

# 基于灰色关联分析的 lasso 法在区域用水经济效益系数中的应用\*

屈国栋<sup>1</sup>, 楼章华<sup>1</sup>, 王磊<sup>2</sup>

(1. 浙江大学水文与水資源工程研究所, 浙江 杭州 310058;

2. 浙江大学电气工程学院, 浙江 杭州 310027)

**摘要:** 水资源优化配置是解决区域水资源短缺问题的一个有效途径, 而进行水资源优化配置的基础是确定区域用水经济效益系数。运用灰色系统理论中的关联分析方法, 分析区域历史用水状况, 对影响用水经济效益的各因素进行分析, 初步区分出主要影响因子。结合 Lasso 套索回归, 挖掘各影响因子与经济总产值之间的关系, 利用该方法重点探讨了周口市工业用水和农业用水经济效益系数的求解。结果表明, 该方法能较为准确的求出研究区域的用水经济效益系数, 为水资源优化配置中经济效益系数的确定提供了一种更加精确和量化的方法。

**关键词:** 灰色关联分析; Lasso 回归; 经济效益系数; 周口市

**中图分类号:** [TV-9] **文献标志码:** A **文章编号:** 0529-6579(2013)02-0019-05

## Research on Regional Water Use Economic Efficiency Coefficient Based on Gray Relational Analysis and Lasso Method

QU Guodong<sup>1</sup>, LOU Zhanghua<sup>1</sup>, WANG Lei<sup>2</sup>

(1. Institute of Hydrology and Water Resources, Zhejiang University, Hangzhou 310058, China;

2. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

**Abstract:** Optimal allocation of water is an effective way to resolve the issue of regional water shortages, while identification of economic efficiency coefficient of the regional water is the basis. Use relational analysis method in gray system theory, the various factors that affect water use economic benefits based on regional historical water use data is analyzed, and the main impact factors is distinguished. Mining the relationship between impact factors and total economic output combined with Lasso regression, and the method is used to solve water use economic efficiency coefficient of industrial water and agricultural water in Zhoukou City. The result shows that the method can calculate the water use economic efficiency coefficient of study area more accurately, and provide a more accurate and quantitative methods in fixing the economic efficiency coefficient of optimal allocation of water.

**Key words:** gray relational analysis; Lasso regression; economic efficiency coefficient; Zhoukou City

水资源优化配置是指在流域或特定区域内, 通过工程或非工程措施对有限的水资源进行空间、时间上的调配, 尽量满足各区域的用水需求, 促进流域及区域经济的持续、稳定发展和生态环境的健康、稳定<sup>[1]</sup>。水资源可持续利用的基本手段之一就是水资源优化配置<sup>[2]</sup>, 水资源优化配置是解决

区域水资源短缺, 提高用水经济和社会效益的重要和有效途径之一, 而各用水户用水经济效益系数的确定则是优化配置的关键。当前, 确定经济效益系数的方法主要有两种: 直接确定法和层次分析法。直接确定法是通过分析经济效益系数的物理意义, 利用现有的一些实际数据直接确定。例如: 工业用

\* 收稿日期: 2012-08-15

基金项目: 国家自然科学基金资助项目(40839902)

作者简介: 屈国栋(1987年生), 男, 博士研究生; E-mail: quguodong1987@163.com

水经济效益系数可取用水定额 (工业万元产值取水量) 的倒数; 农业灌溉用水经济效益系数按灌溉后的农业增产效益乘以水利分摊系数确定<sup>[3]</sup>。赵得军等<sup>[4]</sup>在水资源配置模型中, 把不同形式水源的用水效益系数看作是时间的函数, 利用遗传算法优化效益系数。沈军等<sup>[5]</sup>将水资源优化配置问题模拟为生物进化问题, 通过遗传算法来求解经济效益系数。现有的方法都存在一些问题, 如直接确定法虽然可以在一定程度上反映实际情况, 但是由于需要确定的影响参数较多, 量化比较困难, 很难运用到模型求解中<sup>[6]</sup>。

本文在前人研究的基础上, 首先运用灰色关联分析法的优势, 初步确定影响经济效益系数的主要影响因子, 然后根据历史用水数据, 引入 LASSO 套索回归, 作为量化的统计方法, 对用水经济效益系数进行量化。

## 1 灰色关联分析法辨识经济效益系数影响因子

本文选用灰色关联分析方法, 辨识影响因子, 该方法是在灰色系统理论基础上的一种方法, 可以分析系统中各个因素的关联程度。通过各因子的关联度比较, 可以确定出主次因素, 得出影响某变量的主要因子。

灰色关联分析的基本思想是根据序列曲线几何形状的相似程度来判断其联系是否紧密, 曲线越接近, 相应序列之间的关联度就越大, 反之就越小<sup>[7]</sup>。

各影响因子与需水量的关联系数与关联度计算公式如下<sup>[8]</sup>。

关联系数  $\xi$  :

$$\xi(k) = [\min_i \min_k |X_0(k) - X_i(k)| +$$

$$\rho \max_i \max_k |X_0(k) - X_i(k)|] / [ |X_0(k) - X_i(k)| +$$

$$\rho \max_i \max_k |X_0(k) - X_i(k)| ] \quad (1)$$

关联度:

$$r_i = \frac{1}{N} \sum_{k=1}^N \xi(k) \quad (i = 1, 2, \dots, m) \quad (2)$$

其中:  $X_0(k)$  和  $X_i(k)$ ,  $k = 1, 2, \dots, n$ ,  $i = 1, 2, \dots, m$ , 分别为各个用水部门的用水量时间序列和相应的影响因子时间序列。 $\rho$  为分辨系数, 值处于 0~1 之间, 通常取 0.5。

## 2 lasso 方法介绍

LASSO 套索估计方法用于逼近线形模型,

LASSO 最小化受约束的残留平方和误差, 由于这一约束的本质特性, LASSO 可以使得一些系数严格为零, 从而能够给出可解释的模型<sup>[9]</sup>。Volker<sup>[10]</sup>提出了一个称为广义 LASSO 的稀疏核回归模型, 它是 LASSO 方法的非线性形式。由于解的极度稀疏性, 广义 LASSO 的性能优于支持向量机。

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j)^2 \quad (3)$$

$$s. t. \sum_{j=1}^P |\beta_j| \leq t \quad (4)$$

其中,  $y_i$  是归一化的目标函数真实值,  $\beta_0$  是线性分类函数的截距。 $L_1$  套索罚  $\sum_{j=1}^P |\beta_j|$  的约束使得

解在  $y_i$  上为非线性, 并使用二次规划来计算它们<sup>[11]</sup>。由于该约束的特征, 使得  $t$  充分小将导致某些系数恰好为 0。这样, 套索做了某种连续的子集选择。通过自适应的选择  $t$ , 极小化期望预测误差估计, 得到  $\beta_j$  的估计。

## 3 基于灰色关联分析的 lasso 法求解经济效益系数

### 3.1 研究区域概况

周口市位于河南省东南部, 面积 1.19 万  $\text{km}^2$ , 耕地 82.6 万  $\text{hm}^2$ , 人口 1 090 万人。辖八县一市一区和一个省级经济开发区。周口地势平坦, 土地肥沃, 是国家重要的粮、棉、油、肉、蛋、奶生产基地, 有“中原粮仓”之称; 2010 年粮食产量达 72.35 亿  $\text{kg}$ , 居河南省第一位。周口市多年平均水资源总量为 26.10 亿  $\text{m}^3$ 。全市人均水资源量 244  $\text{m}^3$ , 为河南省人均水资源量 500  $\text{m}^3$  的 48.8%; 亩均水资源量 204  $\text{m}^3$ , 为全省亩均水资源量 395  $\text{m}^3$  的 51.6%。周口市属于水资源严重缺乏地区。

近年来, 由于加快中原经济区建设, 城市化进程不断推进, 周口市水资源短缺问题日益显现, 加之水污染和水环境问题的加剧, 尤其是水质性缺水问题引人注目。由水资源而引发的各种问题, 对经济发展的约束已经凸显, 水资源的数量和质量已经难以承载当地经济社会的可持续发展。因此, 准确的量化用水经济效益系数, 实现合理的配置有限的水资源, 势在必行。

本文根据行政区域, 将周口市十个县市分为十个子区, 根据《周口市水资源公报》、《周口市统计年鉴》等资料得到相关数据, 运用各子区 2010 年的实际用水情况等数据, 对所建模型进行拟合, 确定出周口市各子区工业、农业用水的经济效益系数。

### 3.2 基于灰色关联分析的 lasso 法经济效益系数求解

3.2.1 经济效益系数影响因子灰色关联分析法辨识 分析研究区域实际用水状况，首先定性确定经济效益系数影响因子，依据实际数据，按照前文所述灰色关联分析方法，分别求出各用水部门的实际用水经济效益与各影响因子的关联系数，进一步计算出关联度。通过关联度的排序可初步确定主要影响因子。得出周口市各用水部门实际用水经济效益主要影响因子如表 1 所示。可初步确定与经济效益系数关联性较高的因子。

表 1 周口市各用水部门经济效益系数影响因子  
灰色关联分析法计算结果

Table 1 Results of effecting factors identified by gray relational analysis method foreconomic efficiency coefficient of Zhoukou

用水部门	经济效益系数影响因子	关联度
工 业	火电用水量	0.824 3
	地下水用水量	0.732 4
	万元产值用水量	0.685 6
	国有及规模以上工业用水量	0.842 5
农 业	粮食产量	0.910 5
	小牲畜数量	0.887 6
	菜田用水量	0.923 1
	农田灌溉用水量	0.757 8

3.2.2 基于 lasso 的经济效益系数求解 在运用灰色关联分析法初步确定各子区各部门用水经济效益

系数的主要影响因子后，分析周口市工业、农业两个用水部门的数据，引入 LASSO 套索估计，分别量化每个部门中的各子项的效益系数。以每个行业中总体经济指标为标准，计算出各个影响因素的效益系数。

$$Y = L_0 + \sum_{i=1}^5 k_i E_j \quad (5)$$

其中， $Y$  为总体经济指标， $E_j$  是影响因素， $k_i$  是效益系数， $L_0$  为截距。使用 Matlab (The MathWorks, Inc. R2009b)，利用开源软件包 glmnet package，实现了凸优化方法自适应估计  $K_i$ ，具体算法参见 Koh K 等的文献<sup>[12]</sup>。

根据《周口市统计年鉴》、《周口市水资源公报》等资料，进行数据的初步统计分析，得到 2010 年周口市各行政分区工业、农业经济效益系数主要影响因子的值如表 2、3 所示。

运用前述理论，对表 2、3 中的数据，以工业总产值和农业总产值为总体经济指标，用 lasso 回归分析，得到线性回归函数：

$$\begin{aligned} \text{工业总产值} = & 8.861 + 1.1179 * \text{火电增加值} + \\ & 0.9172 * \text{国有及规模以上增加值} + 0.795 * \text{规模以下增加值} + \\ & 39.4098 * \text{火电用水量} + 9.9909 * \text{地下水用水量} \end{aligned} \quad (6)$$

$$\begin{aligned} \text{农业总产值} = & 7.3411 + 0.0744 * \text{粮食产量} + \\ & 0.1844 * \text{小牲畜数量} + 27.5748 * \text{菜田用水量} + \\ & 1.6395 * \text{农田灌溉用水量} \end{aligned} \quad (7)$$

表 2 2010 年周口市各行政分区工业经济效益系数主要影响因子值

Table 2 The effecting factors value of industrial economic efficiency coefficient of Zhoukou in the year 2010

序号	行政分区	火电用水量 /(亿 m <sup>3</sup> )	火电增加值 /(亿元)	地下水用水量 /(亿 m <sup>3</sup> )	规模以下增加值 /(亿元)	国有及规模以上 增加值/(亿元)	工业总产值 /(亿元)
1	市区	0.068 4	2.400 0	0.195 3	8.041 9	28.590 9	50.746 2
2	项城市	0.105 4	1.500 0	0.457 8	27.048 6	72.286 9	108.962 3
3	扶沟县	0.000 0	0.000 0	0.148 7	6.989 6	23.747 8	35.255 5
4	西华县	0.000 0	0.000 0	0.185 4	6.877 6	31.254 3	43.692 1
5	商水县	0.000 0	0.000 0	0.153 1	9.655 5	15.662 2	29.982 2
6	太康县	0.000 0	0.000 0	0.229 6	12.848 1	23.978 6	42.451 6
7	鹿邑县	0.000 0	0.000 0	0.395 9	14.749 1	54.099 3	75.468 1
8	郸城县	0.000 0	0.000 0	0.401 6	12.629 7	54.983 3	73.359 8
9	淮阳县	0.000 0	0.000 0	0.229 5	15.723 5	26.098 9	48.522 6
10	沈丘县	0.000 0	0.000 0	0.205 6	13.880 4	29.393 4	49.757 3

表 3 2010 年周口市各行政分区农业经济效益系数主要影响因子值

Table 3 The effecting factors value of agricultural economic efficiency coefficient of Zhoukou in the year 2010

序号	行政分区	农业总产值 /(亿元)	粮食产量 /(万 t)	小牲畜 /(万头)	菜田用水量 /(亿 m <sup>3</sup> )	农田灌溉 用水量/(亿 m <sup>3</sup> )
1	市区	7.227 2	9.753 5	8.768 6	0.033 8	0.061 1
2	项城市	33.125 9	74.006 9	66.263 8	0.327 9	0.846 2
3	扶沟县	28.708 4	51.248 7	39.197 5	0.318 9	0.987 6
4	西华县	41.494 8	62.204 8	83.218 6	0.359	1.219 2
5	商水县	45.824 1	102.247 3	86.144 4	0.293	1.116 6
6	太康县	45.316 5	107.002 2	107.314 7	0.376 5	1.756 5
7	鹿邑县	39.544 4	80.681 9	72.267 1	0.351 4	1.253 2
8	郸城县	39.447 6	81.805 4	71.324 4	0.362	1.246 1
9	淮阳县	50.790 4	83.845 4	101.834 3	0.375 3	1.209
10	沈丘县	34.157 3	73.863 6	86.664 9	0.320 6	0.955 3

由式 (6)、(7) 可知, 在以工业、农业总产值最大为目标的情况下, 影响总产值最大的影响因子及其值如表 4 所示。

表 4 2010 年周口市工业、农业用水经济效益影响因子及其系数值

Table 4 The effecting factors of industrial and agricultural water use economic efficiency coefficient and the coefficient of Zhoukou in the year 2010

用水部门	用水经济效益影响因子	效益系数值
工业	火电用水量	39.409 8
	地下水用水量	9.990 9
	火电增加值	1.117 9
	国有及规模以上增加值	0.917 2
	规模以下增加值	0.795
农业	粮食产量	0.074 4
	小牲畜数量	0.184 4
	菜田用水量	27.574 8
	农田灌溉用水量	1.639 5

由表 4 可知, 工业用水方面, 对工业总产值影响较大的有火电用水量和地下水用水量, 因此, 为了提高工业用水经济效益, 可以通过提高火电用水量, 增加工业用水中地下水的比例, 以此提高工业用水经济效益。农业用水方面, 对农业总产值影响较大的是菜田用水量和农田灌溉用水量, 可以通过适当增加菜田种植面积, 以及在可能的情况下, 通过提高农田灌溉用水量来增加农业用水经济效益。根据表 4 中的影响因子及其系数值, 可对区域合理的利用水资源提供参考, 为水资源的合理配置提供对策和建议。

## 4 结 语

本文运用灰色相关分析方法, 初步分析区域用水经济效益系数的主要影响因子, 然后运用 lasso 方法求解, 以区域总产值最大作为目标, 综合考虑各个县市的用水收益情况, 得出各影响因素的效益系数, 这些效益系数可以作为当年和时间较临近的年份的参考依据。其他年份的数据采用本文中的方法, 可以得到相应的效益系数集合。因为每年的效益系数是动态变化的, 所以采用本文中的方法能够动态地计算每年的效益系数, 再据此做进一步的优化。本文为水资源优化配置模型中经济效益系数的确定提供了一种更加精确和量化的方法。为区域水资源管理和水资源的合理配置提供了依据。

## 参考文献:

- [1] 孙志林, 夏珊珊, 许丹, 等. 区域水资源的优化配置模型[J]. 浙江大学学报: 工学版, 2009, 43(2): 344 - 348.
- [2] 王海政, 全允恒. 可持续发展视角下的区域水资源优化配置模型[J]. 清华大学学报: 自然科学版, 2007, 47(9): 1531 - 1536
- [3] 冯耀龙, 韩文秀, 王宏江, 等. 面向可持续发展的区域水资源优化配置研究[J]. 系统工程理论与实践, 2003(2): 133 - 138.
- [4] 赵得军, 洪林, 李远华, 等. 效益系数年际动态变化的水资源合理配置[J]. 中国农村水利水电, 2004, 22(9): 34 - 36.
- [5] 沈军, 刘勇键. 水资源优化配置模型参数识别的遗传算法[J]. 武汉大学学报: 工学版, 2002, 35(3): 13 - 16.

## 4 结 论

以 Bandelet 变换理论和算法研究为主线, 本文从基于 Bandelet 变换的数字水印图像分割和几何信息特征描述表示出发, 建立了一种有效的树结构多级水印数据分配模型。和基于小波变换的单级信息隐藏算法比较, 实验结果说明了本文方法的有效性。未来, 借鉴 Bandelet 变换的多级数据处理, 我们将从指纹水印信息的特征提取、隐藏、检测等算法构造一种密钥密码实现技术, 为数字水印的电子信息技术作为实际应用背景展开深入地研究。

### 参考文献:

- [1] 赵耀. 基于小波变换的抵抗几何攻击的鲁棒视频水印[J]. 中国科学 E 辑, 2006, 36(2): 137 - 152.
- [2] EDWARD J, WONG P W, DITTMANN J, et al. Security, forensics, steganography, and watermarking of multimedia contents X (proceedings volume) [DB/OL]. SPIE Digital Library, April 2008.
- [3] VOLOSHYNOVSKIY S, PEREIRA S, PUN T. Attack modeling: Towards a second generation watermarking benchmark[J]. Signal Processing, Special Issue on Information Theoretic Issues in Digital Watermarking, June 2010, 90(6): 1177 - 1214.
- [4] 赵慧民, 张惠乐, 胡学骏. 一种基于树结构的高斯分布扩频指纹嵌入技术的方法研究[J]. 中山大学学报: 自然科学版, 2006, 45(1): 42 - 45.
- [5] 赵慧民, 康显桂, 黄继武. Laplacian 信源分割及其并行

Gaussian 信道传输的信息隐藏技术研究[J]. 中山大学学报: 自然科学版, 2008, 47(4): 29 - 34.

- [6] DAUBECHIES I, HAN B, RON A, et al. Framelets: MRA based construction of wavelet frames[J]. Applied Comput Harmon Anal, 2004, 14(1): 1 - 46.
- [7] DO M N, VETTERLI M. Contourlets, beyond wavelets [C]. G V Welland ed, Academic Press, 2003.
- [8] PENNEC E L, MALLAT S. Image compression with geometrical wavelets[C]. In Proc. of ICIP2000. Vancouver, Canada, September, 2000: 661 - 664.
- [9] DONOHO D, HUO X. Beamlets and multiscale image analysis, multiscale and multiresolution methods [J]. Springer Lecture Notes in Computational Science and Engineering, 2002(20): 149 - 196.
- [10] PENNEC E L, MALLAT S. Bandelet representations for image compression[C]. Proceedings. 2001 International Conference on Volume 1, 2001: 12.
- [11] PENNEC E L, MALLAT S. Sparse geometric image representations with bandelets [J]. IEEE Trans. Image Processing, 2005, 14(4): 423 - 438.
- [12] 杨岳湘, 罗永, 叶昭辉. 基于 Bandelet 与自适应矩阵范数的全频域无损水印方法[J]. 计算机研究与发展, 2007, 44(12): 1996 - 2003.
- [13] 刘绪崇, 罗永, 王建新, 等. 基于第二代 Bandelet 变换的图像认证水印算法[J]. 通信学报, 2010, 31(12): 123 - 130.
- [14] 赵慧民, 郭一缜, 丁晓艳. 一种数字图像和指纹水印位平面合成的优化方法[J]. 中山大学学报: 自然科学版, 2010, 49(6): 64 - 67.

(上接第 22 页)

- [6] 陈夺峰, 黄岁樑. 改进遗传算法确定水资源优化配置模型经济效益系数[J]. 水资源保护, 2008, 24(2): 6 - 9.
- [7] 刘思峰, 党耀国, 方志耕等. 灰色系统理论及其应用[M]. 北京: 科学出版社, 2005.
- [8] 王艳菊, 王珏, 吴泽宁. 基于灰色关联分析的支持向量机需水预测研究[J]. 节水灌溉, 2010, 10(4): 49 - 52.
- [9] TIBSHIRANI, R. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society.

1996, 58(1): 267 - 288.

- [10] VOLKER R. The generalized LASSO [J]. IEEE Transactions on Neural Networks, 2004, 15(1): 16 - 28.
- [11] 李霞, 刘超. 基于收缩机制的若干回归模型比较研究[J]. 统计与决策, 2008, 5: 30 - 32.
- [12] KOH K, KIM S J, BOYD S. An interior-point method for large-scale L1-regularized logistic regression [J]. Journal of Machine Learning, 2007(8): 1519 - 1555.