i^{(j)})^2}{\sum_{j=1}^{N_{\text{test}}} (y_i^{(j)} - y_i)^2}}, \quad (26)$$

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 j 在目标 i 上的真实值和预测值, y_i 为训练集中样本在第 i 个目标上的平均输出值. aRRMSE 衡量了模型在所有目标输出上的预测结果相对于训练样本均值的均方误差, 其越小说明模型在测试集上的性能表现越好.

4.4 实验结果分析

图 1 显示了不同数据集上不同特征选择方法的 aRRMSE 值. 依据特征选择方法得到的特征排序来

选择特征子集. Atp1d和Atp7d选择60、70、80、90、100、110个特征进行多目标回归建模; Oes10、Rf2、Scm1d选择60、80、100、120、140、160个特征进行回归建模; Oes97选择40、60、80、100、120、140个特征进行建模; Rf1选择10、15、20、25、30、35个特征进行回归建模; Scm20d选择20、25、30、35、40、45个特征建模. Baseline表示采用原始特征进行回归建模.

同时,为了验证本文所提方法和其他对比方法在数据集上的性能结果是否存在显著性差异,依据特征选择方法对不同数据集上的最优aRRMSE值进行排序,进而计算得到不同方法在所有数据集上的平均排名. 然后,通过显著性水平 $\alpha = 0.05$ 的Friedman检

验^[31],其中原假设表示本文所提方法与比较方法之间的性能并无显著性差异. 若拒绝原假设,则进一步利用Bonferroni-Dunn检验^[32]作为后验检验来进一步分析和比较不同算法在排名上的差异性,并计算对应的临界差异(critical difference, CD)作为不同方法之间是否具有差异性的评判标准,其计算公式如下:

$$CD = q_\alpha \sqrt{\frac{n(n+1)}{6T}}. \quad (27)$$

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,对应的 $q_\alpha = 2.498$,因此有 $CD = 2.41(n = 6, T = 8)$. 基于aRRMSE指标所对应的排名如图2所示.

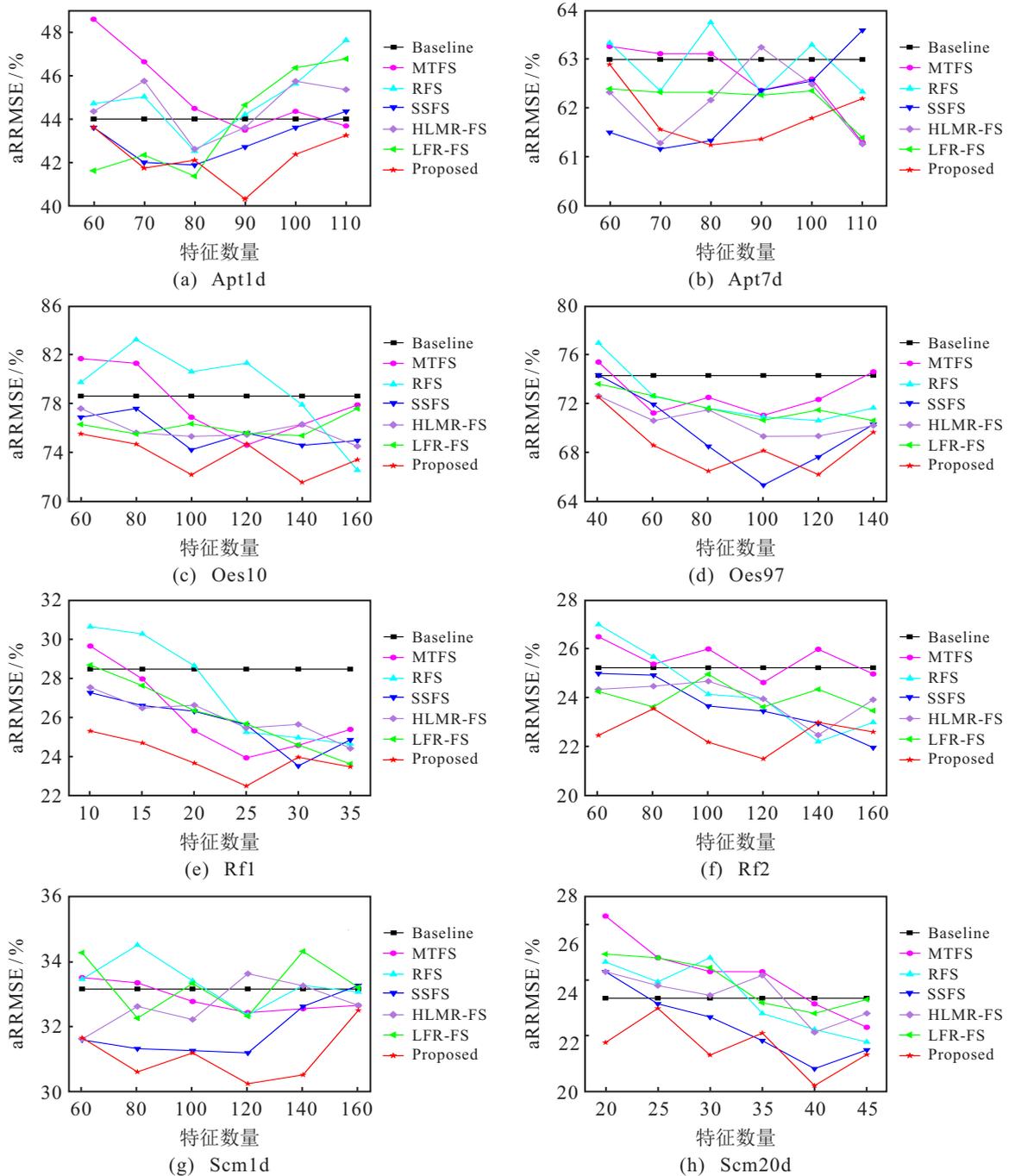


图1 不同数据集上特征选择方法的aRRMSE

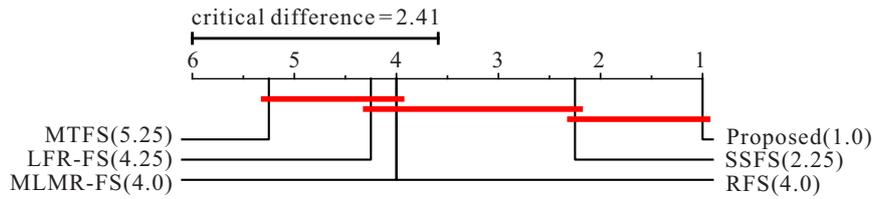


图2 Bonferroni-Dunn 检验下基于 aRRMSE 的算法排名

从图1和图2可以看出,对于不同数据集,选择数量合适的特征子集可以获得较原始特征集更好的效果,这说明对于多目标回归问题,通过有效的特征选择方法不仅可以提升模型的计算效率,同时可以提升模型在不同输出上的综合性能. 随着选择的特征规

模增加,模型的回归性能并不一定会得到提升. 相反,多数情形下,如Oes97、Rf1、Scm20d等数据集,回归性能反而会随着选择特征的数量增加而下降. 这说明在原始特征集中存在一定的冗余特征或噪声,会显著地降低回归建模的性能.

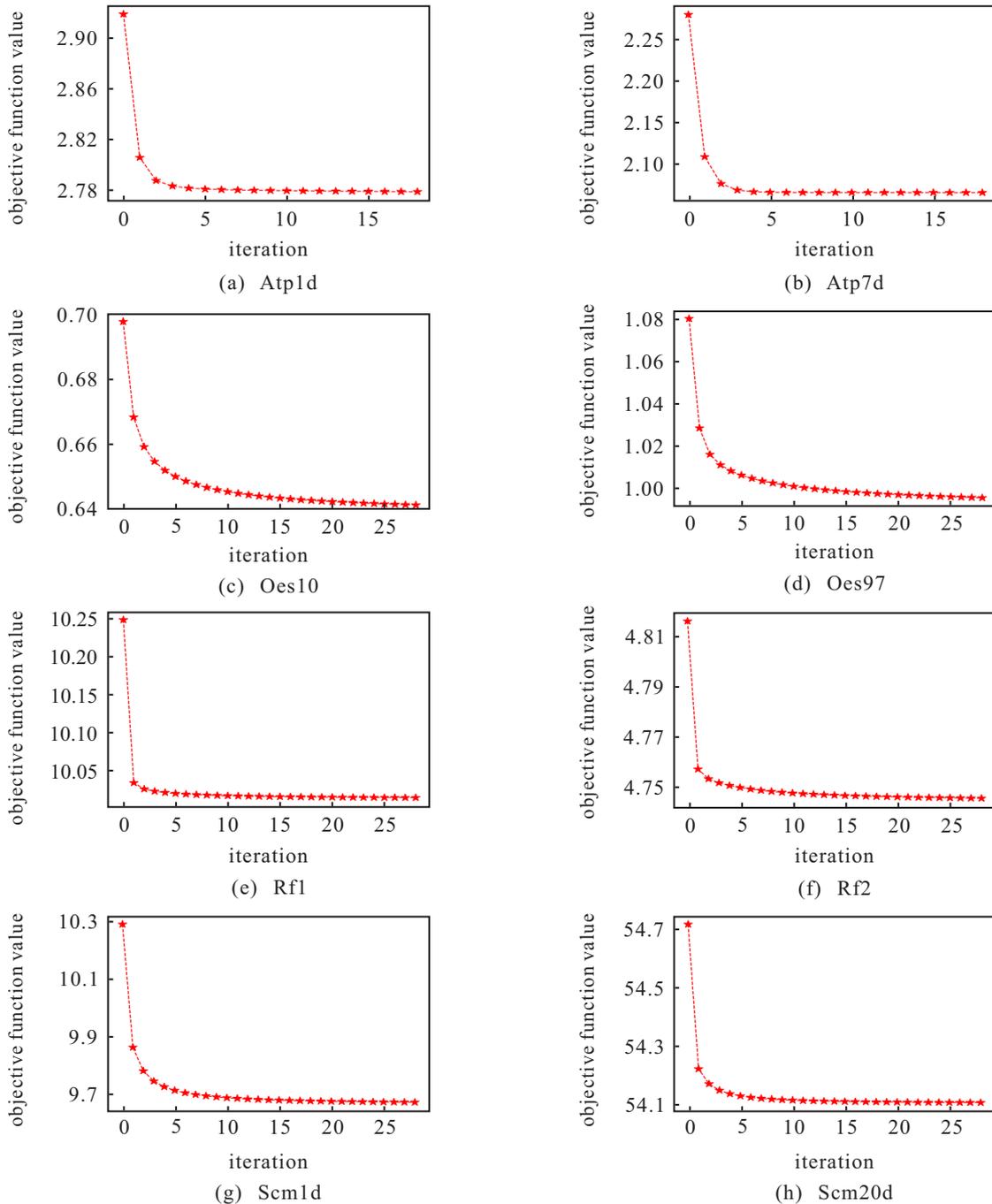


图3 不同数据集下算法的收敛曲线

在多数情形下, SSFS、HLMR-FS、LFR-FS和所提方法较MTFS和RFS可以获得更低的aRRMSE值,这说明对于多目标特征选择而言,利用低秩约束回归系数矩阵可以在一定程度上提升多目标特征选择的性能. 相较于其他方法,本文所提方法能够保证其在多数情形下表现出较好的回归性能. 这得益于所提方法不仅利用低秩约束来降低冗余特征对回归模型的影响,其同时考虑了不同样本在特征输入空间中的结构信息和不同输出之间的相似信息,进而在特征学习的过程中可以更加充分地利用训练集原始特征中的内在信息来提升多目标回归的性能. 所提方法无论特征数量如何变化,其性能可以在多数情形下优于基于完整的原始特征建模的性能. 这说明所提方法在进行特征选择和排序时可以很好地考虑冗余特征所带来的影响,即使有部分冗余特征被纳入选择的特征子集时,在其挑选的特征子集上仍然可以维持较好的性能表现.

4.5 算法收敛性分析

类似于其他方法研究^[15,17-18,33],本文在对模型算法的收敛性进行理论证明后,绘制了算法在不同数据集上进行迭代更新时的目标函数收敛曲线,如图3所示. 可以看出Atp1d、Atp7d和Rfl可以在20次迭代内收敛至最优,其余数据集则可以在30次迭代内收敛,且在前几次迭代中目标函数收敛得非常迅速. 这说明所提交替优化算法在求解目标函数的过程中可以快速收敛和求解,在不同领域的数据集上具有较好的适应性,这进一步说明了所提多目标特征选择方法的有效性.

5 结论

本文提出了一种基于自适应图学习的嵌入式特征选择算法,针对MTR进行有监督特征选择. 为了更加准确地选择出合理的特征子集,本文所提出的特征选择算法同时考虑了原始数据集中的3种关系:样本与样本之间的关系,输入与目标间的关系以及目标之间的关系. 考虑原始数据集中存在特征冗余和噪声,传统基于热核函数度量样本相似度的方法难以学习到正确的样本相似结构,在传统的稀疏低秩MTR模型中引入了基于样本局部结构信息的自适应图学习项;同时,为了更好地探索目标间的相关关系,构建了基于输出相关性的结构矩阵优化项,进而使得回归矩阵在稀疏低秩的学习过程中能够更加充分地考虑目标间相关性. 通过充分的实验,验证了本文所提方法相较于其他主流的多目标特征选择嵌入式算法具有更加优越的性能表现.

参考文献(References)

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Ghosn J, Bengio Y. Multi-task learning for stock selection[C]. Proceedings of the 9th International Conference on Neural Information Processing Systems. Denver, 1996: 946-952.
- [8] Chen B J, Chang M W, Lin C J. Load forecasting using support vector Machines: A study on EUNITE competition 2001[J]. IEEE Transactions on Power Systems, 2004, 19(4): 1821-1830.
- [9] He X, Deng C, Niyogi P. Laplacian score for feature selection[C]. Proceedings of the 18th International Conference on Neural Information Processing Systems. Vancouver, 2005: 507-514.
- [10] Gu Q, Li Z, Han J. Generalized fisher score for feature selection[C]. Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence. Barcelona, 2011: 266-273.
- [11] 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. Berlin, 2014: 225-240.
- [12] 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.
- [13] 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.
- [14] Huang Y Y, Shen Z X, Cai F X, et al. Adaptive graph-based generalized regression model for unsupervised feature selection[J]. *Knowledge-Based Systems*, 2021, 227: 107156.
- [15] Yuan H L, Zheng J J, Lai L L, et al. Sparse structural feature selection for multitarget regression[J]. *Knowledge-Based Systems*, 2018, 160: 200-209.
- [16] Sechidis K, Spyromitros-Xioufis E, Vlahavas I. Information theoretic multi-target feature selection via output space quantization[J]. *Entropy*, 2019, 21(9): 855.
- [17] 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.
- [18] Zhang S C, Yang L F, Li Y G, et al. Low-rank feature reduction and sample selection for multi-output regression[C]. *International Conference on Advanced Data Mining and Applications*. Cham: Springer, 2016: 126-141.
- [19] Hashemi A, Dowlatshahi M B, Nezamabadi-pour H. VMFS: A VIKOR-based multi-target feature selection[J]. *Expert Systems With Applications*, 2021, 182: 115224.
- [20] Petković M, Kocev D, Džeroski S. Feature ranking for multi-target regression[J]. *Machine Learning*, 2020, 109(6): 1179-1204.
- [21] Fan Y L, Chen B H, Huang W Q, et al. Multi-label feature selection based on label correlations and feature redundancy[J]. *Knowledge-Based Systems*, 2022, 241: 108256.
- [22] Xu J H, Liu J L, Yin J, et al. A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously[J]. *Knowledge-Based Systems*, 2016, 98: 172-184.
- [23] Zhang J, Luo Z M, Li C D, et al. Manifold regularized discriminative feature selection for multi-label learning[J]. *Pattern Recognition*, 2019, 95: 136-150.
- [24] Zhu X F, Zhang S C, Hu R Y, et al. Local and global structure preservation for robust unsupervised spectral feature selection[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 30(3): 517-529.
- [25] Zhang S C, Yang L F, Deng Z Y, et al. Leverage triple relational structures via low-rank feature reduction for multi-output regression[J]. *Multimedia Tools and Applications*, 2017, 76(16): 17461-17477.
- [26] He X, Niyogi P. Locality preserving projections[C]. *Proceedings of the 16th International Conference on Neural Information Processing Systems*. Whistler, 2003: 153-160.
- [27] 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.
- [28] Liu J, Ji S W, Ye J P. Multi-task feature learning via efficient $\ell_{2,1}$ -norm minimization[C]. *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. Montreal, 2009: 339-348.
- [29] Nie F P, Huang H, Cai X, et al. Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization[C]. *Proceedings of the 23rd International Conference on Neural Information Processing Systems—Volume 2*. Vancouver, 2010: 1813-1821.
- [30] Shawe-Taylor J, Cristianini N. *Kernel Methods for Pattern Analysis*[M]. Cambridge: Cambridge University Press, 2004.
- [31] Garcia S, Herrera F. An extension on “statistical comparisons of classifiers over multiple data set” for all pairwise comparisons[J]. *Journal of Machine Learning Research*, 2008, 9: 2677-2694.
- [32] Demsar J. Statistical comparisons of classifiers over multiple data sets[J]. *Journal of Machine Learning Research*, 2006, 7: 1-30.
- [33] 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.

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、装备价格管理等研究, E-mail: sxx13397129808@163.com;

梁新(1979—),男,教授,博士生导师,从事系统工程、装备财务管理研究, E-mail: whliangxin@163.com;

谢力(1980—),男,副教授,博士,从事系统工程、装备价格管理等研究, E-mail: xieli_hg@163.com;

张侃(1985—),男,讲师,博士,从事复杂系统分析、建模与仿真研究, E-mail: 719273356@qq.com.