文章编号: 1000-6788(2012)08-1777-07 中图分类号: TP181 文献标志码: A

强化学习在基于多主体模型决策支持系统中的应用 ——— 以湖泊水环境决策支持系统为例

倪建军 [1], 刘明华 [1], 任 黎 [2], 张传标 [1]

(1. 河海大学 计算机与信息学院 (常州), 常州 213022; 2. 河海大学 水文水资源学院, 南京 210098)

摘 要 利用研究复杂系统和多主体 (multi-agent) 建模的相关知识与方法, 将湖泊水环境中的各种实体, 如政府、排污企业以及各种水生生物等抽象为具有一定智能的主体, 建立湖泊水环境智能决策支持系统. 并将强化学习方法应用到智能决策支持系统中, 实现湖泊水污染的智能预测与预警. 最后, 以太湖流域为应用背景, 进行了初步的仿真实验, 实验结果验证了该方法的有效性.

关键词 强化学习; 决策支持系统; 多主体建模; 水污染治理

Reinforcement learning for DSS based on multi-agent model: A case of lake water environment DSS

NI Jian-jun[1], LIU Ming-hua[1], REN Li[2], ZHANG Chuan-biao[1]

(1. College of Computer and Information, Hohai University, Changzhou 213022, China;
2. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

Abstract The lake water environmental problem has been more and more serious. It is a very important subject to find a more effective way of water pollution control. In this paper, the lake water environment decision support system (DSS) is set up, using the knowledge and methods of complex system and multi-agent modeling. The various entities in the lake water environment (such as government, polluting enterprise and a lot of aquatic organisms) are abstracted as the agents, which have some certain intelligence. A method based on reinforcement learning is proposed to achieve the intelligent prediction and warning of the lake water pollution. At last, a preliminary simulation experiment is conducted on the application of Taihu Lake basin. The experiment results show that the proposed method is effective.

Keywords reinforcement learning; decision support system; multi-agent modeling; water pollution control

1 引言

当前湖泊环境问题日益严重, 尤其位于我国东部平原的大部分浅水型湖泊以及城市附近的小型湖泊, 湖泊的水污染问题更加突出, 已成为我国目前急待解决的环境问题之一. 如何建立兼具高效和可操作性的湖泊水污染治理决策支持系统具有十分重要的意义 [1−3].

近年来, 智能决策问题引起了理论研究和实践领域的关注, 实现智能决策的核心问题是建立系统模型. 目前国内外专家在智能决策支持系统领域进行了大量卓有成效的研究工作, 如 Sasikumar 和 Mujumdar[4] 研究了河流水质管理的多目标模糊优化模型, 程春田和欧春平 [5] 针对水库防洪调度系统的特点, 提出了有冲突的多目标协商决策模型, Huang 等 [6] 为农村生态环境管理设计了一个基于农户和作物的两层的智能决策系统, 等. 从现有的研究情况来看, 这些模型和算法虽然各有优缺点, 但针对复杂系统的决策问题, 都存在着模型复杂、数据丰富但知识贫乏等不足, 不能很好实现数据分析、趋势预测以及智能决策的目的. 由于湖泊水环境系统是一个无法重现的复杂系统, 它是由湖泊水环境系统中各实体的相互作用而体现出来的一种水环境

收稿日期: 2010-05-18
资助项目: 河海大学常州校区创新基金 (XZX/09B002-02); 河海大学自然科学基金 (2009423111)
作者简介: 倪建军 (1978–), 男, 安徽黄山人, 博士, 副教授, 硕士生导师, 研究方向: 复杂系统控制与决策, 智能计算; 刘明华 (1986–), 女, 天津人, 硕士研究生, 研究方向: 复杂系统控制与决策; 任黎 (1978–), 女, 江苏常熟人, 博士, 讲师, 研究方向: 湖泊水环境控制与决策, 水环境生态系统评价等; 张传标 (1983–), 男, 江苏徐州人, 硕士研究生, 研究方向: 智能计算.

状态, 运用传统建模的方法很难模拟出实体间相互的关系, 必须应用复杂系统建模的相关理论和方法, 目前在湖泊水环境系统中实现智能决策的研究成果还较少 [7−8].

本文以研究复杂系统的相关理论为基础, 针对湖泊水环境复杂系统, 建立基于多主体 (agent) 模型的智能决策支持系统, 并在该系统中引入强化学习算法, 通过决策类 agent 对环境的感知、分析、学习, 做出水环境演化的趋势预测并给出相关的治理决策. 最后, 以太湖流域水污染治理为例, 进行仿真实验, 仿真结果符合客观实际, 验证了该方法的有效性.

2 基于多 agent 模型的智能决策支持系统

2.1 基于多 agent 的建模理论与方法

目前 agent 已成为一个具有普遍意义的概念, 智能 agent 具有对外界环境的自适应等特性, 以及运用自身知识对问题进行处理、自学习和与外界协同工作等能力. 基于多 agent 系统的建模方法, 是 20 世纪 70 年代随着计算机科学中人工智能的出现而发展起来的, 并随着复杂系统理论研究的深入, 逐渐拓展到对社会、经济和环境问题的研究, 体现了社会经济学、计算机模拟以及基于多 agent 复杂系统理论与技术的融合. 相对于传统的系统建模方法, 基于多 agent 的系统建模有如下突出优势: ①它能更好地表征社会、经济、环境系统的复杂性、适应性. ②它采用自下而上的模拟方法, 注重对系统中微观主体行为的模拟, 通过微观主体价值、决策、行为等驱动, 更接近现实世界. ③它借鉴了计算机科学其他微观模拟方法并有所发展, 具有较强的创新能力. 它还扩充了对主体偏好、决策、计划、作用关系的研究. ④它打破了传统的经济学完全理性的概念, 可以考察不确定条件下具有不完全信息的主体决策特征, 也可考察社会关系和体制在系统发展中的作用 [9−11]. 基于多 agent 的系统模型如图 1 所示.

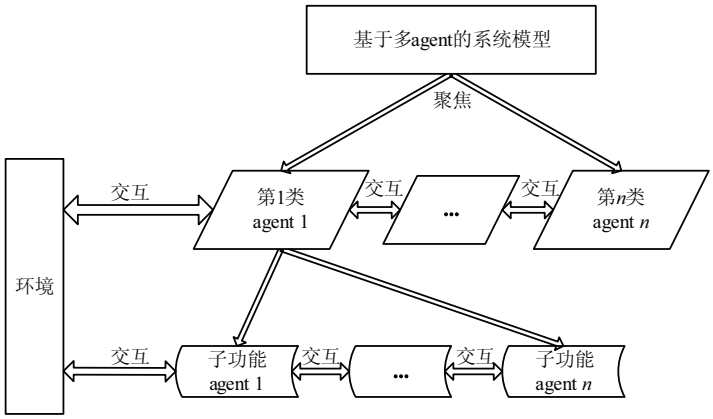

图 1 基于多 agent 的系统模型

根据以上分析可知, 基于多 agent 的建模方法对于那些具有不同行为特征参与者及其关系的系统, 具有较强的解释能力. 目前该方法还主要是用于社会、经济、生态、军事等复杂系统的建模与仿真研究, 将该方法应用于实际的决策支持系统的文献还很少.

2.2 基于多 agent 模型的智能决策支持系统框架

智能决策实质上是一个动态优选过程, 即从若干个备选方案中根据各个备选方案的多个指标数据, 选择最优或最令人满意的解的过程.

一般认为 agent 的结构由模型库、方法库和知识库组成. 在多 agent 的系统模型中, 每一类 agent 都具有决策的功能, 各类 agent 通过其内部结构向环境提取信息, 对信息进行分析综合, 做出决策, 最后将决策作用于环境. 而在基于多 agent 的智能决策支持系统中, 会有一类主决策 agent, 它与其他 agent 不断进行信息交互, 干涉其他 agent 的具体行为. 这类 agent 的决策将代表着决策支持系统的最终决策并对环境产生决定性的影响, 最后, 决策支持系统中的主决策类 agent 将根据环境的变化情况, 对决策方案进行调整. 基于多 agent 模型的智能决策支持系统的一般框架如图 2 所示.

由图 2 可知, 主决策类 agent 从环境中提取信息, 并与其他 agent 进行信息交互, 将有用信息装入知识库中, 建立模型库, 利用方法库对信息进行综合分析, 最终做出决策, 再将决策作用于环境.

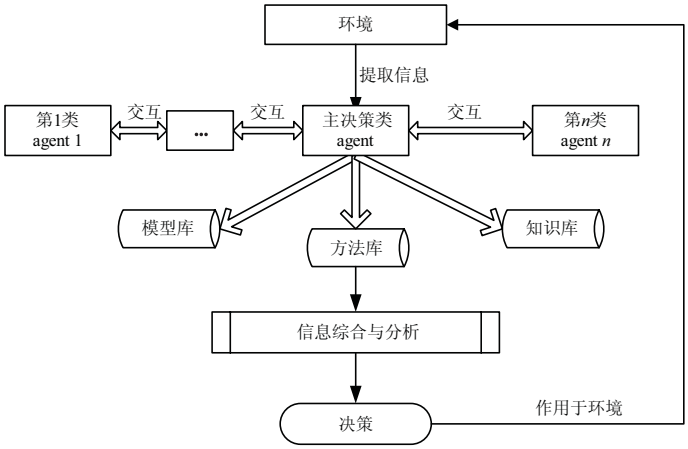

图 2　基于多 agent 模型的智能决策支持系统框架

3　强化学习在智能决策支持系统中的应用

3.1　基于强化学习的 agent 决策结构

强化学习介于监督式学习和无监督式学习之间, 其自学习和在线学习的特点使其成为机器学习研究的一个重要分支. 强化学习方法通过与环境的即时交互来获得环境的状态信息, 并通过反馈信号对所采取的行动进行评价, 通过不断的试错和改进, 从而学习到最优的策略 [12−13].

个体 agent 应用强化学习做出决策的基本原理是如果 agent 的某种策略导致环境的反馈为正反馈, 那么 agent 以后会增大对这种策略选择的概率, 反之则会减少对这种策略选择的概率. 个体 agent 应用强化学习的决策结构主要由感知器 (A)、学习器 (L) 和决策选择器 (P) 三个模块组成 [14]. 个体 agent 通过感知器 (A) 把对环境 (社会环境和自然环境) 当前状态 s 的认知转化为其内部的知识 k; 决策选择器 (P) 根据当前对环境的认识以及所拥有的策略知识, 做出决策 d, 并作用于环境; 环境在决策 d 的作用下, 状态从 s 变化为 s', 并给出反馈 r(即对 agent 的行为做出奖赏或惩罚); agent 学习器 (L) 根据环境的反馈值 r 以及内部知识 k, 更新 agent 的策略知识.

3.2　基于强化学习的智能决策支持系统工作流程

在基于多 agent 模型的智能决策支持系统中, 应用强化学习, 可以实现决策的智能化. 基于强化学习的智能决策支持系统工作流程如图 3 所示.

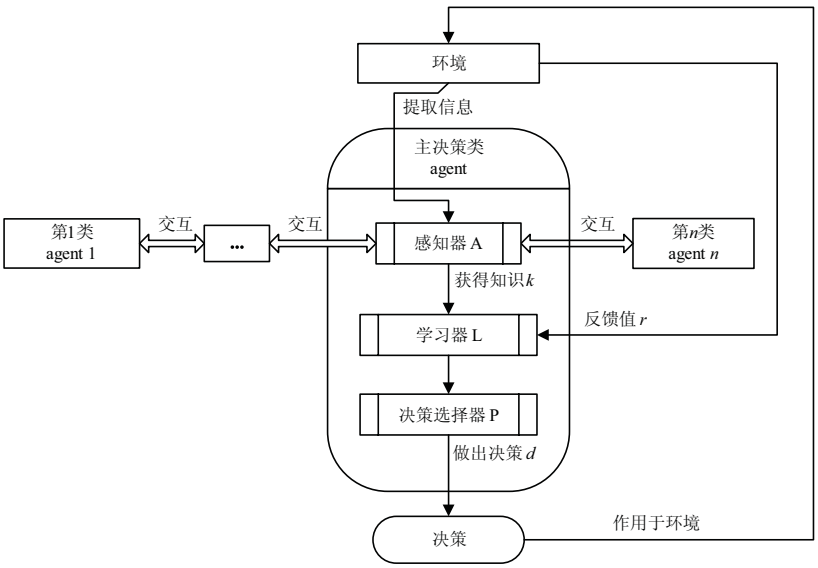

图 3　基于强化学习的智能决策支持系统工作流程

在图 3 中, 感知器 (A) 主要是 agent 对社会环境和自然环境的掌握, 这里用 x 来表示 agent 对环境的总

体认知, 用下式表示:

$$x = \{x_1, x_2, \cdots, x_n\}, \quad n \geq 1 \tag{1}$$

其中 x_n 表示 agent 内部与研究内容相关的具体属性. k 为具体的知识, 可以表示为:

$$k = \{k_1, k_2, \cdots, k_n\}, \quad n \geq 1 \tag{2}$$

学习器 (L) 的学习材料主要来源于两个方面, 包括感知器 (A) 形成的具体知识 k 和环境的反馈值 r, L 可以表示为: L= $\{r, k\}$. 其中环境反馈值 r 表示为:

$$r = \{r_1, r_2, \cdots, r_n\}, \quad n \geq 1 \tag{3}$$

L 形成后又同时影响 agent 的感知器 (A) 与决策选择器 (P).

决策选择器 (P) 主要受到感知器 (A) 和学习器 (L) 的影响, 可表示为: $\mathrm{P} = \{\mathrm{L}, \mathrm{A}\}$. 每个 agent 通过感知和学习后所形成的决策空间用 d 来表示, 它是一系列决策的集合, 表示为:

$$d = \{d_1, d_2, \cdots, d_n\}, \quad n \geq 1 \tag{4}$$

4 仿真实验

本文以太湖流域为研究背景进行仿真实验, 针对太湖流域水污染治理问题建立以多 agent 智能决策支持系统. 将太湖水环境中的各种实体, 如政府、排污企业以及各种水生生物等抽象为具有一定智能的 agent, 将强化学习算法应用到系统之中, 各类 agent 通过与环境以及其他 agent 的交互, 对环境产生影响. 政府 agent 在这个决策支持系统中担当主决策 agent 的角色, 从环境获取信息, 并利用各种模型对水环境系统的演化作出预测, 根据环境的反馈, 给出合理的水污染治理方案.

4.1 太湖水环境多 agent 系统模型

在针对太湖水环境污染治理的问题中, 主要涉及政府、排污企业和水生生物几类对象, 为了简化实验系统的设计, 本文在建立多 agent 系统模型时仅以政府 agent, 排污企业 agent 和水生生物 agent 作为研究对象, 其中政府 agent 为主决策 agent, 对最终决策起决定性作用, 它的知识库数据来源于两个方面, 包括从太湖水环境中提取的信息和与排污企业 agent 和水生生物 agent 交互的信息. 为了确定各 agent 的模型库, 首先对每类 agent 的内部相关属性进行量化处理, 具体量化过程如下 [15]:

1) 对政府 agent 中的湖泊水环境治理政策驱动力属性进行量化处理, 选择该地区具有代表性的 5 个指标为变量因子: 总人口数 x_1、地方财政收入 x_2、第三产业产值 x_3、农渔牧业产值 x_4、政策力度 x_5.

2) 为了便于分析企业行为对湖泊水环境的影响, 需对其内部污水处理损失费用进行量化处理, 这里以印染企业为例, 建立企业环境投入和环境行为影响因素相互关系模型. 选择的指标因子为: 企业体制 x_1、企业规模 x_2、企业利润 x_3、废气排放量 x_4、COD 排放量 x_5、固废排放量 x_6.

3) 为了便于分析水生物对湖泊水环境的影响 (这里主要考虑水质富营养化问题), 需对其内部水生植物中所含矿物质元素浓度进行量化处理, 选择的指标因子为: COD 浓度 x_1、TN 浓度 x_2、TP 浓度 x_3.

4.2 各类 agent 在强化学习算法下的决策过程

在基于多 agent 模型的湖泊水环境智能决策支持系统中, 政府 agent 作为主决策类, 需要重点分析, 同时排污企业 agent 和水生生物 agent 的决策与整个系统的决策过程是紧密联系, 互相影响的, 因此需要对各类 agent 的决策过程进行分析. 基于强化学习的各类 agent 的决策过程如下:

(一)、感知器 (A) 从太湖水环境提取信息得到具体知识 k

不失一般性, 这里简化知识的获取过程, 定义知识 k 的基本公式为: $k = x_1 + x_2 + \cdots + x_n$, 应用到各类 agent 模型中的具体计算公式为:

政府 agent 治理政策驱动力 k_G 计算公式为:

$$k_G = x_0 + ax_1 + bx_2 + cx_3 + dx_4 + ex_5 \tag{5}$$

排污企业 agent 内部污水处理损失费 k_E 计算公式为:

$$k_E = x_0 + ax_1 + bx_2 + cx_3 + dx_4 + ex_5 + fx_6 \tag{6}$$

水生生物 agent 植物富营养化矿物质总浓度 k_A 的计算公式为:

$$k_A = x_1 + x_2 + x_3 \tag{7}$$

在上述各式中 x_0 是修正量, $\{a, b, c, d, e, f\}$ 是各变量的权重.

参考文献 [14], 将有关数据代入上述公式 (5)–(7) 进行仿真实验, 可计算出 1996–2000 年太湖水环境中各类 agent 感知器 (A) 中的 k 值, 如表 1 所示.

(二)、学习器 (L) 从感知器 (A) 中获得信息

由概念模型可知, L= $\{r, k\}$, 其中 r 为反馈值, k 为湖泊多 agent 系统中各类 agent 根据对环境的感知, 得到的具体知识. 反馈值 r 可用湖泊水环境改善的综合效益来计算:

$$r_i = \begin{cases} r_G = y_G/(y_G + y_E + y_A) \\ r_E = y_E/(y_G + y_E + y_A) \\ r_A = y_A/(y_G + y_E + y_A) \end{cases} \tag{8}$$

式中, r 表示某类 agent 对水环境作用后的反馈值; y 为对每类 agent 的量化值进行均一化处理所得到的值, 即: $y_i = k_i/\max(k_i)$, 在实际决策过程中, 反馈值 r 都为正数, 通过比较 r 值的大小来进行决策, r 值越大, 对湖泊水环境中各类 agent 决策的影响作用越大. 由公式 (8) 得出各类 agent 学习器 (L) 中 r 的具体值如表 2 所示.

表 1　1996–2000 年各类 agent 的 k 值

年份	k 值		
	k_G	k_E	k_A
1996	7	0.895	7.43
1997	8	0.618	7.52
1998	15	0.967	6.71
1999	12	0.793	7.69
2000	13	0.612	7.70

表 2　1996–2000 年各类 agent 的 r 值

年份	r 值		
	r_G	r_E	r_A
1996	0.20	0.39	0.41
1997	0.25	0.28	0.47
1998	0.35	0.35	0.3
1999	0.31	0.32	0.37
2000	0.35	0.25	0.40

(三)、决策选择器 (P) 从学习器 (L) 中获取信息

这里将决策选择空间表示为: $d = \{d_G, d_E, d_A\}$, 其中 d_G 起主导作用, 分析 d_G、d_E、d_A 三个值, 如果 d 值降低, 则政府 agent 需要对自身内部相关属性的大小做出相应的调整, 以此对排污企业 agent 和水生生物 agent 进行干涉 (例如调整排污费的收取比例等等), 以便使湖泊水环境得到进一步改善. 由强化学习概念模型中 P = {L, A} 得出最终决策值 d 为: $d_i = (r_i + k_i)/k_i$, 将 k 值统一到相同数量级后, 该公式变为: $d_i = (r_i + y_i)/y_i$. 具体计算公式为:

$$d_i = \begin{cases} d_G = (r_G + y_G)/y_G \\ d_E = (r_E + y_E)/y_E \\ d_A = (r_A + y_A)/y_A \end{cases} \tag{9}$$

由公式 (9) 得出 1996–2000 年各类 agent 决策选择器 (P) 中 d 的具体值, 见表 3 所示, d 值的变化趋势如图 4 所示.

表 3　1996–2000 年各类 agent 的 d 值

年份	d 值		
	d_G	d_E	d_A
1996	1.42	1.41	1.43
1997	1.47	1.44	1.48
1998	1.35	1.35	1.34
1999	1.69	1.38	1.37
2000	1.40	1.39	1.40

基于多 agent 模型的决策支持系统通过政府 agent 根据决策选择器 (P) 中决策值 d 的变化, 对太湖流域排污企业进行管理. 政府 agent 通过感知器 (A) 对现有环境进行感知、学习, 并与其他 agent 交互信息, 利用反馈值 r 不断修正决策值 d, 并给出合理的水污染治理决策方案. 由图 4 可以看出, 1998 年前排污企业 agent 的决策值 d_E 一直略低于其他两类 agent 的决策值, 1998 年后政府 agent 决策值 d_G 大幅度提高, 排污企业 agent 的决策值 d_E 也随之有所提高, 说明政府 agent 通过调整其内部政策驱动力属性值, 加强与其他

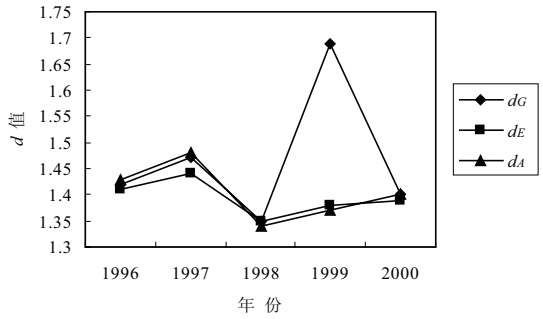

图 4 1996–2000 年各类 agent 的 d 值变化趋势图

agent 信息交互, 对太湖流域排污企业进行有效整治管理. 事实表明, 随着工业生产的迅速发展, 有些企业为追求利润最大化, 对太湖水环境的保护采取消极态度, 以致太湖整体环境遭到破坏. 由于太湖水环境遭受的污染程度愈加严重, 国家环保局 1998 年制订了 "零点行动" 方案, 并关闭了所有向太湖水域排放污水的污染源. 不久 "零点行动" 方案发挥功效, 太湖水环境得到了有效改善. 上述实际的水环境决策与演化过程和本实验结果基本一致, 从而有效说明了该方法适用于湖泊水环境智能决策与管理.

5 结束语

湖泊水环境系统是一个复杂系统, 有企业、政府等社会主体的参与, 是一个社会、经济、环境系统的集合, 对这样系统的建模, 并进行水污染灾害预测与预警, 必须进行跨学科的综合研究. 本文首先通过 agent 概念来抽象现实中繁多的客观主体, 建立基于多 agent 模型的湖泊水环境智能决策支持系统, 使得模型更加简单化、具体化, 可操作性提高, 并将强化学习方法应用到多 agent 决策支持系统中, 可以实现决策值的智能计算, 通过对主决策 agent 决策值的分析, 进行预测预警, 并做出合理决策. 同时可以根据以往的数据推测出现有决策是否合理, 是否需要做出调整, 从而大大提高了决策系统的智能性.

综上所述, 强化学习应用到多 agent 智能决策支持系统是可行的, 将这种方法应用在湖泊水环境智能决策支持系统中具有重要的理论研究意义和实际推广价值, 下一步的工作重点是进一步完善湖泊水环境系统的多 agent 模型, 并对智能决策支持系统进行改进和升级.

参考文献

[1] 刘永, 郭怀成, 范英英, 等. 湖泊生态系统动力学模型研究进展 [J]. 应用生态学报, 2005, 16(6): 1169–1175.
Liu Y, Guo H C, Fan Y Y, et al. Research advance on lake ecosystem dynamic models[J]. Chinese Journal of Applied Ecology, 2005, 16(6): 1169–1175.

[2] Ni J J, Zhang C B, Ren L. An intelligent decision support system of lake water pollution control based on multi-agent model[C]// Proceeding of International Conference on Computational Intelligence and Security, New Jersey: IEEE Computer Society, 2009: 217–221.

[3] 毛国柱, 刘永, 郭怀成, 等. 湖泊富营养化控制技术综合集成方法框架 [J]. 环境工程, 2006, 24(1): 65–67.
Mao G Z, Liu Y, Guo H C, et al. Comprehensive integration of lake eutrophication control technique[J]. Environmental Engineering, 2006, 24(1): 65–67.

[4] Sasikumar K, Mujumdar P P. Fuzzy optimization model for water quality management of a river system[J]. Journal of Water Resources Planning and Management, 1998, 124(2): 79–88.

[5] 程春田, 欧春平. 流域防洪决策支持系统集成管理 [J]. 大连理工大学学报, 2001, 41(1): 108–111.
Cheng C T, Ou C P. Integrated management of decision-support system for flood control of river basin[J]. Journal of Dalian University of Technology, 2001, 41(1): 108–111.

[6] Huang G H, Sun W, Nie X H, et al. Development of a decision-support system for rural eco-environmental management in Yongxin County, Jiangxi Province, China[J]. Environmental Modelling & Software, 2010, 25(1): 24–42.

[7] 王慧敏, 佟金萍, 马小平, 等. 基于 CAS 范式的流域水资源配置与管理及建模仿真 [J]. 系统工程理论与实践, 2005, 25(12): 119–137.
Wang H M, Tong J P, Ma X P, et al. Complex adaptive system (CAS)-based allocation and management of river basin water resource[J]. Systems Engineering — Theory & Practice, 2005, 25(12): 119–137.

[8] Tian J, Wang Y L, Li H Z, et al. DSS development and applications in China[J]. Decision Support Systems, 2007, 42(4): 2060–2077.

[9] 倪建军, 徐立中, 王建颖. 基于 CAS 理论的多 Agent 建模仿真方法研究进展 [J]. 计算机工程与科学, 2006, 28(5): 83–86.
Ni J J, Xu L Z, Wang J Y. Advances in multi-agent modeling and simulation based on the CAS theory[J]. Computer Engineering & Science, 2006, 28(5): 83–86.

[10] Monticino M, Acevedo M, Callicott B, et al. Coupled human and natural systems: A multi-agent-based approach[J]. Environmental Modelling & Software, 2007, 22(5): 656–663.

[11] 廖守亿, 戴金海. 复杂适应系统及基于 Agent 的建模与仿真方法 [J]. 系统仿真学报, 2004, 16(1): 113–117.
Liao S Y, Dai J H. Study on complex adaptive system and agent-based modeling & simulation[J]. Journal of System Simulation, 2004, 16(1): 113–117.

[12] 陈宗海, 杨志华, 王海波, 等. 从知识的表达和运用综述强化学习研究 [J]. 控制与决策, 2008, 23(9): 962–975.
Chen Z H, Yang Z H, Wang H B, et al. Overview of reinforcement learning from knowledge expression and handling[J]. Control and Decision, 2008, 23(9): 962–975.

[13] 高阳, 陈世福, 陆鑫. 强化学习研究综述 [J]. 自动化学报, 2004, 30(1): 86–100.
Gao Y, Chen S F, Lu X. Research on reinforcement learning technology: A review[J]. Acta Automatica Sinica, 2004, 30(1): 86–100.

[14] 王涛, 陈海, 白红英, 等. 基于 Agent 建模的农户土地利用行为模拟研究 —— 以陕西省米脂县孟岔村为例 [J]. 自然资源学报, 2009, 24(12): 2056–2066.
Wang T, Chen H, Bai H Y, et al. Agent-based modeling of simulation on households land-use behavior — A case of Mengcha village of Mizhi County in Shaanxi province[J]. Journal of Natural Resources, 2009, 24(12): 2056–2066.

[15] 黄贤金, 王腊春, 高超, 等. 太湖水资源水环境研究 [M]. 北京: 科学出版社, 2008.
Huang X J, Wang L C, Gao C, et al. Taihu Lake Water Resources and Water Environment Research[M]. Beijing: Science Press, 2008.