

文章编号: 2095-2163(2020)12-0086-05

中图分类号: F205;O213

文献标志码: A

# 基于集对分析—RBF 神经网络的生态文明建设评价指标体系构建

耿 鹏, 郑中团

(上海工程技术大学 数理与统计学院, 上海 201620)

**摘 要:** 本文以长三角 26 个城市的生态文明建设为研究对象, 采用集对分析与 RBF 神经网络相结合的方法, 构建包括人口自然增长率等 30 个指标的长三角城市群生态文明建设评价指标体系。首先, 通过相关性分析, 删除同一系统层中相关系数过大的指标, 避免了指标间的冗余; 其次, 通过集对分析—RBF 神经网络, 删除了对网络输出值影响较小的变量, 保证了筛选出的指标对生态文明建设指数有显著影响。

**关键词:** 生态文明建设; 指标筛选; 集对分析; RBF 神经网络

## Establishment of evaluation index system of ecological civilization construction based on Set Pair Analysis-RBF Neural Network

GENG Peng, ZHENG Zhongtuan

(School of Mathematics, Physics and Statistics, Shanghai University of Engineering and Science, Shanghai 201620, China)

**[Abstract]** Taking the ecological civilization construction of 26 cities in the Yangtze River Delta as the research object, the method of combining pair-pair analysis and RBF neural network is used to construct an evaluation index system of ecological civilization construction in the Yangtze River Delta urban agglomeration including 30 indicators like natural population growth rate. The main points lie in, firstly to delete indicators with excessive correlation coefficients in the unified criterion layer through correlation analysis to avoid redundancy among the indicators. Secondly to delete the variables that have little effect on the network output value through the set pair analysis - RBF neural network, which ensures that the screened indicators have significant influence on the ecological civilization construction index.

**[Key words]** Ecological civilization construction; Index selection; Set pair analysis; RBF neural network

### 0 引 言

生态文明建设是长三角一体化发展战略的重要组成部分。长三角区域多年来快速的经济社会发展, 给该地区资源环境承载造成了相当大的压力, 为实现更高质量的长三角一体化提出了严峻的挑战。无论是 2018 年 6 月长三角三省一市共同签署的《长三角地区一体化发展三年行动计划(2018-2020)》, 还是 2019 年 7 月印发的《长江三角洲区域一体化发展规划纲要》以及三省一市落实规划纲要的实施方案, 均强调要推动“长三角和谐共生绿色发展”, 推进生态文明建设一体化。江浙沪方面明确表示“重点建设长三角生态绿色一体化发展的示范区”, “加强生态环境的共保联治, 共筑绿色美丽长三角”。为了更好地开展长三角城市群生态文明一体化建设, 必须对生态文明建设进行客观合理的评价, 故有必要构建长三角城市群生态文明建设评价指标体系, 对长三角城市群生态文明建设进行综合测度和评价。

国外并没有给出具体的生态文明概念, 相关研究主要集中在可持续发展领域<sup>[1]</sup>, 提出了人类发展指数(HDI)、生态足迹理论、真实储蓄法、能值分析法等一系列构建指标的方法来评价可持续发展水平。在国内, 一是采用单一方法构建指标体系。项贻等通过文献频度统计法从经济、环境、人居、文化和制度 5 方面构建全国生态文明指标体系, 同时采用差异驱动法对江苏和广州各市生态文明建设成效进行综合评价<sup>[2]</sup>; 张钰莹等在总结了已有的评价指标体系后, 基于灰色粗糙集筛选指标体系, 并利用层次分析法与模糊综合评价法对中国生态文明建设水平进行实证研究<sup>[3]</sup>; 二是采用多种方法构建指标体系。钱敏蕾等运用压力-状态-响应(PSR)模型, 结合频度统计与专家咨询等方法, 选取适用于特大型城市的生态文明建设评价指标, 并利用熵权法对上海进行实证研究<sup>[4]</sup>; 侯娜等基于灰色动态聚类与粗糙集相结合的方法, 以内蒙古自治区 12 个盟市的绿

**基金项目:** 全国统计科学研究项目(2018LY16)。

**作者简介:** 耿 鹏(1995-), 女, 硕士研究生, 主要研究方向: 可持续发展的数据挖掘与统计分析; 郑中团(1979-), 男, 博士, 副教授, 主要研究方向: 可持续发展的统计分析。

**通讯作者:** 郑中团 Email: zhongtuanzheng@163.com

**收稿日期:** 2020-09-10

色资源环境的高频指标为筛选对象,构建内蒙古自治区绿色资源环境指标体系<sup>[5]</sup>。

国内外对生态文明建设评价指标体系及评价方法做了广泛深入的研究,这些研究对于推进生态文明建设均具有重要意义。然而,在对生态文明建设指标体系进行构建时,大部分指标体系的构建具有较强的主观性,地区差异性考虑不够,缺乏数据思维<sup>[6]</sup>。本文将针对长三角城市群生态文明建设评价问题,基于“社会—经济—自然”复合生态系统和数据驱动,考虑长三角城市群的资源环境禀赋、土地利用状况、经济发展规模等实际情况,参考国内外生态文明建设指标体系的研究,通过集对分析—RBF 神经网络方法筛选指标,更为科学、客观地构建一套合乎长三角城市群实际情况的生态文明建设评价指标体系。

## 1 指标筛选原理

### 1.1 指标海选思路

在生态文明的理论基础上,将国外权威机构、国家环境保护部、国家统计局等部门出台的环境评价指标体系及相关文献中的高频指标作为海选指标来源,结合长三角城市群生态文明的生态环境特点,将社会、经济与自然作为长三角城市群生态文明建设指标体系的准则层。直接删除不能直接获取或二次计算得到数据的海选指标,对于可获取大部分样本容量观测值的海选指标做多重插补补全。

### 1.2 指标筛选思路

(1) 独立性检验。采用 Spearman 相关性检验对各准则层中的各项指标进行分析,构建全指标要素的相关矩阵,设定指标临界值为 0.9,剔除部分指标,从而得到初选指标体系。

(2) 集对分析。计算初选指标体系的同一度矩阵,经熵权法确定各指标权重后,即可确定各指标的带权同一度矩阵。集对分析中,指标评价集与理想方案的相对贴合度越高,其评价结果则越优。

(3) RBF 神经网络。各初选指标的标准化数据为 RBF 神经网络的输入值矩阵,集对分析中的带权同一度矩阵作为输出值矩阵,依据平均影响值 (MIV) 的累计贡献率大于 85% 进行指标筛选,从而得到二次筛选后的指标体系,即为生态文明建设评价指标体系。

生态文明建设评价指标体系构建原理,如图 1 所示。

## 2 数据来源与指标筛选方法

### 2.1 数据来源与正向化处理

本文依据 2016 年 5 月国务院批准的《长江三角

洲城市群发展规划》,选取长三角城市群 26 个城市作为研究样本实证研究。

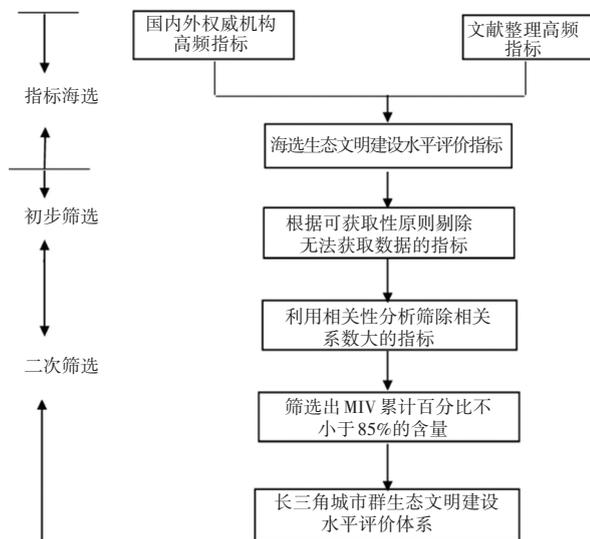


图 1 生态文明建设评价指标体系构建原理

Fig. 1 Principles of constructing evaluation index system for ecological civilization construction

指标数据来自 2018 年长三角城市群 26 个城市的《统计年鉴》、《国民经济与社会发展统计公报》、《水资源公报》、《生态环境质量公报》以及各市的统计局、生态环境局和国家统计局提供的相关资料。

对生态文明建设起明显积极促进作用的指标为正向指标,其值越大越好;相反对生态文明建设起消极阻碍作用的相关指标为逆向指标,其值越小越好,为消除指标量纲单位的差异,采用极差法对各项指标无量纲化处理,处理后的数据区间为 $[0, 1]$ 。

对于正向指标,计算公式(1)为:

$$x'_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \quad (1)$$

对于逆向指标,计算公式(2)为:

$$x'_{ij} = \frac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}} \quad (2)$$

式中,  $x_{ij}$  为指标体系中各指标数据的原始值,  $x'_{ij}$  为极差化后的指标数值。

### 2.2 指标筛选方法

#### 2.2.1 相关性分析

相关性分析是研究变量间关系的一种统计方法,就指标选择而言,同一系统层内的指标相关性应显著,不同系统层的指标相关性应不显著。由于本文中的指标数据不服从正态分布,故采用 Spearman 相关性分析方法进行变量之间的相关性分析。设  $r_{ij}$  表示第  $i$  个指

标与第  $j$  个指标的相关系数,  $x_{ki}$  为第  $k$  个评价对象第  $i$  个指标值,  $\bar{x}_i$  为第  $i$  个指标的平均值, 计算公式(3)如下:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 (x_{kj} - \bar{x}_j)^2}} \quad (3)$$

### 2.2.2 集对分析

集对分析是一种不确定分析方法。本文采用了刘凤朝学者的处理方法, 将不确定性与确定性结合成一个系统, 凭借系统中确定性与不确定性的相互联系及某些条件转换用以评估不确定性<sup>[7]</sup>。集对分析评价模型构建的大致步骤如下:

构造同一度矩阵: 设系统中  $m_1, m_2, \dots, m_n$  共  $n$  个待评选的方案, 每个方案有  $m$  个评价指标  $C = \{C_1, C_2, \dots, C_m\}$ , 每个评价指标均有一个值表示, 记为  $d_{ij}(i = 1, 2, \dots, n; j = 1, 2, \dots, m)$ , 则集对分析的同一度的评价矩阵  $H$  可表示为式(4):

$$H = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{m1} & \cdots & d_{mn} \end{pmatrix} \quad (4)$$

理想方案  $M_0 = (d_{01}, d_{02}, \dots, d_{0j}, \dots, d_{0n})^T$ 。其中,  $d_{0j}$  为  $M_0$  方案第  $j$  个指标的数值, 其大小为  $H$  矩阵中  $j$  个指标的最大值, 对矩阵的指标值  $d_{ij}$  和理想方案  $M_0$  中对应的指标值  $d_{0j}$  进行评价, 将形成被评价对象与理想方案指标无权重的同一度矩阵  $Q$ , 式(5):

$$Q = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad (5)$$

其中,  $a_{ij}$  为被评价指标值  $d_{ij}$  与理想方案  $M_0$  对应指标  $d_{0j}$  的同一度, 式(6):

$$a_{ij} = \frac{d_{ij}}{d_{0j}} \quad (6)$$

指标权重的确定方式大致可分为主观赋权和客观赋权二类, 综合考虑生态文明的特点及避免主观意识产生的偏差, 本文选用熵权法确定各指标的权重。

构造评价模型: 应用之前确定好的权重向量  $W$  与同一度矩阵  $Q$ , 即可确定各评价对象  $M_i$  与理想方案  $M_0$  的加权同一度矩阵  $R$ , 式(7):

$$R = W \times Q = (W_1, W_2, \dots, W_n) \times \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} X = (a_1, a_2, \dots, a_n) \quad (7)$$

矩阵  $R$  中的元素  $a_j(j = 1, 2, \dots, n)$  称为第  $j$  个评价对象与理想方案的同一度。

### 2.2.3 RBF 神经网络

RBF 神经网络对非线性函数具有一致逼近性, 在处理生态文明建设的非线性数据时具有一定的优势, 是典型的 3 层前向网络, 分为输入层、隐藏层和输出层<sup>[8]</sup>。其运行原理为: 隐藏层通过径向基函数(RBF)执行一种非线性变化, 将输入空间映射到另一新的空间, 输出层则在新的空间中线性加权。高斯函数常用作隐藏层的基函数, 其基本形式为式(8):

$$R(x_p - c_i) = \exp\left(-\frac{1}{2\sigma^2} \|x_p - c_i\|^2\right) \quad (8)$$

式中,  $\|x_p - c_i\|$  为欧式范数;  $c_i$  为高斯函数的中心;  $\sigma$  为高斯函数的方差。

输出层第  $j$  个节点的输出为式(9):

$$y_j = \sum_{i=1}^h w_{ij} \exp\left(-\frac{1}{2\sigma^2} \|x_p - c_i\|^2\right), \quad (9)$$

$$j = 1, 2, \dots, n.$$

式中,  $x_p = (x_1^p, x_2^p, \dots, x_m^p)^T$  为第  $p$  个样本输入;  $p = 1, 2, 3, \dots, P$ ,  $P$  为样本总数;  $c_i$  为隐藏层结点的中心;  $w_{ij}$  为隐藏层到输出层的连接权值,  $i = 1, 2, 3, \dots, h$  为隐藏层节点数。

若  $d$  是样本的期望输出值, 则基函数的方差为式(10):

$$\sigma = \frac{1}{P} \sum_j^m \|d_j - y_j c_i\|^2 \quad (10)$$

### 2.2.4 MIV 算法

平均影响值(MIV)反映神经网络中权重矩阵的变动状况<sup>[9]</sup>。具体计算过程: 在对原样本量  $P$  网络训练结束时, 将  $P$  中每一自变量特征相对其原值上下变动 10%, 形成新的输入样本  $P_1$  和  $P_2$ , 再利用已建成的网络进行仿真, 得到仿真结果  $A_1$  和  $A_2$ , 计算  $A_1$  和  $A_2$  的差值, 根据输入神经元数量取平均, 该值就为平均影响值 MIV, 其符号代表相关方向, 绝对值的大小代表重要性的程度, 根据 MIV 的绝对值, 得到各变量对网络输出的相对重要性的顺序表, 从而判断输入变量对网络输出值的影响程度, 最终实现变量的选择。

## 3 实证分析

### 3.1 指标构建基础

以国内外的生态文明建设高频指标为依据, 结合可观测性原则, 建立了生态文明建设海选指标体系, 见表 1。

表 1 生态文明建设评价的海选指标集

Tab. 1 Audition index set of ecological civilization construction evaluation

系统层	指标层	筛选结果	
社会	人口自然增长率	保留	
	人口密度	保留	
	城镇人口登记失业率	定量删除	
	城市恩格尔系数	定量删除	
	农村恩格尔系数	保留	
	城镇化率	保留	
	自来水普及率	定量删除	
	气化率	定量删除	
	每万人在校大学生数	保留	
	每万人在校专任教师数	保留	
	R&D 占 GDP 比例	定量删除	
	每万人医生数	定量删除	
	每万人医院床位数	保留	
	农村居民人均居住面积	保留	
	城镇居民人均居住面积	保留	
	每万人公共交通数	定量删除	
	每万人出租车数	保留	
	职工平均工资	保留	
	人均道路面积	保留	
	居民平均寿命	不可观测	
	互联网普及率	不可观测	
	城市居民文化娱乐消费支出占消费总支出的比重	不可观测	
	城市人均维护建设资金支出	不可观测	
	城镇基本医疗保险参保比例	不可观测	
	教育支出占财政支出比例	不可观测	
	人均地区生产总值	保留	
	人均储蓄存款	保留	
	固定资产投资额年增长率	保留	
	财政收入年增长率	保留	
	进出口总额增长率	定量删除	
	经济	第二产业比重	保留
		第三产业比重	保留
		工业固体废物综合利用率	定量删除
		城镇居民人均可支配收入	保留
		城镇居民人均消费支出	相关性分析删除
		农村居民人均可支配收入	定量删除
		农村居民人均消费支出	相关性分析删除
		单位工业生产总产值能耗	保留
		单位生产总值能耗	定量删除
		工业成本费用利用率	不可观测
高新技术产业增加值占 GDP 比重		不可观测	
人均社会消费品零售总额		不可观测	
工业用水重复利用率		不可观测	
自然		人均能源消费量	保留
	人均电力消费	定量删除	
	人均水资源量	定量删除	
	人均工业烟尘排放总量	定量删除	
	人均工业固体废物产生量	定量删除	
	人均工业废气排放总量	保留	
	人均工业废水排放总量	保留	
	二氧化硫浓度	保留	
	二氧化氮浓度	保留	
	可吸入颗粒物浓度	保留	
	空气质量优良率	定量删除	
	森林覆盖率	保留	
	建成区绿化覆盖率	定量删除	
	人均公园绿地面积	保留	
	城市水质达标率	定量删除	
	河流断面 I-III 比例	保留	
	生活垃圾无害化转化率	定量删除	
	污水处理率	保留	
	湿地覆盖率	不可观测	
	工业粉尘去除率	不可观测	
	工业废水排放达标率	不可观测	
	工业 SO <sub>2</sub> 去除率	不可观测	
	耕地保有量	不可观测	
	环境污染事件	不可观测	
	环境污染治理投入占 GDP 比重	不可观测	
	水土流失强度	不可观测	
	单位种植面积化肥使用量	不可观测	
	每千人拥有环保系统人员数	不可观测	

## 3.2 基于相关性分析的指标筛选

将标准化数据依据系统层代入公式(3),计算得到相关系数,设定临界值为 0.9,结果见表 2。

表 2 相关性系数大于 0.9 的指标筛选结果

Tab. 2 Screening results of indicators with a correlation coefficient greater than 0.9

系统层	保留的指标	删除的指标	相关系数
经济	城镇居民人均可支配收入	城镇居民人均消费支出	0.949
	农村居民人均可支配收入	农村居民人均消费支出	0.962

在经济系统层中,城镇居民人均可支配收入与城镇居民人均消费支出的相关性系数大于 0.9,故筛选除城镇居民人均消费支出。其它指标的筛选与保留见表 2。

## 3.3 基于集对分析—RBF 神经网络的指标筛选

以经济系统层为例,由熵值法计算出指标权重  $W = (0.129, 0.131, 0.073, 0.046, 0.041, 0.071, 0.107, 0.116, 0.133, 0.072, 0.028, 0.053)$ 。根据公式(7)计算出带权同一度矩阵  $R = (0.749, 0.618, 0.694, 0.661, 0.713, 0.508, 0.358, 0.473, 0.534, 0.521, 0.737, 0.674, 0.677, 0.642, 0.684, 0.618, 0.620, 0.609, 0.413, 0.410, 0.420, 0.377, 0.273, 0.259, 0.322, 0.302)$ 。将各系统层的带权同一度矩阵作为期望输出值矩阵输入 RBF 神经网络中,结合 MIV 算法,对各系统层中的指标进行筛选。表 3~5 分别为社会、经济、自然系统层中的输入变量的 MIV 绝对值降序排列。社会系统层中选取排名前 12 的变量,经济系统层中选取排名前 8 的变量,自然系统层中选取排名前 10 的变量。这些变量的累计 MIV 绝对值百分比均不低于 85%。

表 3 社会系统层输入变量 MIV 绝对值排序

Tab. 3 The absolute value sorting of the input variable MIV at the social system layer

排序	指标	MIV 绝对值	MIV 累计百分比
1	人口自然增长率	0.005 2	14.65
2	农村居民人均居住面积	0.004 7	27.89
3	人口密度	0.004 4	40.28
4	每万人在校专任教师数	0.003 4	49.86
5	每万人医院床位数	0.003 2	58.87
6	城镇化率	0.002 4	65.63
7	农村恩格尔系数	0.001 5	69.86
8	人均道路面积	0.001 5	74.08
9	职工平均工资	0.001 4	78.03
10	每万人在校大学生数	0.001 3	81.69
11	城镇居民人均居住面积	0.001	84.51
12	每万人出租车数	0.000 9	87.04
13	每万人医生数	0.000 9	89.58
14	R&D 占 GDP 比例	0.000 9	92.11
15	城镇人口登记失业率	0.000 7	94.08
16	每万人公共交通数	0.000 6	95.77
17	城市恩格尔系数	0.000 6	97.46
18	自来水普及率	0.000 5	98.87
19	气化率	0.000 4	100.00

表4 经济系统层输入变量 MIV 绝对值排序

Tab. 4 The absolute value sorting of the input variable MIV at the economic system layer

排序	指标	MIV 绝对值	MIV 累计百分比
1	固定资产投资额年增长率	0.009 4	17.35
2	第二产业占比	0.009 3	34.51
3	人均储蓄存款	0.007 0	47.43
4	第三产业占比	0.005 8	58.14
5	单位工业生产总值电耗	0.004 9	67.18
6	城镇居民人均可支配收入	0.004 6	75.67
7	财政收入年增长率	0.004 3	83.61
8	人均地区生产总值	0.003 7	90.44
9	进出口总额增长率	0.002 2	94.50
10	农村居民人均可支配收入	0.001 6	97.45
11	单位生产总值能耗	0.001 3	99.85
12	工业固体废物综合利用率	0.000 08	100.00

表5 自然系统层输入变量 MIV 绝对值排序

Tab. 5 The absolute value sorting of the input variable MIV at the nature system layer

排序	指标	MIV 绝对值	MIV 累计百分比
1	二氧化硫浓度	0.006 1	16.44
2	人均工业废水排放总量	0.004 3	28.03
3	人均公园绿地面积	0.004 2	39.35
4	河流断面 I - III 比例	0.003 4	48.52
5	人均能源消费量	0.003 3	57.41
6	森林覆盖率	0.002 9	65.23
7	人均工业废气排放总量	0.002 4	71.70
8	二氧化氮浓度	0.002 0	77.09
9	可吸入颗粒物浓度	0.001 9	82.21
10	污水处理率	0.001 8	87.06
11	人均工业固体废物产生量	0.001 4	90.84
12	人均工业烟尘排放总量	0.001 0	93.53
13	建成区绿化覆盖率	0.000 7	95.42
14	人均电力消费	0.000 6	97.04
15	人均水资源量	0.000 6	98.65
16	空气质量优良率	0.000 3	99.46
17	生活垃圾无害化转化率	0.000 2	100.00
18	城市水质达标率	0	100.00

经初步筛选,二次筛选后,从71个指标中筛选出30个指标构成生态文明建设评价指标体系。

#### 4 结束语

本文结合长三角城市群资源环境禀赋等实际情况,已有的生态文明建设评价体系以及复合生态系统模型,设计形成初选指标体系,并遴选面板数据集。运用相关性分析法与集对分析—RBF神经网络方法对指标优化筛选,构建了包含人口自然增长率、固定资产投资额年增长率、人均工业废水排放总量等30个指标在内的长三角城市群生态文明建设评价差异化指标体系。其要点具体包括:第一,通过相关性分析删除同一系统层中相关系数过大的指标,避免了指标间的冗余;第二,通过集对分析—RBF神经网络删除了对网络输出值影响较小的变量,保证了筛选出的指标对生态文明建设指数有显著影响。

#### 参考文献

- [1] PEARCE D, BARBIER E, MARKANDYA A. Sustainable development: economics and environment in the third world [M]. London: Routledge, 2013.
- [2] 项贇,刘晓文,张剑鸣,等. 我国生态文明建设成效评估指标体系的研究[J]. 生态经济, 2015, 31(8): 14-19.
- [3] 张钰莹,罗洋. 生态文明建设的多层次模糊综合评价[J]. 四川建筑科学研究, 2017, 43(1): 149-154.
- [4] 钱敏蕾,李响,徐艺扬,等. 特大型城市生态文明建设评价指标体系构建—以上海市为例[J]. 复旦学报(自然科学版), 2015, 54(4): 389-397.
- [5] 侯娜,李战江,苏金梅,等. 基于灰色动态聚类—粗糙集的绿色资源环境评价指标构建模型[J]. 统计与信息论坛, 2017, 32(8): 96-103.
- [6] 王汉生. 数据思维:从数据分析到商业价值[M]. 北京:中国人民大学出版社, 2017.
- [7] 刘凤朝,潘雄锋,施定国. 基于集对分析法的区域自主创新能力评价研究[J]. 中国软科学, 2005(11): 83-91.
- [8] 赵海华. 基于灰色 RBF 神经网络的多因素财政收入预测模型[J]. 统计与决策, 2016(13): 79-81.
- [9] 郭中小,宋一凡,廖梓龙,等. 基于 MIV 的遗传神经网络径流预报模型[J]. 人民黄河, 2014, 36(10): 33-35.