

文章编号: 1000-7709(2014)07-0037-04

基于 RAGA 的 PPPCA 模型在水资源配置方案中的应用

王 浩,唐德善,何 格

(河海大学 水利水电学院, 江苏 南京 210098)

摘要: 为了解决日益突出的水资源供需矛盾并得到较为合理的水资源配置方案,将投影寻踪主成分分析模型(PPPCA)与加速遗传算法(RAGA)相结合,即利用 PPPCA 对配置过程中涉及的指标特征变量较多等问题进行降维处理,采用 RAGA 解决高维全局寻优问题,并将其应用于台兰河灌区水资源配置方案中。结果表明,采用加速遗传算法与投影寻踪主成分分析模型得出的配置方案合理、可行。

关键词: 水资源配置; 投影寻踪主成分分析模型; 加速遗传算法; 台兰河灌区

中图分类号: TV213.4

文献标志码: A

1 引言

水资源配置是在水资源评价、水资源利用现状以及需水预测、水资源保护、供水方案生成和供水预测等子规划的基础上进行的,是一个多指标的复杂多目标决策问题,根据水资源配置理论分析计算难以判断水资源配置方案的合理性。目前常用的方案评价方法有模糊综合评判法、层次分析法等^[1,2],这些方法在使用过程中大都存在人为赋权的干扰及等级分辨率较粗的不足。与其他方法相比,投影寻踪主成分分析法(PPPCA)可更客观地确定指标的权重,避免了人为因素产生的误差,通过对原始变量一维投影的研究,找出起主要作用的几个综合指标,这样一来保留了原数据的大部分信息,更具科学性^[3]。但该方法无法处理变量的复杂非线性优化问题。为此,本文将投影寻踪主成分分析模型(PPPCA)与加速遗传算法(RAGA)相结合,即利用投影寻踪主成分分析模型对配置过程中涉及的指标特征变量较多等问题进行降维处理,采用加速遗传算法解决高维全局寻优问题,有效解决了指标间多重相关性问题,可达到在低维空间进行水资源配置决策的目的,并通过实例进行了验证。

2 投影寻踪主成分分析模型(PPPCA)

投影寻踪主成分分析是探索性数据分析方法,它是根据实际问题需要,寻找对数据分类最有效的投影方向,是一种数值方法求极大解的优化方法。其基本思想是若投影指标函数值大于零的部分归因于前 $d(d \leq p)$ 个成分,则这些成分即可取代原来的 p 个特征,且信息完全利用,其具体实现步骤如下。

步骤 1 指标的无量纲处理。首先对水资源配置方案各评价指标进行归一化处理,对于越大越优的指标:

$$x'(i,j) = \frac{x^*(i,j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \quad (1)$$

对于越小越优的指标:

$$x'(i,j) = \frac{x_{\max}(j) - x^*(i,j)}{x_{\max}(j) - x_{\min}(j)} \quad (2)$$

其中 $i=1,2,\dots,n; j=1,2,\dots,p$
式中, $x'(i,j)$ 为指标特征值归一化序列; $x^*(i,j)$ 为指标特征值; $x_{\max}(j)$ 、 $x_{\min}(j)$ 分别为第 j 个指标的最大值和最小值; p 为优化变量的数目。

将 $x'(i,j)$ 进行归一化处理,即:

$$\begin{cases} x(i,j) = \frac{x'(i,j) - \bar{x}'(j)}{S_{x'(j)}} \\ i=1,2,\dots,n; j=1,2,\dots,p \end{cases} \quad (3)$$

式中, $\bar{x}'(j)$ 、 $S_{x'(j)}$ 分别为第 j 个指标的均值和标准差。

步骤 2 设样本 $\{x(i,j) | i=1,2,\dots,n; j=1,2,\dots,p\}$ 是水资源配置指标体系中 p 维观测数

收稿日期: 2013-09-25, 修回日期: 2013-11-18

基金项目: 水利部 2009 年公益性基金项目(200901084)

作者简介: 王浩(1989-),男,硕士研究生,研究方向为水利水电系统规划与工程经济, E-mail: 531225464@qq.com

通讯作者: 唐德善(1955-),男,教授、博导,研究方向为水资源规划与管理, E-mail: tds808@163.com

据,投影寻踪所要解决的是通过研究这些数据的一维投影来分析这些指标体系中的数据结构和特征。设 $a = (a(1), a(2), \dots, a(p))$ 为 p 维单位向量,则数据在 a 方向上的一维投影 $z(i)$ 为:

$$z(i) = \sum_{j=1}^p a(j)x(i,j) \quad i=1,2,\dots,n \quad (4)$$

步骤 3 投影指标即为 $z(i)$ 的函数,记作 $Q(a)$ 。指标值越大越好,投影寻踪即是要求一个单位向量 a_1 ,使得:

$$\begin{cases} Q(a_1) = \max\left(\frac{1}{n-1} \sum_{i=1}^n [z(i) - Ez]^2\right) \\ \sum_{j=1}^p a_1^2(j) = 1 \end{cases} \quad (5)$$

这是一个以 $\{a(j) | j = 1, 2, \dots, p\}$ 为优化变量的复杂非线性优化问题,式(5)中 $\|a_1\| = 1$ 为 a_1 向量的长度,因此在所找的这个方向上 a_1 一定含有数据的结构或特征,从而实现有效特征的提取。显然,这里的 $Q(a)$ 就是主成分分析中的协方差矩阵的最大特征值, a_1 即为主成分分析中的协方差矩阵最大特征值所对应的特征向量。若继续作投影,在与 a_1 垂直的空间里求单位向量 a_2 ,使得:

$$\begin{cases} Q(a_2) = \max\left(\frac{1}{n-1} \sum_{i=1}^n [z(i) - Ez]^2\right) \\ \sum_{j=1}^p a_2^2(j) = 1 \quad a_2 \perp a_1 \end{cases} \quad (6)$$

采用线性代数的方法可证明 a_2 即为主成分分析中的第二大特征向量,如此类推,可得:

$$\begin{cases} Q(a_d) = \max\left(\frac{1}{n-1} \sum_{i=1}^n [z(i) - Ez]^2\right) \\ \sum_{j=1}^p a_d^2(j) = 1 \quad a_d \perp a_1, a_2, \dots, a_{d-1} \end{cases} \quad (7)$$

于是可求出第三、第四主成分等,共提取投影指标函数值的大于零的 $d(d \leq p)$ 主成分。

步骤 4 计算各个主成分。主成分为:

$$\begin{cases} F_{i'} = \sum_{j=1}^p a_{i'}(j)x'(i,j) \\ i=1,2,\dots,n; i'=1,2,\dots,d; d \leq p \end{cases} \quad (8)$$

构造水资源配置方案中各个评价样本的综合评价函数 F_i 为:

$$\begin{cases} F_i = \sum_{i'=1}^d \alpha_{i'} F_{i'} \\ i=1,2,\dots,n; i'=1,2,\dots,d; d \leq p \end{cases} \quad (9)$$

式中, $\alpha_1, \alpha_2, \dots, \alpha_d$ 分别为第一、第二、...、第 d 主成分的贡献率。

3 基于实数编码的加速遗传算法 (RAGA) 的 PPPCA 模型

遗传算法是模拟生物在自然环境中的遗传和进化过程而形成的一种自适应全局优化概率搜索算法^[4]。本文将投影寻踪主成分分析模型 (PPPCA) 与加速遗传算法 (RAGA) 相结合,其具体实现步骤^[5]如下。

步骤 1 建立基于实数编码的加速遗传算法模型。求解最优化问题,即:

$$\begin{cases} \min f(X) \\ \text{s. t. } a_j \leq x_j \leq b_j \end{cases} \quad (10)$$

式中, f 为优化的目标函数。

步骤 2 优化变量的实数编码。为解决投影寻踪主成分分析模型 (式 (5)) 中以 $\{a(j) | j = 1, 2, \dots, p\}$ 为优化变量的复杂非线性优化问题,本文采用线性变换:

$$\begin{aligned} x(j) &= a(j) + y(j)(b(j) - a(j)) \\ j &= 1, 2, \dots, p \end{aligned} \quad (11)$$

将初试变量区间 $[a(j), b(j)]$ 上的第 j 个待优化变量 $x(j)$ 对应到 $[0, 1]$ 区间上的实数 $y(j)$, $y(j)$ 即为 RAGA 中的遗传基因。

步骤 3 计算目标函数值化。将步骤 2 各投影指标优化得到的目标函数值从小到大排列,选择排序最后面的 k 个作为优秀个体,使其直接进入下一代。

步骤 4 计算基于序的评价函数。评价函数用来对种群中的每个染色体设定一个概率,即使每个指标被选择的可能性具有适应性比例。

步骤 5 进行选择操作,检验每一后代可行性,选择可行后代,产生新的种群。

步骤 6 对步骤 5 产生的新种群进行变异操作。

步骤 7 演化迭代。

步骤 8 采用第一次、第二次进化迭代所产生的优秀目标函数值序列作为变量新的初始变化区间,算法进入步骤 2,重新运行步骤 2~7,形成加速运行,则优秀目标函数值序列区间将逐渐缩小,与最优点的距离越来越近,直到最优个体的优化准则目标函数达到预定的加速次数,提取投影指标函数值的大于零的主成分,然后计算各个主成分的百分比,结束整个算法运行。

4 实例应用

4.1 台兰河灌区概况

台兰河灌区位于新疆阿克苏地区温宿县东

部,总面积 $51.46 \times 10^4 \text{ hm}^2$,降水稀少,蒸发较大,多年平均径流量 $7.42 \times 10^8 \text{ m}^3$,径流年际变化不明显,但年内分配极不均匀,易出现季节性干旱或洪涝灾害。该区地下水含量虽丰富,考虑地下水资源可持续利用,应尽量少开采地下水。根据灌区的自然经济、水资源利用概况、水资源供需平衡分析结果,设计了地表水库方案(A)、地表水库+地下水开采方案(B)、地表水库+地下水开采+调节池方案(C)三种水资源配置方案。

4.2 台兰河灌区水资源配置方案评价指标体系及标准

根据科学性、可操作性、可比性、整体性等原则,以 2009 年为基准年,选取 2020 年($P=75\%$)各项指标对三种方案进行综合评价。本文选取生活用水比例(F_1)、工业用水比例(F_2)、工业用水保证率(F_3)、新增单方地表水供水投资(F_4)、新增单方地下水供水投资(F_5)、工程总投资(F_6)、供水工程年运行费(F_7)、地下水增加开采量(F_8)、地下水位埋深(F_9)、地表水利用率(F_{10})、水资源开发利用率(F_{11})、地下水开采率(F_{12})共 12 个指标^[6]。各指标的计算公式与作用见表 1。

表 1 台兰河灌区水资源配置方案评价指标体系
Tab.1 Water resources allocation scheme evaluation index system of Tailanhe irrigation area

评价指标	计算公式	指标作用
$F_1/\%$	生活用水/总供水量	反映生活用水状况
$F_2/\%$	工业用水/总供水量	反映工业用水状况
$F_3/\%$	实际工业供水量/预计工业需水量	反映工业用水满足程度
$F_4/(\text{元} \cdot \text{m}^{-3})$	新增地表水总投资/新增地表水量	反映开采地表水投资状况
$F_5/(\text{元} \cdot \text{m}^{-3})$	新增地下水总投资/新增地下水量	反映开采地下水投资状况
$F_6/\text{亿元}$		反映总投资状况
$F_7/\text{百万元}$		反映供水工程状况
$F_8/10^6 \text{ m}^3$		反映地下水开采状况
F_9/m		反映地下水位状况
$F_{10}/\%$	实际利用地表水量/地表水总量	反映地表水利用状况
$F_{11}/\%$	实际利用水资源总量/水资源总量	反映水资源开发利用状况
$F_{12}/\%$	开采地下水量/地下水总量	反映地下水开采状况

本文将台兰河灌区水资源配置方案评价等级分为差、中、良、好四个评价等级,评判标准^[7]见表 2。

表 2 台兰河灌区水资源配置方案评价分级标准
Tab.2 Water resources allocation scheme evaluation classification standard of Tailanhe irrigation area

评价指标	差	中	良	好
$F_1/\%$	20~30	30~40	40~50	50~60
$F_2/\%$	7~8	8~9	9~10	10~11
$F_3/\%$	80~85	85~90	90~95	95~100
$F_4/(\text{元} \cdot \text{m}^{-3})$	0.45~0.60	0.30~0.45	0.15~0.30	0~0.15
$F_5/(\text{元} \cdot \text{m}^{-3})$	1.13~1.50	0.75~1.13	0.37~0.75	0~0.37
$F_6/\text{亿元}$	24~26	22~24	20~22	18~20
$F_7/\text{百万元}$	40~44	36~40	32~36	28~32
$F_8/10^6 \text{ m}^3$	75~100	50~75	25~50	0~25
F_9/m	3~3.2	2.8~3	2.6~2.8	2.4~2.6
$F_{10}/\%$	68~73	73~78	78~83	83~88
$F_{11}/\%$	73~76	76~79	79~82	82~85
$F_{12}/\%$	62~72	52~62	42~52	32~42

4.3 评价结果分析

采用 Matlab7.0 对投影寻踪主成分分析模型

进行编程,在 RAGA 寻优过程中,选定父代初始种群规模为 $n=400$,交叉概率 $p_c=0.80$,变异概率 $p_m=0.80$,优秀个体数目选定为 20 个, $\alpha=0.05$,加速次数为 20,最终共提取了 3 个主成分,3 个投影指标函数值分别为 $Q(a_1)=7.538$ 、 $Q(a_2)=3.259$ 、 $Q(a_3)=1.253$,各主成分的贡献率分别为 $\alpha_1=62.55\%$ 、 $\alpha_2=27.05\%$ 、 $\alpha_3=10.40\%$ 。

最后通过式(9)求得各个评价样本的综合评价函数值及承载力相对等级见表 3。由表 3 可看出,台兰河灌区水资源配置方案中方案 C 最好,其次是方案 B,最后是方案 A。因此选择方案 C 配置模式,即地表水库+地下水开采+调节池方案进行台兰河灌区水资源的配置。

表 3 各个方案综合评价函数值及评价等级
Tab.3 Each scheme comprehensive evaluation function value and evaluation grades

方案	评价价值	评价等级
A	1.154	差
B	1.844	中等
C	1.876	良好

为了进一步验证基于 RAGA 的 PPPCA 模型(本文方法)计算结果的可行性和合理性,将本文方法计算结果与改进物元法^[7]计算结果进行比较,见表 4。由表 4 可看出,本文方法计算结果与改进物元法计算结果一致,均为方案 C 水资源配置方案良好。由此可见,基于 RAGA 的 PPPCA 模型得出的配置方案合理、可行。

表 4 本文方法与改进物元法结果对比
Tab.4 Comparison of evaluating results of PPPCA method based on RAGA and improved matter element method

方法	方案	评价等级
本文方法	A	差
	B	中等
	C	良好
改进物元法	A	差
	B	中等
	C	良好

5 结论

a. 本文将投影寻踪主成分分析模型(PPPCA)与加速遗传算法(RAGA)相结合,有效解决了指标间多重相关性问题,可达到在低维空间进行水资源配置决策的目的。

b. 台兰河灌区采用方案 C,即地表水库+地下水开采+调节池水资源配置方案有利于水资源可持续利用。

参考文献:

[1] 高波,徐建新,班培莉.基于模糊优选模型的水资源配置方案评价[J].灌溉排水学报,2008,27(6):58-60.

(下转第 32 页)

5 结论

a. 基于区间机会约束的优化模型能够很好地结合区间、机会优化的优点,既能处理区间不确定性又能处理随机不确定性,可获得不同风险水平下的最优区间决策方案。

b. 以交互式运算法则为基础,将优化模型转化为两个确定的符合期望目标下限和上限的子模型,通过求解这两个连续的子模型,可以得到稳定的区间解。

c. 将基于区间机会约束的优化模型应用于滨州新区水资源配置问题中,其结果可为滨州新区水资源优化配置提供支持。但结果的区间范围比

较大,不利于决策者做出准确的决定,未来有待进一步研究。

参考文献:

[1] Huang G H. A Hybrid Inexact-stochastic Water Management Model[J]. European Journal of Operational Research, 1998,107(1):137-158.

[2] Huang G H, Baetz B W, Patry G G. A Grey Linear Programming Approach for Municipal Solid Waste Management Planning under Uncertainty [J]. Civil Engineering Systems, 1992, 9(4): 319-335.

[3] 聂相田,邱林,朱普生. 水资源可持续利用管理不确定性分析方法及应用[M]. 郑州:黄河水利出版社, 1999.

Water Resources Optimization Allocation of Binzhou New District Based on Interval Chance Constrained Programming

LIU Sanxing¹, YIN Jianguang¹, LI Wei¹, HUANG Guohe¹, WANG Zhe^{1,2}

(1. Resources and Environmental Research Academy, North China Electric Power University, Beijing 102206, China; 2. School of Energy and Environment, Inner Mongolia University of Science and Technology, Baotou 014010, China)

Abstract: To deal with the uncertainty and complexity in water resource optimization allocation, an inexact chance constrained programming (ICCP) model is developed for planning the water resource allocation in Binzhou New District. Based on the interactive solution method, it can provide the optimal allocation scheme of water resources in this zone. At the same time, the model can reflect not only uncertainties but also the complexities of water resources systems. The results are valuable for supporting the adjustment of existed development pattern and identifying a desired water-allocation plan under uncertainty.

Key words: interval chance constrained; water resources system; optimization; uncertainty; Binzhou New District

(上接第 39 页)

[2] 金晶,唐德善,李晓英. 基于 AHP-模糊决策模型的水资源配置方案优选[J]. 水电能源科学, 2013, 31(7):33-35.

[3] 付强,赵小勇. 投影寻踪模型原理及其应用[M]. 北京:科学出版社, 2006.

[4] 金菊良,杨晓华,丁晶. 基于实数编码的加速遗传算法[J]. 四川大学学报:工程科学版, 2000, 32(4): 20-24.

[5] 付强,邢桂君,王兆函,等. 基于 RAGA 的 PPC 模型

在节水灌溉项目投资决策中的应用[J]. 系统工程理论与实践, 2003, 23(2): 139-144.

[6] 唐德善,薛文志,张范平,等. 新疆温宿县台兰河灌区综合规划[R]. 新疆:新疆维吾尔自治区水利规划设计局, 2011.

[7] 何格,唐德善. 基于改进物元可拓模型的水资源配置方案评价[J]. 水电能源科学, 2012, 30(12): 24-26, 138.

Water Resources Allocation Schemes Evaluation Based on Projection Pursuit Principal Components Analysis Model

WANG Hao, TANG Deshan, HE Ge

(College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

Abstract: In order to solve the increasingly prominent contradiction between supply and demand of water resources and to get more reasonable water resources allocation schemes evaluation, this paper combines the projection pursuit principal components analysis (PPPCA) model with water resources allocation schemes. By using this model that involves more index characteristic variables to reduce the dimension in the process of water resources allocation problem, real-coded accelerating genetic algorithm (RAGA) is adopted to solve high-dimensional global optimization problems. The model can effectively solve multiple correlation among indices in the problem, which achieves the water resources allocation decisions in low dimensional space. Taking water resources allocation of Tailanhe irrigation area for an example, application results show that the allocation scheme obtained by PPPCA with RAGA is reasonable and feasible.

Key words: water resources allocation; projection pursuit principal components analysis model; accelerating genetic algorithm; Tailanhe irrigation district