



(12)发明专利

(10)授权公告号 CN 106022960 B

(45)授权公告日 2019.07.09

(21)申请号 201610338489.2

G06Q 50/06(2012.01)

(22)申请日 2016.05.19

(56)对比文件

(65)同一申请的已公布的文献号
申请公布号 CN 106022960 A

CN 104462861 A, 2015.03.25,
CN 101714193 A, 2010.05.26,

(43)申请公布日 2016.10.12

丁胜祥. 基于Pareto强度进化算法的供水库群多目标优化调度.《水科学进展》.2008,第679-684页.

(73)专利权人 武汉大学

地址 430072 湖北省武汉市武昌区珞珈山
武汉大学

余胜男. 随机森林在降水量长期预报中的应用.《南水北调与水利科技》.2016,第14卷(第1期),第78-83页.

(72)发明人 郭生练 杨光 李立平 尹家波
刘章君

审查员 朱雪梅

(74)专利代理机构 武汉科皓知识产权代理事务
所(特殊普通合伙) 42222

代理人 魏波

(51)Int.Cl.

G06F 17/50(2006.01)

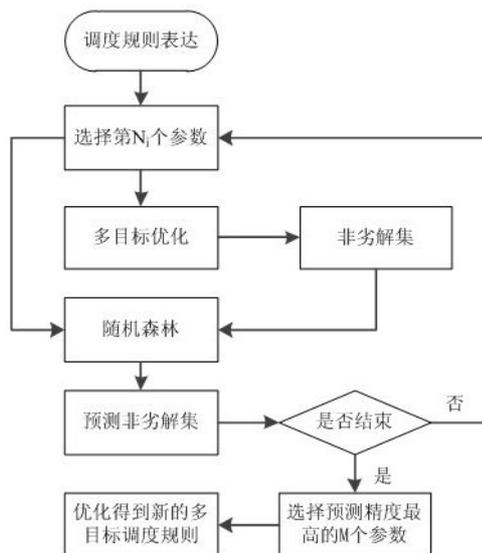
权利要求书2页 说明书5页 附图3页

(54)发明名称

一种基于数据挖掘的水库多目标调度规则
参数优选方法

(57)摘要

本发明公开了一种基于数据挖掘的水库多目标调度规则参数优选方法,采用多目标优化算法得到水库多目标调度非劣解集,并利用随机森林模型建立所有水库调度规则参数与非劣解集的关系,通过评价各参数对非劣解集的预测效果,优选出信息量最大的水库多目标调度规则参数。本发明能有效利用水库优化调度信息,大大减少水库调度规则中的参数个数,提高了水库多目标优化调度模型的求解效率,为水库的科学决策提供更简单且可操作性强的参考依据。



1. 一种基于数据挖掘的水库多目标调度规则参数优选方法,其特征在于,包括以下步骤:

步骤1:收集水库上游站点观测的数据资料,将水库调度规则用数学公式的方式进行描述;具体公式如下:

$$O_t = a_t \cdot (V_t + w \cdot I_t) + b_t;$$

式中: O_t 为水库第 t 时段出库流量,单位为 m^3/s ; w 为单位转换系数; a_t 和 b_t 为调度函数中以年为周期变化的规则参数,即每个月对应不同的值,但每年的同一个月份里 a_t 、 b_t 取同样的值; V_t 为水库第 t 时段初水库可用蓄水量,单位为 m^3/s ; I_t 为水库第 t 时段入库流量,单位为 m^3/s ;

步骤2:采用多目标优化算法,以水库历史入库流量作为输入,对水库调度规则中的参数进行优化,得到一系列能够满足水库多目标利用需求的非劣解集;

所述多目标优化算法,其具体实现包括以下子步骤:

步骤2.1:对水库调度规则中的参数进行编码,对编码的参数进行随机赋值,生成初始种群,对初始种群中每一组编码值进行解码,得到水库调度规则参数,并将历史入库数据作为输入,按规则调度得到最后的调度结果;

步骤2.2:将步骤2.1中产生的调度结果进行筛选排序,并挑选出不被任何其余结果支配的解,作为本次调度产生的非劣解集;

步骤2.3:在步骤2.2中生成的非劣解集中随机抽取预设个数的解,并将这些解对应的参数集作为子代进行编码;

步骤2.4:对步骤2.3生成的子代进行杂交、变异操作,生成新的种群,并解码得到水库调度规则参数,将历史入库数据作为输入,按规则调度得到最后的调度结果;

步骤2.5:重复步骤2.2—步骤2.4,直到本次得到的调度结果与上一次得到的结果之间的差异在预设范围内,将最后生成的一组参数,作为水库多目标优化调度规则,并将对应的调度结果作为能够满足水库多目标利用需求的非劣解集;

步骤3:根据步骤2中得到的非劣解集及其对应的水库调度规则参数,采用随机森林对它们的属性进行分析和归纳,挖掘其间的相互联系,并通过参数预测非劣解集;

所述采用随机森林对它们的属性进行分析和归纳,其具体实现包括以下子步骤:

步骤3.1:将历史资料分为率定期和检验期部分,并将一类水库调度规则参数和对应的非劣解集作为数据集,其中率定期的为训练集,检验期的为测试集;

步骤3.2:在率定期的训练集中进行 n 次重复随机抽样,形成多个训练子集;按照一类规则参数对应一组非劣解集的原则,针对各训练子集分别建立决策树,构成随机森林,以描述调度规则参数与非劣解集的回归关系;

步骤3.3:采用检验期测试集数据,利用步骤3.2中建立的调度规则参数与非劣解集的回归关系,推求测试集中水库调度规则参数对应的非劣解集;

步骤3.4:对步骤3.3中的预测结果进行综合,得到最终的预测结果;

步骤3.5:重复步骤3.1—步骤3.4,直到遍历完所有水库调度规则参数,采用随机森林建立其与非劣解集的关系,并得到每一组水库调度参数对非劣解集的预测结果;

步骤4:统计各水库调度规则参数对非劣解集的预测效果,将预测精度最高的前 M 个数作为描述水库调度规则信息量最大的参数;

其中,将所有水库调度规则参数预测的非劣解集与实际结果进行对比,并采用均方根误差评价所有水库调度规则参数的预测效果,将预测精度最高的前M个参数作为描述水库调度规则信息量最大的参数;

步骤5:采用多目标算法优化步骤4中优选的参数,得到新的水库调度规则。

2.根据权利要求1所述的基于数据挖掘的水库多目标调度规则参数优选方法,其特征在于:步骤1中所述上游站点观测的数据包括入库流量和各水库之间的区间流量过程。

一种基于数据挖掘的水库多目标调度规则参数优选方法

技术领域

[0001] 本发明属于水库调度技术领域,涉及一种基于数据挖掘的水库多目标调度规则参数优选方法。

背景技术

[0002] 水库作为一种径流调节的工具,在使水资源更适合人类社会发展以及维护生态环境方面发挥着重要作用,水库调度规则是一种指导水库运行的方式。随着社会进步和经济发展,人们对水资源利用有了更多需求,一般通过智能多目标优化算法得到满足多种用水需求的水库调度规则。但水库多目标优化问题往往存在着非线性、高维度的特点,增加了智能算法的寻优成本,如何在多个水库调度规则中优选出最有价值的参数,从而降低优化维度,提高寻优效率对水库的运行管理有着重要意义。

[0003] 数据挖掘是指从大量的数据中通过算法搜索隐藏于其中信息的过程,近年来,数据挖掘引起了信息产业界的极大关注,其主要原因是存在大量数据,可以广泛使用,并且迫切需要将这此数据转换成有用的信息和知识。随机森林作为数据挖掘技术中的一种,并以其运算量小、预测精度高等特点而被用于分类和回归分析([文献1])。易尧华等([文献2])提出了一种基于随机森林的卫星遥感影像云量计算方法,通过样本获取、影像回归分类等步骤计算了遥感影像中的云量。随着数据挖掘技术在水库调度领域方面的深入发展,国内学者提出了基于数据挖掘的水库调度方法。例如:张弛([文献3])等利用数据挖掘中的决策树技术对多年水库实际水文数据和调度数据进行分析 and 挖掘,结合预报入流从中挖掘出以调度树形式的水库调度规则。习树峰等([文献4])利用决策树方法利用预报降雨数据、前期土壤含水量、当前库水位等水库调度数据,提取了跨流域引水水库的实时调度规则。在水库多目标优化调度方面,周研来等([文献5])采用自适应遗传算法编制多目标优化调度图,在保证水库防洪安全的前提下最大限度地提高水库调度的综合效益。杨光等([文献6])考虑未来径流变化对水库调度的影响,采用非支配排序遗传算法(NSGA-II)算法推求了兼顾丹江口水库供水和发电的多目标调度规则。然而,他们并没有将数据挖掘技术整合到水库多目标优化调度规则参数的提取中。决策树作为一种数据挖掘技术在提取水库调度规则时可能存在过拟合现象,而随机森林能够通过整合多个决策树而降低过拟合概率。

[0004] 随机森林方法能够从一组无规律的事例中利用信息论原理对大量样本的属性进行分析和归纳,提取隐藏信息,挖掘出数据间的相互关系,寻找出有价值的信息。所以该方法能够从复杂的水库调度规则中找出最有利于多目标优化调度的参数。现有的水库多目标优化模型往往直接对水库调度规则进行优化,没有避开原有规则中的冗余信息,且目前尚未有研究采用随机森林方法优选水库多目标调度规则参数。

[0005] [文献1]李欣海.随机森林模型在分类与回归分析中的应用[J].应用昆虫学报,2013,50(4):1190-1197.

[0006] [文献2]易尧华,袁媛,张宇,等.一种基于随机森林的卫星遥感影像云量计算方法.专利号:CN105260729A.

[0007] [文献3]张弛,周惠成,王本德.决策树技术在水库兴利调度中的应用研究[J].哈尔滨工业大学学报,2007,39(8):1314-1318.

[0008] [文献4]刁树峰,彭勇,梁国华,等.基于决策树方法的水库跨流域引水调度规则研究[J].大连理工大学学报,2012,52(1):74-78.

[0009] [文献5]周研来,郭生练,陈华,等.一种自适应气候变化的水库多目标优化调度图编制方法.专利号:CN103049671A

[0010] [文献6]杨光,郭生练,李立平,等.考虑未来径流变化的丹江口水库多目标调度规则研究[J].水力发电学报,2015,34(12):54-63.

发明内容

[0011] 针对现有技术存在的不足,本发明提供了一种基于数据挖掘的水库多目标调度规则参数优选方法。

[0012] 本发明所采用的技术方案是:一种基于数据挖掘的水库多目标调度规则参数优选方法,其特征在于,包括以下步骤:

[0013] 步骤1:收集水库上游站点观测的数据资料,将水库调度规则用数学公式的方式进行描述;

[0014] 步骤2:采用多目标优化算法,以水库历史入库流量作为输入,对水库调度规则中的参数进行优化,得到一系列能够满足水库多目标利用需求的非劣解集;

[0015] 步骤3:根据步骤2中得到的非劣解集及其对应的水库调度规则参数,采用随机森林对它们的属性进行分析和归纳,挖掘其间的相互联系,并通过参数预测非劣解集;

[0016] 步骤4:统计各水库调度规则参数对非劣解集的预测效果,将预测精度最高的前M个参数作为描述水库调度规则信息量最大的参数;

[0017] 步骤5:采用多目标算法优化步骤4中优选的参数,得到新的水库调度规则。

[0018] 作为优选,步骤1中所述上游站点观测的数据包括入库流量和各水库之间的区间流量过程。

[0019] 作为优选,步骤2中所述多目标优化算法,其具体实现包括以下子步骤:

[0020] 步骤2.1:对水库调度规则中的参数进行编码,对编码的参数进行随机赋值,生成初始种群,对初始种群中每一组编码值进行解码,得到水库调度规则参数,并将历史入库数据作为输入,按规则调度得到最后的调度结果;

[0021] 步骤2.2:将步骤2.1中产生的调度结果进行筛选排序,并挑选出不被任何其他结果支配的解,作为本次调度产生的非劣解集;

[0022] 步骤2.3:在步骤2.2中生成的非劣解集中随机抽取预设个数的解,并将这些解对应的参数集作为子代进行编码;

[0023] 步骤2.4:对步骤2.3生成的子代进行杂交、变异操作,生成新的种群,并解码得到水库调度规则参数,将历史入库数据作为输入,按规则调度得到最后的调度结果;

[0024] 步骤2.5:重复步骤2.2—步骤2.4,直到本次得到的调度结果与上一次得到的结果之间的差异在预设范围内,将最后生成的一组参数,作为水库多目标优化调度规则,并将对应的调度结果作为能够满足水库多目标利用需求的非劣解集。

[0025] 作为优选,步骤3中所述采用随机森林对它们的属性进行分析和归纳,其具体实现

包括以下子步骤:

[0026] 步骤3.1:将历史资料分为率定期和检验期部分,并将一类水库调度规则参数和对应的非劣解集作为数据集,其中率定期的为训练集,检验期的为测试集;

[0027] 步骤3.2:在率定期的训练集中进行n次重复随机抽样,形成多个训练子集;按照一类规则参数对应一组非劣解集的原则,针对各训练子集分别建立决策树,构成随机森林,以描述调度规则参数与非劣解集的回归关系;

[0028] 步骤3.3:采用检验期测试集数据,利用步骤3.2中建立的调度规则参数与非劣解集的回归关系,推求测试集中水库调度规则参数对应的非劣解集;

[0029] 步骤3.4:对步骤3.3中的预测结果进行综合,得到最终的预测结果;

[0030] 步骤3.5:重复步骤3.1—步骤3.4,直到遍历完所有水库调度规则参数,采用随机森林建立其与非劣解集的关系,并得到每一组水库调度参数对非劣解集的预测结果。

[0031] 作为优选,步骤4中,将所有水库调度规则参数预测的非劣解集与实际结果进行对比,并采用均方根误差评价所有水库调度规则参数的预测效果,将预测精度最高的前M个参数作为描述水库调度规则信息量最大的参数。

[0032] 与现有技术相比,本发明具有以下有益效果:

[0033] 该模型合理利用水库初始调度规则参数和非劣解集的信息,通过采用随机森林模型优选出最能反应水库多目标调度规则的参数,大大减少了被优化参数的数量,降低了多目标算法的优化成本,提高了水库多目标优化调度规则的推求效率,提供了更加简单的调度规则,从而大大方便了水库的运行管理。

附图说明

[0034] 图1为本发明实施例的随机森林建立及预测流程图;

[0035] 图2为本发明实施例的多目标优化算法流程图;

[0036] 图3为本发明实施例的多目标非劣解集分布示意图,其中(a)表示三个目标函数情形,(b)表示三个目标函数情形;

[0037] 图4为本发明实施例的水库多目标调度规则参数优选流程图。

具体实施方式

[0038] 为了便于本领域普通技术人员理解和实施本发明,下面结合附图及实施例对本发明作进一步的详细描述,应当理解,此处所描述的实施例仅用于说明和解释本发明,并不用于限定本发明。

[0039] 本发明将水库多目标优化调度模型与随机森林模型相结合,提出了一种基于数据挖掘的水库多目标调度规则参数优选方法,水库多目标调度规则参数优选流程图如图1所示。

[0040] 本发明的具体实现包括以下步骤:

[0041] 步骤1,收集水库上游站点观测的入库流量等数据资料,将水库调度规则用数学公式的方式进行描述;具体公式如下:

[0042]
$$O_t = a_t \cdot (V_t + w \cdot I_t) + b_t;$$

[0043] 式中: O_t 为水库第t时段出库流量(m^3/s); w 为单位转换系数; a_t 和 b_t 为调度函数中

以年为周期变化的规则参数,即每个月对应不同的值,但每年的同一个月份里 a_t (b_t) 取同样的值。 V_t 为安康水库第 t 时段初水库可用蓄水量 (m^3/s); I_t 为安康水库第 t 时段入库流量 (m^3/s)。

[0044] 步骤2,采用多目标优化算法,以水库历史入库流量作为输入,对水库调度规则中的参数进行优化,得到一系列能够满足水库多目标利用需求的非劣解集,多目标优化算法流程图及非劣解分布(示意图)分别如图2和图3所示。

[0045] 步骤2为本领域常规技术,进一步包括以下子步骤:

[0046] (1)对水库调度规则中的参数 a_t 和 b_t 进行编码,对编码的参数进行随机赋值,生成初始种群,对初始种群中每一组编码值进行解码,得到水库调度规则参数 a_t 和 b_t ,并将历史入库数据作为输入,按规则调度得到最后的调度结果。

[0047] (2)将(1)中产生的调度结果进行筛选排序,并挑选出不被任何其余结果支配的解,作为本次调度产生的非劣解集。

[0048] (3)在(2)中生成的非劣解集中挑选结果较优的解,并将这些解对应的参数集作为子代进行编码。

[0049] (4)对(3)生成的子代进行杂交、变异等操作,生成新的种群,并解码得到水库调度规则参数 a_t 和 b_t ,将历史入库数据作为输入,按规则调度得到最后的调度结果。

[0050] (5)重复(2)一(4),直到满足结束条件(本次得到的调度结果与上一次得到的结果之间的差异在预设范围内),将最后生成的一组参数 a_t 和 b_t ,作为水库多目标优化调度规则,并将对应的调度结果作为能够满足水库多目标利用需求的非劣解集。

[0051] 步骤3,根据步骤2中得到的非劣解集及其对应的水库调度规则参数 a_t 和 b_t ,按照一类参数对应一组非劣解集的原则,分别采用随机森林建立每一类水库调度规则参数与所得非劣解集的关系,将它们的属性进行分析和归纳,并通过选定的参数预测非劣解集,随机森林建立及预测流程图如图4所示。

[0052] 步骤3进一步包括以下子步骤:

[0053] (1)将历史资料分为率定期和检验期部分,并将一类水库调度规则参数和对应的非劣解集作为数据集,其中率定期的为训练集,检验期的为测试集。

[0054] (2)在率定期的训练集中进行 n 次重复随机抽样,形成多个训练子集。按照一类规则参数对应一组非劣解集的原则,针对各训练子集分别建立决策树,构成随机森林,以描述调度规则参数与非劣解集的回归关系。

[0055] (3)采用检验期测试集数据,利用(2)中建立的调度规则参数与非劣解集的回归关系,推求测试集中水库调度规则参数对应的非劣解集。

[0056] (4)对(3)中的预测结果进行综合,得到最终的预测结果。

[0057] (5)重复(1)一(4),直到遍历完所有水库调度规则参数,采用随机森林建立其与非劣解集的关系,并得到每一组水库调度参数对非劣解集的预测结果。

[0058] 步骤4,将所有水库调度规则参数预测的非劣解集与实际结果进行对比,并采用均方根误差评价所有水库调度规则参数的预测效果,将预测精度最高的前 M 个参数作为描述水库调度规则信息量最大的参数。

[0059] 步骤5,采用多目标算法优化步骤4中优选的参数,得到新的水库调度规则。

[0060] 本发明能有效利用水库优化调度信息,大大减少水库调度规则中的参数个数,提

高了水库多目标优化调度模型的求解效率,为水库的科学决策提供更简单且可操作性强的参考依据。

[0061] 应当理解的是,本说明书未详细阐述的部分均属于现有技术。

[0062] 应当理解的是,上述针对较佳实施例的描述较为详细,并不能因此而认为是对本发明专利保护范围的限制,本领域的普通技术人员在本发明的启示下,在不脱离本发明权利要求所保护的范围情况下,还可以做出替换或变形,均落入本发明的保护范围之内,本发明的请求保护范围应以所附权利要求为准。

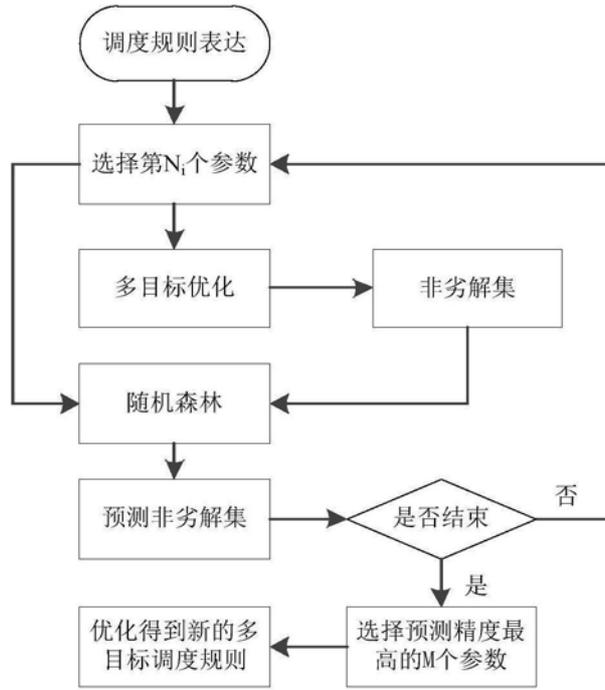


图1

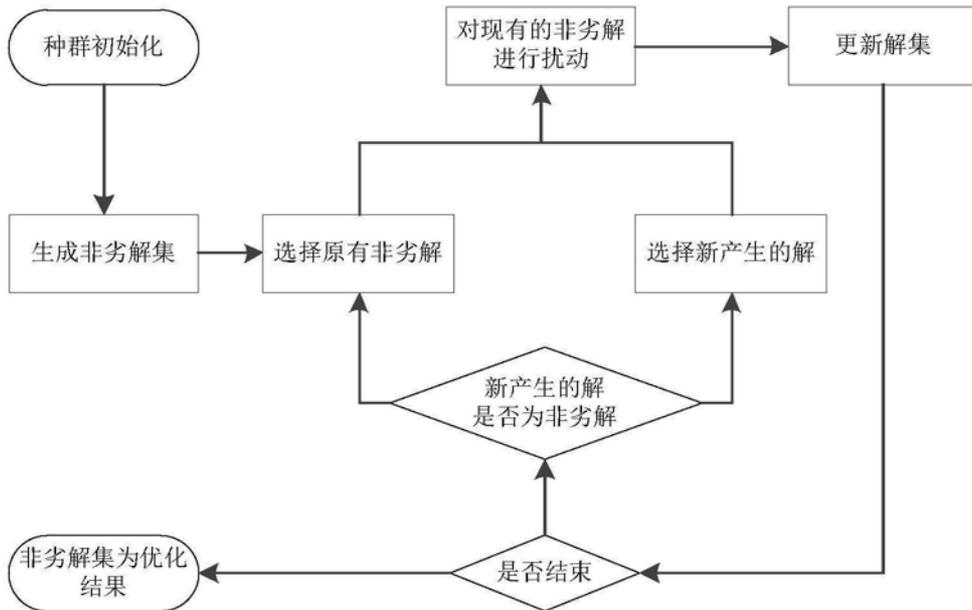
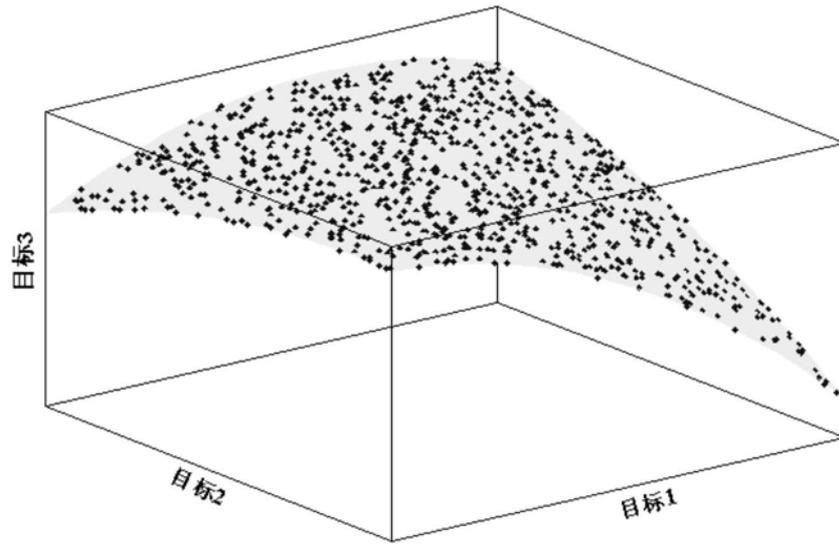
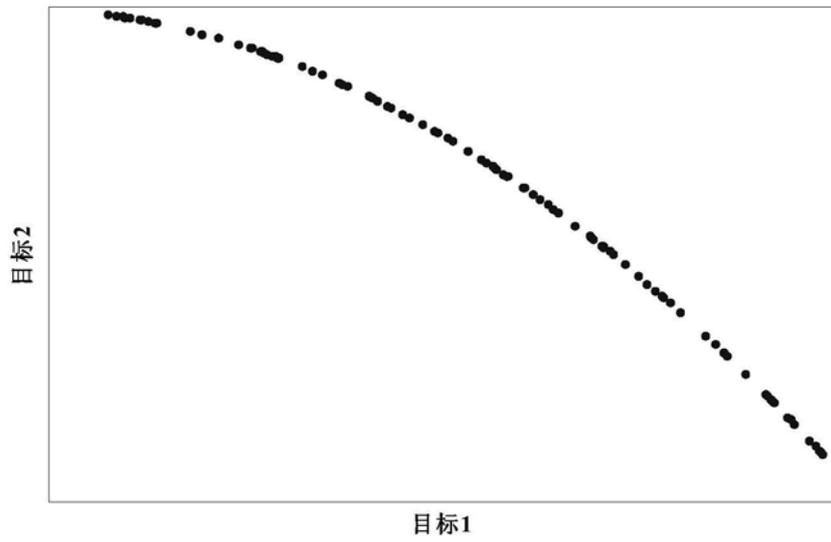


图2



(a) 三个目标函数情形



(b) 二个目标函数情形

图3

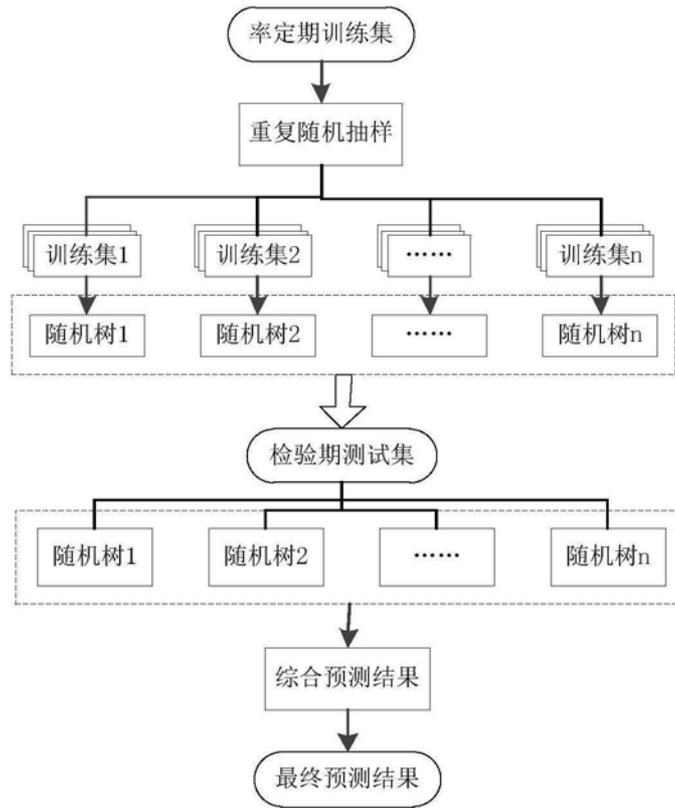


图4