

# 多层前向网络的研究 ——遗传 BP 算法和结构优化策略<sup>1)</sup>

陈 荣 徐用懋 兰鸿森

(清华大学自动化系 北京 100084) (福建炼油厂 福建惠安 362100)

**摘 要** 分析了引起 BP 算法局部收敛现象的原因, 探讨了解决的途径. 通过对遗传算法中的基因群体的实数化, 并将其与 BP 算法有机地集成起来, 提出了遗传 BP 算法. 此外, 在借鉴网络剪枝法思想的基础上, 重新定义网络结构复杂度函数, 并以遗传算法直接求解网络结构优化问题, 给出了网络结构优化策略. 仿真结果和实际应用均表明了上述研究的成效.

**关键词** 人工神经网络, 遗传算法, BP 算法, 结构优化.

## 1 引言

在当前已有的诸多人工神经网络中, 多层前向网络是在实际中应用最为广泛的一类. 关于它的研究有两大方面: 学习算法和结构优化. 其中, BP 算法 (Backpropagation Algorithm) 是目前使用最为普遍的网络学习算法, 但存在着收敛速度慢、可能陷入局部极小点这两个突出弱点. 一般而言, 可以采取动态调整确定学习步幅、自适应改变惯性系数等措施来加快 BP 算法的收敛速度; 而为了克服局部收敛, 则必须彻底摆脱依赖梯度信息来指导权值调整方向这种方式.

长期以来, 网络的结构单凭经验而定, 为保障精度往往偏向于冗余, 这将造成: 第一, 网络训练过程所需时间加长, 增加了学习算法在训练速度上的负担, 不利于网络的在线适应性; 第二, 所得网络的高精度很可能是冗余节点存在的结果, 因而极易具有病态, 表现为对于训练样本之外的数据其精度急剧下降, 网络的泛化能力弱. 所以, 应当科学地确定网络结构.

## 2 遗传 BP 算法

遗传算法 (Genetic Algorithms, 简称 GAs) 是根据达尔文的自然界生物进化思想, 将其灵活运用到优化运算领域而产生的一种寻优算法<sup>[1,3,4]</sup>, 它具有以下优点:

1) 在可行解空间同时由多个起始点开始搜索, 搜索效率高;

<sup>1)</sup> 本文部分结果曾发表于 1994 年第一届中国智能控制与智能自动化学术会议.

- 2)本质上属于随机寻优过程,不存在局部收敛问题;  
3)不要求准则函数可导,可用于求解非连续函数优化问题.

它的缺点在于:

1)通过参数的二进制编码字符串间接运算,人为将连续空间离散化,导致计算精度与字符串长度、运算量之间的矛盾;并且对于那些变量取值范围不明确的问题(例如求取人工神经网络权值的问题)也无从编码.

2)完全依概率随机进行的寻优操作虽避免了陷入局部极小点,但在有限次寻优时一般只能得到全局范围的次优解,不易获取最优解.

针对以上问题本文做如下工作:

1)取消二进制编码,直接将待处理的参数数值(实数)逐位数字地转化为数字字符(形成字符串),之后便可以完全类似于遗传算法对二进制字符串的处理,以相同方式对此实数型字符串进行选择(selection)、交叉(crossover)操作,不同之处在于发生变异(mutation)时,不再是原二进制字符“0”、“1”之间的跳变,而将按模 10 和的运算或 M 序列的产生来完成.例如,对如下两个随机数:3.141592,2.718030,经数字→字符转换,生成第一代基因:

3	.	1	4	1	5	9	2
---	---	---	---	---	---	---	---

和

2	.	7	1	8	0	3	0
---	---	---	---	---	---	---	---

若取第一代基因的后三位进行交叉操作,则演化为第二代基因:

3	.	1	4	1	0	3	0
---	---	---	---	---	---	---	---

和

2	.	7	1	8	5	9	2
---	---	---	---	---	---	---	---

再取第二代基因的第一位按模 10 和进行变异操作,则形成第三代基因:

7	.	1	4	1	0	3	0
---	---	---	---	---	---	---	---

和

8	.	7	1	8	5	9	2
---	---	---	---	---	---	---	---

当然,上述过程只是最简单的示意,具体研究工作请参阅文献[1].

2)将上述实数化之后的遗传算法直接与 BP 算法相集成,以前者的全局寻优能力防止陷入局部极小点,同时依赖后者的梯度下降搜索法保证在有限次搜索后快速找到全局最优解.

据此,本文提出遗传 BP 算法,其具体过程为:

- ① 随机产生  $N$  组在不同实数区间内取值的初始网络权值.
- ② 用 BP 算法对这  $N$  组初始权值分别进行预训练,若经过预训练后,这  $N$  组权值中至少已有一组满足精度要求,则算法结束;否则转入步骤③.
- ③ 分别依据经过预训练的上述  $N$  组权值所对应的上下限确定取值区间,在区间内

随机生成  $r * N$  组新的权值,连同经过预训练的  $N$  组权值一起,构成完整的基因群体,共  $(r+1) * N$  组权值.

④ 对这  $(r+1) * N$  组权值进行选择、交叉、变异等遗传操作.

⑤ 如果经过步骤④的操作已至少得到一组符合精度要求的权值,则算法结束;否则从经过遗传操作的这  $(r+1) * N$  组权值中选出  $N$  组较好的,回复到步骤②.

限于篇幅,关于遗传 BP 算法的详尽描述和算法实现请参阅文献<sup>1)</sup>.

### 3 结构优化策略

关于人工神经网络的结构优化工作,近年才逐渐开展.其中以 Bhat 等人提出的网络剪枝法较为引人瞩目<sup>[2]</sup>,其基本思想是:引入网络结构复杂度函数  $E_{com}$  来定量考察网络结构优化程度,之后,在保证网络学习精度  $E_{bpn}$  的前提下,尽量简化网络的结构复杂度,即减少  $E_{com}$  值,这在数学上表现为求解一个带约束的极值问题.但因为 Bhat 等人以 BP 算法为网络训练手段,难以直接处理这种非连续函数形式,所以不得不借助于罚函数的思想将上述极值问题加以变换,改为求解

$$E = E_{bpn} + \lambda E_{com}, \quad \lambda > 0 \quad (1)$$

一维无约束极值问题.式中,  $E$  为网络综合性能评价指标;

$\lambda$  为比例因子,用以表征  $E_{bpn}$  和  $E_{com}$  的相对重要程度.

由于  $E_{bpn}$  和  $E_{com}$  之间并不存在数量级上的优先次序关系,而  $\lambda$  的不同取值又直接影响着网络结构最终的优化结果与精度,所以确定  $\lambda$  的取值成为一个极为棘手的问题.针对 Bhat 等人工作中的问题,本文提出了基于遗传算法改造后的网络结构优化策略,其具体做法为:

①重新定义网络结构复杂度函数  $E_{com}$  为

$$E_{com} = (2N_{in} + N_{hid} - N_k) \sum_i \sum_j w_{ij}^2 \quad (2)$$

其中,  $N_{in}$  为输入层节点数目;  $N_{hid}$  为隐层节点数目;  $w_{ij}^2$  为从节点  $i$  指向节点  $j$  的权值;  $N_k$  为结构优化指数,初始值置为零,每当删去一个输入节点其值增 2,删去一个隐节点其值增 1.某节点被删除的标准是:当由该节点出发指向下一层节点的所有权值(包括阈值)均落于死区之中时,则该节点被删去(此处死区为预先设定的用来衡量权值对网络精度所起作用显著与否的一个数值区间).

依(2)式重新定义的  $E_{com}$  将能够指导训练过程倾向于优先删除网络节点,实现网络结构的最简化而不仅仅是网络节点间联接的稀疏化.

②根据所需解决问题的难易程度,预先给出一个相对于该问题来说“足够复杂”的多层前向网络.

③以依概率全局寻优的遗传算法来训练步骤②中给出的网络,并直接求解网络结构优化问题的原始形式,即

$$\min E_{com}. \quad (3)$$

约束条件为

$$E_{bpn} < \epsilon, \quad \epsilon > 0 \quad (4)$$

其中,  $E_{bpn}$  为反映网络映射精度的相对均方误差,  $\epsilon$  为所允许的  $E_{bpn}$  最大值, 它根据实际需要而给定.

④ 当遗传算法结束上述寻优过程, 即在  $E_{bpn}$  满足约束条件的前提下,  $E_{com}$  到达收敛点时, 网络结构将已经自动实现优化(最简化).

需要指出的是:

由于以(4)式为约束条件对(3)式进行目标优化, 所以只会删除冗余的网络节点, 而不会导致网络结构的过于简化.

## 4 仿真与实际应用举例

### 4.1 仿真举例

这里我们选取一个输入层节点数为 2、输出层节点数为 1、隐层节点数为 5 的三层前向网络, 分别采取标准 BP 算法、基本遗传算法和遗传 BP 算法作为训练方法, 以求实现如下具有多个极值点的非线性函数:

$$f(x_1, x_2) = \frac{\sin |x_1| + e^{|x_2|}}{3}, \quad (5)$$

其中训练样本数为 30, 检验样本数为 30, 精度要求为  $E_{bpn} = 0.002$ .

表 1.1 三种算法所得网络训练及检验精度

		标准 BP 算法	基本遗传算法	遗传 BP 算法
训 练 过 程	$E_{bpn}$	0.0046	0.0034	0.0018
	$E_{max}$	0.0713	0.0526	0.0216
检 验 过 程	$E_{bpn}$	0.0048	0.0034	0.0019
	$E_{max}$	0.0754	0.0530	0.0283

注: 1.  $E_{bpn}$  为相对均方误差,  $E_{max}$  为最大相对误差;

2. 检验样本与训练样本是内插关系.

表 1.2 三种算法对应的网络训练次数

相对均方误差	标准 BP 算法	基本遗传算法	遗传 BP 算法
0.0050	9,033	4,510	210
0.0046	17,725	—	—
0.0040	$\infty$	8,400	460
0.0034	$\infty$	15,820	—
0.0030	$\infty$	$\infty$	690
0.0018	$\infty$	$\infty$	1,020

注: 1. “ $\infty$ ”表示算法达不到对应精度, “—”表示算法直接跳过了对应精度;

2. 基本遗传算法和遗传 BP 算法对应训练次数系根据所用 CPU 时间折算成标准 BP 算法中的训练次数来表示的.

表 1.1 给出了三种算法在网络的训练和检验过程中的精度比较结果,从中我们可以看出:经过训练,只有遗传 BP 算法达到预先设定的精度要求;从检验结果来看,遗传 BP 算法能够较好地复现学习成果,表明它所取得的较高精度不是病态的.

表 1.2 记录了三种算法在不同精度时所需经历的训练次数,从中可以看出:标准 BP 算法的速度最慢;基本遗传算法在较为迅速地找到次优解后在求取最优解时却发生了困难;相比之下,只有遗传 BP 算法始终保持着快捷的训练速度,迅速地求得了满足所需精度的解.

上述用到的神经网络结构实际上是经结构优化策略处理后而得到的,最初给出的是一个输入层节点数为 3(三个输入分别为  $x_1$ 、 $x_2$ 、 $x_3 = x_1 + 2x_2$ ),输出层节点数为 1,隐层节点数为 10 的三层前向网络,经过网络结构优化,冗余的输入节点  $x_3$  被删去,同时隐层节点也从最初的 10 个减少为 5 个,整个网络结构大为简化.

那么简化之后的网络其映射函数的能力是否也削弱了呢?为此采用相同的样本对结构优化前后的网络分别进行训练和检验,结果如表 2 所示.

表 2 结构优化前后网络训练及检验精度

		未经结构优化的网络	经过结构优化的网络
训练过程	$E_{bpn}$	0.0017	0.0018
	$E_{max}$	0.0207	0.0216
检验过程	$E_{bpn}$	0.0021	0.0019
	$E_{max}$	0.0301	0.0283

从表 2 中不难得出结论:网络经过结构优化之后,其精度不但未受到损害,相反由于其能够将冗余的输入节点和隐层节点剔除,使得它对于所需实现的函数关系的本质认识得更为准确、深入,因此当换用不同的样本来检验时其精度基本维持不变.

## 4.2 实际应用举例

在原油常压蒸馏塔的优化与控制中,由于常压塔各侧线抽出产品的质量信息无法在线获取,给进行闭环控制带来了很大困难,因此利用人工神经网络技术建立常压塔质量估计模型十分必要.而在建立神经网络模型时,网络输入变量的选取和隐层节点数目的多少对于模型的精度及其适应性强弱有着直接的影响,所以这里我们采用上面的网络结构优化策略来确定之.又由于常压塔为典型非线性大型复杂工业对象,样本空间极为复杂,故采用遗传 BP 算法来训练之.

以常一线汽油干点模型的建立为例,结果如图 1—4 所示.图 1 为依据工艺和机理分析而初步确定的网络结构,图 2 为经过结构优化之后得到的网络模型,图 3 和图 4 分别为图 2 所示模型的训练和检验精度,它们均满足误差小于  $4^{\circ}\text{C}$  的工艺要求.而图 1 所示模型的训练精度虽然能满足工艺需要,检验精度却发散了,限于篇幅,文中略去未示.

本文开发了多层前向网络的一种全局快速寻优算法;遗传 BP 算法;同时给出了网络结构的优化策略.上述研究的成效经受了数值仿真和实际工业应用的检验.并且文中所依据的思想并不仅局限于多层前向网络的研究,几乎不需加任何重大改动就可拓展用于其它领域的优化搜索和极值求解.

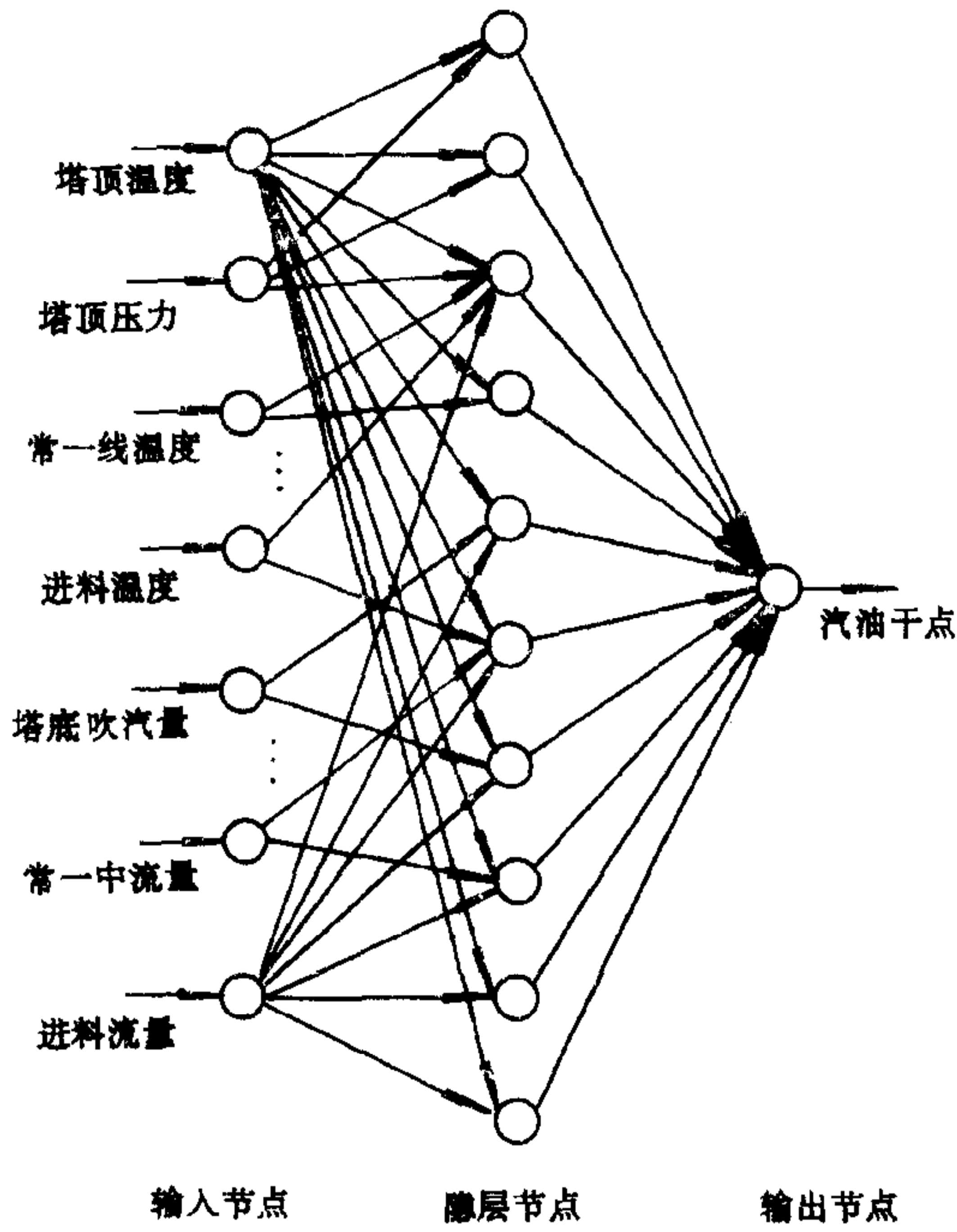


图 1 初始汽油干点网络模型

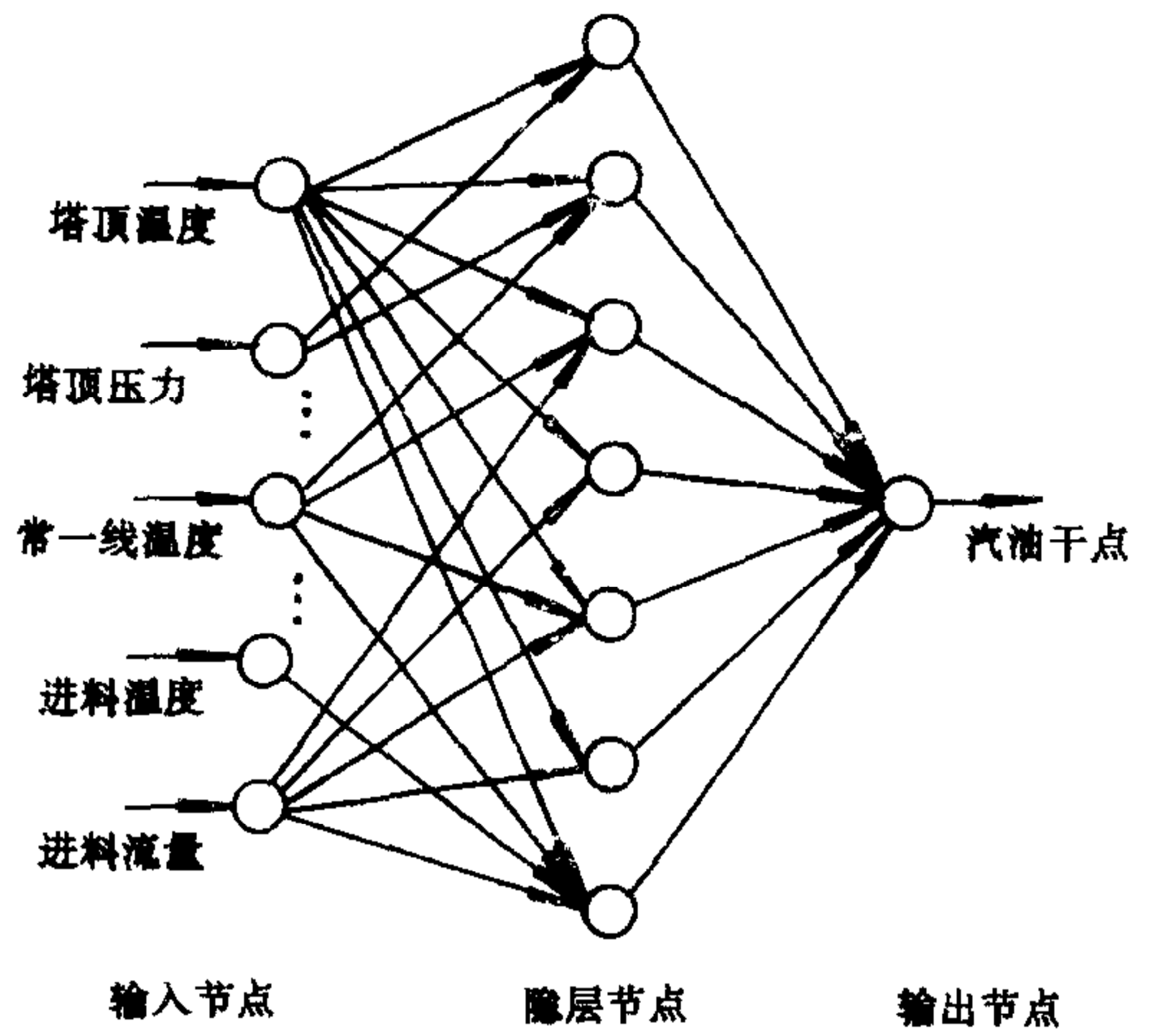


图 2 优化汽油干点网络模型

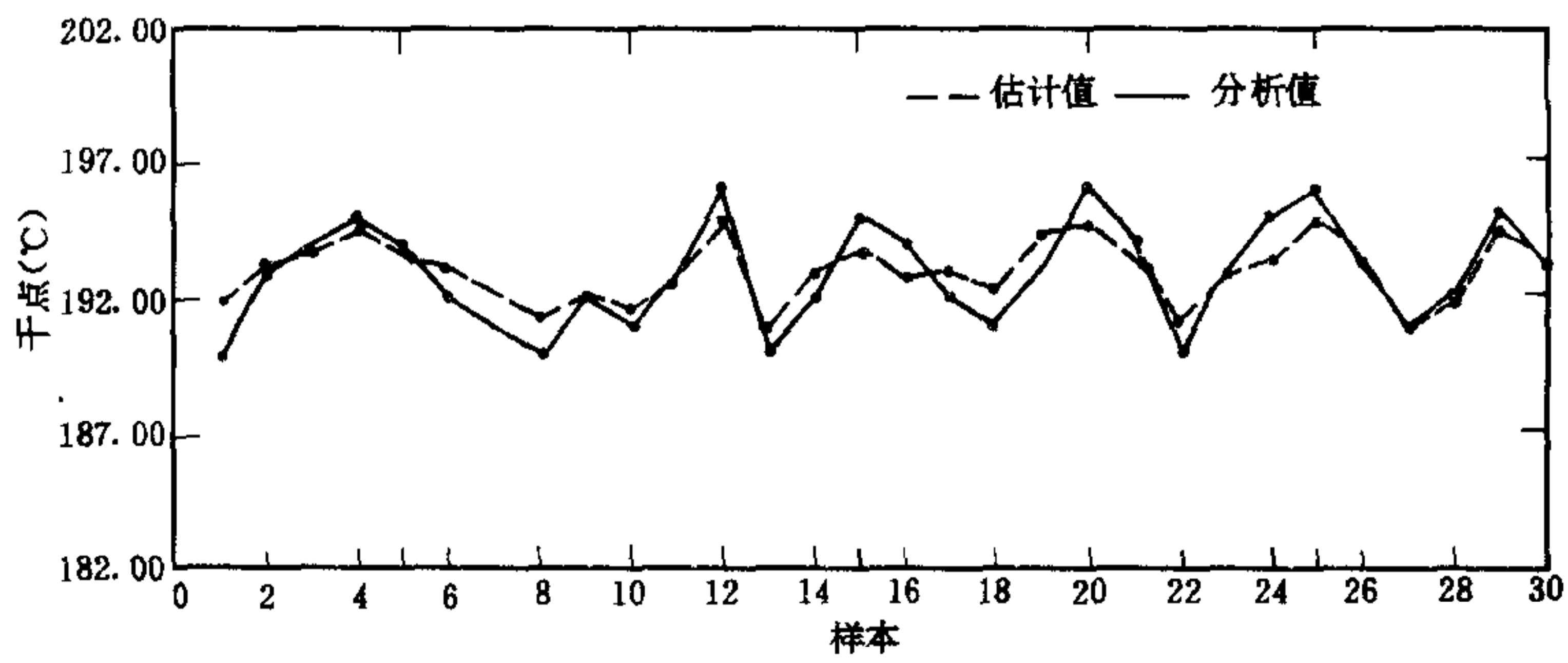


图 3 优化汽油干点网络模型的学习精度

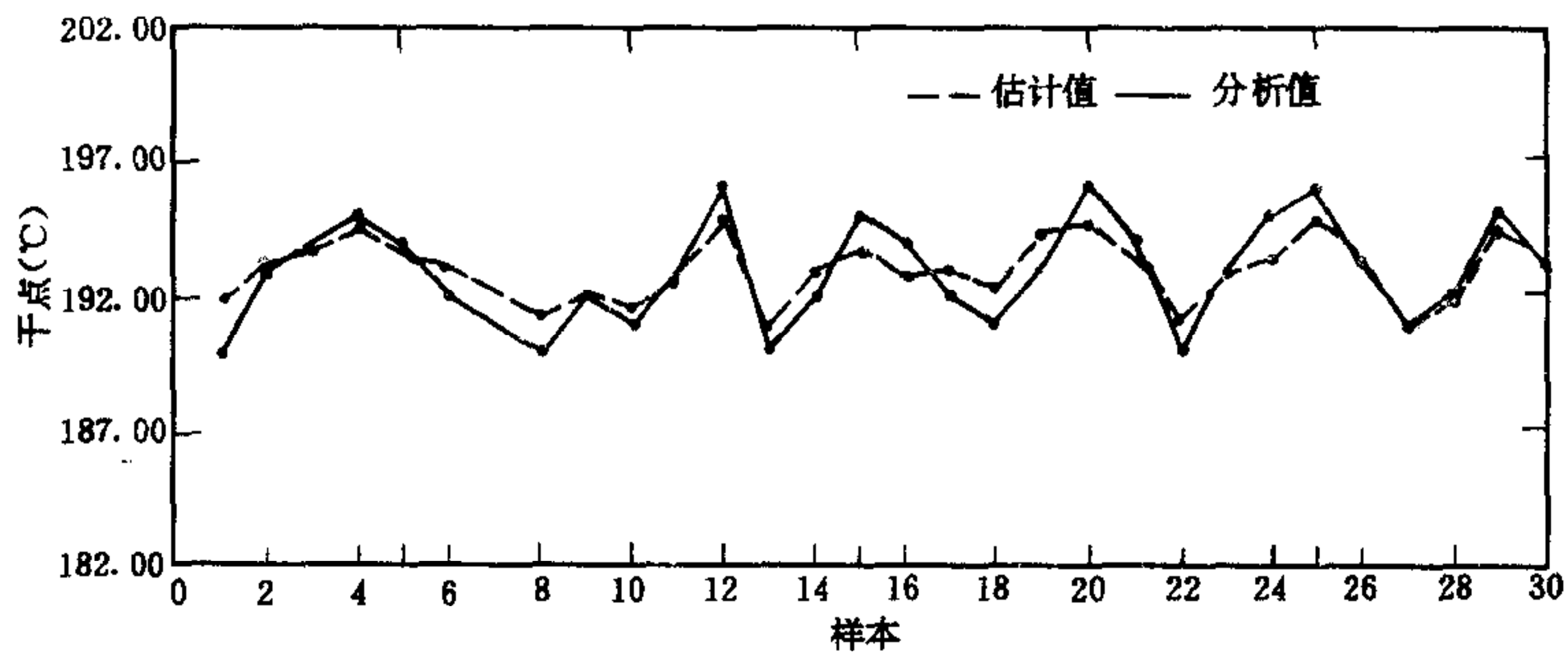


图 4 优化汽油干点网络模型的检验精度

## 参 考 文 献

- [1] Holland J H. Genetic algorithms and the optimal allocations of trials. *SLAM Journal of Computing*, 1973, 2: 88—105.
- [2] Bhat N V, McAvoy T J. Determining model structure for neural models by network stripping. *Computers and Chem. Engng.*, 1992, 16(4): 271—281.
- [3] Goldberg D E. Genetic algorithms in search, optimization and machine learning. Mass: Addison-Wesley, 1989.
- [4] Davis L. Handbook of Genetic Algorithms. London: Pitman, 1991.

## RESEARCH ON MULTILAYER FEEDFORWARD NEURAL NETWORKS ——GENETIC BACKPROPAGATION ALGORITHM & STRUCTURE OPTIMIZATION STRATEGY

CHEN RONG XU YONGMAO

(Department of Automation, Tsinghua University, Beijing 100084)

LAN HONGSEN

(Fujian Refinery, Fujian Hui'an 362100)

**Abstract** In this paper, when classical genetic algorithms are applied to the region of real number and integrated with BP algorithm, a hybrid GA-BP learning algorithm is proposed. Besides, with the aid of genetic algorithms as well as the improvement of network complexity function, a structure optimization strategy which is on the basis of network stripping is proposed. The efficiency of research work mentioned above has been shown by numerical simulations and industrial applications.

**Key words** Artificial neural networks, genetic algorithms, backpropagation algorithm, structure optimization.

**陈 荣** 1969 年生于江苏省南京市, 1992 年获清华大学生产过程自动化学士学位和环境工程学士学位, 1994 年获该校自动控制理论及应用专业硕士学位. 目前在该校自动化系任教, 主要从事人工神经网络、过程建模与优化、连续过程 CIMS 等方面的研究工作.

**徐用懋** 清华大学自动化系教授, 博士生导师, 长期从事过程控制的教学工作, 讲授“过程控制”等课程, 科研方向是工业过程建模、优化及先进控制. 承担了国家多项重点科技攻关课题, 多次获部、委科技进步奖. 近期专著有《模糊理论和神经网络的基础与应用》, 近年来发表论文 40 余篇.