doi:10.11887/j.cn.201905018 http://journal.nudt.edu.cn

交互式 Pareto 前沿可视化决策*

胡佳鑫,杨乐平
(国防科技大学 空天科学学院,湖南 长沙 410073)

摘 要:针对现有高维多目标可视化技术不能有效显示决策偏好信息这一难题,提出一种基于径向轴图的交互式 Pareto 前沿可视化决策方法。进行标准化适应度函数处理,消除量纲差距;根据交互更新的轴向量矩阵建立最小二乘问题模型,通过 Moore-Penrose 广义逆矩阵方法求解投影点,从而获得 Pareto 前沿分布图。该方法能够有效显示目标各维度信息,直观呈现决策偏好,动态表现目标性能趋势。通过实验与传统可视化技术对比分析,验证了该方法的有效性和实用性。

关键词:高维多目标可视化;Pareto 前沿;决策偏好
中图分类号:TP18 **文献标志码**:A **文章编号**:1001-2486(2019)05-128-06

Interactive visualization decision-making method of Pareto front

HU Jiaxin, *YANG Leping*
(College of Aeronautics and Astronautics, National University of Defense Technology, Changsha 410073, China)

Abstract: Considering the inability of current visualization technology of high dimension multi-objective to effectively display preference information of decision-making, a new graphical visualization method of interactive n-dimension Pareto front method based on radial axes plots was proposed. Firstly, the fitness functions were standardized to eliminate dimensional gaps. Secondly, the least square problem model was established according to the interactively updated axes matrix and solved by the Moore-Penrose pseudoinverse method. Sequentially the diagram of dynamic Pareto front distribution was obtained, which effectively shows the data information of each dimension objective, the variation tendency towards performance and the preference information for decision-makers. Finally, the visualization model is devised and the extensive experimental analysis is provided to investigate the Pareto optimal solutions for decision-makers.

Keywords: visualization technology of high dimension multi-objective; Pareto front; preference of decision-making

在高维多目标优化问题中,目标向量构成了多目标优化问题的非劣最优目标域,称为 Pareto 前沿。一般地,各子目标之间存在复杂的冲突性,决策者需要深度发掘 Pareto 前沿特性并结合实际需求作出最终选择。高维多目标可视化技术将 Pareto 前沿投影至低维观测空间,提供用户直观有效的决策辅助,广泛应用于数据挖掘、决策分析、任务规划以及多学科优化设计等领域,成为高维多目标优化问题的研究热点之一[1]。

为辅助决策者正确分析多维目标信息,多目标可视化技术应满足直观、有效以及简单等特点。目前,高维多目标可视化问题的研究成果主要分两类:

一类是完好无损地表现目标各维度信息,保证信息的不缺失。平行坐标系[2]与热图[3]是目前广泛应用的多目标可视化方法,其特点是将所有目标信息通过单一的视图呈现,虽然效果直观,但高维度空间势必会引起视觉混乱。散点图[4-5]通过将目标的各维度信息两两组合成图表进行对比分析,图表的数量随维度增长呈指数级增加,对于决策者在实际应用中十分不便。n 维图表[6]是基于决策偏好的 Pareto 前沿可视化方法,该方法提出一种目标全局信息的共享机制,采用多图表分别绘制权重分配下各维度信息与全局信息的对比结果,能够有效地反映目标信息与决策偏好。但是,该方法的权重分配对于同时多个目标偏好的分层效果不佳。

另一类是对原始数据进行压缩降维,然后投影至低维观测空间进行可视化分析,如主成分图(Principal Component Biplots, PCB)[7]、星图[8]、基于分形的降维方法以及旋转其可视化方法[9]等。PCB 是目前国外对于高维数据可视化非常

* **收稿日期**:2018-05-09
基金项目:国防科技大学科研计划资助项目(ZK18-03-34)
作者简介:胡佳鑫(1989—),男,安徽铜陵人,博士研究生,E-mail:joehu1989@163.com;
杨乐平(通信作者),男,教授,博士,博士生导师,E-mail:ylp_1964@163.com

有效的方法,该方法通过对原始数据矩阵进行奇异值分解,提取高维数据聚类特征后降维映射至低维观测空间完成可视化,具有较高的数据压缩质量。但是,该方法得到静态的轴向量与投影点,不能满足实际应用场景中决策者根据偏好调整参数的交互需求。

文献[11]提出了一种增强高维数据可视化效果的自适应径向轴图方法,根据交互更新的轴向量矩阵建立降维映射模型,对模型求解获得低维投影点,可视化效果直观有效。鉴于此思想,本文利用径向轴图的交互优势,提出一种基于径向轴图的交互式 Pareto 前沿可视化决策方法。首先,介绍了径向轴图的基本原理,径向轴图较好地降低了数据降维映射带来的信息损失。然后,阐述了如何基于径向轴图实现 Pareto 前沿可视化决策,根据用户交互即时更新视图,辅助决策者正确把握目标性能趋势。最后,通过仿真实验以及与传统可视化技术的对比,证明了本文方法能够将决策偏好信息与 Pareto 前沿分布有效地结合显示,辅助决策者正确把握目标性能趋势,筛选出满足实际需求的可行方案。

1　Pareto 前沿可视化问题

设 $\boldsymbol{X}_f$ 为多目标优化问题的可行解集,$\boldsymbol{F}(\boldsymbol{x})=(f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x}))$ 为目标向量,所有 Pareto 非支配解的集合构成该多目标优化问题的 Pareto 非支配解集 $\boldsymbol{P}^*$,如式(1)所示[1]。

$$\boldsymbol{P}^*=\{\boldsymbol{x}\in\boldsymbol{X}_f\mid\exists\boldsymbol{x}'\in\boldsymbol{X}_f:\boldsymbol{x}'\succ\boldsymbol{x}\}\tag{1}$$

其中,$\boldsymbol{x}'\succ\boldsymbol{x}$ 表示 $\boldsymbol{x}'$Pareto 支配 $\boldsymbol{x}$,简称支配。

$\boldsymbol{P}^*$ 的目标向量构成了多目标优化问题的非劣最优目标域,即 Pareto 最优前沿,如式(2)所示。

$$\boldsymbol{PF}=\{f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x})\mid\boldsymbol{x}\in\boldsymbol{P}^*\}\tag{2}$$

Pareto 前沿可视化问题主要解决将 m 维 Pareto 最优前沿 $\boldsymbol{PF}$ 降维投影至观测空间。本文仅考虑 Pareto 前沿在二维平面上的投影问题。

2　径向轴图原理

径向轴图方法通过将高维数据空间向低维观测空间投影变换来提供可视化决策功能。为方便讨论,本文考虑 n 维数据样本投影至 m 维观测空间,$n\geqslant3\geqslant m$,$m=2$。数据样本点表示为 $\boldsymbol{x}^{\mathrm{d}}\in\mathbb{R}^n$,$N$ 个 $\boldsymbol{x}^{\mathrm{d}}$ 组成 $N\times n$ 数据样本矩阵 $\boldsymbol{X}^{\mathrm{d}}$。每一个数据样本对应一个投影点 $p\in\mathbb{R}^m$,$N\times m$ 投影点矩阵 $\boldsymbol{P}$。$\boldsymbol{X}^{\mathrm{d}}$ 由 p 与 n 个 m 维轴向量 $\boldsymbol{v}_i\in\mathbb{R}^m$,$i=1,\cdots,n$表示,$n$ 个 $\boldsymbol{V}_i$ 组成 $n\times m$ 径向轴矩阵 $\boldsymbol{V}$。

对于降维投影方式的高维数据可视化问题,重点在于解决如何减小数据降维映射带来的信息损失。高维数据可视化的降维映射问题可描述为:

$$\min_{\boldsymbol{P}\in\mathbb{R}^{N\times m},\boldsymbol{V}\in\mathbb{R}^{n\times m}}\left\|\boldsymbol{P}\boldsymbol{V}^{\mathrm{T}}-\boldsymbol{X}^{\mathrm{d}}\right\|_F^2\tag{3}$$

其中,F 表示 Frobenius 范数,本文考虑通过欧几里得距离评估数据降维的损失。F - 范数是由所有奇异值组成的向量 2 - 范数,数值等于全部元素平方和的平方根。由于 $N\times n$ 数据矩阵各行数据样本 $\boldsymbol{x}^{\mathrm{d}}$ 是相互独立的,因此问题可转化为:

$$\min_{p\in\mathbb{R}^m}\left\|\boldsymbol{p}\boldsymbol{V}^{\mathrm{T}}-\boldsymbol{x}^{\mathrm{d}}\right\|^2\tag{4}$$

对于上式,径向轴图方法首先考虑提供决策者旋转、缩放径向轴功能,即轴向量的方向和模是任意的。该问题就转换成如何根据特定的轴向量矩阵 $\boldsymbol{V}$ 求解投影点 $\boldsymbol{p}$,使得 $\boldsymbol{p}\boldsymbol{V}^{\mathrm{T}}$ 到 $\boldsymbol{x}^{\mathrm{d}}$ 的欧式距离比到子空间 $\mathrm{span}\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 中其他向量的欧式距离都短,即高维最小二乘问题。

一般情况下,$n\times m$ 轴向量矩阵 $\boldsymbol{V}$ 是列满秩的,采用 Moore-Penrose 广义逆矩阵方法求解该问题,其解的形式如式(5)所示。

$$\boldsymbol{p}=\boldsymbol{V}^{\dagger}\boldsymbol{x}^{\mathrm{d}}=(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{V}^{\mathrm{T}}\tag{5}$$

由上式可知,$\boldsymbol{V}^{\dagger}$ 可视为 $\boldsymbol{V}$ 的线性变换,投影点的求解过程可以在线性时间内完成,满足了交互式决策的实时性要求。

3　交互式 Pareto 前沿可视化决策方法实现

基于径向轴图的交互式 n 维 Pareto 前沿可视化实现需要考虑响应用户交互以及视图刷新问题。现阶段主流编程平台均能满足要求,例如 C ++ 、Java、C#等。由于篇幅有限,本文不赘述详细实现代码,主要给出交互式 n 维 Pareto 前沿可视化方法的具体实现步骤,方法流程如图 1 所示。

1)针对 Pareto 非支配解集 $\boldsymbol{P}^*$ 中可行解各目标之间可能存在较大的数值差距,首先需要对各目标向量 $\boldsymbol{f}_i(\boldsymbol{x})$($i\in\{1,\cdots,r\}$)进行归一化处理,得到 $k_i(\boldsymbol{x})$,如式(6)所示。

$$k_i(\boldsymbol{x})=\frac{f_i(\boldsymbol{x})-f_i^{\min}}{f_i^{\max}-f_i^{\min}}\tag{6}$$

式中,$f_i^{\max}=\max\limits_{\boldsymbol{x}\in\boldsymbol{X}_f}f_i(\boldsymbol{x})$,$f_i^{\min}=\min\limits_{\boldsymbol{x}\in\boldsymbol{X}_f}f_i(\boldsymbol{x})$,$i=1,\cdots,r$。

2)初始化轴向量矩阵 $\boldsymbol{V}$,均匀分布各轴向量初始朝向,初始长度统一设为单位长度,如式(7)所示。

$$v_i=\left\{\cos\left[\frac{2\pi(i-1)}{n}\right],\sin\left[\frac{2\pi(i-1)}{n}\right]\right\}\tag{7}$$

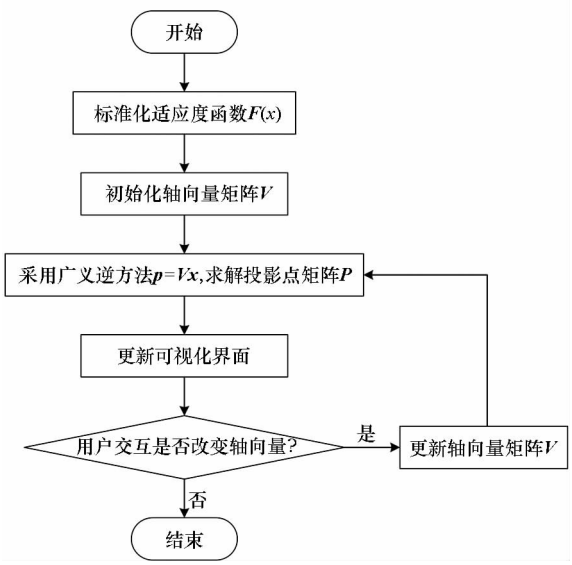


图1 方法流程图

Fig. 1 Method flow diagram

3)设由 $k_i(\boldsymbol{x})=(k_1(\boldsymbol{x}),k_2(\boldsymbol{x}),\cdots,k_n(\boldsymbol{x}))$, $i=1,2,\cdots,N$ 组成了数据矩阵 $\boldsymbol{X}^k$。获得数据矩阵 $\boldsymbol{X}^k$ 与轴向量矩阵 $\boldsymbol{V}$ 后,根据径向轴图原理,采用广义逆矩阵方法求解投影点 $\boldsymbol{p}$,即二维平面投影点坐标。

4)基于通用软件编程平台,根据投影点坐标绘制 n 维 Pareto 前沿信息。

5)根据决策者对可视化视图中各轴向量的交互操作,即对特定轴向量进行旋转、缩放,从而获得新的轴向量矩阵 $\boldsymbol{V}$。返回步骤3重新计算投影点矩阵。

基于径向轴图的交互式 n 维 Pareto 前沿可视化效果如图2所示。

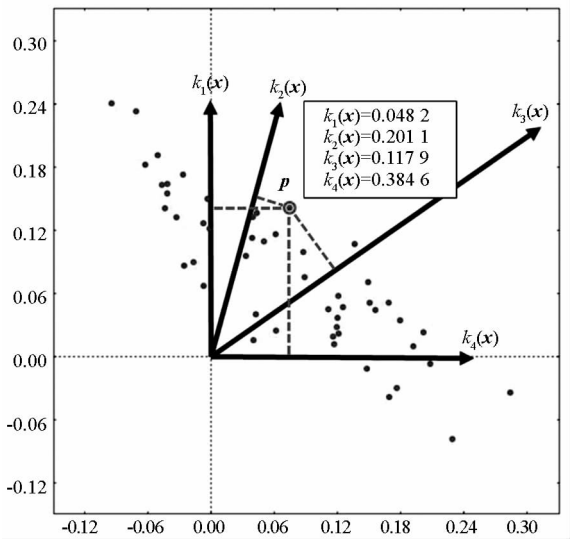


图2 基于径向轴图的交互式 Pareto 前沿可视化效果

Fig. 2 Interactive visualization of pareto front based on radial axes plots

图2表示规模为50的 Pareto 前沿分布。其中,$\boldsymbol{P}$ 点在各轴上的分量表示 $k_i(\boldsymbol{x})$ 子目标估计值。该目标域降维后的总离差等于1.17,估计值较为准确,满足决策分析可行性要求。通过旋转、拉伸或者缩短任一轴向量,视图动态更新投影点分布,直观反映决策偏好。同时,该方法支持决策者点击查看关注目标的各维度信息,准确把握各目标性能趋势,辅助决策者筛选出符合实际需求的最终方案。

4 仿真实验与结果分析

为验证基于径向轴图的交互式高维 Pareto 前沿可视化方法能够满足实际问题中的多目标决策需求。本文采用 DTLZ1 函数进行实验,并与目前主流应用的平行坐标系、散点图以及基于偏好的 n 维图表可视化技术进行对比分析,进一步验证本文方法的有效性以及优越性。DTLZ1 函数为国际通用的无约束多目标优化标准测试函数。最小化目标函数如下:

$$\begin{cases} f_1(\boldsymbol{x})=0.5x_1x_2x_3x_4(1+g(\boldsymbol{x}_M)) \\ f_2(\boldsymbol{x})=0.5x_1x_2x_3(1-x_4)(1+g(\boldsymbol{x}_M)) \\ f_3(\boldsymbol{x})=0.5x_1x_2(1-x_3)(1+g(\boldsymbol{x}_M)) \\ f_4(\boldsymbol{x})=0.5x_1(1-x_2)(1+g(\boldsymbol{x}_M)) \\ f_5(\boldsymbol{x})=0.5(1-x_1)(1+g(\boldsymbol{x}_M)) \\ g(\boldsymbol{x}_M)= \\ 100\left\{|\boldsymbol{x}_M|+\sum_{x_i\in x_M}(x_i-0.5)^2-\cos[20\pi(x_i-0.5)]\right\} \end{cases}$$

式中,$n=9$,$x_i\in[0,1]$。假设通过优化算法得到了 Pareto 非支配解集,种群为300。现采用主流的可视化方法对 Pareto 前沿进行绘制。其中,平行坐标系显示结果如图3所示,散点图显示结果如图4所示,考虑权值为 $\omega_1=0.1$,$\omega_2=0.1$,$\omega_3=0.8$,$\omega_4=0.8$,$\omega_5=0.1$ 的1-w-norm 基准 n 维图表显示结果如图5所示。

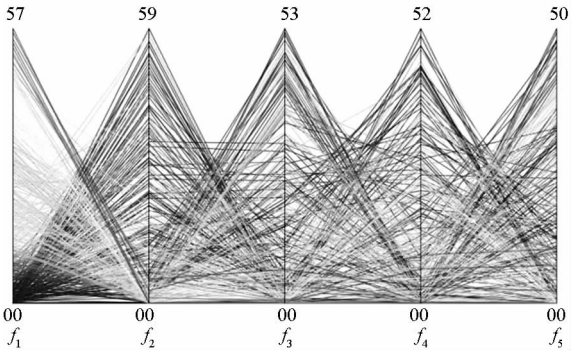


图3 DTLZ1 Pareto 前沿平行坐标系可视化展示

Fig. 3 DTLZ1 Pareto front parallel coordinates plot visualization

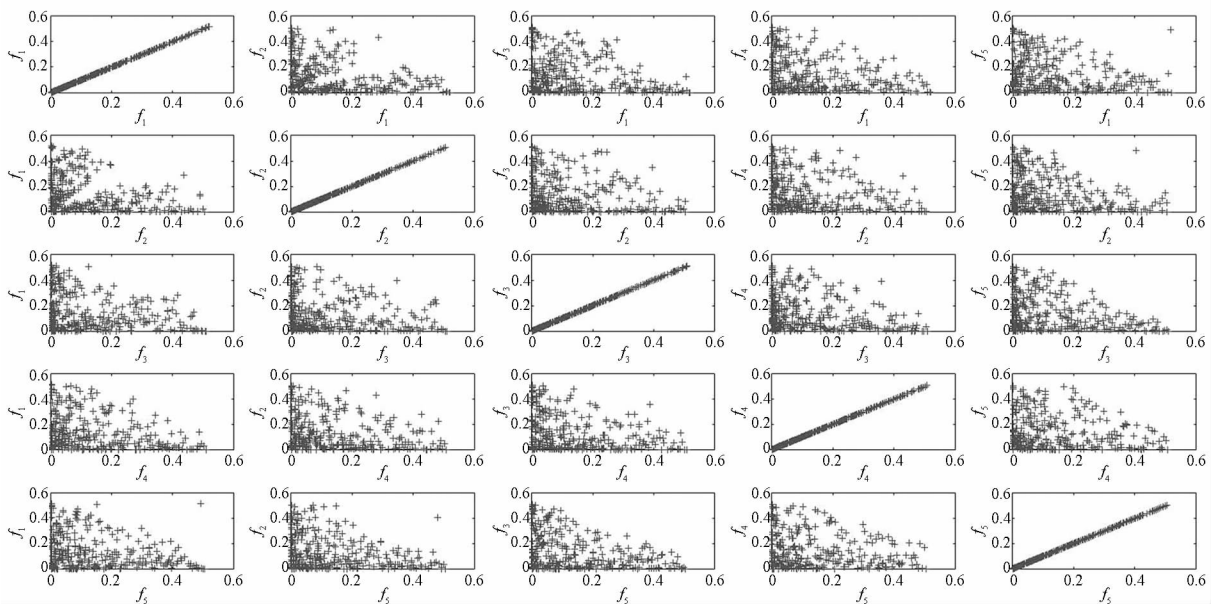


图 4　DTLZ1 Pareto 前沿散点图可视化展示

Fig. 4　DTLZ1 Pareto front scatter diagram visualization

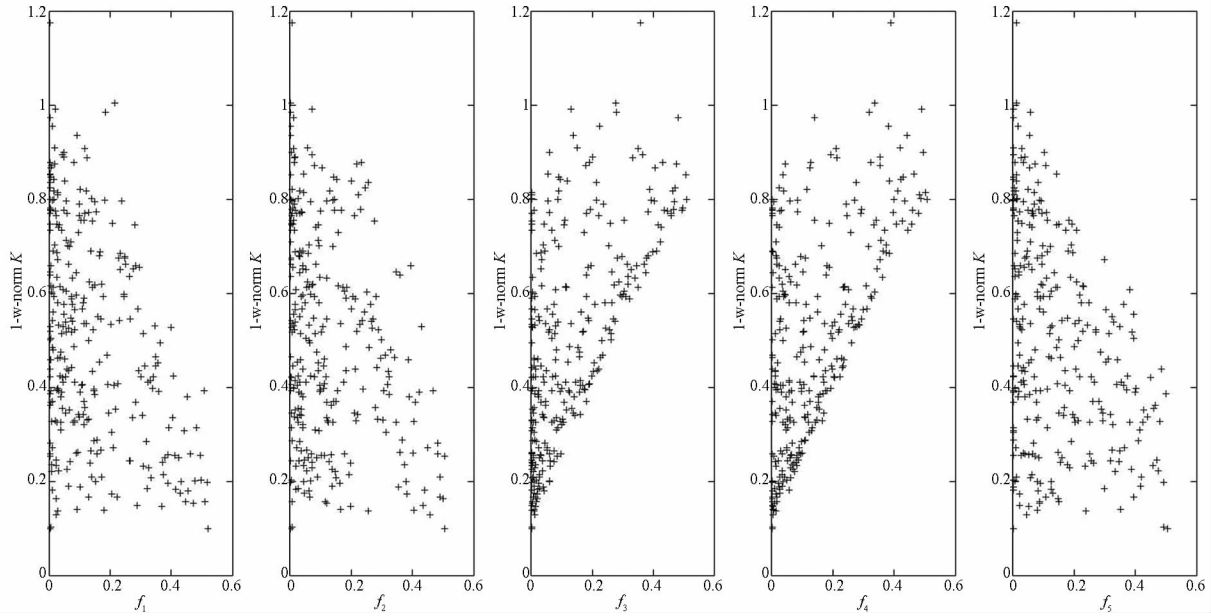


图 5　DTLZ1 Pareto 前沿 1-w-norm 图表可视化展示

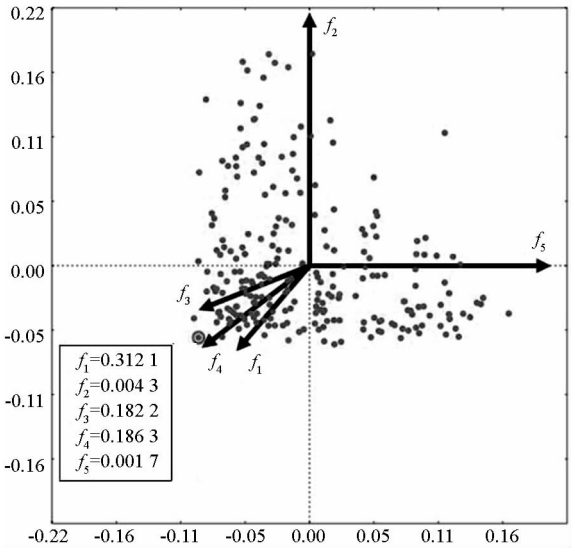
Fig. 5　DTLZ1 Pareto front 1-w-norm diagram visualization

由图 3 可以看出,最优解集的目标向量通过各轴之间的连线表示,决策者能够指定轴进行排序并通过颜色区分,例如 f_1 的取值范围为 0 ~ 0.5,其中浅色部分表示取值在 0 ~ 0.05 范围内分布密集。但是,浅色连线在其他子目标适应度函数上未呈现明显聚类特征。由于各连线交叉重叠,实际应用中决策者很难甄别出各方案的优劣。

图 4 将目标向量的各维度信息两两组合,绘制出 20 个二维图表。决策者需要分析目标向量在所有散点图表中的性能趋势。由于图表数量过多,势必会引起视觉混乱,无法有效辅助决策者进行分析决策。

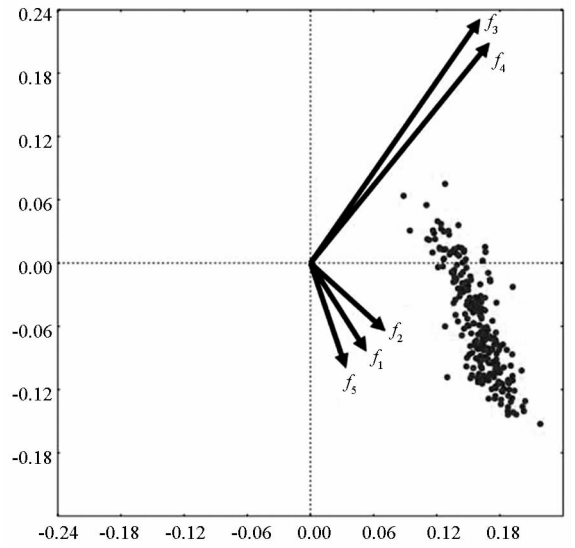
图 5 通过目标函数的共享机制集成了基于权重的决策偏好信息,在某一权重明显大于其他权重时分层效果较为明显。本实验考虑分配权重 $\omega_1 = 0.1, \omega_2 = 0.1, \omega_3 = 0.8, \omega_4 = 0.8, \omega_5 = 0.1$,以测试两类决策偏好共同作用的效果。由图 5 可以看出,f_3、f_4 的取值虽然整体随纵轴递增而增大,但未形成明显的分层效果,无法继续筛选出合适方案。

图6为基于径向轴图的交互式 n 维 Pareto 前沿可视化效果，各径向轴表示实验算例的各子目标函数，在图6中由对应投影点表示，投影点在各轴上的分量表示其子目标估计值。将 f_2、f_5 轴调整为相互正交，并拉伸轴向量长度，表示当前决策偏好于 f_2、f_5。其他轴旋转至其反方向，得到投影点分布情况如图6(a)所示，获得的最左下角点代表某方案 $f_1=0.312\ 1$，$f_2=0.004\ 3$，$f_3=0.182\ 2$，$f_4=0.186\ 3$，$f_5=0.001\ 7$。通过数据校验，该方案确实为最优解集中满足 f_2、f_5 最小的。将轴 f_3、f_4 拉伸并旋转至相近角度，其他轴缩小长度并旋转至相对正交的角度，得到 Pareto 前沿投影点集如图6(b)所示，表示偏好于 f_3、f_4 的最小化取值方



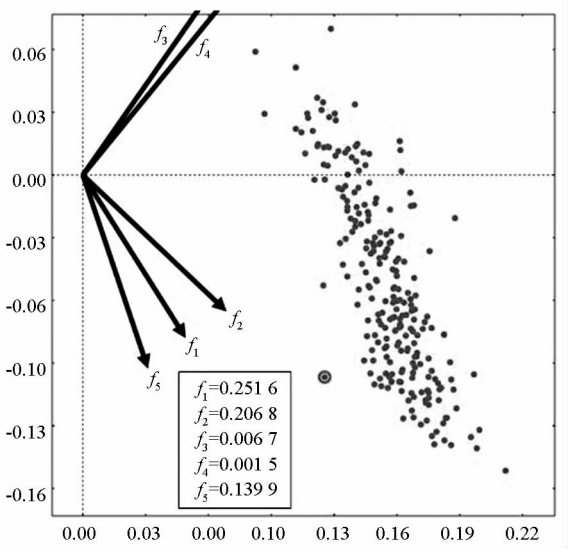
(a) 决策偏好于 f_2 与 f_5 轴的投影点分布

(a) Projection point distribution of decision preference in f_2 and f_5 axises



(b) 拉伸 f_3、f_4 轴后的投影点分布

(b) Projection point distribution after stretching of f_3 and f_4 axises



(c) 图缩放后的投影点分布

(c) Projection point distribution after scaling of figure(b)

图6 基于径向轴图的交互式 n 维 Pareto 前沿可视化效果展示

Fig. 6 Interactive visualization of n-dimensional Pareto front based on radial axes plots

案。由于投影点集分布情况比较密集，通过对整体视图的缩放和平移操作，得到投影点集分布情况如图6(c)所示。沿 f_3、f_4 轴方向寻找最低的一点，表示某方案 $f_1=0.251\ 6$，$f_2=0.206\ 8$，$f_3=0.006\ 7$，$f_4=0.001\ 5$，$f_5=0.139\ 9$。通过数据校验，该方案确实为最优解集中满足 f_3、f_4 最小的。实验证明，基于径向轴图的交互式 n 维 Pareto 前沿可视化决策方法效果直观有效。

5 结论

本文提出了一种基于径向轴图的交互式 n 维 Pareto 前沿可视化决策方法，解决了集成决策偏好的高维多目标可视化问题。通过实验与平行坐标系、散点图以及基于权重的 n 维图表等主流多目标可视化方法进行对比分析，证明了该方法能够以形象直观的方式呈现 Pareto 前沿各维度信息。基于径向轴图的交互式 n 维 Pareto 前沿可视化决策方法不仅有效集成了决策偏好信息，并且能够根据用户交互即时更新视图，辅助决策者正确把握目标性能趋势。并且，该方法易于实现，在数据挖掘、决策分析、任务规划以及多学科优化设计等领域都具有重要的应用价值。

参考文献(References)

[1] 李博. 高维多目标可视化技术的研究及应用[D]. 哈尔滨：哈尔滨工程大学，2013.
LI Bo. Research and application on high-dimensional multi-

objective visualization technology [D]. Harbin Engineering University, 2013. (in Chinese)

[2] Ortiz M C, Sarabia L A, Sánchez M S, et al. Improving the visualization of the pareto-optimal front for the multi-response optimization of chromatographic determinations[J]. Analytica Chimica Acta, 2011, 687(2): 129 – 136.

[3] Pryke A, Mostaghim S, Nazemi A. Heatmap visualization of population based multi objective algorithms [C]// Proceedings of International Conference on Evolutionary Multi-Criterion Optimization, 2006: 199 – 224.

[4] Stump G, Lego S, Yukish M, et al. Visual steering commands for trade space exploration: user guided sampling with example [J]. Journal of Computing and Information Science in Engineering, 2009, 9(4): 044501.

[5] Zou X F, Chen Y, Liu M Z, et al. A new evolutionary algorithm for solving many-objective optimization problems[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2008, 38(5): 1402 – 1412.

[6] 毕晓君, 李博. 基于决策偏好信息的 n 维 Pareto 前沿可视化[J]. 系统工程与电子技术, 2013, 35(3): 557 – 563.
BI Xiaojun, LI Bo. New graphical visualization of n-dimensional Pareto front based on decision preference information[J]. Systems Engineering and Electronics, 2013, 35(3): 557 – 563. (in Chinese)

[7] Gabriel K R. The biplot graphic display of matrices with application to principal component analysis[J]. Biometrika, 1971, 58: 453 – 467.

[8] Rubio-Sánchez M, Sanchez A. Axis calibration for improving data attribute estimation in star coordinates plots[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 2013 – 2022.

[9] Yamamoto M, Yoshikawa T, Furuhashi T. A study on interactive search in MOGA with island model based on user's requirements using visualization in NSP[C]//Proceedings of Second World Congress on Nature and Biologically Inspired Computing, 2010: 490 – 495.

[10] Rubio-Sánchez M, Sánchez A. Adaptable radial axes plots for improved multivariate data visualization [J]. Computer Graphics Forum, 2017, 36(3): 389 – 399.

[11] 毕晓君, 李博, 苍岩. 基于交互式决策的旋转基多目标可视化技术[J]. 计算机辅助设计与图形学学报, 2014, 26(5): 739 – 746.
BI Xiaojun, LI Bo, CANG Yan. Visualization of high dimension multi-objective of rotation basis based on interactive decision making[J]. Journal of Computer Aided Design and Computer Graphics, 2014, 26(5): 739 – 746. (in Chinese)