

基于粗糙集和群决策特征根法的 区域水资源可再生性评价

杨 婷,王志良,谢敏萍,陈海涛
(华北水利水电学院,河南 郑州 450011)

【摘 要】 【目的】基于水资源可再生的特点,将粗糙集理论引入到水资源可再生性评价中,建立区域水资源可再生性评价的准确方法。【方法】针对水资源系统信息的不确定性和不一致性以及水资源中经常遇到的群决策问题,将粗糙集理论和群决策特征根法相结合应用于水资源评价,即将粗糙集作为数据的预处理工具,先构造决策表,再用粗糙集进行属性约简,对约简后的指标集用多属性群决策的定量评价方法进行了评价。【结果】实例计算表明,因粗糙集能较好地处理不确定、不完备、不一致信息,故能得到更为准确、详细的评价结果。【结论】将粗糙集和群决策特征根法相结合用于水资源可再生性评价是可行的。

【关键词】 区域水资源;可再生性评价;粗糙集;群决策特征根法(GEM);属性约简
【中图分类号】 TV213.9;P333.6 【文献标识码】 A 【文章编号】 1671-9387(2008)01-0218-05

Renewable evaluation of regional water resources based on rough set and group eigenvalue method

YANG Ting,WANG Zhi-liang,XIE Min-ping,CHEN Hai-tao
(North China Institute of Water Conservancy and Hydroelectric Power,Zhengzhou,Henan 450011,China)

Abstract: 【Objective】 Based on the renewable characteristic of water resources,rough set theory is introduced and the renewable evaluation accurate methods of regional water resources are established. 【Method】 According to the uncertainty and inconsistency of the information of water resource system and the instance that the problems of group decision-making are always met in this domain, rough set theory and group eigenvalue method are combined to use in the evaluation of water resources. Rough set theory is used as a tool for the data pretreatment, firstly to make a decision table, then to make the reduction of attributes by rough set theory, and finally, to evaluate by the quantitative method of multiple-attribute group decision-making. 【Result】 The example computed indicates that more accurate and detailed evaluation results can be obtained, because rough set can better deal with the data which are uncertain, imperfect and inconsistent. 【Conclusion】 The combined use of rough set theory and group eigenvalue method in the renewable evaluation of water resources is feasible.

Key words: regional water resources; renewable evaluation; rough set; group eigenvalue method (GEM); attributes reduction

随着社会经济的迅猛发展和水资源开发利用规模的不断扩大,以及水资源的严重污染和水资源开

* [收稿日期] 2007-01-19
[基金项目] 河南省高校新世纪优秀人才支持计划项目(2006HANCET-03);河南省科技厅软科学研究项目(0613032000)
[作者简介] 杨 婷(1983—),女,陕西凤翔人,在读硕士,主要从事水资源管理、建模及优化研究。E-mail: yangtingxi520@163.com

采、利用、管理的不科学,使水资源的可持续利用问题成为全球瞩目的焦点。基于水资源可再生的特点,进行区域水资源可再生性评价,能够为水资源的可持续利用提供理论依据和决策支持。但由于水资源系统中存在着大量的不确定、不完备与不一致信息,从而影响了区域水资源的准确评价,而粗糙集在处理这些繁杂信息方面具有不可比拟的优势。粗糙集理论(Rough set theory)是 Pawlak^[1] 教授于 1982 年提出来的,现已在数据挖掘、故障诊断、决策分析等领域得到了成功应用^[2-5],但目前关于粗糙集在水资源评价方面的应用研究还不多见,而粗糙集在水资源可再生性评价方面的应用尚属空白。王霞等^[6]将粗糙集与模糊集理论结合应用于区域水资源的系统评价,虽然该研究引领了粗糙集与其他理论结合应用的新趋势,但该研究未考虑水资源系统是个复杂且不确定的大系统,在处理复杂问题时,群体评价较单人评价更具准确性;且该研究采用可辨识矩阵进行属性约简,因该方法需对评价指标反复进行比较,从而影响了约简效率。本研究将粗糙集理论引入水资源可再生性评价中,尝试将粗糙集和群决策特征根法相结合,充分发挥粗糙集理论无需先验信息及仅依赖原始数据的优势,将其作为数据的预处理工具,去除大量冗余信息,且采用广义信息表的方法^[7-8]进行属性约简以提高约简效率,并充分考虑水资源系统的内在复杂性,采用了群体评价的方式,最后以云南各地区水资源可再生状况的评价为例,进行了区域水资源评价的研究,以期为区域水资源可再生性评价方法的建立提供参考。

1 粗糙集理论

1.1 知识与知识库

粗糙集理论认为,知识就是对论域 U 的一种分类能力,知识用集合表示就是等价关系集 R 对论域 U 的划分结果,记为 U/R 。论域上的 1 族划分构成了 U 的 1 个知识库,可用 1 个关系系统 $K=(U,R)$ 来表示。

1.2 信息系统与决策表

$S=(U,A,V,f)$ 为 1 个信息系统, U 为研究论域。其中 $A=C\cup D,C\cap D=\varnothing,C$ 为条件属性集, D 为决策属性集; $V=\bigcup_{a\in A}V_a,V_a$ 是属性 a 的值域; $f:U\times A\rightarrow V$ 为一个信息函数,由其指定 U 中每个对象的属性值。决策表是一种特殊的信息系统,其“行”对应研究对象,“列”对应属性值,每个对象的 1 个属性值对应 1 个等价关系,所以决策表可看作 1 族等

价关系。

1.3 不可区分关系

粗糙集理论的最基本概念是不可区分关系。令 $\forall P\subseteq A$,定义等价关系 $ind(P)=\{(x,y)|x,y\in U,\forall a\in P,f(x,a)=f(y,a)\}$ 为 P 上的不可区分关系,利用不可区分关系可以对论域 U 进行划分。

1.4 上近似与下近似

粗糙集理论中使用了上近似和下近似两个重要概念。下近似是根据已有知识 R 判断 U 中那些肯定属于 X 的元素组成的集合;上近似是根据已有知识 R 判断 U 中那些可能属于 X 的元素组成的集合。对于给定的知识库 $K=(U,R),X$ 的 R 下近似和 R 上近似集可表示为:

$$\underline{R}X=\{x\in U|[x]_R\subseteq X\};$$
$$\overline{R}X=\{x\in U|[x]_R\cap X\neq\varnothing\}.$$

(1)

(2)

式中: $\underline{R}X$ 为 X 的 R 下近似集, $\overline{R}X$ 为 X 的 R 上近似集, $[x]_R$ 为包含 $x\in U$ 的 R 等价类^[9]。

1.5 知识约简及知识的重要性

知识约简是指在保证正确分类的情况下删除冗余的知识。令 P 为 1 族等价关系,属性 $a\in P$,若 $ind(P)=ind(P-\{a\})$,则 a 为 P 中不必要的属性,称 $P-\{a\}$ 为 P 的 1 个约简。 P 的所有约简的交集组成了 P 的核,记为 $core(P)$ ^[9]。

2 群决策特征根法

群决策问题是一个十分复杂的问题,邱苑华^[10]提出的群决策特征根法(GEM)避免了层次分析法(AHP)需要构造判断矩阵且需判断其一致性的不足,可以直接建立专家评分矩阵,使决策过程更为简洁、方便。因为现实中不存在决策可靠性达 100% 的专家,所以认为与其他专家决策的差异性最小的专家,为现实中的理想专家。群决策特征根法通过求出理想专家的评分向量,进而得到被评价对象的排序。

3 计算步骤

- (1)构建决策表。粗糙集只能处理离散型数据,故需要先将连续指标进行离散,再根据离散结果建立决策表。
- (2)属性约简。属性约简的方法有常规约简法、贪心算法、广义信息表约简法等^[11-13]。当研究对象较多时,可采用广义信息表的方法^[7-8]进行属性约简。
- 广义信息表的定义为:根据信息系统 $S=(U,$

$A, V, f), A = C \cup D$, 构造 $U^* = \{(x_i, x_j) \mid x_i, x_j \in U, i < j; i, j = 1, 2, \dots, |U|\}$ 。广义信息表就是以 U^* 中的所有元素对 (x_i, x_j) 为行, 以条件属性指标

$$\mathbf{M} = (m_{(x_i, x_j), a}) = \begin{cases} 1, f(x_i, a) \neq f(x_j, a), \\ 0, f(x_i, a) = f(x_j, a), \end{cases} \forall (x_i, x_j) \in U^*, \forall a \in C。$$

(3)

式中: $\mathbf{M}$ 为区分矩阵, m 为区分矩阵的元素, $f(x_i, a) (i = 1, 2, \dots, |U|)$ 为被评价对象 x_i 在条件属性 a 下的取值。

从式(3)可以看出, 广义信息表中每一行的“1”的个数表示了能够区分该行元素对中元素的指标数目, 若某行只存在一个“1”, 则其所对应的列属性为“核属性”; 每一列 1 的个数表示了该列属性能够区分的元素对的个数。

利用广义信息表进行属性约简的步骤如下:

①找出所有的“核属性”, 得到核属性集 C_0 。

②去掉“核属性”所在列, 一并去掉“核属性”取值为“1”的所有行, 得剩余表。

③若剩余表中元素不全为 0, 则计算剩余表中各列的和, 将和最大的列对应的属性加入到约简属性集 C_1 (C_1 包括核属性集 C_0) 中去, 同时去掉该属性上取值为“1”的行及该属性所对应的列, 如此循环反复, 直到表中剩余的所有元素为 0^[7-8]。

(3)原始数据标准化。用粗糙集理论约简后得

表 1 专家定性描述的数值定量

Table 1 Quantification of expert qualitative description

定性描述 Qualitative description	定量数值 Value	定性描述 Qualitative description	定量数值 Value	定性描述 Qualitative description	定量数值 Value
很不重要 Very unimportant	0	一般 General	5	重要 Important	9
不重要 Unimportant	1	稍重要 Slightly important	7	很重要 Very important	10
稍不重要 Slightly unimportant	3				

应用线性加权的评价方法, 即可应用公式(5)求得各评价对象的综合评价值, 从而按评价值的大小对其进行优劣排序。

$$y_i = \sum_{j=1}^m \omega_j x_{ij} (i = 1, 2, \dots, n)。$$

(5)

式中: y_i 为第 i 个被评价对象的综合评价值, ω_j 为各指标的权重 ($\sum_{j=1}^m \omega_j = 1, 0 \leq \omega_j \leq 1 (j = 1, 2, \dots, m)$), x_{ij} 为第 i 个被评价对象在第 j 个指标下的属性值。

4 实例计算

现以云南省各地区水资源可再生性问题为例, 利用粗糙集和群决策特征根法对其水资源可再生性进行评价分析。所需数据资料根据文献[14]整理而成。为便于比较, 本研究选取与文献[14]相同的评价指标, 具体见表 2。

为列所构成的矩阵, 矩阵元素为 0 或 1。广义信息表可表示为矩阵:

到的指标, 对评价目标而言都是非常重要的, 需要进行这些指标的原始数据进行标准化处理, 以便后续计算。对于逆指标, 可采用取倒数的方法转变成正指标, 然后再进行标准化处理。数据标准化公式为:

$$Z_{ij} = (x_{ij} - \bar{x}_j) / S_j。$$

(4)

式中: Z_{ij} 为第 i 个被评价对象的第 j 个指标经标准化处理后的值, x_{ij} 为第 i 个被评价对象的第 j 个指标的原始值, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$, $S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ 。

(4)群决策特征根法评价。邀请几位专家直接给出各评价指标的重要性评分, 若专家给出的是定性描述, 可根据表 1 将其定量化, 从而得到专家的评分矩阵 $\mathbf{X}$ 。将 $\mathbf{X}$ 转置自乘记为矩阵 $\mathbf{A}$, 即 $\mathbf{A} = \mathbf{X}^T \mathbf{X}$, 矩阵 $\mathbf{A}$ 的最大特征根对应的特征向量就是理想专家的评分向量, 即为各评价指标的重要性排序值^[10]。将理想专家的评分向量归一化后即可得到各评价指标的权重值。

4.1 构建决策表

本文采用等距分割法进行离散(因篇幅所限, 不再列出离散化后的结果), 将单位面积年水资源量、单位面积年地表水资源量、单位面积年地下水资源量、降水水资源量、工业耗水率、农田灌溉耗水率、耕地灌溉率、农业总产值增长率、GDP 年增长率分别用 C_1 、 C_2 、 C_3 、 C_4 、 C_5 、 C_6 、 C_7 、 C_8 、 C_9 表示, 其中 C_9 为决策属性, 其余为条件属性。分别用 A_1 、 A_2 、 A_3 、 A_4 、 A_5 、 A_6 、 A_7 、 A_8 、 A_9 、 A_{10} 、 A_{11} 、 A_{12} 、 A_{13} 、 A_{14} 、 A_{15} 和 A_{16} 表示昆明、曲靖、玉溪、昭通、楚雄、红河、文山、思茅、西双版纳、大理、保山、德宏、丽江、怒江、迪庆和临沧地区, 根据离散化后的结果, 可得到水资源可再生性评价决策表, 如表 3 所示。

4.2 属性约简

因为表 3 中涉及的地区较多, 故本文采用广义

信息表方法进行属性约简。属性约简结果为：核属性为 C_6 ，冗余属性为 C_2 和 C_4 ，约简后的指标集为 $\{C_1, C_3, C_5, C_6, C_7, C_8, C_9\}$ ，并且得知 A_8 、 A_{16} 属于同一分类，即思茅和临沧两地区水资源的可再生能力相当。

表 2 云南省各地区水资源评价指标及其基本资料
Table 2 Evaluation indicators and basic data of water resources in areas of Yunnan province

地区 Areas	单位面积 年水资源量/ ($\text{m}^3 \cdot \text{m}^{-2} \cdot \text{a}^{-1}$)	单位面积年 地表水资源量/ ($\text{m}^3 \cdot \text{m}^{-2} \cdot \text{a}^{-1}$)	单位面积年 地下水资源量/ ($\text{m}^3 \cdot \text{m}^{-2} \cdot \text{a}^{-1}$)	降水水 资源量/mm	工业耗水率/ ($\text{m}^3 \cdot \text{万元}^{-1}$)	农田灌溉 耗水率/ ($\text{m}^3 \cdot \text{万元}^{-1}$)	耕地灌 溉率/%	农业总产值 增长率/%	GDP 年增 长率/%
	Annual water resources quantity per unit area	Annual surface water resources quantity per unit area	Annual underground water resources quantity per unit area	Rainfall amount	Consum ptionrate of industrial water	Consumption rate of farm irrigation water	Farm irrigati- on rate	Growth rate of gross agricultural output value	Annual growth rate of GDP
昆明 Kunming	0.321	0.302	0.100	914.707	216	661	25.3	5.4	8.9
曲靖 Qujing	0.508	0.449	0.121	1 056.499	270	683	18.9	7.0	8.5
玉溪 Yuxi	0.334	0.283	0.108	1 052.562	225	649	31.5	7.9	6.1
昭通 Zhaotong	1.097	0.922	0.365	1 058.751	300	668	10.3	1.0	1.0
楚雄 Chuxiong	0.292	0.214	0.054	1 007.111	259	636	30.4	15.0	10.5
红河 Honghe	0.677	0.636	0.215	1 334.898	315	648	18.4	4.4	11.9
文山 Wenshan	0.426	0.485	0.162	1 054.968	302	633	9.4	7.1	11.7
思茅 Simao	0.793	0.694	0.277	1 658.887	321	620	10.9	6.6	6.6
西双版纳 Xishuangbanna	0.600	0.512	0.231	1 594.800	270	635	19.0	8.5	6.1
大理 Dali	0.452	0.329	0.120	1 164.744	309	651	39.7	5.4	10.0
保山 Baoshan	0.020	0.801	0.302	1 755.159	280	625	27.3	6.5	7.2
德宏 Dehong	1.323	1.254	0.446	2 193.249	299	620	29.3	10.5	9.2
丽江 Lijiang	0.512	0.391	0.134	1 175.671	280	689	27.0	5.8	7.1
怒江 Nujiang	1.819	1.619	0.436	2 416.294	300	677	14.5	5.9	5.9
迪庆 Diqing	0.688	0.531	0.191	1 261.382	328	680	18.9	5.1	6.2
临沧 Lincang	0.913	0.693	0.246	1 644.373	350	625	11.7	6.1	7.6

表 3 云南省各地区水资源可再生性评价决策表
Table 3 Decision table of water resources renewable evaluation in areas of Yunnan province

地区 Areas	属性 Attributes								
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
A_1	1	1	1	1	1	2	2	1	3
A_2	1	1	1	1	2	3	1	2	3
A_3	1	1	1	1	1	2	3	2	2
A_4	2	2	3	1	2	3	1	1	1
A_5	1	1	1	1	1	1	3	3	3
A_6	2	1	2	1	3	2	1	1	3
A_7	1	1	1	1	2	1	1	2	3
A_8	2	2	2	2	3	1	1	2	2
A_9	1	1	2	2	2	1	1	2	2
A_{10}	1	1	1	1	3	2	3	1	3
A_{11}	1	2	2	2	2	1	2	2	2
A_{12}	3	3	3	3	2	1	2	3	3
A_{13}	1	1	1	1	2	3	2	2	2
A_{14}	3	3	3	3	2	3	1	2	2
A_{15}	2	1	2	1	3	3	1	1	2
A_{16}	2	2	2	2	3	1	1	2	2

4.3 原始数据标准化

在粗糙集约简后得到的指标集中， C_5 、 C_6 即工业耗水率和农田灌溉耗水率为逆指标，其余 5 个指标均为正指标，故需先将 C_5 、 C_6 两指标取倒数转化为正指标，随后再用公式(4)对原始数据进行标准化处理（因篇幅所限，不再列出标准化处理后的结

果）。

4.4 群决策特征根法评价

笔者咨询了在水资源领域里卓有成绩的 4 位专家，让其对约简后的指标的重要性进行评分，因这些专家对水资源可再生问题颇为熟悉，故其给出的评分矩阵具有较高的可信度。

根据表 1 将专家的定性描述定量化后,得到的专家评分矩阵为:

$$\mathbf{X}=\begin{pmatrix} 3 & 1 & 5 & 5 & 7 & 9 & 10 \\ 5 & 3 & 5 & 5 & 7 & 7 & 9 \\ 5 & 5 & 3 & 3 & 7 & 10 & 9 \\ 3 & 3 & 5 & 7 & 9 & 7 & 9 \end{pmatrix}。 \tag{6}$$

用 matlab 编程计算矩阵 $\mathbf{A}=\mathbf{X}^T\mathbf{X}$ 的特征根,求得最大特征根 $\rho_{\max}=1.121\ 1$,其对应的特征向量即理想专家的评分向量为:

$$x^*=(0.238\ 0,0.179\ 2,0.268\ 6,0.299\ 2,0.448\ 6,0.493\ 1,0.552\ 5)。 \tag{7}$$

由式(7)可知评价指标的重要性排序为: $C_9>C_8>C_7>C_6>C_5>C_1>C_3$ 。将向量 x^* 作归一化处理后可得到各指标的权重,再利用式(5)计算得到云南省 16 个地区水资源可再生能力的排序为:楚雄>德宏>大理>玉溪>昆明>红河>保山>文山>西双版纳>怒江>曲靖>丽江>临沧≈思茅>迪庆>昭通。与文献[14]结果基本一致。

5 讨论与结论

文献[14]采用多级模糊综合评价的方法,对云南 16 个地区的水资源可再生能力进行了评价,但结果只是将这些地区划分为五类,并未进行排序。本研究采用粗糙集和群决策特征根法相结合的评价方法,对这些地区的水资源可再生性进行了排序,排序结果与文献[14]的结果较为接近,仅局部地区有出入。文献[14]构造隶属函数时依据的“综合评判指标分级值表”是参考文献[15]建立的评价标准而确定的,由于部分指标极难获得数据或数据不够,文献[15]的“评价标准(以全国为基准)”是参考黄河流域及西北地区研究数据而建立的,因此用基于该评价标准建立的隶属函数,来评价云南各地区水资源的可再生性,其准确性将受到一定的影响。且文献[14]直接给出了第一层准则间的权重,其主观性较强。本研究将粗糙集引入到水资源可再生性的评价中,并尝试将粗糙集与群决策特征根法结合应用,既充分利用了粗糙集仅依赖于原始数据的优势,又符合水资源工作中常需进行群决策的实际情况,所得结论应该更为准确,且更具实用性。但研究中还发现,如何提高属性的约简效率和准确性,还应进行更为深入的研究。

[参考文献]

[1] Pawlak Z. Rough set-theoretical aspects of reasoning about da-

ta[M]. Dordrecht: Kluwer Academic Publishers,1991.

[2] 崔广才. 基于粗糙集的数据挖掘方法研究[D]. 吉林长春: 吉林大学计算机科学技术学院,2004.

Cui G C. Research on data mining methods based on Rough set [D]. Jilin, Changchun: Department of Computer Science and Technology of Jilin University,2004. (in Chinese)

[3] 朱根标,张凤鸣,王金干. 基于粗糙集的管理人才甄选方法研究[J]. 工业工程,2006,9(2):95-97.

Zhu G B,Zhang F M,Wang J G. Selection of managers based on Rough set[J]. Industrial Engineering Journal,2006,9(2):95-97. (in Chinese)

[4] 韩敏,孙林夫,赵慧娟. 基于粗糙集理论的设备资源优化配置[J]. 计算机工程,2005,31(17):197-199.

Han M,Sun L F,Zhao H J. Equipment resources optimizing based on Rough set theory[J]. Computer Engineering,2005,31(17):197-199. (in Chinese)

[5] 张文修,吴伟志. 粗糙集理论介绍和研究综述[J]. 模糊系统与数学,2000,14(4):1-12.

Zhang W X,Wu W Z. An introduction and a survey for the studies of Rough set theory[J]. Fuzzy Systems and Mathematics,2000,14(4):1-12. (in Chinese)

[6] 王霞,唐德善. 基于粗糙模糊集的区域水资源系统的评价方法[J]. 水利规划与设计,2006(1):31-33;57.

Wang X,Tang D S. Evaluation method of regional water resources based on Rough-fuzzy set[J]. Water conservation planning and design,2006(1):31-33;57. (in Chinese)

[7] 陈自洁. 一种改进的粗集综合评价方法[J]. 海南师范学院学报:自然科学版,2005,18(4):324-329.

Chen Z J. Improved method for synthetical evaluation based on Rough set[J]. Journal of Hainan Normal University: Natural Science Edition,2005,18(4):324-329. (in Chinese)

[8] 刘文军,王加银,冯艳宾,等. 一种求粗糙集中最小属性约简的新算法[J]. 北京师范大学学报:自然科学版,2004,40(1):8-12.

Liu W J,Wang J Y,Feng Y B, et al. A new minimal attribute reduction algorithm[J]. Journal of Beijing Normal University: Natural Science Edition,2004,40(1):8-12. (in Chinese)

[9] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北京:科学出版社,2001:1-20.

Zhang W X,Wu W Z,Liang J Y, et al. Rough set theory and methods[M]. Beijing:Science Press,2001:1-20. (in Chinese)

[10] 邱苑华. 管理决策与应用熵学[M]. 北京:机械工业出版社,2001:285-289.

Qiu W H. Management decision-making and applied entropy study[M]. Beijing:China Machine Press,2001:285-289. (in Chinese)

[11] 姚明臣,孟凡超. 属性约简的一种贪心算法[J]. 佳木斯大学学报:自然科学版,2003,21(3):307-311.

Yao M C,Meng F C. A greedy algorithm for attribute reduction[J]. Journal of Jiamusi University: Natural Science Edition,2003,21(3):307-311. (in Chinese)