

DOI:10.3880/j.issn.1004-6933.2017.06.22

基于改进的投影寻踪-云模型的农业灌溉水质综合评价

于嘉骥^{1,2}, 张慧妍^{1,2}, 王小艺^{1,2}, 许继平^{1,2}, 王立^{1,2}

(1. 北京工商大学计算机与信息工程学院, 北京 100048;

2. 北京工商大学食品安全大数据技术北京市重点实验室, 北京 100048)

摘要:提出基于投影寻踪函数和云模型的水质综合评价模型,选取太湖流域20个样本盐度、氯化物、氨氮、溶解性固体4类具有代表性的农业灌溉水质监测数据,在综合其投影值及隶属度基础上,计算农业灌溉水质的等级区分粒度。结果表明,投影寻踪模型计算值平均绝对误差仅为0.1252级,达到了较好的水质评价精确度,同时利用云模型计算各个监测指标得到的最大综合确定度所属级别与经验等级一致。

关键词:水质评价;农业灌溉;投影寻踪;云模型;等级区分粒度

中图分类号:X824

文献标志码:A

文章编号:1004-6933(2017)06-0142-05

Comprehensive evaluation of agricultural irrigation water quality based on modified projection pursuit-cloud model

YU Jiaji^{1,2}, ZHANG Huiyan^{1,2}, WANG Xiaoyi^{1,2}, XU Jiping^{1,2}, WANG Li^{1,2}

(1. School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China; 2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China)

Abstract: This paper proposed a comprehensive evaluation model of water quality based on projection pursuit function and cloud model, 4 representative water quality monitoring data of agricultural irrigation water collected from 20 samples in Taihu River Basin were selected, covering salinity, chloride, ammonia nitrogen and dissolved solids. On the basis of synthesizing their projection value and membership degree, the grade discrimination granularity of the agricultural irrigation water quality was calculated, the results show that the average absolute error of the calculated value of the projection pursuit model was only 0.1252, which means the accuracy of water quality evaluation has been achieved preferably. Meanwhile, the maximum comprehensive certainty degree of each monitoring index calculated by cloud model was consistent with the experience grade.

Key words: water quality evaluation; agricultural irrigation; projection pursuit; cloud model; grade distinguish granularity

农业用水安全是农作物生产的源头保证,是决定农产品质量安全的重要因素之一。当前我国的农业相关水质标准分别由国家、行业及地方制订,涉及众多评价指标,如何有效协调、相互补充,对于防范水污染所造成的农业生态环境破坏,具有重要意义。

目前的综合评价方法主要有模糊综合评价法^[1-2]、投影寻踪模型^[3-4]、水质标识指数法^[5-6]等。近年来,水质评价中不确定因素的存在逐渐受到重视。余勋等^[7]针对水质评价过程中模型结构和参数的不确定性,建立了基于三角模糊数的贝叶斯模

基金项目:北京市属高校创新能力提升计划(PXM2014_014213_000033);北京市教委科技计划重点项目(KZ201510011011)

作者简介:于嘉骥(1992—),男,硕士研究生,研究方向为水环境监测与治理。E-mail:njc-jiaji@126.com

通信作者:张慧妍,副教授。E-mail:zhanghuiyan369@126.com

糊综合水质评价模型,对不确定性的刻画更为全面、符合实际。巩奕成等^[8]为解决评价中的模糊性与指标不相容性,引入萤火虫算法,优化寻求最佳投影方向,实现合理评价水质的目的。梁中耀等^[9]针对水质不确定性可能导致的决策风险,利用二项分布检验法对水质达标情况进行了评价,使评价结果更具鲁棒性。

本研究根据农业灌溉用水水质特征,在参考文献[10-11]基础上,结合监测数据混有噪声、划分等级存在中间过渡状态等实际情况,建立基于数据驱动的投影寻踪函数,运用多维正态云模型中的隶属度判据,并计算等级区分粒度,对农业灌溉用水水质进行综合评价。

1 原理及方法

水质综合评价实际上是一个确定性与不确定性相结合的分析过程,其评价指标和评价标准是确定的,而评价因子及其浓度变化过程又具有不确定性^[12-13],应从确定性和不确定性两方面对水质进行综合评价。

1.1 投影寻踪函数

投影寻踪函数^[14-16]是通过将高维数据投影到低维空间,以此来揭示高维数据中的隐含规律和结构。水质综合评价就是要将多维指标综合为一个评价结果,通过寻找出使投影指标函数达到最优的投影值,在此基础上构建反映水质类别与投影值之间的映射关系,清晰、有效地表达水质综合评价结果。

选取 p 个水质评价指标,首先进行标准化处理以消除量纲的影响,得到水质指标值矩阵 $\mathbf{X}$ 。经验灌溉水质等级 $\mathbf{y} = (y_1, y_2 \cdots, y_n)^T$,其中 n 为水质样本个数。通过投影方向向量 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2 \cdots, \alpha_p)$ 将 p 维水质指标数据 x_i 变换为一维投影值,即 $z = \mathbf{x}_i \boldsymbol{\alpha}^T$,则指标投影函数为

$$Q(\boldsymbol{\alpha}) = S(z)R(z\mathbf{y}) \tag{1}$$

其中

$$S(z) = \sqrt{\frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n}}$$

$$R(z\mathbf{y}) = \left| \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (z_i - \bar{z})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right|$$

式中: $Q(\boldsymbol{\alpha})$ 为投影函数; z_i 为根据各组样本数据计算的一维投影值; $\bar{z}$ 为一维投影值的平均值; y_i 为经验灌溉等级; $\bar{y}$ 为经验灌溉等级均值。

在综合投影时应使 $S(z)$ 越大越好,以尽可能提取监测数据中的变异信息,同时要求 z 与 $\mathbf{y}$ 的相关

系数的绝对值 $R(z\mathbf{y})$ 尽量大,即寻求最佳投影向量 $\boldsymbol{\alpha}^*$ 使投影函数 $Q(\boldsymbol{\alpha})$ 达到最大。根据获得的最佳投影向量 $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^* \cdots, \alpha_p^*)$,可得到第 i 个水样的最优投影值 $z_i^* = \sum_{j=1}^p \alpha_j^* x_{ij}$,其中 $i = 1, 2, \cdots, n$ 。根据 $z^* \sim \mathbf{y}$ 即可建立相应的数学模型,采用逻辑斯谛曲线^[2]作为水质评价模型:

$$y_i^* = \frac{N}{1 + e^{c_1 - c_2 z_i^*}} \tag{2}$$

式中: y_i^* 为第 i 个样本水质评价等级的计算值; N 为最大水质等级; c_1 、 c_2 为待定参数。该函数可变换为求 y_i^* 与 $\mathbf{y}$ 最佳拟合的 c_1 、 c_2 值。

由以上分析可知,投影寻踪函数的优化求解实质上是一个非线性优化问题。本文采用免疫遗传算法,对 p 维非线性优化问题进行优化求解,算法流程见图1。

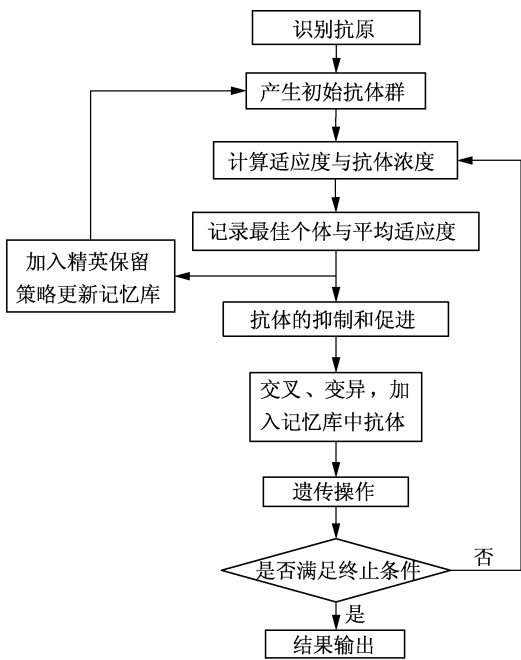


图1 免疫遗传算法流程

考虑抗体浓度表示解群体的多样性,在此对浓度计算方法进行改进,通过抗体与抗原亲和度的计算,综合亲和度和浓度评价抗体优秀程度,得出繁殖概率,记录当代最佳个体和种群平均适应度,同时更新记忆库,加入精英保留策略,保留优秀解及其相关参数,最终经过交叉变异产生新种群。此种方法可以有效防止最优解的丢失,同时加快收敛速度并防止算法陷入局部最优。

1.2 云模型

云模型是一种定性知识描述和定性概念与其定量数值表示之间的不确定性转换模型,能够充分体现语言概念的随机性和模糊性^[17]。每一个正态云模型均可由数字特征值 E_x 、 E_n 、 H_e 唯一确定,其中

$E_x = \frac{1}{n} \sum_{i=1}^n x_i$, 表示样本期望,是最能够代表定性概念的点; $E_n = \sqrt{\frac{\pi}{2} \frac{1}{n} \sum_{i=1}^n |x_i - E_x|}$, 表示熵,是定性概念不确定性的度量; $H_e = \sqrt{S^2 - E_n^2}$, 为超熵,表示云模型的凝聚度,反映评价的不确定性,其中 S^2 为样本方差。

在确定了评价指标及其标准的基础上,利用以下方法确定云模型的 3 个数字特征:

$$\begin{cases} E_x = \frac{B_{\max} + B_{\min}}{2} \\ E_n = \frac{B_{\max} - B_{\min}}{6} \\ H_e = k \end{cases} \quad (3)$$

式中: $B_{\max}$ 、 $B_{\min}$ 分别为变量的最大与最小边界;超熵 H_e 可根据变量的模糊阙度进行调整,通常根据经验取为某一经验值 k 。

由于各评价指标对水质等级分类评价的贡献度不同,因此权重也不相同。利用投影寻踪法确定评价指标权重,设评价指标权重为 $\omega = (\omega_1, \omega_2, \dots, \omega_p)$, 将获得的最佳投影向量 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_p^*)$ 代入式(4),则权重 ω_j 为

$$\begin{cases} \omega_j = \frac{\alpha_j^*}{\sum_{j=1}^p \alpha_j^*} & 0 \leq \omega_j \leq 1 \\ \sum_{j=1}^p \omega_j = 1 \end{cases} \quad (4)$$

基于云模型理论确定的云参数 E_x 、 E_n 、 H_e , 依据云发生器的算法,计算某一指标数据 $X_j(j = 1, 2, \dots, p)$ 的等级隶属度,结合各评价指标的权重,得到综合确定度 U :

$$U = \sum_{j=1}^p (\mu_j \omega_j) \quad (5)$$

式中: μ_j 为确定度。根据综合确定度值,每组监测数据的水质评价等级判别式为

$$C = \max(U_1, U_2, \dots, U_p) \quad (6)$$

式中, C 为最大隶属度。

1.3 等级区分粒度计算

利用免疫遗传算法改进的投影寻踪函数得到水质等级的计算值,比较计算等级与经验等级的差值,可以判断计算结果的准确性。云模型通过云发生器计算每个水质样本对应于每个等级的隶属度来确定最终的等级归属,通过计算各隶属度差值来判断分类是否准确。两种模型结果相互支撑,但是在判断综合模型的评价效果时,为了简单有效,综合投影寻踪函数的模型计算值 y_i^* 与云模型的综合确定度 U ,

引入等级区分粒度的概念,计算公式为

$$\delta_i = \frac{1}{2} \left| \frac{U_{i\max} - U_{i\sec}}{U_{i\max}} + e^{-|y_i^* - y_i|} \right| \quad (9)$$

式中: δ_i 为等级区分粒度值; $U_{i\sec}$ 为次大隶属度。

通过分析可知,当 $|y_i^* - y_i| \geq 0.5$ 时,投影寻踪函数评价结果错误,因此 $y_i^* - y_i$ 允许的最大差值为 0.5,此时该综合评价模型处于临界状态。当 $U_{i\sec}$ 无限趋近 $U_{i\max}$ 、 $|y_i^* - y_i| = 0.5$ 时, $\lim \delta_i = 0.3033$,此时评价模型没有正确区分出水质等级,评价结果失效。当 $U_{i\sec} = 0$ 、 $U_{i\max} = 1$, y_i^* 与 y_i 无限接近时, $\lim \delta_i = 1$,此时评价模型水质等级区分明显,评价结果最佳。因此当 $\delta_i \in (0.3033, 1)$, δ_i 越趋近于 1,表明该评价结果与实际情况越接近,对应的评价模型区分度越大,效果越好。

2 实例分析

选取太湖流域监测的 20 个水质样本,盐度 (U_1)、氯化物 (U_2)、氨氮 (U_3)、溶解性固体 (U_4) 4 个监测指标,其中盐度包含硫酸盐、硝酸盐,可以反映水体的盐碱化程度;氯离子主要与钙离子、镁离子结合,反映了水体的矿化度;氨氮反映了灌溉过程中肥料使用情况;溶解性固体可以直观反映水体污染程度。参考国家农田灌溉水质标准,并结合太湖流域的水域背景值并进行适当调整,将水质等级分为 I 级(好水)、II 级(中等水)、III 级(差水)、IV 级(劣质水)(表 1),利用投影寻踪-云模型进行评价,结果见表 2。

表 1 水质评价指标

评价等级	ρ (盐度)/ ($\text{mg} \cdot \text{L}^{-1}$)	ρ (氯化物)/ ($\text{mg} \cdot \text{L}^{-1}$)	ρ (氨氮)/ ($\text{mg} \cdot \text{L}^{-1}$)	ρ (溶解固体)/ ($\text{g} \cdot \text{L}^{-1}$)
I 级	[0,10)	[0,3.5)	[0,0.02)	[0,1.5)
II 级	[10,20)	[3.5,8)	[0.02,0.08)	[1.5,3)
III 级	[20,30)	[8,12)	[0.08,0.15)	[3,4.5)
IV 级	[30,∞)	[12,∞)	[0.15,∞)	[45,∞)

图 2 为模型计算等级与经验等级的拟合图。从图 2 可以看出,评价结果基本集中在经验等级值左右。个别样本的计算等级与经验等级相差较大,例如第十组样本,其等级区分粒度值比较小,说明此种情况下的评价结果不是很理想,但仍大于 0.3033。这种评价过程中的不确定性没有改变最终评价结果,对水质等级评价影响较小。20 个水质样本的投影寻踪函数计算等级平均绝对误差仅为 0.1252 级,达到了较好的水质评价精确度。

表 2 太湖流域农业灌溉水质等级评价结果

水样	U_1	U_2	U_3	U_4	z^*	y	y^*	δ_i
1	0.8153	0.2518	0.1742	0.0324	-1.4536	1	1.0475	0.822396
2	0.7040	0.1907	0.1528	0.0407	-1.5020	1	1.0087	0.860249
3	0.6668	0.2616	0.1270	0.0309	-1.8944	1	0.7302	0.685616
4	0.8498	0.2582	0.1116	0.0309	-1.5017	1	1.0089	0.843653
5	0.7065	0.2387	0.0989	0.0309	-2.0120	1	0.6594	0.686736
6	0.0967	0.6398	0.2456	0.0036	-0.5510	2	1.9115	0.765720
7	0.1570	0.6787	0.2857	0.0311	-0.5830	2	1.8780	0.732099
8	0.0649	0.7708	0.0373	0.0911	-0.2301	2	2.2471	0.831435
9	0.1293	0.8291	0.1584	0.0314	-0.4654	2	2.0013	0.903845
10	0.1059	0.6603	0.2429	0.0044	-0.1224	2	2.3576	0.665754
11	0.0661	0.0688	0.7712	0.2417	0.5639	3	2.9875	0.837108
12	0.1137	0.0251	0.6652	0.2385	0.4164	3	2.8661	0.758051
13	0.0749	0.0572	0.7575	0.1073	0.3316	3	2.7923	0.835413
14	0.1044	0.0299	0.5441	0.3493	0.5873	3	3.0060	0.676007
15	0.0824	0.0454	0.6094	0.2858	0.6012	3	3.0169	0.757133
16	0.0836	0.0441	0.0233	0.5568	2.5625	4	3.8404	0.851168
17	0.0837	0.0438	0.0290	0.4414	2.1024	4	3.7475	0.793624
18	0.1070	0.0274	0.0264	0.7221	2.3690	4	3.8062	0.837817
19	0.1278	0.0210	0.0233	0.6567	7.2310	4	3.9988	0.902078
20	0.1201	0.0230	0.0296	0.3491	10.8230	4	4.0000	0.827972

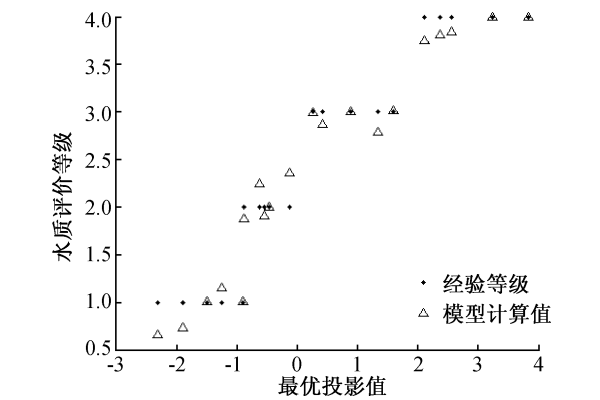


图 2 模型计算等级与经验等级拟合散点图

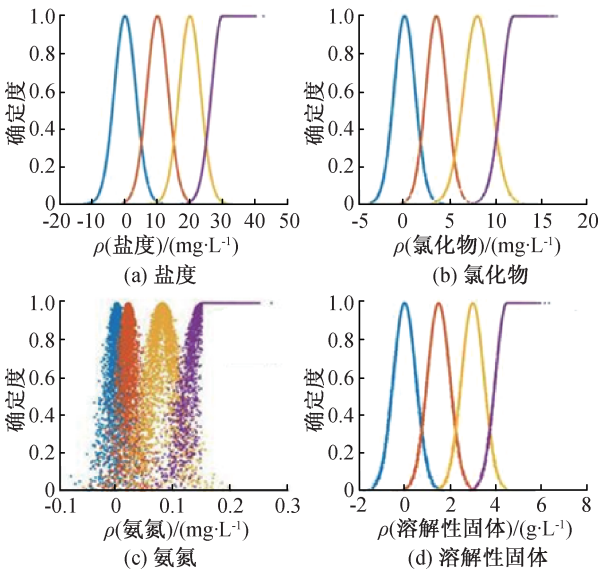


图 3 各指标的云模型

利用云发生器分别对评价指标盐度、氯化物、氨氮、溶解性固体 4 个指标生成云模型(图 3),每个子图中从左到右均是 I ~ IV 级对应的云,各监测指标的最大隶属度所在的级别即为该水质样本的评价级别。利用云模型计算各个监测指标得到的最大综合确定度所属级别与经验等级一致。

3 结 语

结合投影寻踪函数和云模型的优势,提出了一种定量、定性融合的双模型综合评价法。选取太湖流域 20 个样本盐度、氯化物、氨氮、溶解性固体 4 类具有代表性的农业灌溉水质监测数据,利用寻踪投影函数,计算投影向量,反映各个指标的水质类别与投影值之间的非线性关系;同时利用改进的免疫遗传优化算法求解投影寻踪模型,实现云模型的客观权重分配,后续可以准确获取监测数据波动情况下的水质评价等级。此外,通过引入等级区分粒度概念,不仅可以反映模型评价结果等级间的区分程度,还可以防止细微噪声存在可能带来的等级误判,从而提升评价模型准确性。

参考文献:

[1] 潘峰,付强,梁川. 模糊综合评价在水环境质量综合评价中的应用研究[J]. 环境工程,2002,20(2):58-60. (PAN Feng, FU Qiang, LIANG Chuan. Applying fuzzy synthesize judgement in the study of water environment

- quality evaluation [J]. Environmental Engineering, 2002, 20(2):58-60. (in Chinese))
- [2] MEI Y, YE J, ZENG Z. Entropy-weighted ANP fuzzy comprehensive evaluation of interim product production schemes in one-of-a-kind production [J]. Computers & Industrial Engineering, 2016(100):144-152.
- [3] 黄显峰,贾永乐,方国华. 基于投影寻踪法的城市水生态文明建评价[J]. 水资源保护, 2016, 32(6):117-122. (HUANG Xianfeng, JIA Yongle, FANG Guohua. Evaluation of urban aquatic ecological civilization construction based on projection pursuit method[J]. Water Resources Protection, 2016, 32(6):117-122. (in Chinese))
- [4] 黄健元,金广宇,于彦博. 投影寻踪模型在行蓄洪区运用风险评价中的应用[J]. 水利经济, 2016, 34(5):60-63. (HUANG Jianyuan, JIN Guangyu, YU Yanbo. Application of projection pursuit model to risk assessment of flood detention zones[J]. Journal of Economic of Water Resource, 2016, 34(5):60-63. (in Chinese))
- [5] 刘玲花,吴雷祥,吴佳鹏,等. 国外地表水水质指数评价法综述[J]. 水资源保护, 2016, 32(1):86-90. (LIU Linghua, WU Leixiang, WU Jiapeng, et al. Review on surface water quality index assessment method in foreign countries[J]. Water Resources Protection, 2016, 32(1):86-90. (in Chinese))
- [6] MA X, SHANG X, WANG L, et al. Innovative approach for the development of a water quality identification index; a case study from the Wenruitang River watershed, China [J]. Desalination and Water Treatment, 2014, 55(5):1-11.
- [7] 余勋,梁婕,曾光明,等. 基于三角模糊数的贝叶斯水质评价模型[J]. 环境科学学报, 2013, 33(3):904-909. (YU Xun, LIANG Jie, ZENG Guangming et al. An integrated fuzzy-Bayesian water quality assessment model based on triangular fuzzy numbers [J]. Acta Scientiae Circumstantiae, 2013, 33(3):904-909. (in Chinese))
- [8] 巩奕成,张永祥,丁飞,等. 基于萤火虫算法的投影寻踪地下水水质评价方法[J]. 中国矿业大学学报, 2015, 44(3):566-572. (GONG Yicheng, ZHANG Yongxiang, DING Fei, et al. Projection pursuit model for assessment of groundwater quality based on firefly algorithm[J]. Journal of China University of Mining & Technology, 2015, 44(3):566-572. (in Chinese))
- [9] 梁中耀,张雨宇,钱松,等. 基于二项分布检验法的水质达标评价方法研究[J]. 环境科学学报, 2017, 37(1):339-346. (LIANG Zhongyao, ZHANG Yuyu, QIAN Song, et al. A Statistical method with water quality compliance assessment based on binomial test [J]. Acta Scientiae Circumstantiae, 2017, 37(1):339-346. (in Chinese))
- [10] 魏博文,黄海鹏,徐镇凯. 基于云模型和组合赋权的岩体质量二维评价模型[J]. 岩石力学与工程学报, 2016, 35(增刊1):3092-3099. (WEI Bowen, HUANG Haipeng, XU Zhenkai. Two-dimensional evaluation model of rock mass based on combination weighting and cloud model [J]. Chinese Journal of Rock Mechanics and Engineering, 2016, 35(sup1):3092-3099. (in Chinese))
- [11] 李如琦,苏浩益,凌武能,等. 基于云模型和灰色关联投影法的烟气脱硫技术优选[J]. 中国电机工程学报, 2012, 32(23):21-26. (LI Ruqi, SU Haoyi, LING Wuneng, et al. Optimization of flue gas desulphurization technologies based on cloud model and grey relation projection method [J]. Proceedings of the Chinese Society for Electrical Engineering, 2012, 32(23):21-26. (in Chinese))
- [12] 范丽丽,邱利,田威,等. 基于马尔科夫模型的太湖水质综合评价[J]. 水资源保护, 2015, 31(2):50-54. (FAN Lili, QIU Li, TIAN Wei, et al. Comprehensive evaluation of water quality in Taihu Lake based on Markov model[J]. Water Resources Protection, 2015, 31(2):50-54. (in Chinese))
- [13] 龙珂良,陈星,崔广柏,等. 基于马尔科夫模型下的水质评价[J]. 水资源保护, 2016, 32(2):133-138. (LONG Keliang, CHEN Xing, CUI Guangbo, LIN Song, WANG Yiyi, FAN Yunzhu, et al. Water quality evaluation based on Markov model [J]. Water Resources Protection, 2016, 32(2):133-138. (in Chinese))
- [14] 许准,郭晓亮,徐昕,等. 生物群体智能优化的投影寻踪模型在灌区水资源综合效益研究中的应用[J]. 水资源保护, 2016, 32(3):38-43. (XU Zhun, GUO Xiaoliang, XU Xin, et al. Application of projection pursuit classification model basing on biological swarm intelligence in comprehensive benefit research of water resources in irrigation district [J]. Water Resources Protection, 2016, 32(3):38-43. (in Chinese))
- [15] 薛联青,朱祎真,王思琪,等. 基于投影寻踪的湖泊生态健康评价[J]. 水资源保护, 2016, 32(2):67-72. (XUE Lianqing, ZHU Yizhen, WANG Siqi, et al. Lake ecosystem health assessment based on projection pursuit [J]. Water Resources Protection, 2016, 32(2):67-72. (in Chinese))
- [16] 崔东文. 正弦余弦算法-投影寻踪水污染物总量分配模型[J]. 水资源保护, 2016, 32(6):75-81. (CUI Dongwen. Sine cosine algorithm-projection pursuit model for water pollutant load allocation [J]. Water Resources Protection, 2016, 32(6):75-81. (in Chinese))
- [17] 张雪洁,王志坚,张伟建. 云计算中服务质量预测数据的信心建模[J]. 河海大学学报(自然科学版), 2015, 43(6):588-593. (ZHANG Xuejie, WANG Zhijian, ZHANG Weijiang. Confidence model of QoS prediction data in cloud computing [J]. Journal of Hohai University (Natural Sciences), 2015, 43(5):588-593. (in Chinese))

(收稿日期:2017-01-06 编辑:王 芳)