

模糊规则的自动提取与优化¹

顾林跃 张立明

(复旦大学电子工程系 上海 200433)

摘 要 本文提出了一种自动提取模糊规则集的新方法。本方法首先利用 CCM 分析闭环模糊系统在某一规则集控制下的全局动态行为, 根据控制目标对该规则集进行评价得其性能指标。然后利用 GA 根据这个性能指标对规则集进行逐代更新和优化, 最后得到一个满足要求的规则集。本文以复杂非线性系统——倒摆系统为例, 介绍用本文的方法提取模糊规则的具体过程, 仿真结果表明了本方法的有效性和通用性。

关键词 遗传算法, 单元映射, 染色体, 适合度

中图分类号 TP872

1 引 言

自 L. A. Zadeh 提出模糊集后, 模糊控制器得到了广泛应用, 已有众多成功的实例。在多数情况下, 模糊控制器优于传统控制器, 但有缺点: (1) 没有一种解析的方法来分析闭环模糊控制系统的全局动态行为。(2) 很难获得一个较优的规则集, 模糊规则的获取已成为模糊控制器能得到广泛应用的一个瓶颈。目前, 一般是根据专家的经验知识形成一些规则, 然后人工逐步修正, 使之趋于较优, 此过程十分烦琐、费时。文献 [1,2] 提出了利用聚类算法和神经网络来自动提取或优化规则。这些方法遇到的主要障碍是很难获得一个能真正反映系统的动态特性的训练样本集。

遗传算法^[3] (Genetic Algorithm, GA) 是一种基于自然选择和群体遗传机理的搜索算法。它模拟了自然选择和自然遗传过程中发生的繁殖、交配和突变现象。GA 对于复杂的优化问题无需建模和进行复杂的运算, 只要用 GA 的三种算子 (选择、交叉、突变) 就能寻找到优化的解, 至今已得到广泛的应用^[4-7]。本文利用 GA 搜索在某一控制目标下为最优的规则集。

为了用 GA 提取模糊规则集, 首先必须根据控制目标对模糊规则集进行评价。从非线性理论中获得的基于相空间的分析方法——单元映射 (Cell to Cell Mapping, CCM)^[8-10], 已成为分析模糊系统的有效方法。它用离散的单元状态空间来近似系统的连续相空间。利用 CCM 分析模糊系统能直接获得有关系统全局动态行为的信息, 因而能用它评价模糊规则集的优劣。

¹ 1994-04-21 收到, 1995-10-16 定稿

上海市科委重点基金项目和攀登计划认知科学 (神经网络) 重大关键项目的资助课题

本文综合了 GA 和 CCM 的优点, 利用 CCM 分析系统的全局动态行为对模糊规则集进行评价, 确定一个适合度值, GA 利用适合度值的信息来分区域搜寻较优的规则集。最终得到一个全局性能较优的规则集。

2 基本概念

2.1 闭环模糊控制系统

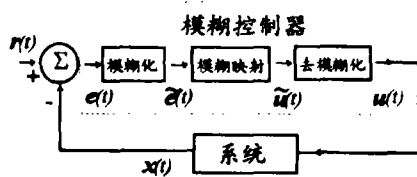


图 1 闭环模糊控制系统

图 1 为闭环模糊控制系统, $x(t)$ 为被控系统在 t 时刻的状态, $r(t)$ 为期望的状态。虚线框内为一模糊控制器, 其中 $e(t) \in R^n$, $\bar{e}(t) \in R^n$, $u(t) \in R^m$, $\bar{u}(t) \in R^m$ 。在模糊控制器的第 j 个输入变量 e_j 的论域 U_j 上定义 p_j 个模糊子集 $A_j^i, i = 1, 2, \dots, p_j$ 。在模糊控制器的第 j 个输出变量 u_j 论域

V_j 上定义 q_j 个模糊子集 $B_j^i, i = 1, 2, \dots, q_j$ 。这些模糊子集可以用语言变量 NL、NM、NS、Z、PS、PM、PL 等表示。图中的模糊映射由“ If ”、“ Then ”构成的模糊规则所决定, 规则的形式为

If \bar{e}_1 is A_1^i and \bar{e}_2 is A_2^k and \dots and \bar{e}_n is A_n^l , Then $\bar{u}_1 = FM(\bar{e}_1, \bar{e}_2, \dots, \bar{e}_n)$ is B_1^k 。

其中 $1 \leq l \leq m, 1 \leq k \leq q_1$ 。十分明显, 总共存在 $(q_1 \times q_2 \times \dots \times q_m)^{(p_1 \times p_2 \times \dots \times p_n)}$ 个可能的规则集。本文利用 GA 从这些可能的规则集中搜寻一个在某一控制目标下性能指标较好的规则集。

2.2 遗传算法

2.2.1 编码 在应用遗传算法之前, 首先必须将待优化问题的参数集用有限的字符编码成有限长度的字符串, 该字符串称为染色体 (Chromosome)。这里经编码后的规则集作为一个染色体。

2.2.2 适合度 由于 GA 只利用群体中每个个体的优劣信息来进行搜索, 因而必须根据预定的控制目标对每个染色体确定一个适合度值。由染色体、性状 (表示的实际参数) 以及适合度构成一个个体 (Individual)。

2.2.3 GA 的三种基本算子 (Operator)

(1) 复制 (Reproduction) 复制就是依据每个个体的适合度值来复制新个体的过程。适合度越大, 被复制的可能性越大, 其子孙在下一代的个数则越多。复制算子模拟了生物圈的自然选择现象。

(2) 交叉 (Crossover) 首先在新复制的群体中随机选两个个体, 然后沿着这两个个体均匀随机地取某个位置, 最后两者互换从该位置起的末尾部分。例如, 有两个用二进制编码的个体 A 和 B, 长度 L 为 5: $A=01101, B=11000$ 。随机选择一个整数 $k \in [1, L-1]$, 假设 $k=4$, 经交叉后产生两个新个体 A' 和 B' , $A'=01100, B'=11001$ 。

(3) 突变 (Mutation) 突变就是以很小的概率随机地改变染色体某个位置的字符。在二进制编码中, 突变算子随机地将某个位置的'1'变成'0', 或者'0'变成'1'。突变本身是一种随机搜索, 然而与复制、交叉算子结合在一起, 就能避免由复制和交叉算子而引起的某些信息的永久性丢失, 保证了 GA 的有效性。

本文采用了较复杂的复制方法: 无替代剩余随机抽样方法, 并且采用了多点交叉的方法。同时, 对适合度进行了比例变换, 以加快 GA 搜索的速度, 详细情况可参考文献 [3]。

2.3 单元映射 (CCM)^[8-10]

单元相空间和单元映射是用于分析非线性系统的全局动态行为的有效方法。

以图 1 的闭环系统为例。首先, 将系统的连续相空间 \mathbf{x} 离散化为单元相空间, 其中每一维子空间 $x_i \in [bl_i, bu_i]$ 。根据精度要求, 将 x_i 分割成长度为 s_i 的等分, 每个等分用整数 z_i 来表示, 属于 z_i 等分的所有 x_i 满足: $bl_i + z_i s_i \leq x_i \leq bl_i + (z_i + 1) s_i$ 。相空间划分后所得的每个超立方体称为单元, 可用一个整数矢量 $Z = [z_1, z_2, \dots, z_n]^T$ 来表示。至此已将连续相空间离散化为一单元所构成的单元相空间, 每个单元中心点 \mathbf{x}^c 的动态行为代表了该单元的行为。

定义 1 单元相空间中包含 $r(t)$ 的单元称为目标单元 (Target Cell), 用 Z^T 表示。

然后, 将点映射转化为单元映射。假设被控系统的下一个状态仅由当前状态和模糊控制器的输出确定, 即: $\mathbf{x}(t_{k+1}) = h(\mathbf{x}(t_k), \mathbf{u}(t_k))$, 其中 t_k 为离散采样时刻。又设 $\mathbf{u}(t_k) = g(\mathbf{x}(t_k))$, g 为一个非线性函数, 则 $\mathbf{x}(t_{k+1}) = h(\mathbf{x}(t_k), g(\mathbf{x}(t_k))) = f(\mathbf{x}(t_k))$ 。

定义 2 单元映射 $G: Z \rightarrow Z$: 对所有单元 Z , 在 t_k 时刻满足: 若 $\mathbf{x}(t_{k+1}) = f(\mathbf{x}^c(t_k)) \in Z^*$, 则 $G(Z) = Z^*$, 其中 $\mathbf{x}^c(t_k)$ 为单元 Z 的中心点。

定义 3 若单元 Z 满足: $G(Z) = Z$, 则称 Z 为单元映射 G 的稳定单元。

由于控制的目标是使得系统在某个规则集的控制下所有单元 (包括目标单元本身) 最终都能映射到目标单元, 因此目标单元必须是稳定单元, 即 $G(Z^T) = Z^T$ 。

3 本文算法

下面以倒摆系统为例详细介绍本文算法。如图 2 所示, 倒摆系统由可在水平方向移动的小车和可在垂直平面里运动的杆组成。本文只研究杆在垂直平面里的动态特性, 而不考虑小车在水平方向的运动。杆的动态特性可由杆偏离垂直方向的角度 $\theta(t)$ 以及杆的角速度 $\dot{\theta}(t)$ 描述, 其运动方程如下:

$$\ddot{\theta}(t) = \frac{mg \sin(\theta(t)) - \cos(\theta(t)) [F(t) + m_p l \dot{\theta}^2(t) \sin(\theta(t))]}{(4/3)ml - m_p l \cos^2(\theta(t))} \quad (1)$$

取杆质量 $m_p = 0.1 \text{ kg}$, 车与杆总质量 $m = 1.1 \text{ kg}$, 杆长 $l = 0.5 \text{ m}$ 。重力加速度 $g = 9.8 \text{ m/s}^2$, 并且规定 $\theta(t) \in [-\pi/2, \pi/2]$, $\dot{\theta}(t) \in [-4, 4]$, 采样时间 $t_s = 0.01 \text{ s}$ 。控制目标为: 通过给小车加力 $F(t) \in [-20, 20]$ 使杆在任何初始状态下都能以最短的时间恢复到目标状态 $\theta = 0, \dot{\theta} = 0$, 并保持动态平衡。对于倒摆系统, 有 $n = 2, m = 1, \mathbf{x} = [\theta, \dot{\theta}]^T, \mathbf{r} = [0, 0]^T, \mathbf{e} = [\theta, \dot{\theta}]^T, \mathbf{u} = [F]$ 。

步骤 1 首先对输入变量和输出变量的论域进行模糊化 对于倒摆系统, 模糊控制器有两个输入 $\theta, \dot{\theta}$ 和一个输出 F 。取 $p_1 = p_2 = q_1 = 7$, 将每个论域均模糊化成 7 个模糊集, 并使相邻模糊集有 20% ~ 50% 的交叠。这样既能满足精度要求和一定的鲁棒性, 又不至于使运算太过复杂。经模糊化后得一控制规则表, 目标状态位于控制规则表上的 (Z, Z) 位置, 区域 0 如图 3 所示。

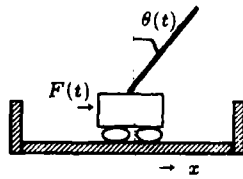


图 2 倒摆系统

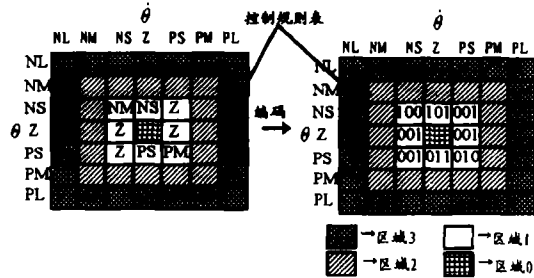


图 3 规则集的编码过程

步骤 2 将倒摆系统的连续相空间离散化 根据精度要求, 现将相空间离散化为 71×71 个单元。易知 $(35, 35)$ 单元为目标单元。

定义 4 在某个规则集的控制下, 在单元相空间中最终不能映射到目标单元的单元称为不可控单元; 反之, 称为可控单元。

步骤 3 对控制规则表进行区域划分 对于倒摆系统, 整个搜索空间中包含有 $7^{(7 \times 7)} = 7^{49}$ 个可能规则集, 而较优的规则集只占极小一部分。如果利用 GA 从整个搜索空间来寻找较优的规则集, 则搜索时间必然很长。为了加快算法的搜索速度, 本文对控制规则表进行区域划分, 分四步来确定控制规则表上的所有规则。

定义 5 在控制规则表上包含目标状态 ($\theta = 0, \dot{\theta} = 0$) 的位置称为目标区域, 记为区域 0。区域 0 周围的位置所构成的区域称为区域 1, 依次类推定义区域 2、区域 3……, 直至它们的并构成了整个控制规则表, 如图 3 所示。

首先利用专家知识或 GA 确定区域 0 中的规则, 然后提取区域 1 中的规则, 依次类推, 最终确定控制规则表上的所有规则, 即得到一个优化的规则集。分区域提取规则可有效地缩短搜索的时间, 这是因为 GA 分别在四个空间中寻找最优解, 而每个搜索空间中包含的规则集总数分别为: $7^1, 7^8, 7^{16}, 7^{24}$, 远远小于整个搜索空间中的规则集数 7^{49} 。

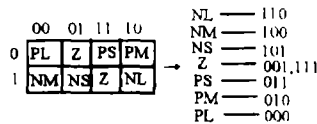


图 4 模糊集的编码过程

步骤 4 对模糊集进行编码 对每个论域上的七个模糊集只需用 3 位二进制数进行编码, 如图 4 所示。考虑到这七个模糊集的内在联系和相关, 使相邻模糊集的编码只有一位发生变化, 即相邻模糊集的编码也相邻。

步骤 5 对整个规则集进行编码 在对模糊集编码的基础上, 对规则集进行编码。下面只以提取区域 1 中的 8 条规则为例来介绍规则集的编码过程, 如图 3 所示。考虑到区域 1 中 8 条规则之间的联系和相关, 从位置 (NS, NS) 开始顺时针对区域 1 中的 8 条规则进行编码, 最终得表示区域 1 规则集的字符串为: 100101001001010011001001。

步骤 6 适合度函数的选取 遗传算法的每种算子都是根据适合度的信息对个体进行操作的, 适合度函数的选取直接影响到遗传算法收敛速度的快慢以及成败。以下介绍如何利用 CCM 为每个规则集(即染色体) 确定一个适合度值。

从倒摆系统的控制目标可知, 它属于时间最优化控制问题。首先按步骤 2 将倒摆系统的连续相空间离散化为单元相空间。由于是时间最优问题, 然后只要把所有单元归结为两类, 一类为可控单元, 一类为不可控单元。利用分组算法^[8-10] 很容易确定在某个规则集控制下的所有单元的分类, 同时确定每个可控单元 Z_i 到达目标单元的最大时间步长 MS_{Z_i} , 以及实际时间步长 S_{Z_i} , 对于不可控单元, 则令 $S_{Z_i} = MS_{Z_i}$ 。最后利用每个单元的时间步长信息, 便可以确定某个规则集的适合度。本文采用的适合度函数如下:

$$Ft = 1 - \frac{1}{N} \sum_{o=1}^{N_c} \frac{S_{Z_j}}{MS_{Z_j}}, \quad (2)$$

其中 N_c 为单元相空间中的单元总数。(2) 式充分利用了单元空间中所有单元的步长信息, GA 利用上式确定的适合度值就能搜索到满足控制目标的规则集, 使得在该规则集控制下所有单元到达目标单元的时间总和为最短。由于分组算法可依据各种优化目标计算单元变迁(即状态从一个单元变化到另一个单元) 所付出的代价, 因而可依据各种控制目标确定适合度函数。

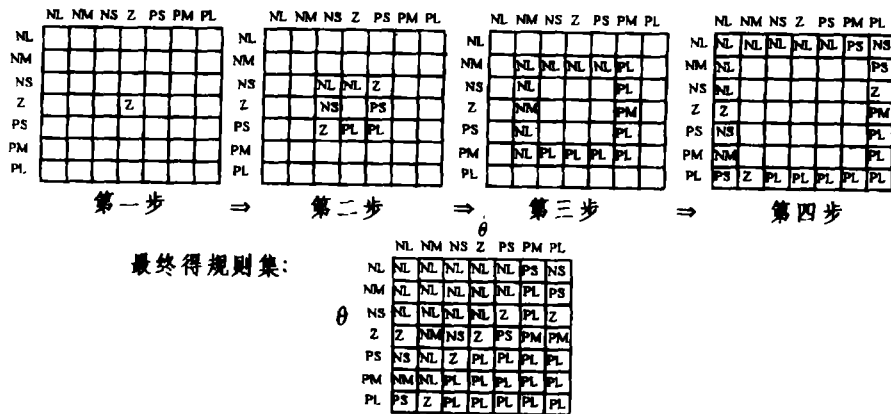


图 5 分四步提取控制规则表上的所有规则

步骤 7 利用 GA 分四步确定控制规则表上的所有规则, 如图 5 所示 (1) 令控制规则表上的区域编号 A 为 0。(2) 根据专家的经验知识或随机地形成 50 个染色体(即表示区域 A 上的规则集的字符串), 编码如步骤 4、5 所述。令迭代次数 G 为 0。(3) 利用 (2) 式对每个染色体作出评价, 确定一个适合度值。若适合度大于预定的值或 G 大于 200, 则转到 (7)。(4) 利用无替代剩余随机采样 (Remainder Stochastic Sampling without Replacement) 选择方法依据每个规则集的适合度值预复制 50 个个体。(5) 对预复制的个体进行交叉、突变, 形成新的个体, 并保持个体数量不变。(6) $G = G + 1$, 转到 (3)。(7)

若 $A < 3$, 则 $A = A + 1$, 跳转到 (2); 否则选取适合度最大的染色体作为最终的解 (规则集), 整个算法结束。

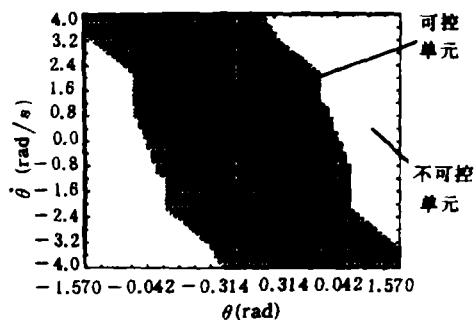


图 6 倒摆系统单元相空间中的可控单元和不可控单元

4 仿真结果

图 6 给出了在所提取的规则集控制下系统单元相空间中的可控单元和不可控单元, 其中阴影部分为可控单元。可控区域的大小与力 F 的范围直接有关。图 7(a), 7(b) 为不同初始状态下倒摆系统的响应曲线。仿真结果表明, 用本文方法提取的规则集具有良好的控制效果, 因而验证了本文方法的有效性。

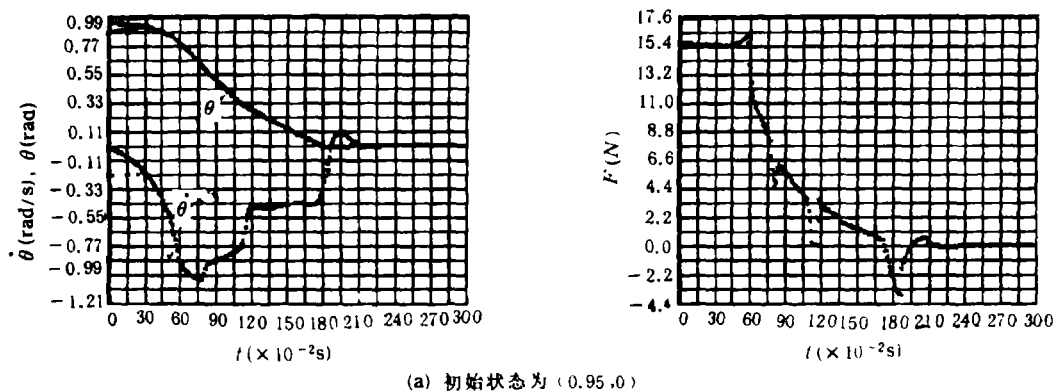


图 7 倒摆在不同初始状态下的响应

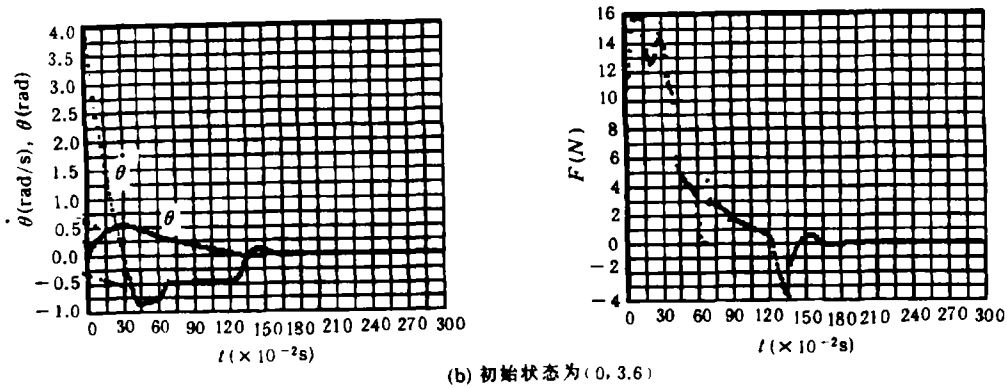


图 7 倒摆在不同初始状态下的响应

5 结束语

本文给出了一种利用 GA 和 CCM 方法来自动提取和优化规则的方法, 克服了传统方法由于样本集的不完整而导致规则集性能较坏的缺点, 而且能广泛适用于线性或非线性系统、连续或离散系统的各种优化目标下的规则提取。但是由于本文方法是在各模糊隶属度函数确定的基础上提取规则集的, 因而具有一定的人为因素。另外, 由于采用 CCM 来评价规则集的优劣, 从而影响了提取整个规则集的速度。如果能找到一种新的评价规则集的快速算法, 则整个算法将大大加快。本文主要提出了一种利用 GA 来自动提取和优化规则的全新思路, 许多细节还有待进一步研究, 如规则之间的相关性, 分步来提取规则集的理论依据, 如何选取各论域模糊隶属度函数等。

参 考 文 献

- [1] Bart Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice-Hall Intelligence, Inc, 1992, 299-337.
- [2] Sulzberger M, Tschichold-Gurman N, Vestli J. FUN: Optimization of Fuzzy Rule Based Systems Using Neural Networks. *IEEE ICNN, Vol.1, San Francisco: March 28-April 1, 1993, 312-322.*
- [3] Goldberg David E. *Genetic Algorithms in Search, Optimization & Machine Learning*, Reading, MA: Addison Wesley Publishing Company, Inc., 1989, Chap3, 4.
- [4] Karr L, Gentry J. *IEEE Trans. on FS, 1993, FS-1(1): 46-53.*
- [5] McInerney, Dhawan P. Use of Genetic Algorithms with Back Propagation in Training of Feed-Forward Neural Networks. *IEEE ICNN, Vol.1, San Francisco: March 28-April 1, 1993: 203-208.*
- [6] Koza R, Keane A, Rice P. Performance Improvement of Machine Learning via Automatic Discovery of Facilitating Functions as Applied to a Problem of Symbolic System Identification. *IEEE ICNN, Vol.1, San Francisco: March 28-April 1, 1993: 191-198.*
- [7] Bornholdt Stefan, Graudenz Dirk. *Neural Networks, 1992, 5(2): 327-334.*
- [8] Yung Yaw Chen, Tsu Chin Tsao. *IEEE Trans. on SMC., 1989, SMC-19(4): 745-755.*
- [9] Smith Samuel M, Comer David J. *IEEE Control Systems Magazine, 1991, 11(5): 18-28.*

[10] Hsu C S. *Journal of Optimization Theory and Application*, 1985, 46(4): 547-567.

AUTOMATICALLY EXTRACTING AND OPTIMIZING FUZZY RULE BASE

Gu Linyue Zhang Liming

(*Electronic Engineering Department, Fudan University, Shanghai 200433*)

Abstract A method of automatically extracting and optimizing fuzzy rule base is presented. Firstly, it applies the method of CCM (Cell-to-Cell Mapping) to analyze the evolving trend and global behavior of the closed-loop fuzzy dynamical system based on a cell state space. Secondly, each rule base is evaluated to determine a performance index according to the control objective. Thirdly, GA (Genetic Algorithm) updates the old population of coded rule bases using the performance index of each rule base generation by generation. A complicated nonlinear system—an inverted pendulum is simulated to illustrate the details of the algorithm and the results demonstrate the validity of the method.

Key words GA (Genetic Algorithm), CCM (Cell-to-Cell Mapping), Chromosome, Fitness

顾林跃：男，1970年生，硕士，从事人工智能、神经网络、模式识别、图象处理等方面的研究工作。

张立明：女，1943年生，教授，从事神经网络、模式识别、人工智能、图象处理和压缩、非线性理论等方面的研究工作。