基于等价分量交叉相似性的 Pareto 支配性预测

郭观七 [1] 尹 呈 [1,2] 曾文静 [1,3] 李 武 [1] 严太山 [1]

摘 要 研究用最近邻分类预测多目标优化问题 Pareto 支配性的相似性测度方法. 在分析决策分量对各目标分量贡献率的基础上定义决策向量的等价子向量, 等价子向量由贡献率相同的决策分量所组成. 提出基于等价子向量的最小交叉距离加权和相似性测度方法. 对每个目标分量, 独立评价待测数据与 N 个已知样本的相似度, 每个样本按其相似度值的升序赋予 [0: $N-1$] 之间的序号, 按各目标上的序号之和最小准则确定最近邻样本. 等价子向量最小交叉距离加权和相似性测度以及多目标最近邻搜索方法在确定决策向量相似性时, 引入了决策空间到目标向量空间的映射知识, 使决策变量相似性测度更真实地反映目标向量相似性. 对典型多目标优化问题的 Pareto 支配性最近邻分类实验结果表明, 提出的方法可显著地提高分类准确性.

关键词 多目标优化, Pareto 支配性, 等价分量, 交叉相似性, 最近邻分类

引用格式 郭观七, 尹呈, 曾文静, 李武, 严太山. 基于等价分量交叉相似性的 Pareto 支配性预测. 自动化学报, 2014, **40**(1): 33−40

DOI 10.3724/SP.J.1004.2014.00033

Prediction of Pareto Dominance by Cross Similarity of Equivalent Components

GUO Guan-Qi[1] YIN Cheng[1, 2] ZENG Wen-Jing[1, 3] LI Wu[1] YAN Tai-Shan[1]

Abstract This study investigates the similarity measurement of nearest neighbor classification for Pareto dominance prediction in multi-objective optimization. The equivalent components of a decision vector are defined by analyzing the contribution rate of each decision component to each objective component. For each objective component, the decision vector is divided into a group of equivalent sub-vectors, each consisting of the equivalent components with the same contribution rate. The distance between two equivalent sub-vectors is computed by minimizing the cross distance among the equivalent components. The similarity of two decision vectors is measured by the weighted sum of the minimized cross distances (WMCDs) corresponding to each equivalent sub-vector, where the weight takes the corresponding contribution rate. For each objective component, after evaluating WMCDs between an observed data and N samples, each sample is assigned a sequential number in [0: $N-1$] according to the ascending order of WMCDs. The nearest sample is the one with the minimal sum of the sequential numbers on all objective components. For WMCDs and multi-objective nearest neighbor searching introduce the mapping information from the decision space to the objective space, the similarity measurement in the decision space reflects the similarity in the objective space more seemingly. The experiments on tested problems show that the proposed algorithm remarkably improves the prediction performance of nearest neighbor classifiers.

Key words Multi-objective optimization, Pareto dominance, equivalent components, cross similarity, nearest neighbor classification

Citation Guo Guan-Qi, Yin Cheng, Zeng Wen-Jing, Li Wu, Yan Tai-Shan. Prediction of Pareto dominance by cross similarity of equivalent components. *Acta Automatica Sinica*, 2014, **40**(1): 33−40

大量研究表明进化算法是解决多目标优化问题 (Multi-objective optimization problems, MOPs) 的有效方法[1−2]. 但对昂贵多目标优化问题, 一次目标函数或约束函数的评估需耗时数小时甚至数天[3−4], 称之为计算成本灾难问题. 若采用基于目标向量和约束向量直接评估计算的多目标优化进化算法 (Multi-objective evolutionary algorithms, MOEAs) 来解决此类问题势必导致难以接受的计算开销和时间效率, 因此, 解决计算成本灾难问题成为多目标优化工程实践的研究重点之一[5−7].

为了克服计算成本灾难问题, 减少对原始目标函数评价次数, 近年来, 基于代理模型的多目标优化

收稿日期 2012-08-29 录用日期 2013-05-02
Manuscript received August 29, 2012; accepted May 2, 2013
国家自然科学基金 (60975049), 湖南省自然科学基金 (11JJ2037), 湖南省高校科技创新团队支持计划资助
Supported by National Natural Science Foundation of China (60975049), Natural Science Foundation of Hunan Province (11JJ2037), and Aid Program for Science and Technology Innovative Research Team in Higher Educational Instituions of Hunan Province
本文责任编委 陈杰
Recommended by Associate Editor CHEN Jie
1. 湖南理工学院信息与通信工程学院 岳阳 414006 2. 湘潭大学信息工程学院 湘潭 411105 3. 中南大学信息科学与工程学院 长沙 410083
1. College of Information and Communication Engineering, Hunan Institute of Science and Technology, Yueyang 414006 2. Institute of Information Engineering, Xiangtan University, Xiangtan 411105 3. College of Information Science and Engineering, Central South University, Changsha 410083

进化算法备受学术界关注[8−9], 涌现出许多优秀的代理模型, 其中, 最具代表性的有多项式模型[10]、高斯过程 (即 Kriging 模型)[11]、人工神经网络[12] 和径向基函数[13]. 此类算法的主要思想就是对一组样本点 (样本点由决策向量及其所对应的真实目标向量组成) 进行学习训练而拟合出原始目标函数的近似模型, 并将此模型与原始目标函数交替运用于整个进化过程.

基于代理模型的 MOEAs 在一定程度上减少了对原始目标函数的评价次数, 但选择代理模型及逼近技术[14] 需要较多的先验知识, 每个目标函数都需要单独建模, 模型类型和参数精度直接影响估计的准确性. 尽管改善预测搜索[15] 以及采用融合建模[16] 和自适应建模[17] 等技术能弥补现有基于代理模型 MOEAs 的一些不足, 但优化模型参数本身就需要较多的样本数据和优化成本, 建模对 MOPs 本身是一种挑战, 代价高昂.

为此, Guo 等[18] 提出了用分类器预测候选解间 Pareto 支配性的模式分类方法. 在假设类条件概率密度函数服从正态分布的前提下, Guo 初步实现了基于统计学习理论的贝叶斯分类器, 该分类器用于 SCH[19] 问题时, 取得了可接受的预测精度. 仿真优化表明, 预测结果可用于基于 Pareto 支配性的多目标优化算法, Pareto 支配性模式分类为解决复杂多目标优化的计算成本灾难和建模难题提供了新的有效方法.

然而, 对大多数多目标优化问题, 类条件概率密度函数服从正态分布的假设不一定成立, 基于这一假设的贝叶斯分类器难以达到满意的预测准确率. 为提高分类预测的准确性并进一步推广预测方法, Guo 等深入研究了 Pareto 支配性最近邻分类方法, 结合多目标优化的自身特点, 相继提出了决策空间中基于二进制位串加权和[20] 及同维距离序号和[21] 的相似性测度方法, 它们的共同特点是将各维决策分量转换到同一坐标刻度来度量候选解的相似性. 与基于欧氏距离的相似性测度方法相比, 对于决策分量定义域存在数量级差异的问题, 二者均显著地提高了最近邻分类的精度, 且对不同类型的多目标优化问题有较强的鲁棒性.

作者的研究表明, 由于多目标优化 Pareto 支配性是根据两个候选解的目标向量确定的, 各目标可能相互矛盾, 且从决策空间到目标空间的映射是非线性的和多维的, 在有限样本集条件下, 孤立地测度决策向量的相似性难以真实地反映目标向量的相似性, 这严重地制约了最近邻分类方法的应用. 本文分析多目标优化问题决策分量 (Pareto 支配性空间的特征分量) 对目标分量的贡献程度, 基于决策分量对各目标分量的贡献率分析, 在决策空间中提出基于等价子向量最小交叉距离的相似性测度方法和多目标最近邻分类算法, 并给出相关的实验结果.

1 Pareto 支配性最近邻预测

对具体的 MOPs, 任意两个候选解 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 间的 Pareto 支配性关系记为 $\omega(\boldsymbol{x}_i, \boldsymbol{x}_j)$, 这里 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 表示候选解的决策向量. 根据 Pareto 最优性概念, 可将 $\omega(\boldsymbol{x}_i, \boldsymbol{x}_j)$ 粗略地分为 3 类: 支配 ($\prec$), 被支配 ($\succ$) 和不可比 ($\sim$). 给定两个规模为 N 的候选解样本集 S_1 和 S_2, 采用计算和比较目标向量的方法能确定任意 $\omega(\boldsymbol{s}_i \in S_1, \boldsymbol{s}_j \in S_2)$, 从而构造出规模大小为 $N \times N$ 的 Pareto 支配关系样本集 $\{\omega(\boldsymbol{s}_i, \boldsymbol{s}_j)\}$, 其中, $i, j = 1, 2, \cdots, N$. 对候选解二元组 $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, 预测其 Pareto 支配性的最近邻分类方法基本框架如下:

1) 在决策向量定义域内, 随机产生规模大小为 N 的两个候选解集 S_1 和 S_2;

2) 计算和比较目标向量, 确定 $\omega(\boldsymbol{s}_i \in S_1, \boldsymbol{s}_j \in S_2), i, j = 1, 2, \cdots, N$;

3) 从 S_1 中找出与 $\boldsymbol{x}_1$ 最相似的候选解 $\boldsymbol{s}_1^*$;

4) 从 S_2 中找出与 $\boldsymbol{x}_2$ 最相似的候选解 $\boldsymbol{s}_2^*$;

5) 令 $\omega(\boldsymbol{x}_1, \boldsymbol{x}_2) = \omega(\boldsymbol{s}_1^*, \boldsymbol{s}_2^*)$.

2 基于决策变量等价分量交叉距离的相似性测度

2.1 等价分量最小交叉距离加权和

如何在决策空间为待测样本找到与之具有相似目标向量的样本, 是提高 Pareto 支配性最近邻预测精度的关键. 由于决策空间和目标空间的多维性和映射的非线性, 在有限样本条件下, 决策向量间的简单距离测度难以真实反映目标向量的相似性. 为了提高相似性测度的真实性, 在决策向量的相似性测度时, 引入决策空间到目标空间的映射信息, 使决策向量的相似性测度尽量包含目标空间的信息.

定义 1. 设 $y = f(\boldsymbol{x})$ 表示从空间 X 到 Y 的映射函数, $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 为 n 维向量, y 为标量. 将分量 x_i 在点 $\boldsymbol{x}$ 对 y 的贡献率 g_i 定义为

$$g_i = \frac{f(x_1, \cdots, x_i + \Delta x_i, \cdots, x_n) - f(\boldsymbol{x})}{\Delta x_i} \tag{1}$$

其中, $i = 1, 2, \cdots, n$, Δx_i 为分量 x_i 的增量. 计算 g_i 时, Δx_i 可取其绝对值足够小的正数或负数, 需保证 $x_i + \Delta x_i$ 不超出 x_i 的定义域. 对任意 $i \neq j$, 当 $x_i = x_j$ 且 $\Delta x_i = \Delta x_j$ 时, 若 $g_i = g_j$, 称 x_i 和 x_j 为等价分量. 将向量 $\boldsymbol{x}$ 按各分量对函数的贡献率值分成若干个子向量, 使同一子向量的各分量有相同贡献率, 这样的子向量称为等价子向量.

推论 1. 给定 $\boldsymbol{x}_1=(x_1,\cdots,x_i,\cdots,x_j,\cdots,x_n)$, 交换 x_i 和 x_j 的值, 得到向量 $\boldsymbol{x}_2=(x_1,\cdots,x_j,\cdots,x_i,\cdots,x_n)$, 若 x_i 和 x_j 是等价分量, 则当 x_i 和 x_j 取任意值时, $f(\boldsymbol{x}_1)=f(\boldsymbol{x}_2)$.

证明. 当 x_i 和 x_j 取相同值时, 推论显然成立. 当 $x_i\neq x_j$ 时, 令

$$d_1=\frac{f(\boldsymbol{x}_1)-f(x_1,\cdots,x_j,\cdots,x_j,\cdots,x_n)}{\Delta x}=\frac{f(x_1,\cdots,x_i,\cdots,x_j,\cdots,x_n)}{\Delta x}-\frac{f(x_1,\cdots,x_j,\cdots,x_j,\cdots,x_n)}{\Delta x}\tag{2}$$

$$d_2=\frac{f(\boldsymbol{x}_2)-f(x_1,\cdots,x_j,\cdots,x_j,\cdots,x_n)}{\Delta x}=\frac{f(x_1,\cdots,x_j,\cdots,x_i,\cdots,x_n)}{\Delta x}-\frac{f(x_1,\cdots,x_j,\cdots,x_j,\cdots,x_n)}{\Delta x}\tag{3}$$

不失一般性, 可令 $\Delta x=x_i-x_j$, 那么

$$d_1=\frac{f(x_1,\cdots,x_j+\Delta x,\cdots,x_j,\cdots,x_n)}{\Delta x}-\frac{f(x_1,\cdots,x_j,\cdots,x_j,\cdots,x_n)}{\Delta x}=g_i\tag{4}$$

$$d_2=\frac{f(x_1,\cdots,x_j,\cdots,x_j+\Delta x,\cdots,x_n)}{\Delta x}-\frac{f(x_1,\cdots,x_j,\cdots,x_j,\cdots,x_n)}{\Delta x}=g_j\tag{5}$$

因 x_i 和 x_j 是等价分量, $g_i=g_j$, 则必有:

$$f(x_1,\cdots,x_j+\Delta x,\cdots,x_j,\cdots,x_n)=f(x_1,\cdots,x_j,\cdots,x_j+\Delta x,\cdots,x_n)\tag{6}$$

将 $\Delta x=x_i-x_j$ 代入式 (6) 即得 $f(\boldsymbol{x}_1)=f(\boldsymbol{x}_2)$. □

对于存在等价分量的问题, 按式 (1) 计算贡献率向量, 将决策向量划分为 k 个等价子向量, 每个等价子向量由若干具有相同贡献率的等价分量组成. 上述推论说明, 给定两个决策向量, 二者并不只是当同维分量相似时, 相应的目标函数值才相似; 即使同维分量不相似, 如果等价分量交叉相似, 二者的目标函数值也相似. 基于等价分量的这一性质, 我们提出通过交叉配对等价分量来计算等价子向量距离的方法, 配对的准则是使各等价分量交叉距离之和最小化. 两个决策向量的距离定义为 k 个等价子向量距离的加权和, 贡献率为加权因子. 按此方法计算的决策向量距离隐含了通过计算贡献率所获得的目标空间信息, 从而使决策向量的相似性测度更真实地表达目标函数的相似性, 达到提高最近邻分类准确性的目的.

根据等价分量的定义, 将决策向量 $\boldsymbol{x}$ 分为 k 个等价子向量, 即 $\boldsymbol{x}=(\boldsymbol{x}(1),\boldsymbol{x}(2),\cdots,\boldsymbol{x}(k))$, 其中, $\boldsymbol{x}(i)$ 表示第 i 个子向量. 第 i 个子向量仅包含贡献率为 g_i 的等价分量, 令 n_i 表示第 i 个子向量的等价分量个数 $|\boldsymbol{x}(i)|$, $G=\sum_{i=1}^{k}|g_i|$, $p_i=|g_i|/G$, 决策向量 $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ 的距离 $D(\boldsymbol{x}_1,\boldsymbol{x}_2)$ 用 k 个等价子向量的最小交叉距离加权和表示. 计算 $D(\boldsymbol{x}_1,\boldsymbol{x}_2)$ 的算法 1 描述如下:

算法 1. 计算 $D(\boldsymbol{x}_1,\boldsymbol{x}_2)$

```
    // 初始化
    D(x1, x2) = 0;
// 计算 k 个等价子向量最小交叉距离加权和
    For i = 1 : k
// 第 i 个等价子向量的最小交叉距离
      d(x1(i), x2(i)) = 0;
// 构造第 i 个等价子向量的交叉距离矩阵 A,
// a_pq = |x_1p(i) − x_2q(i)|
      A = (a_pq)_{n_i×n_i};
// max{a_pq} 为 A 的最大元素
      maximum = max{a_pq} + 1;
// 计算第 i 个等价子向量的最小交叉距离
      For loops = 1 : n_i
// min{a_pq } 为 A 的最小元素
        d(x1(i), x2(i)) = d(x1(i), x2(i)) + min{a_pq};
        将矩阵 A 最小元素所在行和列的元素置为 maximum;
      Endfor
// 加权计入第 i 个子向量的最小交叉距离
      D(x1, x2) = D(x1, x2) +d(x1(i), x2(i)) × (1 + p_i);
    Endfor
```

2.2 基于等价子向量最小交叉距离加权和相似性测度的多目标最近邻搜索

对于具体的 MOPs, 要使两个候选解的相似性最大, 要求多个目标尽可能都有最大相似度. 因此, 在进行最近邻搜索前, 需要对各个目标分量计算贡献率向量, 根据各目标的贡献率向量确定相应的等价子向量. 设 MOPs 的目标向量为 m 维, 决策向量为 n 维, 样本集 S的大小为 N, $\boldsymbol{x}$ 表示待测样本, $\boldsymbol{s}_j$ 表示第 j 个已知样本. c_{ij} 表示 $\boldsymbol{x}$ 在第 i 个目标上与 $\boldsymbol{s}_j$ 的等价子向量最小交叉距离加权和相似度, 每个 $\boldsymbol{s}_j$ 被赋予一个按 N 个 c_{ij} 升序排序后的序号 r_{ij}, $r_{ij}\in\{0:N-1\}$. D_j 为 $\boldsymbol{x}$ 在 m 维目标上与 $\boldsymbol{s}_j$ 的总相似度, 在样本集中搜索 $\boldsymbol{x}$ 的最相似样本的算法描述如下:

1) 对每个目标分析决策分量的贡献率确定等价子向量;

2) 对待测样本 $\boldsymbol{x}$, 根据算法 1 求矩阵 $C=(c_{ij})_{m\times N}$, $c_{ij}=D(\boldsymbol{x},\boldsymbol{s}_j)$;

3) 对矩阵 C 的第 i 行，每个 c_{ij} 按 N 个 c_{ij} 的升序赋予 $[0, N-1]$ 间的一个序号值, 并将该序号值存入序号矩阵 $R=(r_{ij})_{m\times N}$;

4) 分别计算序号矩阵 R 每一列的和;

5) 和数最小的列下标值所对应的已知样本即为 $\boldsymbol{x}$ 的最近邻.

此算法在相似性测度时, 利用了决策空间到目标空间的映射关系. 这一先验知识的引入, 充分考虑了决策变量对多个目标值的影响, 其思想也可用于其他应用的最近邻分类预测中. 当决策分量的贡献率互不相等时, 决策向量中不存在等价分量, 该算法蜕变成同维距离按贡献率加权相似性测度的多目标最近邻搜索. 当决策分量的定义域不同时, 只需在各分量定义域相同的假设下分析等价子向量, 算法 1 和算法 2 仍然适用.

3 Pareto 支配性预测实验

为验证基于等价子向量最小交叉距离加权和相似性测度的 Pareto 支配性最近邻分类算法 (Nearest neighbor classification based on equivalent similarity, ESNNC) 的效果, 本文选择常用多目标优化测试问题 FON[22] ($n=3, 10$, $m=2$, 定义域 $[-4, 4]$), KUR[23] ($n=3, m=2$, 定义域 $[-5, 5]$), ZDT6[24] ($n=10$, $m=2$, 定义域 $[0, 1]$), DTLZ1, DTLZ2, DTLZ4 [25] ($n=10$, $m=3$, 定义域 $[0, 1]$) 进行实验, 表 1 列出了这些测试问题的定义. 由于这些测试问题的解析模型均已知, 对各测试问题分别令各决策分量相等, 各决策分量的值取其定义域范围内的同一随机数, 其增量均取其值的 1%, 根据式 (1) 可计算出决策分量在该点上对各目标函数的贡献率, 将具有相同贡献率的等价分量归于同一子向量. 虽然贡献率在决策向量的不同点上是不同的, 但根据等价分量的定义, 只要当各决策分量取相同值且增量也相等时, 其等价子向量的划分结果是相同的. 因此, 本文实验中各决策分量在其定义域内取相同的随机值, 且仅在最近邻搜索前计算一次. 各测试问题的等价子向量划分如表 2 所示. 由此可见, 虽然这些测试问题的形态和复杂度不同, 但对于不同的目标函数, 决策向量均存在等价子向量.

表 1 多目标优化测试问题

Table 1 Multi-objective optimization test problems

Problem	n, m	Domain	Objective function
FON	$n=3, 10$ $m=2$	$[-4, 4]$	$\min f_1(\boldsymbol{x}) = 1-\exp(-\sum_{i=1}^{n}(x_i-\frac{1}{\sqrt{n}})^2)$ $f_2(\boldsymbol{x}) = 1-\exp(-\sum_{i=1}^{n}(x_i+\frac{1}{\sqrt{n}})^2)$
KUR	$n=3$ $m=2$	$[-5, 5]$	$\min\ f_1(\boldsymbol{x}) = \sum_{i=1}^{n-1}(-10\mathrm{e}^{-0.2\sqrt{{x_i}^2+{x_{i+1}}^2}})$ $f_2(\boldsymbol{x}) = \sum_{i=1}^{n}(\lvert x_i\rvert^{0.8}+5\sin(x_i)^3)$
ZDT6	$n=10$ $m=2$	$[0, 1]$	$\min\ f_1(\boldsymbol{x}) = 1-\exp(-4x_1)\sin^6(6\pi x_1)$ $f_2(\boldsymbol{x}) = g(\boldsymbol{x})(1-(\frac{f_1}{g(\boldsymbol{x})})^2)$ $g(\boldsymbol{x}) = 1+9\left(\frac{\sum_{i=2}^{n}x_i}{n-1}\right)^{0.25}$
DTLZ1	$n=10$ $m=3$	$[0, 1]$	$\min\ f_1(\boldsymbol{x}) = (1+g(\boldsymbol{x}))x_1\frac{x_2}{2}$ $f_2(\boldsymbol{x}) = (1+g(\boldsymbol{x}))x_1\frac{1-x_2}{2}$ $f_3(\boldsymbol{x}) = (1+g(\boldsymbol{x}))\frac{1-x_1}{2}$ $g(\boldsymbol{x}) = 100(8+\sum_{i=3}^{n}(x_i-0.5)^2-\cos(20\pi(x_i-0.5)))$
DTLZ2	$n=10$ $m=3$	$[0, 1]$	$\min\ f_1(\boldsymbol{x}) = (1+g(\boldsymbol{x}))\cos(\frac{x_1\pi}{2})\cos(\frac{x_2\pi}{2})$ $f_2(\boldsymbol{x}) = (1+g(\boldsymbol{x}))\cos(\frac{x_1\pi}{2})\sin(\frac{x_2\pi}{2})$ $f_3(\boldsymbol{x}) = (1+g(\boldsymbol{x}))\sin(\frac{x_1\pi}{2})$ $g(\boldsymbol{x}) = \sum_{i=3}^{n}(x_i-0.5)^2$
DTLZ4	$n=10$ $m=3$	$[0, 1]$	$\min\ f_1(\boldsymbol{x}) = (1+g(\boldsymbol{x}))\cos(\frac{{x_1}^{100}\pi}{2})\cos(\frac{{x_2}^{100}\pi}{2})$ $f_2(\boldsymbol{x}) = (1+g(\boldsymbol{x}))\cos(\frac{{x_1}^{100}\pi}{2})\sin(\frac{{x_2}^{100}\pi}{2})$ $f_3(\boldsymbol{x}) = (1+g(\boldsymbol{x}))\sin(\frac{{x_1}^{100}\pi}{2})$ $g(\boldsymbol{x}) = \sum_{i=3}^{n}(x_i-0.5)^2$

表 2　测试问题在定义域内随机给定点上的等价子向量

Table 2　Equivalent sub-vectors on randomly given point in domains of test problems

Problem		与各目标分量对应的等价子向量划分		
		f_1	f_2	f_3
FON	$n=3$	(x_1,x_2,x_3)	(x_1,x_2,x_3)	
	$n=10$	$(x_1,x_2,\cdots,x_{10})$	$(x_1,x_2,\cdots,x_{10})$	
KUR	$n=3$	$(x_1,x_3),(x_2)$	(x_1,x_2,x_3)	
ZDT6	$n=10$	$(x_1),(x_2,\cdots,x_{10})$	$(x_1),(x_2,\cdots,x_{10})$	
DTLZ1	$n=10$	$(x_1,x_2),(x_3,\cdots,x_{10})$	$(x_1),(x_2),(x_3,\cdots,x_{10})$	$(x_1),(x_2),(x_3,\cdots,x_{10})$
DTLZ2	$n=10$	$(x_1,x_2),(x_3,\cdots,x_{10})$	$(x_1),(x_2),(x_3,\cdots,x_{10})$	$(x_1),(x_2),(x_3,\cdots,x_{10})$
DTLZ4	$n=10$	$(x_1,x_2),(x_3,\cdots,x_{10})$	$(x_1),(x_2),(x_3,\cdots,x_{10})$	$(x_1),(x_2),(x_3,\cdots,x_{10})$

对每一个测试函数, 在定义域范围内随机产生两个规模为 200 的候选解集, 计算并比较其所对应的目标向量, 构造规模为 200×200 的 Pareto 支配性样本集, 结合分类算法预测随机产生的规模为 40×40 的待测样本集中候选间 Pareto 支配性. 采用简单欧氏距离相似性测度 (Nearest neighbor classification based on Euclidian distance measurement, ENNC)[21]、同维距离序号和相似性测度 (Nearest neighbor classification based on ranking the dimentional distances, RNNC)[21]、ESNNC 三种最近邻分类算法. 为了与典型的代理模型比较, 在已知样本集规模为 200 的条件下, 我们用高斯过程模型 (Gaussian process model, GPM)[11] 分别建模各目标函数, 并用模型的输出测试 Pareto 支配性分类准确性. 分别对每个测试问题进行 100 次实验, 每次实验都重新生成已知样本集, 待测样本集保持不变. 已知样本集 Pareto 支配性类分布 100 次实验的平均类比例如表 3 所示, 待测样本集各类平均预测准确率如表 4 所示.

表 3　100 次实验中样本集各类的平均类比例

Table 3　Average class proportion over 100 random sample sets

Problem	n,m	$\prec$ (%)	$\succ$ (%)	$\sim$ (%)
FON	$n=3,m=2$	34.51	34.20	31.29
	$n=10,m=2$	16.87	16.57	66.58
KUR	$n=3,m=2$	29.74	29.67	40.59
ZDT6	$n=10,m=2$	23.61	24.47	51.91
DTLZ1	$n=10,m=3$	3.67	4.41	91.91
DTLZ2	$n=10,m=3$	3.41	3.43	93.16
DTLZ4	$n=10,m=3$	11.90	11.79	76.31

表 4 的数据表明, 采用简单欧氏距离相似性测度的 ENNC 和采用同维距离序号和相似性测度的 RNNC 具有基本相同的预测准确率, 而 ESNNC 的预测准确率一致地高于 ENNC、RNNC 和 GPM. 随着测试问题决策向量维数的增加, 各目标函数的崎岖度 (峰密度) 显著增大, ENNC 和 RNNC 的预测性能急剧地退化, 但 ESNNC 仍能达到较高的预测准确率. 从表 3 可以看出, 二目标的 FON 问题仅存在一个等价子向量, 当决策向量为 3 维时, 三种最近邻分类算法的预测准确率差异并不明显; 但当决策向量增大到 10 维时, ESNNC 的预测准确率并未发生明显退化, 且 ENNC 和 RNNC 的预测准确率仅能达到 ESNNC 的 60% 左右. 在实验中我们修改了对 ZDT6 的定义, 令决策向量的维数逐渐增加, 但维持各目标的等价子向量个数不变, 仅增加等价子向量中等价分量的个数, 我们发现 ESNNC 对各类预测的准确率基本维持不变, 但 ENNC 和 RNNC 的预测性能急剧地退化, 对 ZDT1[24] 问题也表现了类似的现象. 这一现象表明决策向量的等价子向量个数越少, 等价子向量中的等价分量越多, ESNNC 效果越明显.

二目标 KUR 和 ZDT6 的各目标均存在等价子向量, ESNNC 均表现了一致的显著优越性. 这一现象说明采用等价子向量的最小交叉距离加权和测度相似性的 ESNNC, 不但改善了最近邻分类的准确率, 而且对决策向量的维数变化有更强的鲁棒性. 这是因为采用单纯的决策向量距离测度相似性, ENNC 和 RNNC 只能在样本集中找到与待测样本决策向量间距离最短的最近邻; 但由于目标向量和决策向量的多维性以及从决策空间到目标空间映射的非线性, 在有限样本条件下, 决策空间的最近邻未必就是目标空间的最近邻, 所以导

致采用纯决策向量距离测度相似性的最近邻分类难以达到准确率足够高的 Pareto 支配性预测. 通过对决策分量在各目标上的贡献率分析和等价子向量的划分, 且采用贡献率对各等价子向量的最小交叉距离加权求和, 使得在采用决策空间的距离测度相似性时, 引入了从决策空间到目标空间的映射知识, 从而使决策空间的相似性测度尽可能真实地反映目标空间的相似性, 最终达到提高 Pareto 支配性预测准确性的目的. 而等价分量交叉相似性测度在一定程度上起到了对决策向量降维处理的作用, 从而提高了最近邻分类对维数变化的鲁棒性.

从表 3 可以看出, 三目标 DTLZ1、DTLZ2、DTLZ4 问题的样本集类比例很不平衡. 实验表明 4 种算法对少数类的预测准确率均低于对多数类的预测准确率, 鉴于 ENNC 和 RNNC 对 DTLZ1 和 DTLZ4 的预测性能与 DTLZ2 相似, 为节省篇幅, 表 4 未列出其具体数据. 表 4 表明, ESNNC 对少数类的预测准确率明显地高于其他三种算法, 且 GPM 也优于 ENNC 和 RNNC. 对类比例不平衡度稍轻的 DTLZ4, 通过适当增大样本集的方法可使 ESNNC 的性能得到一定程度的改善, 但对于样本类比例极不平衡的 DTLZ1 和 DTLZ2 问题, 改善效果不明显, 这有待研究更有效的解决方法.

表 4 ENNC、RNNC、GPM 和 ESNNC 的平均预测精度

Table 4 Average prediction accuracy of ENNC, RNNC, GPM, and ESNNC

Problem	n, m	Algorithm	Average accuracy rate (%)		
			$\prec$	$\succ$	$\sim$
FON	$n=10, m=2$	ENNC	81.64	78.48	69.09
		RNNC	72.20	81.81	65.40
		GPM	32.09	32.09	51.44
		ESNNC	89.59	88.06	80.52
	$n=10, m=2$	ENNC	46.26	48.50	53.35
		RNNC	50.16	49.28	50.26
		GPM	32.91	32.91	33.10
		ESNNC	73.07	74.83	86.95
KUR	$n=3, m=2$	ENNC	74.23	72.95	61.73
		RNNC	70.72	72.50	66.29
		GPM	59.16	59.16	51.87
		ESNNC	79.34	81.83	72.65
ZDT6	$n=10, m=2$	ENNC	39.54	42.54	52.36
		RNNC	36.02	39.85	53.79
		GPM	60.03	60.03	57.39
		ESNNC	70.51	69.71	74.27
DTLZ2	$n=10, m=3$	ENNC	11.94	11.13	93.44
		RNNC	11.90	15.93	94.25
		GPM	22.27	22.27	97.90
		ESNNC	37.61	37.61	95.14
DTLZ1	$n=10, m=3$	GPM	11.23	11.23	95.03
		ESNNC	19.64	19.64	94.38
DTLZ4	$n=10, m=3$	GPM	34.17	34.17	75.27
		ESNNC	45.24	45.24	80.41

4　结论

基于 Pareto 最优性的多目标优化需要确定候选解间的 Pareto 支配性关系, 对昂贵多目标优化问题, 面临着大量目标函数计算和比较所带来的计算成本灾难问题. 研究表明采用最近邻分类方法对候选解 Pareto 支配性进行预测, 可以有效克服计算成本灾难问题. 但现有的相似性测度方法都忽视了决策空间到目标空间的映射关系这一先验知识, 导致分类错误率高. 本文通过决策分量对各目标分量的贡献率分析, 在相似性测度时引入了决策空间到目标空间的映射知识, 提出了基于决策向量等价子向量最小交叉距离加权和的相似性测度方法, 并结合多目标问题的固有特性, 在多个目标上同时评价和搜索待测数据的最近邻样本, 使决策空间的相似性测度更真实地反映目标空间的相似性. 对几个典型多目标优化问题进行测试, 实验结果表明, 基于决策向量等价分量交叉相似性测度的多目标最近邻搜索分类效果, 显著地优于现有的最近邻分类算法和典型代理模型的预测结果. 当然, 在多目标问题解析模型未知的情况下, 如何确定决策分量的贡献率是有待进一步深入研究的问题.

References

1 Deb K. Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction, KanGAL Report 2011003, Indian Institute of Technology Kanpur, India, 2011

2 Zhou A M, Qu B Y, Li H, Zhao S Z, Suganthan P N, Zhang Q F. Multiobjective evolutionary algorithms: a survey of the state of the art. *Swarm and Evolutionary Computation*, 2011, **1**(1): 32−49

3 Nain P K S, Deb K. A Multi-objective Search and Optimization Procedure with Successive Approximate Models, KanGAL Report 2004012, Indian Institute of Technology Kanpur, India, 2004

4 Jin Y C, Sendhoff B. A systems approach to evolutionary multiobjective structural optimization and beyond. *IEEE Computational Intelligence Magazine*, 2009, **4**(3): 62−76

5 Knowles J. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 2006, **10**(1): 50−66

6 Schmidt M D, Lipson H. Coevolution of fitness predictors. *IEEE Transactions on Evolutionary Computation*, 2008, **12**(6): 736−749

7 Samad A, Kim K Y, Goel T, Haftka R T, Shyy W. Multiple surrogate modeling for axial compressor blade shape optimization. *Journal for Propulsion and Power*, 2008, **24**(2): 302−310

8 Shi L, Rasheed K. A survey of fitness approximation methods applied in evolutionary algorithms. *Computational Intelligence in Expensive Optimization Problems, Adaptation Learning and Optimization*. Berlin Heidelberg: Springer, 2010. **2**: 3−28

9 Jin Y C. Surrogate-assisted evolutionary computation: recent advances and future challenges. *Swarm and Evolutionary Computation*, 2011, **1**(2): 61−70

10 Goel T, Vaidyanathan R, Haftka R T, Shyy W, Queipo N V, Tucker K. Response surface approximation of Pareto opimal front in multi-objective optimization. *Computer Methods in Applied Mechanics and Engineering*, 2007, **196**(4−6): 879−893

11 Zhang Q F, Liu W D, Tsang E, Virginas B. Expensive multiobjective optimization by MOEA/D with Gaussian process model. *IEEE Transactions on Evolutionary Computation*, 2010, **14**(3): 456−474

12 Deb K, Nain P K S. An evolutionary multi-objective adaptive meta-modeling procedure using artificial neural networks. *Evolutionary Computation in Dynamic and Uncertain Environments*. Berlin, Germany: Springer-Verlag, 2007. **51**: 297−322

13 Marjavaara B D, Lundström T S, Goel T, Mack Y, Shyy W. Hydraulic turbine diffuser shape optimization by multiple surrogate model approximations of Pareto fronts. *Journal of Fluids Engineering*, 2007, **129**(9): 1228−1240

14 Lim D, Ong Y S, Jin Y C, Sendhoff B. A study on metamodeling techniques, ensembles, and multi-surrogates in evolutionary computation. In: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation. New York, USA: ACM, 2007. 1288−1295

15 Lim D, Jin Y C, Ong Y S, Sendoff B. Generalizing surrogate-assisted evolutionary computation. *IEEE Transactions on Evolutionary Computation*, 2010, **14**(3): 329−355

16 Acar E, Rais-Rohani M. Ensemble of metamodels with optimized weight factors. *Structural and Multidisciplinary Optimization*, 2009, **37**(3): 279−294

17 Shi L, Rasheed K. ASAGA: an adaptive surrogate-assisted genetic algorithm. In: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation. New York, USA: ACM, 2008. 1049−1056

18 Guo G, Li W, Yang B, Li W, Yin C. Predicting Pareto dominance in multi-objective optimization using pattern recognition. In: Proceedings of the 2012 International Conference on Intelligent System Design and Engineering Application (ISDEA). Sanya, China: IEEE, 2012. 456−459

19 Schaffer J D. Multiple objective optimization with vector evaluated genetic algorithms. In: Proceedings of the 1st International Conference on Genetic Algorithms. New Jersey, USA: Lawrence Erlbaum, 1987. 93−100

20 Guo G Q, Yin C, Yan T S, Li W B. Binary nearest neighbor classification of predicting Pareto dominance in multi-objective optimization. In: Proceedings of the 2012 International Conference on Swarm Intelligence, Part I, LNCS 7331. Berlin: Springer-Verlag, 2012. 537−545

21 Guo G Q, Yin C, Yan T S, Li W. Nearest neighbor classification of Pareto dominance in multi-objective optimization. In: Proceedings of the 5th IEEE International Conference on Advanced Computational Intelligence. Nanjing, China: IEEE, 2012. 328−331

22 Fonseca C M, Fleming P J. Multiobjective optimization and multiple constraint handling with evolutionary algorithm Part II: Application example. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 1998, **28**(1): 38−47

23 Kursawe F. A variant of evolution strategies for vector optimization. *Parallel Problem Solving from Nature*, Schwefel I H P, Manner R (eds.), Berlin, Germany: Springer, 1990. 193−197

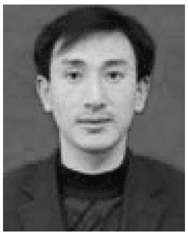
24 Zilter E, Deb K, Thiele L. Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation*, 2000, **8**(2): 173−195

25 Deb K, Thiele L, Laumanns M, Zitzler E. Scalable multi-objective optimization test problems. In: Proceedings of the 2002 Congress on Evolutionary Computation (CEC′02). Honolulu, HI: IEEE, 2002. 825−830

郭观七 湖南理工学院信息与通信工程学院教授. 2003 年获得中南大学信息科学与工程学院博士学位. 主要研究方向为计算智能, 多目标优化, 模式识别. 本文通信作者. E-mail: gq.guo@163.com
(**GUO Guan-Qi** Professor at the College of Information and Communication Engineering, Hunan Institute of Science and Technology. He received his Ph. D. degree from Central South University in 2003. His research interest covers computational intelligence, multi-objective optimization, and pattern recognition. Corresponding author of this paper.)

尹 呈 湘潭大学信息工程学院硕士研究生. 主要研究方向计算智能, 多目标优化. E-mail: yincheng886337@163.com
(**YIN Cheng** Master student at the Institute of Information Engineering, Xiangtan University. His research interest covers computational intelligence and multi-objective optimization.)

曾文静 中南大学信息科学与工程学院硕士研究生. 主要研究方向为多目标优化, 进化算法.
E-mail: zengwenjingzhongnan@126.com
(**ZENG Wen-Jing** Master student at the College of Information Science and Engineering, Central South University. Her research interest covers multi-objective optimization and evolutionary algorithm.)



李 武 湖南理工学院信息与通信工程学院副教授. 2009 年获得华中科技大学控制科学与工程系博士学位. 主要研究方向为决策分析, 复杂系统建模与优化.
E-mail: liwu0817@163.com
(**LI Wu** Associate professor at the College of Information and Communication Engineering, Hunan Institute of Science and Technology. He received his Ph. D. degree from Huazhong University of Science and Technology in 2009. His research interest covers decision analysis, model and optimization of complex systems.)

严太山 湖南理工学院信息与通信工程学院副教授. 2010 年获得西安理工大学自动化学院博士学位. 主要研究方向为进化计算, 智能优化, 神经网络.
E-mail: yantaishan163@163.com
(**YAN Tai-Shan** Associate professor at the College of Information and Communication Engineering, Hunan Institute of Science and Technology. He received his Ph. D. degree from Xi′an University of Technology in 2010. His research interest covers evolutionary computation, intelligent optimization, and neural network.)