

基于可拓数据挖掘的黄河三角洲土地利用评价

王建林^{1,2}, 杨印生¹, 王学玲²

(1. 吉林大学 生物与农业工程学院, 长春 130022; 2. 滨州学院 计算机科学技术系, 山东 滨州 256600)

摘要:以黄河三角洲地区的滨州市为例,从土地资源、生态环境和社会经济三个方面选择评价指标建立了针对土地利用的可拓评价模型,并在此基础上利用可拓数据挖掘的理论和方法分析该地区20年来的土地利用方式及结构变化对社会经济及生态环境造成的影响,为该区域今后土地资源的优化配置及科学的土地利用管理提供决策参考意见。

关键词:农业工程; 土地利用; 可拓评价模型; 可拓数据挖掘; 黄河三角洲

中图分类号:S289 文献标志码:A 文章编号:1671-5497(2012)Sup. 1-0479-05

Evaluation of land use in Yellow river delta based on extension data mining

WANG Jian-lin^{1,2}, YANG Yin-sheng¹, WANG Xue-ling²

(1. College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China; 2. Dept of Computer Science, Binzhou University, Binzhou 256600, China)

Abstract: A new evaluation of land use method is proposed in this paper based on extension data mining. Firstly, taking yellow river delta as an example and selecting index from database of land resources, environment and social-economic development, and then establishing the Extension evaluation model of land use. Secondly, using the extension evaluation model and theory of extension data mining to analysis the effect of land use change on environment and social-economic in the past twenty years. Finally, results showed that this research could provide the theoretical and reference basis for optimization of resources allocation and scientific land utilization management.

Key words: agricultural engineering; land use; extension evaluation model; extension data mining; Yellow river delta

土地利用方式和结构变化不仅表现为空间概念上的分布形式,也应反映社会、经济、生态环境等各方面因素共同作用的结果。蔡运龙^[1]指出土地利用研究要把社会经济信息纳入进来,并寻求新的综合分析途径。人类发展对经济利益的追求驱动着土地利用方式和结构的转变,而土地利用方式和结构的转变又反作用于人类生存的社会环境和自然环境。因此,为保证经济效益,社会效益

和生态效益的协调发展及土地资源的可持续利用,综合考虑自然资源状况和人类活动的影响来评价土地利用状况,能得到更科学全面的结果^[2-4]。计算机技术的快速发展,使得社会经济管理的各个部门数据库中都积累了大量的统计数据,如何利用这些数据更好的为社会经济发展服务,可拓数据挖掘技术^[5-6]为问题的解决指明了方向。本文将可拓学的理论与数据挖掘技术结合来

收稿日期:2012-03-30.

基金项目:国家自然科学基金项目(60673089);山东省自然科学基金项目(ZR2009GL001);山东省科学计划技术发展计划项目(2010RKGB1143).

作者简介:王建林(1971-),男,副教授,博士研究生. 研究方向:系统工程. E-mail:bzxywjl@163.com

通信作者:杨印生(1963-),男,教授,博士生导师. 研究方向:系统工程. E-mail:yys@jlu.edu.cn

构建土地利用可拓评价模型。该评价模型从社会统计数据中选择涉及土地资源、生态环境和社会经济的多个指标进行评价,以黄河三角洲地区的滨州市为例,分析了该地区近 20 年来的土地利用方式和结构变化对社会经济及生态环境造成的影响,为今后该地区的经济社会、人口、土地资源和生态环境的协调发展提供科学的指导。

1 可拓学及可拓数据挖掘技术

1.1 可拓学基本思想

可拓学^[5]理论主要包含物元理论和可拓集合理论。为了形式化描述事物而建立的物元模型,是以事物、特征及事物关于该特征的量值所组成的三元组,描述为: $M = (\text{事物}, \text{特征}, \text{量值}) = (O, C, X)$ 。可拓集是为了描述事物的动态分类而提出的,它可以定量化的描述事物的量变和质变过程。可拓集合是在经典集和模糊集的基础上发展起来的另一集合概念。设 U 为论域, u 为论域 U 中的任一元素, k 为 U 到实数域 R 的一个映射, $T = (T_U, T_k, T_u)$ 为给定变换, 称 $E(T) = \{(u, y, y') | u \in T_U, y = k(u) \in R, y' = T_k(T_u u) \in R\}$ 为论域 U 上的可拓集。其中, $y = k(u)$ 为可拓集上的关联函数。 $y' = T_k(T_u u)$ 为可拓集上的可拓函数, T_U, T_k, T_u 分别为对论域中 U 、关联函数 k 和元素 u 的变换。物元可拓集因能够合理地描述自然和社会现象中各事物的内部结构、彼此关系及事物变化状态从而成为描述事物可变性的有效工具。

1.2 可拓数据挖掘技术

可拓数据挖掘克服了传统数据挖掘^[7]只能挖掘不变条件下知识的弊端。它利用物元模型描述数据库或者数据仓库中的数据,形成适合于“变换”的数据表示方法,然后利用关联函数评价的结果,可以挖掘出“变换”对于数据的作用规律(简称可拓知识)。可拓数据挖掘的算法描述如图 1 所示。

2 面向土地利用的可拓评价模型

2.1 构建物元模型

由于土地利用系统是由不同属性的子系统相互交织、相互作用、相互渗透而构成的具有特定结构和功能的开放的复杂系统。某地区的土地利用系统受土地资源、土地经济、生态环境和人口数量的影响最大,为了科学、客观评价该地区的土地利

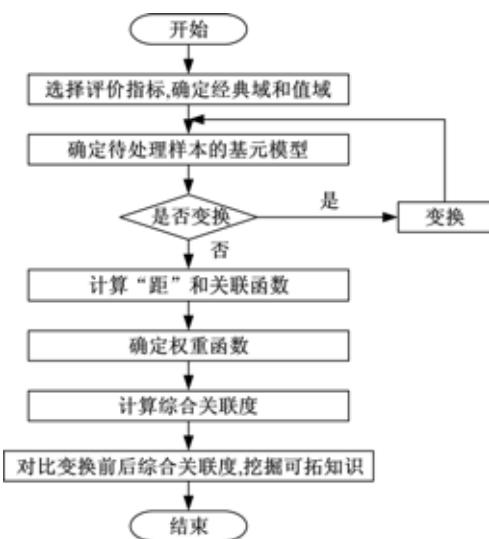


图 1 可拓数据挖掘算法流程

Fig. 1 The flowchart of extension data mining algorithm
用状况,本文将从土地资源、生态环境和社会经济三个方面选择评价指标。

选择评价指标并在此基础上构建物元模型的过程为:首先,从社会经济管理部门积累了大量的统计数据中抽取出反映土地资源状况的数据:建国后耕地面积、土地总面积、人口数量;反映社会经济发展状况的数据:农业产值、粮食作物播种面积、粮食总产;反映生态环境状况的数据:化肥总用量,之所以选择化肥用量作为评价指标是因为农田过量施用化肥已是造成农业非点源污染的主要原因^[8]。其次,在此基础上进行数据转换,选择耕地面积变化、人口密度、农业产值增长、粮食单产、每公顷耕地化肥用量 5 个评价指标。为了反映工业经济发展对生态环境的影响情况,在上面 5 个评价指标的基础上增加了反映工业污染状况的大气 SO_2 含量评价指标。最后建立评价信息元表,即把数据库中的评价数据用可拓学的物元模型表示

为: $B = \begin{pmatrix} O & C_1 & x_1 \\ \vdots & \vdots & \vdots \\ C_m & x_m \end{pmatrix}$, 其中, O 为土地利用状

况, C_i 为某个评价指标, x_i 为该评价指标的量值。

2.2 关联函数计算

建立关联函数,首先要确定特征变量的值域区间 $X = \langle c, d \rangle$ (节域)和对事物的满意区间 $X_0 = \langle a, b \rangle$ (经典域)。文中节域与经典域的确定采用统计的方法从数据库中获取或者来自文献资料如山东省 1999 年平均每公顷耕地粮食产量、2008 年全国平均每公顷耕地化肥施用量、GB/T3095—1996 中 SO_2 平均浓度标准等数据确定。设 X

$= (x_1, x_2, \dots, x_m)$ 为物元 B 关于特征 C 的量值描述, 可拓数学中提出的“距”概念能够定量、客观的描述事物具有某种性质的程度及其量变与质变的过程。关联函数表示为

$$k(x) = \frac{\rho_r(x, x_0, X_0)}{D(x, x_0, X)} \quad (1)$$

由于评价指标的不同, 耕地面积变化、农业产值增长、粮食单产 3 个评价指标的期望值越大越好(选用右侧距)。人口密度、每公顷耕地化肥用量、大气 SO_2 含量 3 个评价的期望值则越小越好(选用左侧距)。文中计算关联函数选用的侧距和位置公式如下:

右侧距:

$$\rho_r(x, x_0, X_0) = \begin{cases} a - x, & x \leqslant x_0 \\ \frac{a - x_0}{b - x_0}(b - x), & x \in (x_0, b) \\ x - b, & x \geqslant b \end{cases} \quad (2)$$

左侧距:

$$\rho_l(x, x_0, X_0) = \begin{cases} a - x, & x \leqslant a \\ \frac{b - x_0}{a - x_0}(x - a), & x \in (a, x_0) \\ x - b, & x \geqslant x_0 \end{cases} \quad (3)$$

位置:

$$D(x, x_0, X_0) = \begin{cases} \rho(x, X) - \rho(x, X_0), & x \in X_0 \\ \rho(x, X) - \rho(x, X_0) + a - b, & x \notin X_0 \end{cases} \quad (4)$$

2.3 根据权重计算综合关联度

权重的确定是土地利用评价的关键, 由于评价的侧重点不同, 采用专家定性后层次分析是目前较为可行的方法, 具体方法见文献[9]。确定评价目标的权重 $a_i, i = 1, 2, \dots, n$, 且满足 $\sum_{i=1}^n a_i = 1$ 后, 利用关联函数, 计算出待评价事物 N 关于对于评价指标 c 的关联度为:

$$K(N) = \sum_{i=1}^n a_i k(c_i) \quad (5)$$

3 评价实例及应用

对于土地系统这样一个复杂系统, 土地利用涉及社会生活的方方面面, 本文从滨州市的社会统计数据中选择涉及土地资源、生态环境和社会经济发展方面的 6 个指标建立可拓评价模型, 利用可拓数据挖掘的方法对该地区的土地利用状况

进行分析评价。

3.1 评价过程

本文首先在 Microsoft SQL Server 2005 数据库系统下对数据库中的数据抽取、转换, 并依此建立了评价信息元表, 然后, 在此环境下设计了计算关联函数及综合关联度的相关存储过程, 最后, 利用 Microsoft Visual Studio 2008 集成开发环境来展示挖掘的可拓知识。

变换前后信息元表分别为 1988~1997 年间和 1998~2007 年间滨州市 6 县 1 区耕地面积变化、人口密度、农业产值增长、粮食单产、每公顷耕地化肥用量评价, 变换前后的大气 SO_2 含量指标分别取自环保局 1996 年数据和 2010 年数据。变换后数据见表 1。本文建立关联函数采用的经典域和节域区间见表 1 中描述。表 1 中各指标具体含义如下:

指标 1: 耕地变化(1998~2007 年平均耕地面积占建国初面积的百分比)

指标 2: 人口密度(2008 年平均每平方公里人口数量)

指标 3: 农业产值增长(1998~2007 平均产值相对 1978~1987 平均产值的增长率)

指标 4: 粮食单产(1998~2007 每公顷耕地粮食均产)

指标 5: 化肥用量(1998~2007 每公顷耕地化肥平均用量/t)

指标 6: 2010 大气 SO_2 含量(mg/m^3)

表 1 评价信息元表

Table 1 The table of evaluation message

县市名称	指标 1	指标 2	指标 3	指标 4	指标 5	指标 6
滨城区	0.563	577	10,239	5421.5	0.681	0.094
惠民县	0.735	458	10,778	5925.3	0.667	0.082
阳信县	0.749	568	10,836	5536	0.608	0.058
无棣县	0.878	219	14,801	4267.3	0.498	0.038
沾化县	0.808	180	14,907	4616.7	0.434	0.051
博兴县	0.669	533	11,294	6383	0.646	0.065
邹平县	0.704	579	10,712	6360.7	0.643	0.123
经典域	0.75, 1	100,428	5,15	4310,7348	0,0.43	0,0.06
节域	0.5, 1	100,600	1,15	2000,7348	0,0.7	0,0.15

3.2 评价结果及分析

通过计算关联函数并进行规范化后, 就可以通过查询关联函数来对数据进行初步分析, 举例见图 2、图 3。根据需求设置不同的权重系数计算综合关联度后的实验结果见图 4~图 7。

关联函数分类查询				
选择查询条件: 自然资源变化 确定				
县市名称	变换前耕地	变换后耕地	变换前人口密度	变换后人口密度
滨城区	-1	-1	-0.94489...	-0.98675...
惠民县	-0.0906192...	-0.081724...	-0.17322...	-0.19867...
阳信县	0.15326554...	-0.006851...	-0.98425...	-0.92715...
无棣县	0.47141055...	0.3412036...	0.582362...	0.476132...
沾化县	0.10366628...	0.1546509...	0.682582...	0.56498...
博兴县	-0.5027419...	-0.4927095...	-0.58267...	-0.4927095...

图 2 关联函数条件查询

Fig. 2 The Criteria query of correlation function

关联函数分类查询				
选择查询条件: 生态环境变化 确定				
县市名称	变换前化肥	变换后化肥	变换前二氧化硫	变换后二氧化硫
滨城区	-0.2840...	-1	0.731707...	-0.53968251...
惠民县	-0.1644...	-0.9507...	0.902438...	-0.34920539...
阳信县	-0.0595...	-0.7389...	0.604877...	0.019047621...
无棣县	0.13627...	-0.3499...	1	0.209523804...
沾化县	0.03068...	-0.1213...	0.707316...	0.085714279...
博兴县	-1	-0.8740...	0.780487...	-0.079365069...

图 3 关联函数条件查询

Fig. 3 The Criteria query of correlation function

综合关联度评价					
耕地减少	人口密度	经济发展	粮食单产	化肥施用	二氧化硫
0.6	0.4	0	0	0	0

变换前后的综合关联度及可拓分类类别

县市名称	变换前	变换后	变换后强度值	可拓分类
滨城区	-0.977952774	-0.994702013	-0.016749039	变差
惠民县	-0.123662677	-0.126505181	-0.004842304	变差
阳信县	-0.301741457	-0.374971916	-0.07320458	变差
无棣县	0.5157912361	0.3951751683	-0.120616067	变差
沾化县	0.3352328529	0.3187825692	-0.016450283	变差
博兴县	-0.534717222	-0.537583055	-0.002866043	变差
邹平县	-0.358807000	-0.549034039	-0.190167038	变差

图 4 土地资源综合关联度

Fig. 4 The degree of comprehensive association about land resources

综合关联度评价					
耕地减少	人口密度	经济发展	粮食单产	化肥施用	二氧化硫
0	0	0	0	0.5	0.5

变换前后的综合关联度及可拓分类类别

县市名称	变换前	变换后	变换后强度值	可拓分类
滨城区	0.2238422041	-0.765941258	-0.993653452	变差
惠民县	0.3690062841	-0.6499669892	-1.016974869	变差
阳信县	0.3746782722	-0.359953125	-0.734631397	变差
无棣县	0.56813960798	0.070189512	-0.636328120	变差
沾化县	0.3689992519	-0.0178077964	-0.366807218	变差
博兴县	-0.109756153	-0.476731092	-0.264974938	变差
邹平县	0.3005757560	-0.906621751	-1.233203507	变差

图 5 环境污染综合关联度

Fig. 5 The degree of comprehensive association about environment

由可拓学的理论可知,关联函数能够定量、客观的描述事物具有某种性质的程度及其量变与质变的过程,规范的关联函数取值为[-1,1],值越大说明评价指标越好。例如,从图 2 查询出的关

设置权重系数					
耕地减少	人口密度	经济发展	粮食单产	化肥施用	二氧化硫
0.15	0.1	0.3	0.15	0.15	0.15

变换前后的综合关联度及可拓分类类别

县市名称	变换前	变换后	变换后强度值	可拓分类
滨城区	-0.255699590	0.5101803467	1.2364266258	变好
惠民县	-0.225446848	0.6420009394	0.8977004202	变好
阳信县	-0.225446848	0.5897779671	0.8162048154	变好
无棣县	-0.123730646	0.6719965371	0.7987271831	变好
沾化县	-0.2564484739	0.7443849663	1.0006697061	变好
博兴县	0.0492647428	0.7447144142	0.6954496713	变好
邹平县	0.0262100370	0.7003511371	0.6721417001	变好

图 6 农业经济发展综合关联度

Fig. 6 The degree of comprehensive association about social-economic

设置权重系数					
耕地减少	人口密度	经济发展	粮食单产	化肥施用	二氧化硫
0.15	0.1	0.3	0.15	0.15	0.15

变换前后的综合关联度及可拓分类类别

县市名称	变换前	变换后	变换后强度值	可拓分类
滨城区	-0.460039787	-0.2405059029	0.2394807578	变好
惠民县	-0.028366057	0.0647236795	0.0930917365	变好
阳信县	-0.0810265069	0.06370582193	0.1291322892	变好
无棣县	0.2112255804	0.3642707554	0.1300451750	变好
沾化县	0.0631573906	0.3965457536	0.3333680630	变好
博兴县	-0.143269861	0.0831767537	0.2075466149	变好
邹平县	0.0149833383	-0.0958986033	-0.110661371	变差

图 7 土地利用综合关联度

Fig. 7 The degree of comprehensive association of land use 联函数结果可知,就耕地面积减少和人口密度这两项涉及土地资源的评价指标来讲,只有滨海的沾化县和无棣县的关联函数都在正值域,说明这两个县的人均土地占用量多,利用率低。虽然滨州沿海区这两个县的这一特点通过其他方式也可以得到,但是只有利用关联函数,才能实现定量的描述事物在多个评价指标情况下的“变换”规律。再如图 4 所示,近 10 年来滨州市 6 县 1 区的土地资源状况与上世纪 90 年代相比均向差的方面发展,尽管这种变化在工业化的进程中不可避免,但是,怎样调整产业结构,尽量减轻工业化革命带来影响,可以根据综合关联度量化的变化情况不同来决策。

分析滨州市北部沿海具有土地资源优势的无棣县、沾化县情况。该区土地利用程度低,土地盐碱多,丰富的资源使得该区域发展林牧渔业的潜力巨大。文中对统计数据的分析显示沾化县在土地资源(图 4)、环境污染方面(图 5)变差的幅度都比无棣县小,而农业经济增长情况则好于无棣县(图 6),图 7 中的综合关联度评价也验证了这个结果。这就显示出沾化县最近十年的农业发展模式远好于无棣县,如沾化县大力发展的冬枣种植产业,就在近些年取得良好的经济和社会效益。

分析地处黄河以南的邹平、博兴两县情况。由于长期以来的高度开发利用加上工业化的影

响,使得原本滨州市土地经济效益最高、土地质量最好的两个县农业经济发展综合关联度增加值远低于其他县市(图 6),邹平在土地资源(图 4)和环境污染(图 5)方面变差的程度则位列全区之首。图 7 的综合关联度评价也显示邹平县的土地利用状况在变差。这正是工业化、城镇化进程的不断加快及人口的不断增长,资源环境的过渡开发利用,人地矛盾日益加剧的一个体现。控制工业污染、科学施用化肥,使现在的“资源型”农业模式尽快向“知识型”农业模式转移对邹平县来讲已是刻不容缓。

分析地处黄河沿岸的惠民县、阳信县和滨城区。该区域土地利用程度较高,水资源丰富。统计数据分析显示惠民县、阳信县具有一定的土地资源优势,滨城区则由于是市政府所在地,城市的发展和人口的快速增长使得该区原有的土地资源优势消失殆尽(图 4)。该区农业经济发展较快,其中滨州市的增长最快(图 6),但是环境污染尤其是化肥的过量施用状况严重(图 3、图 5)。以上信息表明近年来该区域农业经济发展速度较快,尤其是滨城区在近年的农业发展中围绕城市需要,开展的高产、高效商品农业方面取得了很大成绩,但是在注重经济效益的同时怎样兼顾生态效益是必须要解决的一个难题。

4 结 论

本文应用可拓学的理论并结合数据挖掘技术,构建了针对土地利用的可拓评价模型。该评价模型从社会统计数据中选择涉及土地资源、生态环境和社会经济的多个指标进行评价,并在实例应用取得了良好效果。该方法主要优点为:①从社会、经济、生态环境等不同方面选取评价指标,解决了复杂系统难评价的问题。②根据不同的评价指标定义关联函数,不仅客观真实地反映了各个评价指标的性质,而且解决了不同量纲的指标的综合评价问题。③计算机技术实现的评价过程不仅可以形象、直观的观察评价结果,而且可以根据需要来调整评价指标的权重。

参考文献:

[1] 蔡运龙. 土地利用/土地覆被变化研究:寻求新的综

- 合途径[J]. 地理研究, 2001, 20(6): 645-652.
- Cai Yun-long. A study on land use/cover change: the need for a new integrated approach[J]. Geographical Research, 2001, 20(6): 645-652.
- [2] 刘荣霞,薛安,韩鹏,等. 土地利用结构优化方法述评[J]. 北京大学学报:自然科学版, 2005, 41(4): 655-662.
- Liu Rong-xia, Xue An, Han Peng, et al. Optimization of land use structure: commentary of methods [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2005, 41(4): 655-662.
- [3] 赵军,杨凯. 生态系统服务价值评估研究进展[J]. 生态学报, 2007, 27(1): 346-356.
- Zhao Jun, Yang Kai. Valuation of ecosystem services: characteristics, issues and prospects[J]. Acta Ecologica Sinica, 2007, 27(1): 346-356.
- [4] 刘海,陈晓玲,黄蓉,等. 基于土地利用信息的区域可持续发展评价方法[J]. 农业工程学报, 2011, 27(11): 306-312.
- Liu Hai, Chen Xiao-ling, Huang Rong, et al. Evaluation method of regional sustainable development based on land use information[J]. Transactions of the CSAE, 2011, 27(11): 306-312.
- [5] 蔡文,杨春燕,陈文伟,等. 可拓集与可拓数据挖掘[M]. 北京:科学出版社, 2008.
- [6] Yang Chun-yan. Extension classification method and its application based on extensible set[C]// Proceedings of 2007 International Conference on Wavelet Analysis and Pattern Recognition. Beijing, 2007, 11: 819-824.
- [7] Witten I H, Frank E. Data Mining: Practical Machine Learning Tools and Techniques[M], 2nd Edition. Morgan Kaufmann, San Francisco, 2005.
- [8] 郭鸿鹏,朱静雅,杨印生. 农业非点源污染防治技术的研究现状及进展[J]. 农业工程学报, 2008, 24(4): 290-294.
- Guo Hong-peng, Zhu Jing-ya, Yang Yin-sheng. Research status and development of technologies for controlling the agricultural non-point source pollution[J]. Transactions of the CSAE, 2008, 24(4): 290-294.
- [9] 杨印生,李洪伟. 管理科学数量分析方法选讲[M]. 长春:吉林大学出版社, 2008.