DOI:10.3969/j.issn.1003-9511.2013.01.009

# 基于熵权的集对分析在地下水质量评价中的应用

鲁婷婷,俞 淞,王红瑞,丁建新

(北京师范大学水科学研究院,北京 100875)

**摘要**:为解决地下水水质评价中评价因子与评价标准之间模糊、非线性的复杂关系,将集对分析理论应用于地下水水质评价中。为避免评价过程中个人赋予评价指标权重的主观性,引用信息论中的熵权理论,建立基于熵权的集对分析模型。并将其评价结果与模糊综合评价结果比较,评价结果表明,基于熵权的集对分析方法,不仅原理易懂、计算简便,而且能充分利用所有数据,减少误差,反映水体综合质量,为地下水质量评价提供了一种合理可行的方法。

**关键词**:集对分析;地下水;熵权;质量评价

**中图分类号**:X824 **文献标识码**:A **文章编号**:1003-9511(2013)01-0034-04

地下水是水资源的重要组成部分,对经济和社会的可持续发展具有不可替代的作用,特别是在地表水缺少或污染严重的地区,地下水常常被用作人们的饮用水源。地下水污染趋势日益严重,正面临着由点到面、由浅到深、由城市到农村不断扩展的趋势[1]。2011 年,中国工程院与中国环保部联合发布的中国环境宏观战略研究报告指出,中国地下水污染严重,其中 118 个大中城市地下水已普遍遭到污染,严重污染的城市占 64%,平原区约有 54% 的地下水不符合生活饮用水水质标准。

近几年,随着科学技术的不断发展,地下水质量综合评价方法也越来越多,主要包括单因子评价法[2]、灰色聚类法[3]、神经网络法[4]和支持向量机[5]等,这些方法都取得了良好的效果。在 1994 年实施的 GB/T 14848—93《地下水质量标准》中明确给出地下水质量分类指标,但由于影响地下水质量的因素众多,评价指标与评价因子之间仍存在模糊、非线性的复杂关系,使得各种方法在应用中有其优点和弊端。例如单因子评价法,评价方法简单、明了,能清晰的判断出主要污染因子及其分布区域,但不能全面反应水质的整体状况。模糊综合评价法能充分考虑边界的模糊性,但是它的计算比较复杂,在评价面积较大、评价对象数量较多的情况下,应用比较困难,需要借助计算机辅助编程完成[6]。支持向量机虽然评价精度高,能简化评价过程,但在相关参数选择方面有待完善[7]。为解决水质评价中评价指标与评价因子之间模糊、非线性的复杂关系,将赵克勤[8]提出的集对分析理论应用于地下水水质评价中。

## 1 基于集对分析的地下水质量评价模型的建立

### 1.1 集对分析基本原理

集对分析是赵克勤于 1989 年在包头召开的全国系统理论会上提出的一种新型的处理模糊和不确定知识的系统分析方法,目前已广泛应用于地表水环境质量评价中[9],其核心思想是对不确定性系统中有一定联系的 2 个集合构建集对,再对集对的特性做同一性、差异性、对立性分析,并加以度量刻画,建立集对的同异反联系度表达式。

设有联系的集合 $X$ 和 $Y$,$X$ 有 $n$ 项表征其特征,即 $X=(x_1,x_2,\cdots,x_n)$,$Y$ 亦有 $n$ 项表征其特征,即 $Y=(y_1,y_2,\cdots,y_n)$。$X$ 和 $Y$ 构成集对 $H(X,Y)$。描述 $H(X,Y)$ 间关系的联系度定义为

$$\mu_{X\sim Y}=\frac{S}{n}+\frac{FI}{n}+\frac{PJ}{n} \tag{1}$$

式中:$\mu_{X\sim Y}$为集合 $X$ 和 $Y$ 的联系度,$S$ 为集对 $H(X,Y)$ 中特性同一的个数;$F$ 为特性差异的个数;$P$ 为特性对立的个数;$S+F+P=n$。$I$ 为差异度系数,取值 $[-1,1]$。当 $I$ 取 1 时,差异性转化为同一性;当 $I$ 取 $-1$ 时,差异性转化为对立性;当 $I$ 在 $(-1,1)$ 取值时,

**基金项目**:国家科技支撑计划(2012BAB02B04)

**作者简介**:鲁婷婷(1989—),女,山东威海人,硕士研究生,从事水文水资源研究。

反映了同一性与对立性所占的比例，$I$ 越趋于0，不确定性越明显。$J$ 为对立度系数，规定其恒取值为-1。

记 $a=\frac{S}{n}, b=\frac{F}{n}, c=\frac{P}{n}$，则式(1)可写成

$$\mu_{X\sim Y}=a+bI+cJ \tag{2}$$

式中：$a,b,c$ 满足 $a+b+c=1$。$a,b,c$ 分别为集对 $H(X,Y)$ 的同一度、差异度和对立度。当 $a$ 越接近于1时，表示这2个集合的特性越倾向于同一；当 $c$ 越接近于1时，表示这2个集合的特性越倾向于对立；当 $b$ 越接近于1时，表示这2个集合的特性越倾向于差异(既不同一也不对立)。

在地下水质量评价中，地下水质量分类及分类指标是确定的，实际采用的评价因子是不确定的，基于集对分析的原理，将实际采用的评价因子与地下水评价标准之间构成一个集对，通过计算2个集合的联系度，评判地下水的环境质量。

### 1.2 联系度的确定

联系度是对不确定系统中有一定联系的2个集合关系的量化，能清晰定量的显示2个集合之间的复杂关系。在确定联系度之前，首先要进行单因子评价，将样本的实测值与地下水质量标准进行对比，分别得出每个指标所属的水质级别，然后再计算单个指标所处的水质级别与其他水质级别的联系度。确定准确的联系度是进行集对分析的关键，具体确定方法是：将单因子评价的结果与5类水质标准进行对比，若单因子评价结果与该评价级别相同，联系度为1；当单因子评价结果与该评价级别相隔，联系度为-1；当单因子评价结果与该评价级别相邻，根据评价指标与评价标准距离远近确定联系度。具体量化公式为

**a.** 样本 $i$ 第 $j$ 项评价指标与Ⅰ级水质联系度

$$u_{ij1}=\begin{cases}1 & x_{ij}\in[0,s_{1j}]\\ 1-\dfrac{2(x_{ij}-s_{1j})}{s_{2j}-s_{1j}} & x_{ij}\in(s_{1j},s_{2j}]\\ -1 & x_{ij}\in(s_{2j},+\infty)\end{cases} \tag{3}$$

式中：$x_{ij}$ 为样本 $i$ 第 $j$ 项评价指标实测值；$s_{1j},s_{2j}$ 为各评价标准中分类指标的限值。

**b.** 样本 $i$ 第 $j$ 项评价指标与Ⅱ级水质联系度

$$u_{ij2}=\begin{cases}1+\dfrac{2(x_{ij}-s_{1j})}{s_{1j}} & x_{ij}\in[0,s_{1j}]\\ 1 & x_{ij}\in(s_{1j},s_{2j}]\\ 1-\dfrac{2(x_{ij}-s_{2j})}{s_{3j}-s_{2j}} & x_{ij}\in(s_{2j},s_{3j}]\\ -1 & x_{ij}\in(s_{3j},+\infty]\end{cases} \tag{4}$$

式中：$s_{3j}$ 为评价标准中分类指标的限值。

**c.** 样本 $i$ 第 $j$ 项评价指标与Ⅲ级水质联系度

$$u_{ij3}=\begin{cases}-1 & x_{ij}\in[0,s_{1j}]\\ 1+\dfrac{2(x_{ij}-s_{2j})}{s_{2j}-s_{1j}} & x_{ij}\in(s_{1j},s_{2j}]\\ 1 & x_{ij}\in(s_{2j},s_{3j}]\\ 1-\dfrac{2(x_{ij}-s_{3j})}{s_{4j}-s_{3j}} & x_{ij}\in(s_{3j},s_{4j}]\\ -1 & x_{ij}\in(s_{4j},+\infty)\end{cases} \tag{5}$$

式中：$s_{4j}$ 为评价标准中分类指标的限值。

**d.** 样本 $i$ 第 $j$ 项评价指标与Ⅳ级水质联系度

$$u_{ij4}=\begin{cases}-1\ x_{ij} & \in[0,s_{2j}]\\ 1+\dfrac{2(x_{ij}-s_{3j})}{s_{3j}-s_{2j}}\ x_{ij} & \in(s_{2j},s_{3j}]\\ 1\ x_{ij} & \in(s_{3j},s_{4j}]\\ 1-\dfrac{2(x_{ij}-s_{4j})}{s_{5j}-s_{4j}}\ x_{ij} & \in(s_{4j},s_{5j}]\\ -1\ x_{ij} & \in(s_{5j},+\infty)\end{cases} \tag{6}$$

**e.** 样本 $i$ 第 $j$ 项评价指标与Ⅴ级水质联系度

$$u_{ij5}=\begin{cases}-1 & x_{ij}\in[0,s_{3j}]\\ 1-\dfrac{2(s_{4j}-x_{ij})}{s_{4j}-s_{3j}} & x_{ij}\in(s_{3j},s_{4j}]\\ 1 & x_{ij}\in(s_{4j},+\infty)\end{cases} \tag{7}$$

式中：$s_{5j}$ 为各评价标准中分类指标的限值，是一个不确定的值，为了便于计算样本单因子位于Ⅴ级时与Ⅳ级评价标准之间的联系度，将 $s_{5j}$ 赋值为 $2s_{4j}-s_{3j}$。

### 1.3 确定指标权重的熵权法

在信息论中，熵是对不确定性的一种度量，信息量越大，不确定性就越小，熵也就越小；信息量越小，不确定性越大，熵也就越大。根据熵的特性，可以根据评价指标值构成的矩阵判断各指标的权重，避免人为给予权重的主观性。主要的计算步骤如下：

**a.** 假定有 $m$ 个评价对象，每个评价对象有 $n$ 个评价指标，构成矩阵 $\boldsymbol{R}$

$$\boldsymbol{R}=(C_{ij})_{m\times n}$$
$$i=(1,2,\cdots,m),j=(1,2,\cdots,n) \tag{8}$$

式中：$C_{ij}$ 是第 $i$ 个评价对象的第 $j$ 个评价指标的实测值。

**b.** 将判断矩阵 $\boldsymbol{R}$ 归一化，得到归一化矩阵 $\boldsymbol{B}$，归一化公式：

$$b_{ij}=\begin{cases}\dfrac{C_{ij}-C_{\min}}{C_{\max}-C_{\min}} & \text{越大越优型}\\ \dfrac{C_{\max}-C_{ij}}{C_{\max}-C_{\min}} & \text{越小越优型}\end{cases} \tag{9}$$

式中：$b_{ij}$ 为矩阵 $\boldsymbol{B}$ 中的元素；$C_{\max},C_{\min}$ 分别为同指标下，不同方案中最满意者和最不满意者的值(本文评价为越小越满意)。

**c.** 根据传统的熵的概念可以定义各评价指标的熵为

$$H_j = -\frac{1}{\ln m}\sum_{i=1}^{m}(f_{ij}\ln f_{ij})$$
$$(i = 1,2,\cdots,m;j = 1,2,\cdots,n) \qquad (10)$$

式中 $f_{ij} = \dfrac{b_{ij}}{\sum_{i=1}^{m} b_{ij}}$

显然,当 $f_{ij}=0$ 时,$\ln f_{ij}$ 无意义,因此对 $f_{ij}$ 的计算加以修正,将其定义为 $f_{ij} = \dfrac{1 + b_{ij}}{\sum_{i=1}^{m}(1 + b_{ij})}$。

**d.** 计算各指标的熵权

$$W = (w_j)_{1\times n} \qquad (11)$$

其中 $w_j = \dfrac{1 - H_j}{n - \sum_{j=1}^{n} H_j}$,且满足 $\sum_{t=1}^{n} w_t = 1$

式中:$w_j$ 为第 $j$ 个评价指标的权重;$w_t$ 为第 $t$ 个评价指标的权重。

### 1.4 地下水环境质量等级的确定

根据式(3)~(7),计算样本与各评价等级之间的联系度,并加权处理:

$$u_{ik} = \sum_{j=1}^{n} w_j u_{ijk} \qquad (12)$$

式中:$u_{ik}$ 为样本 $i$ 对评价等级 $k$ 级之间的联系度;$u_{ijk}$ 为样本 $i$ 中评价指标 $j$ 对 $k$ 级评价标准的联系度。

依据集对分析基本原理,选取联系度最大的评价等级为该样本地下水环境质量等级,即

$$u_p = \max_{l=1}^{k}\{u_{il}\} \quad (p \text{ 在 } 1 \sim k \text{ 之间取值}) \qquad (13)$$

判定地下水水质级别为 $p$ 级。

## 2 实例应用

为了验证集对分析方法的合理性,本文采用山东省聊城市东昌府区 5 个办事处地下水监测数据[10],以铁、锰、总硬度、氯化物、高锰酸盐指数和硫酸盐为评价指标,以 GB/T14848—93《地下水水质标准》为水质分级原则,进行地下水水质评价。地下水环境质量标准见表 1。

**表 1 地下水环境质量标准** mg/L

| 评价指标 | Ⅰ类 ($s_1$) | Ⅱ类 ($s_2$) | Ⅲ类 ($s_3$) | Ⅳ类 ($s_4$) | Ⅴ类 | $s_5$ |
|---|---|---|---|---|---|---|
| 铁 | 0.10 | 0.20 | 0.30 | 1.5 | >1.50 | 1.7 |
| 锰 | 0.05 | 0.05 | 0.10 | 1.00 | >1.00 | 1.90 |
| 总硬度 | 150.00 | 300.00 | 450.00 | 550.00 | >550.00 | 650.00 |
| 氯化物 | 50.00 | 150.00 | 250.00 | 350.00 | >350.00 | 450.00 |
| 高锰酸盐指数 | 1.00 | 2.00 | 3.00 | 10.00 | >10.00 | 17.00 |
| 硫酸盐 | 50.00 | 150.00 | 250.00 | 350.00 | >350.00 | 450.00 |

现以东昌府区新区办事处为例,根据水质监测数据和地下水环境质量标准,确定单因子评价结果和联系度,见表 2 和表 3。

**表 2 新区办事处单因子评价结果**

| 评价指标 | 评价等级 | 所属区间 |
|---|---|---|
| 铁 | Ⅳ | $(s_3,s_4]$ |
| 锰 | Ⅲ | $(s_2,s_3]$ |
| 总硬度 | Ⅴ | $(s_4,s_5]$ |
| 氯化物 | Ⅱ | $(s_1,s_2]$ |
| 高锰酸盐指数 | Ⅰ | $(0,s_1]$ |
| 硫酸盐 | Ⅱ | $(s_1,s_2]$ |

**表 3 各评价因子与评价等级之间的联系度**

| 评价指标 | Ⅰ | Ⅱ | Ⅲ | Ⅳ | Ⅴ |
|---|---|---|---|---|---|
| 铁 | -1 | -1 | -0.85 | 1 | 0.85 |
| 锰 | -1 | -1 | 1 | 1 | -1 |
| 总硬度 | -1 | -1 | -1 | 0.42 | 1 |
| 氯化物 | -1 | 1 | 1 | -1 | -1 |
| 高锰酸盐指数 | 1 | 0.2 | -1 | -1 | -1 |
| 硫酸盐 | -0.16 | 1 | 0.16 | -1 | -1 |

根据熵权计算公式,得出各指标在地下水环境质量评价中的权重分别为(0.20,0.18,0.13,0.16,0.15,0.18)。由式(12)计算出新区办事处地下水水质与Ⅰ、Ⅱ、Ⅲ、Ⅳ、Ⅴ 5 类评价等级之间的综合联系度分别为(-0.55,-0.14,-0.08,-0.06,-0.37),由式(13)可判定该地区地下水水质属于Ⅳ类。同理可得北城办事处、柳园办事处、凤凰办事处、湖西办事处地下水水质与各评价等级之间的综合联系度分别为(-0.61,-0.11,-0.05,-0.60,-0.34)、(-0.47,-0.21,0.10,-0.06,-0.63)、(-0.34,-0.17,-0.03,0.05,-0.63)、(0.31,0.37,-0.31,-0.85,-1)。相应的地下水水质等级分别为:Ⅲ类,Ⅲ类,Ⅳ类,Ⅱ类。

## 3 评价结果的比较与分析

根据上述方法,将所得的结果与文献[11]中模糊综合评价结果进行比较,评价结果见表 4.

**表 4 集对分析方法与模糊综合评价法评价结果比较**

| 评价方法 | 水质等级 | | | | |
|---|---|---|---|---|---|
| | 北城办事处 | 新区办事处 | 柳园办事处 | 凤凰办事处 | 湖西办事处 |
| 集对分析方法 | Ⅲ | Ⅳ | Ⅲ | Ⅳ | Ⅱ |
| 模糊综合评价法 | Ⅴ | Ⅳ | Ⅲ | Ⅲ | Ⅱ |

从表 4 可以看出,新区办事处、柳园办事处和湖西办事处 2 种评价方法的评价结果相同,凤凰办事处水质评价结果相差一级,北城办事处评价结果相差较大。产生这种差异的原因可能有 2 个方面:第一方面是 2 种评价方法之间的差异。2 种方法都是通过建立函数进行质量评价,但模糊综合评价法是通过计算隶属度确定评价等级,隶属度在[0,1]取值,没有对水质等级之间的过渡性进行详细描述,而

集对分析方法是通过计算联系度确定评价等级,不仅将跨越区间扩大至[-1,1],而且从同一性、对立性、差异性3个方面进行分析,特别是将差异度引入函数中,描述水质等级之间的模糊过渡性,使评价结果更加精细、合理。另一方面的原因是评价过程中权重确定的方法不同。在模糊综合评价过程中,权重的求法是根据该地区指标的实测值与该指标标准限值的算术平均值的比值求得。在计算过程中,权重的大小只与该指标实测值有关,这就会出现因为某一个指标实测值偏高而导致评价结果偏高的现象,如北城办事处铁值明显高于其他地区,最后的结果也使模糊综合评价法的结果高于集对分析评价方法结果。而且模糊综合评价法只是根据一个单一的数值确定指标的权重,误差较大。本文在确定权重的过程中引进信息论中熵的概念,根据评价指标值构成的矩阵来判断各指标的权重,权重的大小与每个地区的实测值都存在关系,既减少了单一数值确定权重产生的误差,又能够反应各指标的集中趋势。

## 4 结 语

本文通过构建评价因子的实测值与水质评价标准之间的集对关系计算联系度来确定地下水的质量等级。在计算过程中,引用信息论中的熵值理论确定各个评价指标的权重,避免人为赋予权重的主观性,使评价结果更趋合理性。集对分析方法的主要优点在于不仅能够通过构建同、异、反联系度将评价过程量化,而且评价的原理易懂、计算简便、结果直观,适合大部分地区的地下水的质量评价。但是集对分析方法在评价过程中也有不成熟的地方,例如水质分级时 $s_5$ 的确定,本文虽然根据水质分级的一般规律将 $s_5$ 赋值为 $2s_{4j}-s_{3j}$,但水质分级的指标限值并不是标准线性关系,在确定过程中存在一定的主观性。这也需要对集对分析方法不断改进优化,使其能客观有效地对地下水质量进行评价。

**参考文献:**


[1] 罗兰. 我国地下水污染现状与防治对策研究[J]. 中国地质大学学报:社会科学版,2008,8(2):72-75.

[2] 胡浩云,王海峰,刘新霞. 邯郸市地下水水质现状分析[J]. 人民长江,2005,36(10):31-32.

[3] 张祖亮. 灰色聚类法在地下水质量评价中的应用[J]. 云南环境科学,2004,23(1):60-62.

[4] 罗定贵,王学军,郭青. 基于MATLAB实现ANN方法在地下水质评价中的应用[J]. 北京大学学报:自然科学版,2004,40(2):296-302.

[5] 徐劲力. 支持向量机在水质评价中的应用[J]. 中国农村水利水电,2007(3):7-9.

[6] 张新钰,辛宝东,刘文臣,等. 三种地下水水质评价方法的应用对比分析[J]. 城市地质,2011,6(1):30-36.

[7] CHAPELLE O, VAPNIK V. Choosing multiple parameters for support vector machines [J]. Machine Learning, 2002, 46(1):131-159.

[8] 赵克勤. 集对分析及其初步应用[J]. 大自然探索,1994,13(47):67-72.

[9] 高军省. 湖泊富营养化综合评价的五元联系数法[J]. 人民长江,2010,41(21):81-84.

[10] 彭小金,张艳红,李辉辉. 模糊综合评价在地下水质评价中的应用[J]. 河南水利与南水北调,2009(1):34-35,37.


(收稿日期:2012-09-04 编辑:方宇彤)

·简讯·

## 国务院办公厅印发《实行最严格水资源管理制度考核办法》

国务院办公厅2013年1月2日印发《实行最严格水资源管理制度考核办法》(以下简称《办法》)。这是落实《国务院关于实行最严格水资源管理制度的意见》的一项重要举措,对应对我国水资源严峻形势、解决水资源问题、实现水资源可持续利用具有重要意义。

《办法》明确,国务院将对各省、自治区、直辖市最严格水资源管理制度落实情况进行考核,水利部会同有关部门成立考核工作组,具体实施。考核结果将作为干部主管部门对各省、自治区、直辖市人民政府主要负责人和领导班子综合考核评价的重要依据。《办法》规定,各省、自治区、直辖市人民政府是实行最严格水资源管理制度的责任主体,政府主要负责人对本行政区域水资源管理和保护工作负总责。考核内容为最严格水资源管理制度目标完成、制度建设和措施落实情况。目标为各省、自治区、直辖市"三条红线"控制指标及阶段性管理目标。制度建设和措施落实情况包括用水总量控制制度、用水效率控制制度、水功能区限制纳污制度、水资源管理责任和考核制度建设及相关措施落实情况。

《办法》明确了考核方式,即考核评定采用评分法,考核结果划分为优秀、良好、合格、不合格4个等次。考核与国民经济和社会发展五年规划相对应,每5年为一个考核期,采用年度考核和期末考核相结合的方式。年度考核对各省、自治区、直辖市人民政府上年度目标完成、制度建设和措施落实情况进行考核,考核结果纳入期末考核。

《办法》要求,对在考核工作中瞒报、谎报的地区,予以通报批评,对有关责任人员依法依纪追究责任。水利部会同有关部门组织制定实行最严格水资源管理制度考核工作实施方案。各省、自治区、直辖市人民政府应根据本办法,结合当地实际,制定所辖区域内实行最严格水资源管理制度考核办法。 (本刊编辑部供稿)