

熵权和属性识别模型在水利现代化评价中的应用

牟萍艾萍

(河海大学水文水资源学院, 江苏 南京 210098)

摘要 :为科学便捷地评价不同地区的水利现代化综合发展水平,提出应用属性识别模型和熵值法确定样本属性值和不同指标之间权重的方法,并用该评价方法对全国及 4 个典型省份的水利现代化综合发展水平进行评价。评价结果与实际大致相符,也与模糊聚类迭代等方法的评价结果基本一致,验证了该方法的合理性。

关键词 水利现代化;评价指标;属性识别;熵权

中图分类号:F407.9

文献标识码:A

文章编号:1003-9511(2011)05-0001-04

水利现代化是指由传统水利向现代水利转变的动态、渐进的发展过程。目前对于水利现代化的评价主要是采用评价指标体系,傅春等^[1]提出了建设水利现代化的评价指标体系,尹豪等^[2]运用统计学和计量经济学方法,采用三级指标体系,设计了具有检验功能的水利现代化评价模型,孟祥礼等^[3]基于层次分析法构建的水利现代化评价指标体系,提出了运用目标距离法原理来确定指标权重的新方法,欧建峰等^[4]采用主成分分析法对江苏省 13 个省辖市的水利现代化水平进行了综合评价与排序,张海涛等^[5]基于水利现代化评价指标体系建立模糊聚类迭代模型,并采用基于重要性互补的二元一致性方法确定权重。利用指标体系进行评价时,关键问题是样本的属性值及指标之间权重的确定。程乾生^[6]提出的属性识别模型在解决有序分割问题上具有显著的优势,目前已广泛应用于水资源评价、水质评价、空气质量评价等^[7-9]多个领域。笔者拟将属性识别模型和熵值法应用于水利现代化评价。

科学设置评价水利现代化水平的指标体系,是

客观反映水利现代化水平的重要依据^[1],同时对水利现代化的发展具有导向作用。我国已有不少学者对水利现代化水平的评价指标体系进行了研究。张海涛等^[5]构建了水利现代化评价指标体系,共有 13 项指标,见表 1。笔者拟在此评价指标体系的基础上,应用属性识别模型^[6]对我国水利现代化综合发展水平进行评价。

1 水利现代化综合发展水平的评价

1.1 指标体系

用属性识别模型进行评价时,需要建立水利现代化综合发展水平分级标准,文献^[5]在以往水利现代化研究成果的基础上,选取了代表目前国外发达国家水利先进水平的参考值,见表 1。对于越大越优的指标,笔者分别选取参考值、0.85 倍的参考值、0.7 倍的参考值和 0.6 倍的参考值来划分 I 级、II 级、III 级、IV 级,对于越小越优的指标,分别选取参考值、1.2 倍的参考值、1.3 倍的参考值和 1.5 倍的参考值来划分 I 级、II 级、III 级、IV 级,见表 1。

表 1 水利现代化评价指标参考值及分级标准

	万元 GDP 用水量/ (m ³ ·万元 ⁻¹)	城乡生活供水普及率/ %	防洪能力指数/ %	水旱灾害损失率/ %	水功能区水质达标率/ %	水土流失治理率/ %	平原区地下水超采率/ %	水利信息化指数/ %	城市节水器具普及率/ %	灌溉水利用系数	大专以上管理人员比例/ %	法规制度完善率/ %	水利管理机构建设完善率/ %
参考值	48.80	100	90.00	0.20	95.00	95.00	1.00	95.00	98.00	0.80	98.00	98.00	98.00
I	48.80	100	90.00	0.20	95.00	95.00	1.00	95.00	98.00	0.80	98.00	98.00	98.00
II	58.56	85.00	76.50	0.24	80.75	80.75	1.20	80.75	83.30	0.68	83.30	83.30	83.30
III	63.44	70.00	63.00	0.26	66.50	66.50	1.30	66.50	68.60	0.56	68.60	68.60	68.60
IV	73.20	60.00	54.00	0.30	57.00	57.00	1.50	57.00	58.80	0.48	58.80	58.80	58.80

基金项目 国家自然科学基金重大研究计划培育项目(9092402)

作者简介 牟萍(1986—),女,山东德州人,硕士研究生,主要从事水文水资源信息理论与技术研究。

1.2 基于熵权的属性识别模型

1.2.1 属性空间矩阵的建立

设 X 为研究对象的全体 称为对象空间。为了评价不同区域的水利现代化综合发展水平 在 X 中取 n 个样本 X_1, X_2, \dots, X_n 对每个样本采用 m 个指标 V_1, V_2, \dots, V_m 来进行评价。第 i 个样本 X_i 的第 j 个评价指标 V_j 的测量值为 x_{ij} 因此 第 i 个样本 X_i 可以表示为一个向量 $\bar{X}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ $1 \leq i \leq n$ 。

属性识别模型在进行评价时 需要对各项评价指标进行分级 假设在进行水利现代化评价时评价等级分为 C_K 级 并且 $C_1 > C_2 > \dots > C_K$ 或 $C_1 < C_2 < \dots < C_K$ 则每项评价指标 V_j 与每个分级 C_K 构成了分级标准判断矩阵 $\{a_{jk}\}_{m \times K}$ 且满足 $a_{j1} > a_{j2} > \dots > a_{jK}$ 或 $a_{j1} < a_{j2} < \dots < a_{jK}$ ($j = 1, 2, \dots, m$)。

1.2.2 样本的测度及属性识别

根据属性识别模型^[6]的基本原理 计算样本的属性测度 步骤如下:

a. 计算第 i 个样本的第 j 个指标的测量值 x_{ij} 具有属性 C_k 的属性测度 $\mu_{ijk} = \mu(x_{ij} \in C_k)$ 。不妨假设 $a_{j1} < a_{j2} < \dots < a_{jK}$

当 $x_{ij} \leq a_{j1}$ 时 取 $\mu_{ij1} = 1, \mu_{ij2} = \dots = \mu_{ijK} = 0$;
当 $x_{ij} \geq a_{jK}$ 时 取 $\mu_{ijK} = 1, \mu_{ij1} = \dots = \mu_{ijK-1} = 0$;

(1)

当 $a_{jl} \leq x_{ij} \leq a_{j(l+1)}$ 时 取 $\mu_{ijl} = \frac{|x_{ij} - a_{j(l+1)}|}{|a_{jl} - a_{j(l+1)}|}, \mu_{ij(l+1)} = \frac{|x_{ij} - a_{jl}|}{|a_{jl} - a_{j(l+1)}|}, \mu_{ijk} = 0, k < l$ 或 $k > l + 1$ 。

b. 样本的属性识别。根据置信度准则^[6] 当置信度为 λ (λ 的取值范围通常为 $0.5 < \lambda < 1$) 时 分别对 n 个样本的 m 个指标进行属性识别。当 $C_1 > C_2 > \dots > C_K$ 时,

$$h_0 = \min \left\{ h : \sum_{k=1}^h \mu_{ijk} \geq \lambda, 1 \leq h \leq K \right\} \quad (2)$$

当 $C_1 < C_2 < \dots < C_K$ 时,

$$h_0 = \max \left\{ h : \sum_{k=h}^K \mu_{ijk} \geq \lambda, 1 \leq h \leq K \right\} \quad (3)$$

则认为 μ_{ijk} 属于 C_{h_0} 类。

c. 计算样本指标的属性测度综合值。

$$\mu_{ij} = \mu(x_{ij} \in C_k) =$$

$$\sum_{k=1}^K \gamma_k \mu_{ijk}, 1 \leq i \leq n, 1 \leq j \leq m. \quad (4)$$

在水利现代化综合发展水平评价中 评价标准的分级是按照参照值乘以不同的系数来划分的。假设参照值服从正态分布 与参照值相距越远的等级 其可信程度越低 因此 取 $\gamma_k = (K + 1) - k, k = 1, 2, \dots, K$ 。

水利现代化综合发展水平评价中的指标共有 m 个 每个指标的重要性可能相同 也可能不相同 因此 要考虑指标的权重 $(\omega_1, \omega_2, \dots, \omega_m), \omega_j \geq 0$,

$$\sum_{j=1}^m \omega_j = 1.$$

1.2.3 熵值法确定指标权重

确定指标权重的方法有专家赋值、AHP 方法、Delphi 方法等 这些方法的主观性比较强 易受个人知识水平及偏好的影响 因此评价结果容易形成偏差。信息熵反映信息的无序化程度 信息熵越小 系统无序度越小 信息的效用值越大 信息熵越大 系统无序度越大 其信息的效用值越小。对于水利现代化评价中的 m 个指标 可以用信息熵来度量其信息效用值。因此 利用熵值法计算各指标的权重 信息的效用值越高 其对评价的重要性越大 权重值越大^[7]。同时熵值法从实测数据入手 充分利用实测数据的自身特点来计算权重 因而尽可能减少了人为主观因素的影响。

a. 构建判断矩阵。构建 n 个样本的 m 个评价指标的判断矩阵:

$$R_{ij} = (x_{ij})_{nm}$$

$$(i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (5)$$

b. 将判断矩阵归一化(标准化)处理。为了消除指标间量纲的不同所带来的不可公度性 需要对指标进行归一化处理 得到归一化后的判断矩阵:

$$B_{ij} = (b_{ij})_{nm}$$

$$(i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$

$$\begin{cases} b_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} & (\text{递增型指标}) \\ b_{ij} = \frac{x_{\max} - x_{ij}}{x_{\max} - x_{\min}} & (\text{递减型指标}) \end{cases} \quad (6)$$

式中 x_{\max}, x_{\min} 分别为同一指标下不同样本中的最大值和最小值。

c. 定义熵。C. E. Shannon 熵的数学描述为

$$H = -k \left(\sum_{i=1}^n p_i \ln p_i \right), p_i \text{ 为各事件的概率。由 C. E. Shannon 熵的定义 定义评价指标的熵:}$$

$$H_j = -k \left(\sum_{i=1}^n f_{ij} \ln f_{ij} \right) \quad (j = 1, 2, \dots, m) \quad (7)$$

式中 $k = \frac{1}{\ln n}$; f_{ij} 为一种概率 在水利现代化综合发

展水平评价中 以频率代替概率 定义 $f_{ij} = \frac{b_{ij}}{\sum_{i=1}^n b_{ij}}$ 。

为使 $\ln f_{ij}$ 有意义 一般需要假定当 $f_{ij} = 0$ 时 $f_{ij} \ln f_{ij} = 0$ 。但当 $f_{ij} = 1$ 时 $f_{ij} \ln f_{ij} = 0$ 这显然不切合实际 与熵的含义相悖 故需对 f_{ij} 加以修正 将其定义为^[7-8]

$$f_{ij} = \frac{1 + b_{ij}}{\sum_{i=1}^n (1 + b_{ij})} \quad (8)$$

d. 计算评价指标的熵权 ω 。由于信息熵越大,系统无序度越大,其信息的效用值越小,权重值越小,同时结合实际数据指标和数据自身的特点,定义熵权^[7-8]:

$$\omega_j = \frac{1 - H_j}{\sum_{i=1}^m (1 - H_j)} \quad (9)$$

$$\sum_{j=1}^m \omega_j = 1$$

1.2.4 综合评价

通过上述方法,计算出样本指标的属性测度综合值和各指标的权重系数,就可以进行评价样本的比较分析了。

评分准则^[6]:若 $q_{x_i} > q_{x_j}$,则认为 q_{x_i} 比 q_{x_j} 强,记为 $x_i > x_j$ 。按照评分准则,计算

$$q_{x_i} = \sum_{j=1}^m \omega_j \mu_{ij} \quad (10)$$

则可根据 q_{x_i} 的大小对 x_i 进行比较和排序。

2 应用实例

2.1 样本的选取

结合水资源条件和水利建设状况,在东、中、西和东北四大区域各选择 1 个有代表性的省份和全国平均水平为样本^[5]。为了与文献^[5]的评价结果比较,选取与其一致的数据,如表 2 所示。

2.2 评价计算过程

第 1 步 根据式(1)计算全国及 4 省的 13 个指标的测量值在 I ~ IV 分级标准下的属性测度。

第 2 步 取 $\lambda = 0.7$,根据式(2)或式(3)对全国及 4 省的 13 个指标进行属性识别。结果如表 3 所示。

第 3 步 根据式(4)计算各样本指标的属性测度综合值,见表 4。

第 4 步 将表 2 中的数据按式(5)(6)进行归一化处理。

第 5 步 根据式(7)(8)(9)计算各评价指标的熵和熵权,结果见表 5。

表 2 2007 年全国和典型省份水利现代化评价指标数值

评价区域	万元 GDP 用水量/ ($m^3 \cdot \text{万元}^{-1}$)	城乡生活供水普及率/%	防洪能力指数/%	水旱灾害损失率/%	水功能区水质达标率/%	水土流失治理率/%	平原区地下水超采率/%	水利信息化指数/%	城市节水器具普及率/%	灌溉水利用系数	大专以上学历管理人员比例/%	法规制度完善率/%	水利管理机构建设完善率/%
全国	229.00	65.34	41.00	0.46	41.60	43.00	5.90	70.00	30.00	0.48	36.20	76.00	61.20
浙江省	106.00	85.00	95.00	0.86	62.00	63.00	18.03	75.00	68.00	0.54	60.00	78.00	70.00
辽宁省	127.00	71.05	60.00	0.04	45.00	35.57	0.51	70.00	75.00	0.50	37.00	75.00	80.00
湖北省	287.00	47.72	41.00	0.71	69.90	80.15	0.33	65.00	30.00	0.46	21.78	70.00	80.00
甘肃省	442.00	41.06	34.00	0.56	50.00	58.00	11.50	60.00	40.00	0.48	18.00	65.00	65.00

表 3 评价等级

评价区域	万元 GDP 用水量	城乡生活供水普及率	防洪能力指数	水旱灾害损失率	水功能区水质达标率	水土流失治理率	平原区地下水超采率	水利信息化指数	城市节水器具普及率	灌溉水利用系数	大专以上学历管理人员比例	法规制度完善率	水利管理机构建设完善率
全国	IV	IV	IV	IV	IV	IV	IV	III	IV	IV	IV	III	IV
浙江省	IV	II	I	IV	IV	IV	IV	III	III	III	IV	III	III
辽宁省	IV	III	IV	IV	IV	IV	IV	III	III	IV	IV	III	II
湖北省	IV	IV	IV	I	III	II	I	III	IV	IV	IV	III	II
甘肃省	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV

表 4 样本指标的属性测度综合值

评价区域	万元 GDP 用水量	城乡生活供水普及率	防洪能力指数	水旱灾害损失率	水功能区水质达标率	水土流失治理率	平原区地下水超采率	水利信息化指数	城市节水器具普及率	灌溉水利用系数	大专以上学历管理人员比例	法规制度完善率	水利管理机构建设完善率
全国	1.00	1.53	1.00	1.00	1.00	1.00	1.00	2.25	1.00	1.00	1.00	2.50	1.24
浙江省	1.00	3.00	4.00	1.00	1.53	1.63	1.00	2.60	1.94	1.75	1.12	2.64	2.10
辽宁省	1.00	2.07	1.67	4.00	1.00	1.00	1.00	2.25	2.44	1.25	1.00	2.44	2.78
湖北省	1.00	1.00	1.00	1.00	2.24	2.96	4.00	1.84	1.00	1.00	1.00	2.10	2.78
甘肃省	1.00	1.00	1.00	1.00	1.00	1.11	1.00	1.32	1.00	1.00	1.00	1.63	1.63

表 5 各评价指标的熵和熵权

评价指标	万元 GDP 用水量	城乡生活供水普及率	防洪能力指数	水旱灾害损失率	水功能区水质达标率	水土流失治理率	平原区地下水超采率	水利信息化指数	城市节水器具普及率	灌溉水利用系数	大专以上学历管理人员比例	法规制度完善率	水利管理机构建设完善率
熵	0.9833	0.9811	0.9784	0.9825	0.9788	0.9820	0.9815	0.9842	0.9722	0.9825	0.9808	0.9831	0.9777
熵权	0.0663	0.0750	0.0856	0.0693	0.0842	0.0715	0.0736	0.0629	0.1103	0.0694	0.0761	0.0672	0.0886

第 6 步,计算各地的综合得分。根据式(10)计算评价样本的综合得分,应用评分准则,对样本进行排序,如表 6 所示。

表 6 得分及排序结果

评价区域	分数	排序
全 国	1.2412	4
浙江省	1.9688	1
辽宁省	1.8529	2
湖北省	1.7489	3
甘肃省	1.1260	5

2.3 评价结果分析

由表 6 的评价结果可以看出,在 4 个评价省份中浙江省的水利现代化综合发展水平最高,辽宁省的水利现代化综合发展水平排名第 2,湖北省排名第 3,甘肃省的水利现代化综合发展水平较低。这与 4 个省的社会经济发展状况一致,且与文献[5]中的评价结果基本一致。

3 结 语

采用指标体系法对我国的水利现代化综合发展水平进行评价,并利用属性识别模型来识别样本的属性测度,同时应用熵值法确定指标之间的权重。熵值法是根据信息论中利用信息熵来度量信息效用的原理来确定权重的,同时充分利用实测数据自身的特点,能够尽可能地减少人为主观因素的干扰。在文献[5]研究的指标体系基础上,利用该方法对 2007 年全国及部分省份的水利现代化综合发展评价指标数值进行了评价,其评价结果与模糊聚类迭代模型^[5]的评价结果基本一致。

参考文献:

[1] 傅春, 杨志峰, 刘昌明. 水利现代化的内涵及评价指标体系的建立[J]. 水科学进展, 2002, 13(4): 502-506.
 [2] 尹豪, 章仁俊. 水利现代化评价模型及其应用[J]. 农业现代化研究, 2005, 26(5): 393-396.
 [3] 孟祥礼, 张玉福. 水利现代化评价指标体系的赋权方法研究[J]. 沈阳农业大学学报, 2009, 40(2): 248-250.

[4] 欧建锋, 叶健, 程吉林. 主成分分析法在江苏水利现代化评价中的应用[J]. 人民长江, 2010, 41(2): 97-100.
 [5] 张海涛, 谢新民, 杨丽丽. 水利现代化评价指标体系与评价方法研究[J]. 中国水利水电科学研究院学报, 2010, 8(2): 107-113.
 [6] 程乾生. 属性识别理论模型及应用[J]. 北京大学学报: 自然科学版, 1997, 33(1): 12-20.
 [7] 张先起, 梁川, 刘慧卿. 基于熵权的属性识别模型在地下水水质综合评价中的应用[J]. 四川大学学报: 工程科学版, 2005, 37(3): 28-31.
 [8] 刘慧卿, 张先起. 空气质量综合评价的基于熵权的属性识别模型[J]. 环境科学与技术, 2008(7): 114-143.
 [9] 王丽琼. 基于信息熵的水资源可持续利用综合评价分析[J]. 泉州师范学院学报: 自然科学版, 2007, 25(4): 73-77.
 (收稿日期 2011-06-01 编辑 陈玉国)

《水利水电科技进展》征订启事

(邮发代号: 28-244, 双月刊, A4 开本, ISSN1006-7647, CN32-1439/TV)

《水利水电科技进展》由河海大学主办,是全国中文核心期刊,中国科技核心期刊,全国水利系统优秀期刊,华东地区优秀期刊,江苏省优秀期刊。主要刊登水科学、水工程、水资源、水环境、水管理方面的科技论文,主要栏目有水问题论坛、研究探讨、工程技术、水管理、专题综述、国外动态等,适合与水利、水电、水科学、水工程、水资源、水环境有关的科研、工程、管理人员以及大专院校师生阅读。

本刊由邮局发行,邮发代号:28-244,2012 年每期定价 12 元,全年 6 期共计 72 元。可在全国各地邮局订阅,也可直接向编辑部订阅。

编辑部地址 210098 南京市西康路 1 号

电话/传真 025-83786335

E-mail: jz@hhu.edu.cn

http://kbb.hhu.edu.cn/web/indexjz.asp?id=5