

改进型 Elman 神经网络发酵过程建模研究

刘尧猛, 马永军, 杨美艳

LIU Yao-meng, MA Yong-jun, YANG Mei-yan

天津科技大学 计算机科学与信息工程学院, 天津 300222

School of Computer Science and Information Engineering, Tianjin University of Science & Technology, Tianjin 300222, China

E-mail: liuym@tust.edu.cn

LIU Yao-meng, MA Yong-jun, YANG Mei-yan. Research on fermentation process modeling of improved Elman neural network. Computer Engineering and Applications, 2009, 45(32): 240-243.

Abstract: According to the characteristics of fermentation process and the dynamic modeling theory of modified Elman neural network, a new batch training algorithm for fermentation process modeling is proposed. The training and testing of fermentation process simulation experiments show that modified Elman neural network modeling algorithm has specialties that convergence speed is faster and generalization is better, compared with traditional BP modeling algorithm. Furthermore, the software based on the modeling algorithm can be embedded into fermentation process control system, realizes fermentation process online modeling and state parameter online estimating.

Key words: Elman; neural network; batch training; fermentation process; online modeling; online estimating

摘 要: 依据发酵过程的机理和改进的 Elman 神经网络动态建模原理, 提出了一个新的发酵过程建模分批训练算法。通过发酵过程仿真实验, 与传统的 BP 建模算法比较, 改进的 Elman 神经网络建模算法具有收敛速度快、泛化能力强等特点。此外, 利用该算法编制的软件可以内嵌到发酵过程监控系统中, 实现发酵过程在线建模与状态参量的在线预估。

关键词: Elman; 神经网络; 分批训练; 发酵过程; 在线建模; 在线预估

DOI: 10.3778/j.issn.1002-8331.2009.32.074 **文章编号:** 1002-8331(2009)32-0240-04 **文献标识码:** A **中图分类号:** TP273

1 引言

由于发酵过程具有高度的时变性、非线性等原因, 利用发酵过程的机理建立模型比较困难, 加上大多发酵过程生物量不能在线测量, 导致实现发酵过程优化控制困难。神经网络由于逼近任意连续非线性映射的特性而广泛应用于系统的辨识与控制, 其中应用于发酵过程建模和发酵参量软测量已经有相当多的应用与研究^[1-4]。

神经网络发酵过程建模普遍采用传统的 BP 神经网络算法。在利用传统的 BP 算法进行发酵过程建模研究中发现, 若引入前一时刻或前前时刻的状态参量作为输入进行训练得到的模型其泛化性能有一定的改善, 这主要是因为发酵过程具有高度的非线性和时变性, 每一时刻菌体状态不仅与当前的物理参量、生物参量和化学参量有关, 而且还与其历史的状态参量有关, 利用一个静态多层前馈神经网络对一个动态的发酵过程进行辨识, 实际上是将动态时间建模问题变为一个静态空间建模问题, 必然会出现很多问题。对于动态过程的辨识, 改进的 Elman 神经网络为动态建模提供了有效的方法^[5]。

该文利用改进的 Elman 神经网络建模原理结合发酵过程

的机理, 为发酵过程建模提出了一种新的 Elman 网络分批训练算法。

2 改进的 Elman 神经网络

Elman 网络是具有一个或多个反馈环的神经网络, 由于增加了层间或层内的反馈联结, 使得其能够表达输入与输出之间在时间上的延迟, 因此需要用动态方程来描述, 而前馈型网络仅实现非线性映射。正是因为这种反馈, 使得网络具有了记忆功能, 所以其在序列分析、系统辨识、控制等领域获得了极为广泛的应用^[6-8]。

实际应用中, 改进的 Elman 神经网络拓扑设计如图 1 所示, 神经网络设计成 3 层网络, 包括输入层、隐层和输出层, 并设计一个反馈结构层, 即隐层前一时刻的输出 $h(t-1)$ 作为输入层部分输入 $x(t)$ 。此外, 为了使 Elman 神经网络能逼近高于一阶的动态系统, 在反馈部分增加了自反馈连接, 其系数为 α , 因此 $x_l^c(t) = \alpha x_l^c(t-1) + h_l(t-1)$, 其中 $l=1, 2, \dots, q$ 。改进的 Elman 神经网络输入层到隐层的连接权为 W_{jl}^2 , 隐层到输出层的连接权为

基金项目: 天津市科技攻关重点项目 (the Tianjin Science and Technology Research Key Projects No.04310951R); 天津科技大学自然科学基金资助项目 (the Natural Science Foundation on Tianjin University of Science & Technology No.20040228)。

作者简介: 刘尧猛 (1972-), 男, 讲师, 主要研究领域为嵌入式系统应用、系统辨识; 马永军 (1971-), 男, 博士, 副教授, 主要研究领域为系统辨识、图形与图像处理; 杨美艳 (1979-), 女, 讲师, 主要研究领域为神经网络应用。

收稿日期: 2009-06-03

修回日期: 2009-07-28

W_{kj}^3 , 反馈连接层的权为 W_{jl}^1 , 其中 $j=0, 1, \dots, q, i=0, 1, \dots, m$ 。这里当 $i=0, j=0$ 且输入为 -1 时, 分别表示隐层和输出层的各神经元的阈值。

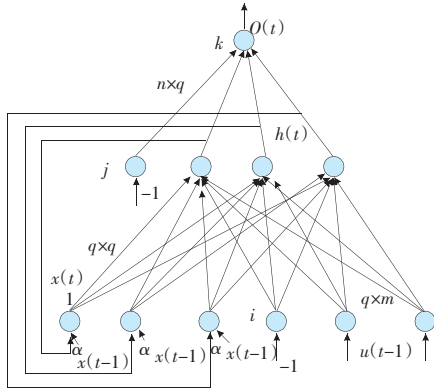


图1 动态递归神经网络的结构图

3 改进的 Elman 神经网络训练算法

3.1 Elman 神经网络算法的推导

3.1.1 隐层输出和输出层输出

$$\begin{aligned} net_h_j(t) &= \sum_{i=0}^m w_{ji}^2 u_i(t-1) + \sum_{l=1}^q w_{jl}^1 x_l^c(t) \\ h_j(t) &= f[net_h_j(t)] \end{aligned} \quad (1)$$

其中, $i=0, 1, 2, \dots, m, i=0$ 时表示隐层的阈值, $l=1, 2, \dots, q, h_j(t)$ 为隐层的输出, $x_l^c(t) = \alpha x_l^c(t-1) + h_l(t-1)$, α 为自反馈连接增益。

$net_o_k(t) = \sum_{k=0}^n w_{kj}^3 h_j(t)$, 其中 $k=0, 1, 2, \dots, n, k=0$ 时表示输出层的阈值。

$$y_k(t) = f[net_o_k(t)] \quad (2)$$

3.1.2 代价函数

$$e_k(t) = \frac{1}{2} (d_k(t) - y_k(t))^T (d_k(t) - y_k(t))$$

其中, $d_k(t)$ 和 $y_k(t)$ 分别表示 t 时刻系统的实际输出和 Elman 神经网络计算输出, 从时间步 1 工作到时间步 N , 则每个周期的总误差函数为 (N 为每个训练周期的学习步数):

$$E = \sum_{t=1}^N e_k(t) \quad (3)$$

3.1.3 隐层与输出层连接 w_{jk}^3 调整

$$w_{kj}^3(t+1) = w_{kj}^3(t) + \eta (m_c \frac{\partial E(t)}{\partial w_{kj}^3} + (1-m_c) \frac{\partial E(t-1)}{\partial w_{kj}^3}) \quad (4)$$

式(4)中的 η 为学习效率, m_c 为动量因子。根据连锁法则可知:

$$\frac{\partial E(t)}{\partial w_{kj}^3} = \frac{\partial E(t)}{\partial y_k(t)} \frac{\partial y_k(t)}{\partial net_o(t)} \frac{\partial net_o(t)}{\partial w_{kj}^3}$$

$$\text{而 } \frac{\partial E(t)}{\partial y_k(t)} = -(d_k(t) - y_k(t)), \frac{\partial net_o(t)}{\partial w_{kj}^3} = h_j(t), \frac{\partial y_k(t)}{\partial net_o(t)} =$$

$f'(net_o(t))$, 选择激发函数为 $f(x) = \frac{1}{1+e^{-x}}$, 因此

$$f'(net_o(t)) = f(net_o_k(t))(1-f(net_o_k(t)))$$

$$\frac{\partial E(t)}{\partial w_{kj}^3} = -(d_k(t) - y_k(t)) y_k(t) (1 - y_k(t)) h_j(t)$$

设 $\delta_k^o(t) = (d_k(t) - y_k(t)) y_k(t) (1 - y_k(t))$, 则

$$\frac{\partial E(t)}{\partial w_{kj}^3} = -\delta_k^o h_j(t) \quad (5)$$

3.1.4 输入层与隐层的权值 w_{ji}^2 调整

$$w_{ji}^2(t+1) = w_{ji}^2(t) + \eta (m_c \frac{\partial E(t)}{\partial w_{ji}^2} + (1-m_c) \frac{\partial E(t-1)}{\partial w_{ji}^2}) \quad (6)$$

依据隐层与输出层求梯度一样的方法,

$$\frac{\partial E(t)}{\partial w_{ji}^2} = \frac{\partial E(t)}{\partial h_j(t)} \frac{\partial h_j(t)}{\partial net_h_j(t)} \frac{\partial net_h_j(t)}{\partial w_{ji}^2}$$

$$\frac{\partial E(t)}{\partial h_j(t)} = -\sum_{k=1}^n \delta_k^o(t) w_{kj}^3$$

$$\frac{\partial h_j(t)}{\partial net_h_j(t)} = f'(net_h_j) = h_j(t)(1-h_j(t))$$

由于 $net_h_j(t) = \sum_{i=1}^m w_{ji}^2 u_i(t-1) + \sum_{l=1}^q w_{jl}^1 x_l^c(t)$, 所以

$$\frac{\partial net_h_j(t)}{\partial w_{ji}^2(t)} = \sum_{i=1}^m u_i(t-1)$$

由链式法则可知

$$\frac{\partial h_j(t)}{\partial w_{ji}^2(t)} = h_j(t)(1-h_j(t)) \sum_{i=1}^m u_i(t-1) \quad (7)$$

$$\frac{\partial E(t)}{\partial w_{ji}^2} = -\sum_{k=1}^n (\delta_k^o(t) w_{kj}^3) h_j(t)(1-h_j(t)) \sum_{i=1}^m u_i(t-1)$$

3.1.5 结构层与隐层的连接权 w_{jl}^1 调整

$$w_{jl}^1(t+1) = w_{jl}^1(t) + \eta (m_c \frac{\partial E(t)}{\partial w_{jl}^1} + (1-m_c) \frac{\partial E(t-1)}{\partial w_{jl}^1}) \quad (8)$$

根据链式规则知道 $\frac{\partial E(t)}{\partial w_{jl}^1} = -\sum_{k=1}^n (\delta_k^o(t) w_{kj}^3) \frac{\partial h_j(t)}{\partial w_{jl}^1}$, 关键是

如何求出 $\frac{\partial h_j(t)}{\partial w_{jl}^1}$ 呢? 因为计算 $\frac{\partial h_j(t)}{\partial w_{jl}^1}$ 要涉及到承接层的输入。

因为 $net_h_j(t) = \sum_{i=1}^m w_{ji}^2 u_i(t-1) + \sum_{l=1}^q w_{jl}^1 x_l^c(t)$, 所以

$$\frac{\partial net_h_j(t)}{\partial w_{jl}^1} = x_l^c(t) + \sum_{i=1}^q w_{ji}^1 \frac{\partial x_l^c(t)}{\partial w_{jl}^1}$$

若不考虑 $x_l^c(t)$ 与 w_{jl}^1 的关系, 则 $\frac{\partial x_l^c(t)}{\partial w_{jl}^1} = 0$, 则

$$\frac{\partial h_j(t)}{\partial w_{jl}^1} = \frac{\partial h_j(t)}{\partial net_h_j(t)} \cdot \frac{\partial net_h_j(t)}{\partial w_{jl}^1} = \frac{\partial h_j(t)}{\partial net_h_j(t)} x_l^c(t) \quad (9)$$

将 $x_l^c(t) = \alpha x_l^c(t-1) + h_l(t-1)$ 代入式(9), 得

$$\begin{aligned} \frac{\partial h_j(t)}{\partial net_h_j(t)} x_l^c(t) &= \frac{\partial h_j(t)}{\partial net_h_j(t)} h_l(t-1) + \alpha \frac{\partial h_j(t)}{\partial net_h_j(t)} x_l^c(t-1) \\ \frac{\partial h_j(t)}{\partial w_{jl}^1} &= h_j(t)(1-h_j(t)) h_l(t-1) + \alpha \frac{\partial h_j(t-1)}{\partial w_{jl}^1} \end{aligned} \quad (10)$$

式(10)就构成了动态的递归关系, 其中 $\frac{\partial h_j(0)}{\partial w_{jl}^1} = 0, l=0, 1, 2, \dots,$

$q, j=1, 2, \dots, q$ 。

3.2 改进的 Elman 神经网络建模软件实现

发酵过程一个生长周期通过实验获取到一批数据, 一批数

据内构成一个时间序列,批与批之间没有对应的时间序列关系。依据于 Elman 动态建模原理,在利用改进的 Elman 神经网络进行发酵过程建模训练时是按照批来训练的。

每一个批次的不同时刻都利用到了前面不同时刻的参量进行训练,每一批不同时刻都对网络的权值与阈值进行修正,并在一个发酵过程生长周期内累计均方误差,判断该误差是否小于设定的值,若是,则退出,否则,对选下一批样本进行训练。若所有批次训练结束,依然达不到所设定的值,则启动下一次训练。直到均方误差和小于给定值或训练次数大于给定最大训练次数就结束训练。具体分批训练流程如图 2 所示。

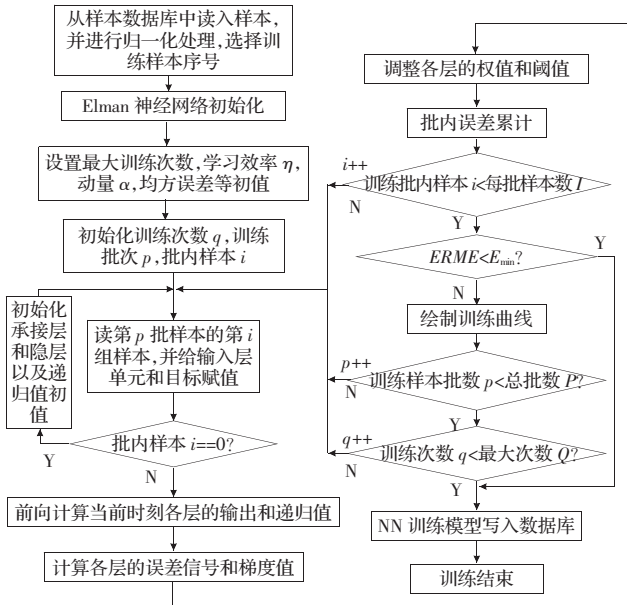


图 2 改进的 Elman 神经网络训练流程图

为了编写的建模软件具有通用性,训练与测试样本以数据库保存,即只需把不同发酵过程样本导入到样本数据库,选择不同预估参量序号,就能得到不同发酵过程对象的 Elman 神经网络模型。此外,由于编程时涉及到多维变量,考虑到初始化困难,多维变量采用多维指针变量定义。

4 仿真实验

4.1 输入与输出样本处理

样本的选择主要考虑因素是根据发酵过程的机理,确定哪些参数作为输入,哪些参数作为输出,从而确定神经网络的输入节点数和输出节点数。对于一般的发酵过程,除了一般的物理参数外,重点是化学参数和生物参数,仿真实验以聚赖氨酸发酵过程建模为例,根据发酵过程的机理,时间 t 、pH 值溶解氧 DO、菌体浓度、底物浓度 S 和产物浓度 P 可做网络的输入参量;要预估参数菌体浓度 X 、底物浓度 S 、产物浓度 P 可作为网络的输出参量。但是为了减少网络节点数,减少计算量,提高神经网络的收敛性和泛化性,每个预估参量可以单独建立一个神经网络模型,也就是对多输出网络分割,Elman 神经网络输入结点是 6 个,而输出结点是 1。编程时可以根据选择要训练的预估参量确定输出参量得到不同的预估模型,然后依据不同模型对不同参量进行预估。以聚赖氨酸发酵过程为例,主要预估参数包括菌体浓度、底物浓度、产物浓度等。隐层结点的个数由经验来确定。

样本包括训练样本和测试样本,样本来源于发酵工程实验中心,共 6 批次实验数据,选取 4 批次作为训练样本,2 批次作为测试样本。由于聚赖氨酸发酵过程周期为 72 h,采样时间为 4 h,每批次样本数为 19 个。

4.2 Elman 网络模型与 BP 网络模型比较分析

4.2.1 训练参数设定

利用改进的 Elman 神经网络和改进的 BP 神经网络算法实现发酵过程建模时,主要的初始设定参数如表 1 所示,其中学习效率以及改进的 Elman 所使用的自增益参数 α 等参量可以在具体的训练中多次修正,使训练误差和测试误差更小。

表 1 初始设定参数表

参数设置	Elman 网络	BP 网络	参数设置	Elman 网络	BP 网络
输入层 m	6	6	训练次数	1 500	5 000
隐层 n	9	9	设定误差	0.005	0.005
输出层 q	1	1	自增益 α	0.2	无
承接层 l	9	无	动量项	0.2	0.2
学习效率	0.2	0.2			

4.2.2 测试结果比较

通过编制软件,经过输入学习参数以及不同输出参量的模型训练得到不同的改进的 Elman 模型和改进的 BP 神经网络模型,图 3 是改进的 Elman 神经网络聚赖氨酸发酵模型测试曲线图,其中横坐标为发酵周期 0~72 h,纵坐标表示菌体浓度、底物浓度、产物浓度预估曲线;同理,图 4 为改进的 BP 神经网络聚赖氨酸发酵模型测试曲线。训练均方误差与测试均方误差比较如表 2 所示。

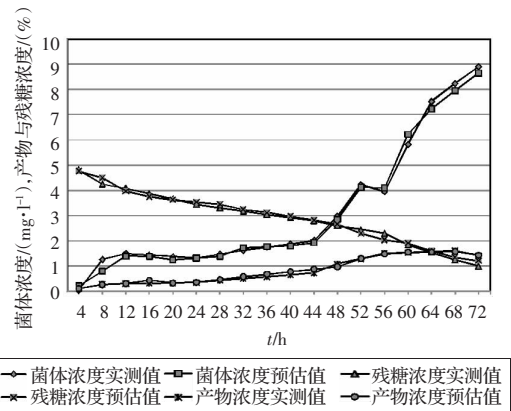


图 3 改进的 Elman 神经网络聚赖氨酸发酵模型测试曲线

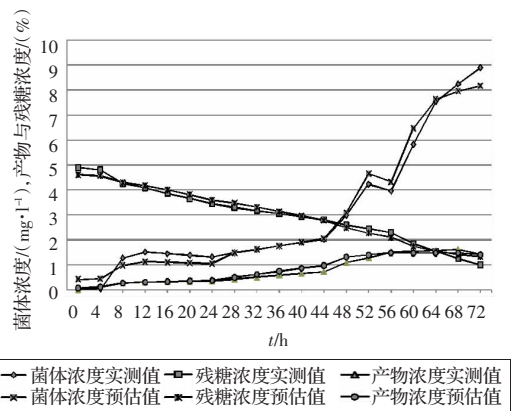


图 4 改进的 BP 神经网络聚赖氨酸发酵模型测试

表2 Elman 与 BP 网络训练均方误差与测试均方误差比较

输出变量	Elman 网络		BP 网络	
	训练均方误差	测试均方误差	训练均方误差	测试均方误差
产物浓度/(%)	0.004 123	0.041 871	0.001 072	0.056 679
残糖浓度/(%)	0.001 906	0.016 236	0.001 602	0.019 631
菌体浓度/(mg·l ⁻¹)	0.002 727	0.002 936	0.001 260	0.006 822

5 结论

针对发酵过程的复杂性,提出了基于改进的 Elman 神经网络实现发酵过程建模的分批训练算法,通过利用 VC++ 编制相应的软件,可以实现不同发酵过程在线建模和不同参量的在线预估。经过具体发酵过程的仿真以及改进的 Elman 模型与改进的 BP 模型比较,改进的 Elman 网络发酵过程模型具有收敛速度快、泛化能力强等特点,能够准确地模拟发酵过程的动态行为。与 MATLAB 实现比较,该动态建模软件可以更好嵌入发酵过程监控系统实现状态参量的在线预估和软测量。

(上接 199 页)

控仿真系统。该系统基于非平稳的泊松随机过程,模拟了大楼人流交通情况。仿真实验验证了在不同交通情况下该系统的真实合理性,证明了该系统能够作为其他调度算法的测试平台。研究的下一步工作准备在考虑建立大楼内人流情况预测模型的基础上,提出更具实用性的电梯群控调度算法。

致谢 该文设计的电梯群控仿真系统参加第三届全国计算仿真比赛并获得一等奖的好成绩,在此仅对参赛期间陈士钦、朱小菲、李婷婷同学的辛勤付出给予致谢。

参考文献:

- [1] Shunji T, Yukihiro U, Mituhiko A. Dynamic optimization of the operation of single-car elevator system with destination hall call registration[J]. *European Journal of Operational Research*, 2005, 167: 550-587.

(上接 239 页)

表示机器人重新规划出一条局部避碰路径,对原有路径进行了实时修正,并按修正后的路径继续行进。

6 结论

讨论了在环境信息部分未知的动态环境下的机器人路径规划问题,从系统的观点提出了一种新方法,并得到了有效的仿真实验结果,其主要特点为:

(1) 采用栅格法建模,使得建模容易,并且使算法简单易实现。

(2) 提出一种较新的机器人全局路径规划算法,以机器人的起始点为根节点,逐渐增加的叶节点中包含了目标节点,从起始点到目标点之间的一条以随机树的边组成的路径就是目标路径。

(3) 在局部路径规划中,采用了基于滚动窗口的规划思想,将预测控制与反馈机制有效结合起来,对动态不确定环境具有

参考文献:

- [1] 张会清,王普,高学金,等.生物发酵参数相关性及其菌体浓度软测量技术的研究[J]. *计算机工程与应用*, 2006, 42(30): 215-217.
- [2] 杨强大.基于改进 BP 神经网络的菌体浓度软测量[J]. *控制与决策*, 2008, 23(8): 869-873.
- [3] 马勇.动态软测量建模方法初探[J]. *化工学报*, 2005(8): 1516-1519.
- [4] Tholuder A, Ramirez W F. Neural network modeling and optimization of induced foreign protein production[J]. *AIChE Journal*, 1999, 45(8).
- [5] 孙增圻. *智能控制理论与技术*[M]. 北京:清华大学出版社, 1997: 191-194.
- [6] Gao X Z, Ovaska S J. Genetic algorithm training of Elman neural network in motor fault detection[J]. *Neural Comput & Applic*, 2002, 11: 37-44.
- [7] 时小虎. Elman 神经网络与进化算法的若干理论研究及应用[D]. 长春:吉林大学, 2006: 10-25.
- [8] 高钦和,王孙安.基于 Elman 神经网络的非线性动态系统辨识[J]. *计算机工程与应用*, 2007, 43(31): 87-89.

- [2] 罗飞,许玉格,曹建忠.基于目的层预约的电梯群控系统建模与控制[J]. *控制与决策*, 2006, 21(10): 1159-1162.
- [3] Cho Y C, Gaogov Z, Kwon H. Elevator group control with accurate estimation of hall call waiting times[C]// *Proceedings of 1999 IEEE International Conference on Robotics and Automation*, 1999, 1: 447-452.
- [4] 曾国伟,赵国军,邢海潇,等.基于模糊控制技术的电梯群控系统设计[J]. *浙江工业大学学报*, 2009, 37(1): 115-118.
- [5] 张昆,段其昌,张从力.基于模糊控制的多目标电梯群控技术[J]. *仪器仪表学报*, 2004, 25(4): 248-251.
- [6] 李中华,朱燕飞,李春华,等.基于人工免疫聚类算法的电梯交通流分析[J]. *华南理工大学学报*, 2003, 31(12): 26-29.
- [7] 许玉格,罗飞.新型电梯群控系统交通模式识别方法[J]. *控制理论与应用*, 2005, 22(6): 900-904.
- [8] 李中华,朱燕飞,李春华,等.面向对象的电梯群控系统仿真平台开发与实现[J]. *计算机仿真*, 2005, 22(2): 156-159.

良好的适应性。

参考文献:

- [1] 陈根社,陈新海.遗传算法的研究与进展[J]. *信息与控制*, 1994(8): 215-220.
- [2] D'Amico A, Ippoliti G, Longhi S. A radial basis function networks approach for the tracking problem of mobile robot[C]// *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2001, 1: 498-503.
- [3] Trovato K I, Dorst L. Differential A*[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2002, 14(6): 1218-1229.
- [4] LaValle S M. Resolution complete rapidly-exploring random trees[C]// *Proceedings of the 2002 IEEE International Conference on Robotics & Automation*, Washington, DC, May 2002.
- [5] 席裕庚,张纯刚.一类动态不确定环境下机器人的滚动路径规划[J]. *自动化学报*, 2002, 28(2): 161-175.