

基于数据挖掘的生态公益林效益评价决策支持系统研究

彭佳红^{1,2}, 邹冬生¹

(¹湖南农业大学生态学研究所, 长沙 410128; ²湖南农业大学信息科学与技术学院, 长沙 410128)

摘要:生态公益林效益评价是生态公益林系统管理与可持续发展的重要内容。在生态公益林效益评价的主要内容与价值量评估公式、参数设置及方法的基础上,给出了基于数据挖掘的生态公益林效益评价决策支持系统的功能结构与体系结构;设计了以模型库与知识库为中心,辅以数据库、方法库和计算机网络技术的生态公益林效益评价决策支持系统;利用软件工程方法、J2EE 框架技术、MVC 设计模式和数据库访问技术实现系统开发。系统利用数据挖掘中聚类分析技术,能进行生态公益林资源分类和关联规则技术挖掘出公益林不同效益指标之间的内在联系。系统具有较强的分析、预测与决策、可视化输出功能,辅助生态公益林管理与决策者提高决策水平和质量。

关键词:决策支持系统;数据挖掘;公益林;效益评价;J2EE

中图分类号:S718, TP301

文献标志码:A

论文编号:2010-2515

Research on Decision Support System of Benefit Evaluation of Non-Commercial Forest Based On Data Mining

Peng Jiahong^{1,2}, Zou Dongsheng¹

(¹Institute of Ecology, Hunan Agricultural University, Changsha 410128;

²College of Information Science and Technology, Hunan Agricultural University, Changsha 410128)

Abstract: The benefit evaluation of non-commercial forest is the main part of non-commercial forest system management and sustainable development. On the basis of the main part of benefit evaluation of non-commercial forest, value magnitude formula, preferences and method, it showed the functional structure and architecture of decision support system of benefit evaluation of non-commercial forest based on data mining, and designed a decision support system of benefit evaluation of non-commercial forest centered on model base and knowledge base, assisting with data base, method base and computer network technology, using methodology of software engineering, J2EE framework, MVC design pattern and database access technique to realize system development. The system can classify non-commercial forest sources by using cluster analysis and find out the internal relation among the performance indicators of non-commercial forest with association rules. This system has comparatively strong functions of analyzing, predicting, decision-making and visual output to assist non-commercial forest management and decision makers to enhance the level and quality of decision making.

Key words: decision support system; data mining; non-commercial forest; benefit evaluation; J2EE

0 引言

生态公益林效益评价是多目标、多因素、多层次和多指标的综合评价,数据多,计算量大。遥感(RS)、全球定位(GPS)与地理信息系统(GIS)技术的应用大大

提高了工作效率。但目前,对生态公益林效益研究大多停留在理论与机制的探讨上,缺乏具体性、时效性与系统性地研究,缺乏应用新技术、新方法对其功能状态的科学定性定量分析。如何利用现代生态学理论、

基金项目:湖南省科技厅科技计划项目(2010GK3037);湖南省教育厅科学研究项目(10C0789)。

第一作者简介:彭佳红,女,1962年出生,湖南永州人,教授,在读博士,研究方向:数据挖掘、人工智能、系统生态学与生态系统评价。通信地址:410128 湖南农业大学信息学院, E-mail: pjh719@163.com。

收稿日期:2010-08-25, **修回日期:**2010-10-22。

系统分析原理和信息技术有效地进行生态公益林效益评价,创造一个稳定的、结构合理、功能优化、高生态功能、高社会效益的生态公益林系统,是科学研究迫切需要解决的技术问题。

作为决策支持系统(DSS)的一个重要组成部分,数据挖掘技术已经越来越成为近年来国内农林业部门关注的焦点之一^[1]。基于数据挖掘的决策支持系统作为一种新兴的信息技术,为决策者提供分析问题、建立模型、模拟决策过程和方案的环境,调用各种信息资源和分析工具,减轻管理者从事低层次信息处理、分析而专注于决策知识、经验的工作^[2]。生态公益林管理信息系统向决策支持系统演变是未来发展趋势。因此,本研究拟构建生态公益林效益评价指标体系,利用信息技术将生态公益林信息数字化,设计与实现生态公益林效益评价决策支持系统,并将数据挖掘技术应用其中,辅助林业部门对生态公益林经营、管理与效益补偿,提供科学的决策。

1 生态公益林效益评价

生态公益林是为维护和改善生态环境,保持生态平衡,保护生物多样性等满足人类社会的生态、社会需求和可持续发展为主体功能,主要提供公益性、社会性产品或服务的森林、林木、林地。Constanza^[3]提出的从气体调节到休闲、文化等17个生态系统服务类型,这是目前最有影响力和生态服务研究的成果^[4],实际上已经构成了生态公益林效益的评价指标体系。但在具体的评估过程中,价值评价指标体系上仍有一定的差异。在区域生态系统效益评价的实际应用上,不同的学者在估算生态公益林效益时,从不同的角度涉及到效益评价指标体系中的不同部分^[5-8]。综合起来,生态公益林效益评价的主要内容包括:

- (1)涵养水源效益。调节水量、净化水质等。
- (2)保育土壤效益。固土、保肥等。

- (3)固碳释氧效益。固碳、释氧等。
- (4)营养物质循环效益。林木营养积累。
- (5)净化大气环境效益。提供负离子、吸收污染物、降低噪音、滞尘、灭菌等。
- (6)调节小气候效益。气温、大气相对湿度、蒸散力、太阳辐射等。
- (7)公益林防护效益。抵御大风、地震等自然灾害。
- (8)生物多样性保护效益。物种保育。
- (9)公益林游憩效益。旅游、休憩和保健等。
- (10)社会效益。教育、文化、美化环境、维护人们沟通、稳定人际关系等。

上述10项效益评价,各国学者采取了各自的计量模型与方法计算其效益。笔者主要采用国家林业局发布的LY/T 1721—2008森林生态系统服务功能评估规范中的森林生态系统服务功能价值量评估公式、参数设置及方法^[9]。其中,森林游憩和社会效益评价的方法主要采用市场价值法和条件价值法。

2 系统功能与体系结构设计

2.1 系统功能设计

基于数据挖掘的生态公益林效益评价DSS主要对生态公益林效益评价相关的历史数据、空间数据,利用现代生态学理论、系统分析原理、数据挖掘和决策支持系统技术等有效地解决生态公益林效益评价问题,主要包括公益林林分知识管理、数据挖掘、决策、输入输出、统计等功能。系统功能结构如图1所示。

2.1.1 系统管理模块 用户分为系统管理员和普通用户。系统管理员有分配权限,也能对登录用户的权限进行控制;系统可以设置多种不同权限的用户进行登陆和操作。普通用户一般为知识工程师或数据录入者,对数据有增、删、查、改等操作权限。

2.1.2 数据管理模块 效益评价相关的历史数据、空间

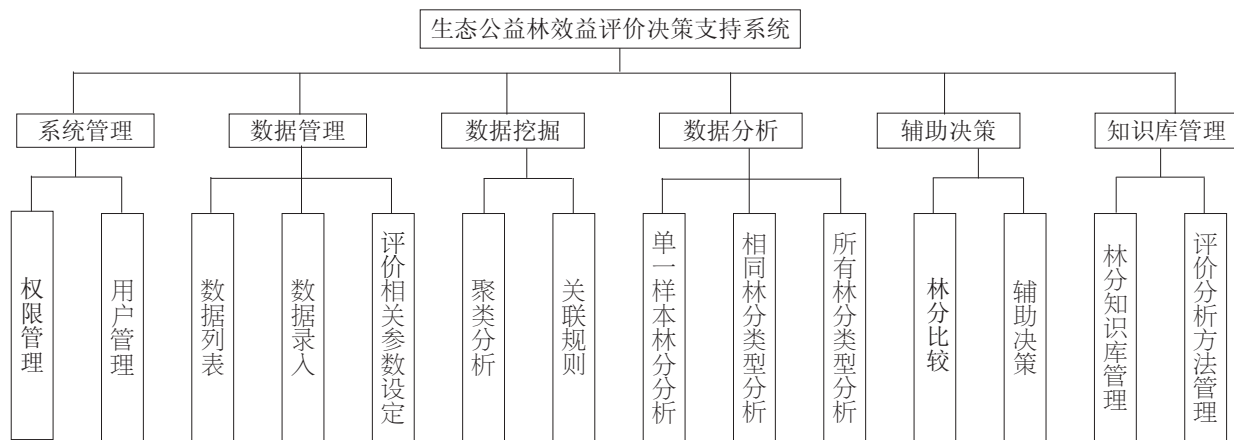


图1 系统功能结构图

数据、指标与参数,主要包括降水量、林分蒸散量、地表径流量、水库建设单位库容投资(占地拆迁补偿、工程造价、维护费用等)、水的净化费用、林分面积、无林地土壤侵蚀模数、林地土壤侵蚀模数、挖取和运输单位体积土方所需费用、林地土壤容重、林分土壤平均含氮量、林分土壤含钾量、林分土壤平均含磷量、林分土壤有机质含量、磷酸二铵化肥含氮量、磷酸二铵化肥含磷量、氯化钾化肥含钾量、磷酸二铵化肥价格、氯化钾化肥价格、有机质价格、林分净生产力、单位面积林分土壤年固碳量、固碳价格、CO₂中碳的含量(为27.27%)、氧气价格、林木含钾量、林木含氮量、林木含磷量、林分负离子浓度、负离子生产费用、负离子寿命、林分高度、二氧化硫治理费用、氟化物治理费用、单位面积林分年吸收二氧化硫量、单位面积林分年吸收氟化物量、单位面积林分年吸收重金属量、森林面积折合为隔音墙的公里数、单位面积林分年滞尘量、林地所处地农作物、牧草等价格、由于农田防护林、防风固沙等森林存在增加的单位面积农作物、牧草等年产量、单位面积年物种损失的机会成本、采用市场价值法和条件价值法计算森林游憩和社会效益等各种数据录入数据库,同时设定评价指标参数,并进行相应的数据管理。

2.1.3 数据分析模块 生态公益林效益评价主要是完成各个效益指标价值评估和不同林分类型的效益整体功能指标的价值评估,包括各类数据的分析工作。该模块主要实现单个林分类型分析、相同林分类型分析、所有林分类型分析、比较排序等功能。根据输入的数据和相关参数进行统计分析,其分析结果呈现的方式可以是文本、表格、饼图等形式,查询条件和组合方式灵

活,为分析决策人员提供多角度的数据可视化分析界面^[10]。

2.1.4 数据挖掘模块 根据生态公益林效益的特点,对数据分析后的数据进行数据挖掘中聚类分析和关联规则分析以便获取知识,提供决策支持。通过聚类分析进行森林资源的分类和利用关联规则技术挖掘生态公益林不同效益指标之间的内在联系。

2.1.5 辅助决策模块 根据生态公益林效益评价复杂的特点,采用产生式规则对获取的知识和数据进行表示、组织和存储;结合数据库技术,建立系统的知识库和规则库,完成对生态公益林效益决策支持系统辅助决策功能。该模块包含:一是生态公益林效益评价能力比较,比如调节水源、保育水土、固土、保肥、固碳、释氧、营养积累、吸收污染物、滞尘、森林防护、生物多样性保护、游憩、教育、文化、美化环境等能力指标比较;二是种植辅助决策,如需要达到某种效益的效果,根据现有的数据分析结果以及希望的效益水平具体值进行生态公益林的林种类型以及种植面积的推荐,辅助决策。

2.1.6 知识管理模块 此模块对管理林分知识库,进行林分知识的添加、删除和修改,以及评价模型名称、表达式的设定。

2.2 系统体系结构设计

基于数据挖掘的生态公益林效益评价 DSS 采用数据挖掘与三库衔接的 DSS 系统结构。数据挖掘的核心模块(数据挖掘引擎)独立于模型库,这有利于充分发挥数据挖掘的功能与作用,也有利于保持现有系统的稳定^[11]。由人机交互界面、解释机、模型库、知识库、数据挖掘引擎、数据库等组成,其体系结构如图 2

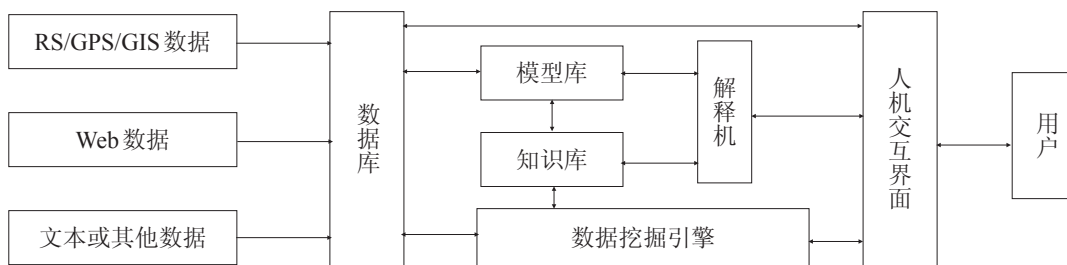


图2 基于数据挖掘决策支持系统体系结构图

所示。

2.2.1 模型库设计 模型库存储辅助决策所需的各种模型,包含生态公益林效益评价的各种计算模型、单一样本效益评价模型、相同林分类型效益评价模型、所有林分类型效益评价模型、林分类型生态效益能力比较模型等。对影响生态公益林效益各指标能力及生态公益林效益因素进行综合分析,帮助决策者了解和预测现

有林分类型效益和所体现的价值。

(1)生态公益林效益评价计算模型。采用国家林业局发布的LY/T 1721—2008森林生态系统服务功能评估规范中的森林生态系统服务功能价值量评估计算公式、参数设置及方法,即 $U_{调}$ 、 $U_{水质}$ 、 $U_{肥}$ 、 $U_{固土}$ 、 $U_{碳}$ 、 $U_{氧}$ 、 $U_{营养}$ 、 $U_{负离子}$ 、 $U_{二氧化硫}$ 、 $U_{氟}$ 、 $U_{氮氧化物}$ 、 $U_{重金属}$ 、 $U_{噪音}$ 、 $U_{滞尘}$ 、 $U_{保护}$ 、 $U_{生物}$ 、 $U_{游憩}$ 、 $U_{社会}$ 等各种计算模型。其中,森林游憩

和社会效益评价的方法主要采用市场价值法和条件价值法。

(2)单一样本效益评价模型。单个林分数据的各个效益指标价值总和： $U_{单}=U_{调}+U_{水质}+U_{肥}+U_{固土}+U_{碳}+U_{氧}+U_{营养}+U_{负离子}+U_{二氧化硫}+U_{氟}+U_{氮氧化物}+U_{重金属}+U_{噪音}+U_{滞尘}+U_{保护}+U_{生物}+U_{游憩}+U_{社会}$ ， $U_{单}$ 为单个样本生态公益林效益总价值，单位：元/年；每个生态效益价值在单个样本所提供的生态效益价值中所占的比例值，例如： $h_{调}=U_{调}/U_{单}$ ， $h_{调}$ 即为单一样本调节水量价值在样本所提供的生态效益价值中所占的比例。

(3)相同林分生态效益价值评价模型。对相同林分类型的各个生态效益指标及所具备的生态效益价值评价模型：

$U_{同} = \sum_{i=1}^n U_{i同}$ ， $U_{同}$ 为相同林分类型生态效益总价值，单位：元/年。

(4)所有林分生态效益评价子模型。对所有林分类型的各个生态效益指标及所具备的生态效益价值进行评价分析：

$U_{总} = \sum_{i=1}^n U_i$ ， $U_{总}$ 为所有林分类型森林生态系统服务功能总价值，单位：元/年。

(5)生态公益林效益林种比较模型。每个公益林林种所具备生态效益能力不一样，对每个公益林林种的各项生态效益指标进行排序比较观测。首先，对每个公益林林种所具备的生态效益进行评价，即计算 $U_{调}$ 、 $U_{水质}$ 、 $U_{肥}$ 、 $U_{固土}$ 、 $U_{碳}$ 、 $U_{氧}$ 、 $U_{营养}$ 、 $U_{负离子}$ 、 $U_{二氧化硫}$ 、 $U_{氟}$ 、 $U_{氮氧化物}$ 、 $U_{重金属}$ 、 $U_{噪音}$ 、 $U_{滞尘}$ 、 $U_{保护}$ 、 $U_{生物}$ 、 $U_{游憩}$ 、 $U_{社会}$ 等，然后进行比较排序。例如：调节水源能力：马尾松>竹林>阔叶林>...，固土能力：竹林>阔叶林>马尾松>...。

(6)公益林辅助种植模型。根据生态公益林效益的知识特点，将生态公益林效益知识分为常识性知识和专家经验知识。通过对这两类知识的分析，建立生态公益林效益辅助种植决策的基本模型。首先，找到数据库中达到目标生态效益相对应的最高林种类型，然后计算出所需要种植的面积，即种植面积=所需要达到的生态效益期望值/所选择林种的生态效益能力。表示为： $A=S/Q_{MAX}$ ， S 为所需要达到的生态效益期望值， Q_{MAX} 为所选择生态效益指标下能力值最高的林种， A 为达到生态效益期望值的种植面积。

2.2.2 知识库设计 知识库是决策支持系统的核心之一。系统知识库中知识以知识单元的形式为主，规则体由IF...THEN...产生式规则构成。为了表达知识的

模糊性与不确定性，引进了模糊逻辑，对于每一个结论，针对不同情况设计了不同的隶属关系，系统可根据用户要求给出相应的隶属度。

(1)知识来源。系统的知识源主要包括4个方面：国内外公益林的历史文献及相关资料等进行归纳、总结、整理形成的知识；领域专家的知识 and 经验；参与公益林实际数据测绘、调查、研究并与其他专家知识源进行补充、修改形成的知识；通过数据挖掘技术对数据库中进行挖掘获得知识。

(2)数据挖掘获取知识。a.聚类分析。将数据库中所包含所有林分类型数据放入数据挖掘引擎模型中，进行聚类分析^[12]，然后得出林分类型的聚类结果。其聚类分析过程如下：

CLUSTER (2-1)

$Q_{调}$ ， $Q_{生}$ ， $Q_{固土}$ ， $Q_{固碳}$ ， $Q_{氧}$ ， $Q_{营养}$ ， $Q_{二氧化碳}$ ， $Q_{滞尘}$ ， $Q_{氟化物}$ ， $Q_{氮氧化物}$

MISSING=LISTWISE (2-2)

CRITERIA=CLUSTER(4)MXITER(x)

CONVERGE(0) (2-3)

METHOD=KMEANS(NOUPDATE) (2-4)

PRINT ID(AB002)INITIAL (2-5)

过程解释：(2-1)调用聚类分析，根据 $Q_{调}$ ， $Q_{生}$ ， $Q_{固土}$ ， $Q_{固碳}$ ， $Q_{氧}$ ， $Q_{营养}$ ， $Q_{二氧化碳}$ ， $Q_{滞尘}$ ， $Q_{氟化物}$ ， $Q_{氮氧化物}$ 对所有林分类型进行聚类分析；(2-2)剔除分析变量中任一有缺失值的观测量；(2-3)要求迭代结果将观测值分为4类，最大迭代值为 x (x 为林分类型数量)，收敛判断为0。即迭代到 x 次或收敛到0次，迭代停止；(2-4)使用给定聚类分析方法进行聚类，但聚类过程中不改变类中心；(2-5)制定按指标变量为变量(AB002)输出聚类结果。系统将聚类结果调入每个林分类型的知识库中，供管理者决策时参考。

b.关联规则挖掘。将数据库中所包含所有林分类型数据放入数据挖掘引擎模型中，进行关联规则挖掘，然后得出生态公益林效益能力值(即系统中设定的 Q 值)之间的关联规则结果。例如：分析所有林分类型的 $Q_{调}$ 和 $Q_{固土}$ 之间的关联规则，给定最小支持度 $\alpha=0.3$ ，最小置信度 $\beta=0.6$ ，采用基于Apriori的模糊关联规则算法^[12]对 $Q_{调}$ 和 $Q_{固土}$ 的数值进行关联规则数据挖掘，得出支持度 support 和置信度 confident。如果得出的支持度 support>0.3，置信度 confident>0.6，即可以得出结论 $Q_{调}$ 和 $Q_{固土}$ 之间存在关联，否则 $Q_{调}$ 和 $Q_{固土}$ 之间不存在关联。同样，可以得出 $Q_{调}$ ， $Q_{生}$ ， $Q_{固土}$ ， $Q_{固碳}$ ， $Q_{氧}$ ， $Q_{营养}$ ， $Q_{二氧化碳}$ ， $Q_{滞尘}$ ， $Q_{氟化物}$ 等生态公益林效益各指标的能力值之间的关联规则。将得出的各条关联规则结论以知识

的形式存储到知识库中。

c.知识表现方式。通过数据挖掘对数据库中林分数据挖掘获得的知识用知识的框架表示法表示;使用辅助决策解决问题时,使用产生式规则的知识表示法。

2.2.3 解释机与推理机设计

(1)解释机。解释机将模型库中的模型及知识库中知识解释为用户能够理解的自然语言。

(2)推理机。智能系统的推理包括2个基本策略:推理方法和推理的控制机制。

系统主要采用基于规则的演绎推理方法和正向推理机制^[13]。系统中公益林辅助种植推理的基本推理形式如图3所示。

首先,通过人机对话确定选择生态效益中的指标类别选项以及输入需要达到的目标值,然后根据数据库中的数据得到数据库中该指标类别下能力最强的林

分类型,以及达到该目标值需要种植的指标类别能力最强林种的面积,并从林分类别知识库中调出该林分类型的相关文本资料及图片。

2.2.4 数据库设计 数据库数据为RS/GPS/GIS数据、Web数据以及其他效益评价相关的历史数据、空间数据和相关指标与参数,包括林分生态指标数据库、评价参数数据库、林分品种数据库、地点数据库等标准库和用于储存推理过程中产生的中间结果和最终结果数据库等。系统采用关系型数据库Oracle,将数据库表示为表的集合,包括林分基本信息数据结构表、生态效益评价指标参数表、林分知识管理表、生态效益评价方法表等。

模型库与数据库有密切的关系。模型所需的数据来自于数据库,而模型设置的值和输出的数据将返回数据库中。在DSS中所有的模型利用同样的数据库,

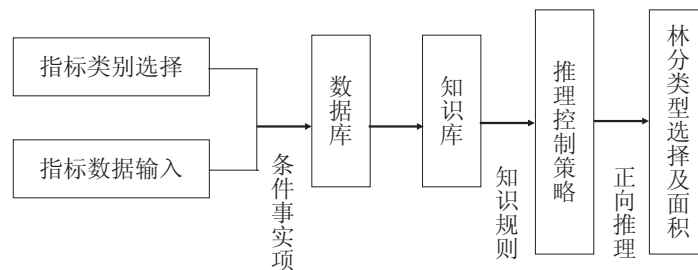


图3 生态公益林效益辅助种植正向推理

以保证存取数据和更新的一致性^[14]。

2.2.5 人机界面设计 系统采用从左到右多级菜单形式,包括系统启动登录界面、用户管理界面、数据录入界面、生态效益评价相关参数录入界面、单一样本分析界面、相同林分类型分析界面、所有林分类型分析界面、聚类分析界面、关联规则挖掘界面、林分生态服务功能指标能力比较界面、种植辅助决策界面、模型管理界面、林分类型知识库管理界面等设计。

3 系统实现

系统开发工具为Eclipse和应用服务器Tomcat,采用MVC模式,基于J2EE平台上的Struts框架+JavaBean+DAO的体系架构,分为如图4所示的5个层次^[15]。

此架构便于系统管理,同时通过分层次设计,降低了数据在各层之间的冗余度和数据耦合性,提高了系统的可维护性。此外,由于Java的跨平台特性,使得系统具有很强的平台通用性,实现了在线人机交互、数据分析、数据挖掘、辅助决策与可视化输出等功能。

4 结论

生态公益林效益评价需要科学复杂的统计计算方

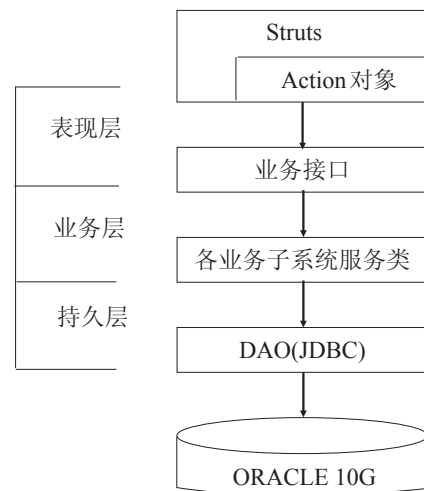


图4 系统架构图

法以及长期的定位监测,难度较大,其研究是一项极其复杂的系统工程。科学的生态公益林效益评价决策支持系统在林业管理中有着极为重要的作用,有利于提高林业管理决策的有效性和科学化。系统将数据挖掘技术融入到生态公益林效益评价决策支持系统中,实现了以模型库与知识库为中心,辅以数据库、方法库与

现代网络技术的生态公益林效益评价决策支持系统,具有较强的分析、统计、预测与决策功能,为生态公益林管理者和决策者提供可靠的信息技术支持、为生态公益林效益评价决策支持提供了参考,以适应生态公益林管理的发展需要。

5 讨论

本研究中生态公益林效益评价指标体系的构成借鉴了 Constanza 提出的 17 个生态系统服务类型,体现生态公益林的生态、社会需求和可持续发展的主体功能的 10 项效益评价;同时,各项效益评价的计量模型与方法主要采用 LY/T 1721—2008 的评估公式、参数设置及方法。其中,森林游憩和社会效益评价的方法主要采用市场价值法和条件价值法。但由于公益林生态系统服务功能内涵的广泛性、服务功能的间接性、多样性和动态性,同时生态公益林效益评价需要系统、可靠的评价基础数据,因此在具体的区域生态公益林系统效益评价的实际应用中,研究者可能从不同的角度构建不同的效益评价指标体系,以及不同的效益评价的计量模型与方法。此时,只需要对本系统中的模型库的生态公益林效益评价计算模型进行相应设计即可。

参考文献

[1] 黎军.基于 GIS 的区域生态环境建设规划决策支持系统的研究[D].广西:广西大学,2002:2-13.

- [2] 邓苏,张维明,黄宏斌.决策支持系统[M].北京:电子工业出版社,2009:1-87.
- [3] Costanza R. The value of the World's ecosystem services and natural capital[J].Nature,1997,387:253-260.
- [4] Lubchenco J. Entering the century of the environment: A new social contract for science[J].Science,1998,279:279-491.
- [5] 陈建平,陈根年.陕西佛坪自然保护区森林生态服务价值测评[J].陕西师范大学学报:自然科学版,2004,32(3):107-111.
- [6] 周晓峰,张洪军.森林生态系统服务功能[A].见:生态系统服务功能研究[C].北京:气象出版社,2000:34-66.
- [7] 袁正科,林柏.湖南鹰嘴界自然保护区综合科学考察报[R].湖南省会同县人民政府,湖南林业科学院,2003:52-53.
- [8] Chavas, Jean-Paul. On the commodity value of Travel time in reactional activities[J].Applied Economics,1989,21:711-722.
- [9] 国家林业局.中华人民共和国林业行业标准 LY/T 1721—2008 森林生态系统服务功能评估规范[S].2008:4.
- [10] 丁建华,张平,邵文照,等.网络化大樱桃智能管理专家系统的建立与应用[J].计算机与农业,2003(8):9-11.
- [11] 蒲晓湘,刘文才.联机分析挖掘技术现状和发展[J].重庆大学学报,2004,27(3):36-38.
- [12] Jiawei Han.数据挖掘:概念和技术[M].范明,孟小峰译.北京:机械工业出版社,2007,3:146-173,251-299.
- [13] 张仰森,黄改娟.人工智能教程[M].北京:高度教育出版社,2008,3:72-116.
- [14] 李化,杨盘洪.玉米专家系统知识库的构建[J].太原理工大学学报,2003(3):116-118.
- [15] 梁立新.项目实践精解:基于 Struts-Spring-Hibernate 的 Java 应用开发[M].北京:电子工业出版社,2006:36-105.