

# 数据融合算法在农药残毒检测系统中的应用<sup>①</sup>

邵金涛, 秦宁宁, 钱菲菲, 徐保国

(江南大学 物联网工程学院, 无锡 214122)

**摘要:** 农药残毒检测在食品安全监管的过程中具有重要意义。通过自行设计的微型农药残毒快速检测仪对不同浓度的敌百虫溶液进行了多组测量。试验结果表明, 大于最低检出值的溶液浓度与发光强度之间呈线性关系。在概率融合意义下定义模糊支持区间和模糊支持概率, 将一种适合于单传感器重复采样的数据融合方法应用于农药残毒检测数据的处理中, 克服了传统数字滤波方法在抗干扰方面的不稳定性, 达到快速、准确检测的目的。

**关键词:** 农药残毒; 单传感器; 数据融合; 稳定性; 抗干扰

## Data Fusion Algorithm in Pesticide Residue Testing System

SHAO Jin-Tao, QIN Ning-Ning, QIAN Fei-Fei, XU Bao-Guo

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

**Abstract:** The examination of agricultural chemicals and residual toxicity play an important role in food safety control. Different concentration of dipterex were measured with the mini agricultural chemicals and residual toxicity fast examination instrumentation designed by our group. The result indicated that the linear relationship between the concentration of tested agricultural chemicals and luminous intensity value were established. Defines the fuzzy support sector and the fuzzy support probability in the probability fusion significance, and applied one kind in the single sensor redundant sampling data fusion method in agricultural chemicals residual poison examination data processing, which successfully overcome the tradition digital filtering method in the aspect instability, and realize the aim of fast and accurate test the residual toxicity in samples.

**Key words:** agricultural chemicals residual poison; single sensor; data fusion; stability; resistance to interference

## 1 引言

在解决农药残毒导致生态系统的结构和功能严重失衡的问题中, 农药残毒快速检测具有重要的意义<sup>[1-3]</sup>。近几年, 国内测定果蔬农药残毒含量普遍使用质谱和色谱检测器, 检测精度较高, 但前期准备时间长、测量过程繁琐, 很难实现在线检测, 也给实现检测仪的微型化带来困难<sup>[4]</sup>。本文以 MSP430 单片机为核心芯片, 自行研制出微型农药残毒快速检测仪测量了果蔬中敌百虫等多种有机磷类农药的残留含量, 结果表明, 该方法具有较好的抗干扰能力。

在对采集信息的处理过程中, 进行检测的仪器、环境、同一水平下的检测次数以及人员操作流程等都

不能保证检测数据能够完全反映农药残留的真实状态和数量关系, 因而会造成信息的不确定性。而利用一种具有较高稳定性的适合智能仪器检测的数字滤波算法对农药残毒检测数据的处理有很重要的意义。本文用单传感器(光电倍增管)重复采样, 依据传感器的有限次检测数据, 在概率融合意义下定义模糊支持区间和模糊支持概率, 得到一种改进的数据融合算法, 快速、稳定的计算出数据相对误差最小的检测结果。

## 1 检测原理与试验方法

### 1.1 检测系统的设计原理

农药残毒检测系统利用“光转化—化学发光法”对

① 收稿时间:2011-05-07;收到修改稿时间:2011-05-30

农药残毒进行检测, 可将测得的发光强度信号转化成电信号<sup>[5,6]</sup>, 从而间接得到农药残毒含量。图 1 是依据上述原理而设计的农药残毒检测系统的原理图。由微通道将预处理的样品引入进样阀, 经过分离柱分离, 流出液与另一通道加入的发光体系试剂混合, 经光电倍增管实现光电转换, 电信号通过 A/D 转换电路后进入单片机, 再经串口电路将转化后的农药残毒信息送入计算机进行数据融合处理。

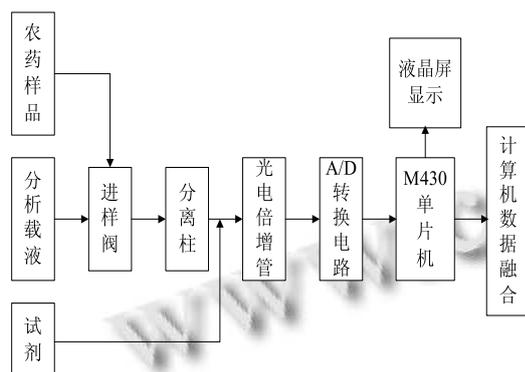


图 1 农药残毒检测系统原理图

## 1.2 光反应原理与试验方法

由 1.1 节提出的利用“光转化-化学发光法”对农药残毒进行检测, 其检测系统的光反应原理为: 乙酞胆碱酯酶对鲁米诺-过氧化氢化学发光体系具有催化作用, 加入有机磷农药会对其产生抑制作用, 使鲁米诺-过氧化氢-乙酞胆碱酯酶化学发光体系产生的光信号降低。基于上述原理可以对有机磷农药进行残留含量检测。论文选用敌百虫农药作为被检测的样品, 配备成不同浓度的敌百虫工作溶液 (0mg/L, 0.05mg/L, 0.07mg/L, 0.09mg/L, 0.15mg/L, 0.25mg/L, 0.30mg/L, 0.35mg/L) 进行检测。

在实验开始前, 由于环境温度、热辐射和其他因素的影响, 即使没有光信号的输入, 光电倍增管上电以后仍然会产生暗电流。为消除暗电流的影响, 需进行农药残毒快速检测仪的光电转换部分电压调零。调零完成后, 将配备好的不同浓度的敌百虫溶液 2mL 通过分离柱, 取鲁米诺-过氧化氢体系 3mL 与分离后的敌百虫溶液混合, 经光电倍增管测出光强 (电压信号) 的变化。在每种浓度下进行 6 次重复采样, 样品浓度与光强 (电压信号) 之间的输入—输出特性如表 1 所示。

表 1 样品浓度与光强 (电压信号) 输入—输出

样品浓度 (mg/L)	检测次数					
	1	2	3	4	5	6
0	1.17	1.19	1.19	1.15	1.20	1.19
0.05	1.17	1.20	1.13	1.19	1.21	1.17
0.07	1.20	1.17	1.19	1.20	1.15	1.15
0.09	1.04	1.00	1.01	1.05	1.05	1.08
0.15	0.85	0.83	0.88	0.81	0.85	0.85
0.25	0.53	0.50	0.49	0.47	0.50	0.51
0.30	0.30	0.29	0.25	0.29	0.32	0.29
0.35	0.07	0.04	0.05	0.04	0.06	0.07

## 2 数据融合处理流程

在检测仪对敌百虫等有机磷农药残毒信号进行检测时, 重复采样得到一检测列, 可计算出农药残毒含量的算术平均值。当检测列中包含海量数据时, 检测列的算术平均值是最理想的检测结果<sup>[7]</sup>。但在线测量要求检测过程具有快速性, 需通过较少的检测次数得到被测样品的结果。然而当检测次数较少时, 部分检测数据 (不能完全反映样品的真实状态和数量关系) 存在不确定性; 算术平均值的准确性和重复性受到影响, 不再是最理想的检测结果。

论文利用统计学理论的数据分组和适合于单传感器重复采样的数据融合方法, 首先将通过光电倍增管检测得到的同一浓度水平下的数据进行分组, 然后定义模糊支持区间计算出数据间的综合支持度, 进而得到融合后的数据。此融合方法可获得更加精确的检测结果。

### 2.1 基于统计学理论的数据分组

分别为每种浓度的敌百虫工作溶液做 6 次重复检测, 对历史数据累计分组。第一组为前两次检测得到的数据  $[x_1, x_2]$ , 将第三次检测得到的数据  $x_3$  与历史数据  $[x_1, x_2]$  合并后作为第二组数据  $[x_1, x_2, x_3]$ , 以此类推, 在同一浓度水平可以得到 5 组数据。分别对五组数据进行数据融合处理, 得到的 5 个融合后的新数据作为新的一组再次进行数据融合, 获得精确度较高的检测结果。

### 2.2 数据的相互支持度

在农药残毒检测数据  $x_1, x_2, \dots, x_n$  中, 如果  $x_i$  的检测精确度越高, 则  $x_i$  被其他数据所支持的程度也越高。假如  $w_{ij}$  表示  $x_i$  被  $x_j$  支持的程度, 则  $|x_i - x_j|$  越小, 支持度  $w_{ij}$  越大。当  $w_{ij} = 1$  或  $w_{ij} = 0$  时, 数据间的相互支持程度容易区分; 而当  $w_{ij}$  处于两者之间时, 较难区分数据间的相互支持程度<sup>[8,9]</sup>。

针对敌百虫农药残毒检测数据的处理, 设定参数  $\varepsilon_1$  和  $\varepsilon_2$ , 其中  $0 \leq \varepsilon_1 < \varepsilon_2$ 。如果  $|x_i - x_j| \leq \varepsilon_1$ , 则  $x_i$  认为完全被  $x_j$  支持,  $w_{ij} = 1$ ; 如果,  $|x_i - x_j| \geq \varepsilon_2$  则认为  $x_i$  完全不被  $x_j$  支持,  $w_{ij} = 0$ ; 如果,  $\varepsilon_1 < |x_i - x_j| < \varepsilon_2$  则认为  $x_i$  部分  $x_j$  被支持,  $w_{ij} = f(|x_i - x_j|)$ 。由此可得  $x_i$  被  $x_j$  支持的程度为:

$$\begin{cases} w_{ij} = 0, |x_i - x_j| > \varepsilon_2, \\ w_{ij} = f(|x_i - x_j|), \varepsilon_1 \leq |x_i - x_j| \leq \varepsilon_2, \\ w_{ij} = 1, |x_i - x_j| < \varepsilon_1, \end{cases}$$

其中,  $i, j = 1, 2, \dots, n$ 。  $f(|x_i - x_j|)$  为严格单调下降的连续函数, 当  $x \in [\varepsilon_1, \varepsilon_2]$  时  $0 \leq f(|x_i - x_j|) \leq 1$ ,  $f(\varepsilon_1) = 1$ ,  $f(\varepsilon_2) = 0$ 。如果能够确定  $f(|x_i - x_j|)$  则可以确定相互支持度  $w_{ij}$ , 确定  $f(|x_i - x_j|)$  的一般方法为:

- 1) 确定  $\varepsilon_3 \in (\varepsilon_1, \varepsilon_2)$ , 使得  $f(\varepsilon_3) = 0.5$ ;
- 2) 若  $x \in (\varepsilon_1, \varepsilon_3)$ , 则  $f(x) = 0.5 - (x - \varepsilon_3) / 2(\varepsilon_3 - \varepsilon_1)$ ;

若  $x \in (\varepsilon_3, \varepsilon_2)$ , 则  $f(x) = 0.5 - (x - \varepsilon_3) / 2(\varepsilon_2 - \varepsilon_3)$ 。

从而得到  $x_i$  被  $x_j$  的支持度  $w_{ij}$ , 定义矩阵  $W = (w_{ij})$  为  $x_1, x_2, \dots, x_n$  的相互支持矩阵。

### 2.3 数据的综合支持度

$x_i$  的可信程度并不是由  $w_{ij}$  单纯决定的, 而是  $w_{i1}, w_{i2}, \dots, w_{in}$  的综合。定义  $w_i$  为  $x_i$  的综合支持程度, 则  $w_i$  越大,  $x_i$  的可信程度越高。因此, 数据融合的结果将由  $w_i$  所决定。  $w_i$  需要满足以下两个条件:

- 1)  $0 \leq w_i \leq 1$  且  $\sum_{i=1}^n w_i = 1$ ;
- 2) 由概率源合并理论, 需要存在一组非负数  $\alpha_1, \alpha_2, \dots, \alpha_n$ , 使得  $w_i = \alpha_1 w_{i1} + \alpha_2 w_{i2} + \dots + \alpha_n w_{in}$ , 其中  $i = 1, 2, \dots, n$ 。

将  $w_i$  的关系式改写为矩阵形式  $w = W\alpha$ , 其中  $w$  为  $w_1, w_2, \dots, w_n$  的列向量,  $\alpha$  为  $\alpha_1, \alpha_2, \dots, \alpha_n$  的列向量。因为  $w_{ij}$  为非负数, 所以  $W$  是一个非负矩阵, 存在  $W$  的最大模特征值  $\lambda$  使得  $W\alpha = \lambda\alpha$ ,  $\alpha$  为  $\lambda$  所对应的特征向量。从而  $\lambda\alpha$  可以作为农药残毒检测数据的衡量标准, 令  $w_i = \alpha_i / (\alpha_1 + \alpha_2 + \dots + \alpha_n)$ ,  $w_i$  即为  $x_i$  与所有检测数据之间的综合支持度。

通过计算出的数据间综合支持度, 对于  $n$  个检测数据可知其融合数据为

$$x = w_1 x_1 + w_2 x_2 + \dots + w_n x_n \quad (1)$$

### 2.4 融合数据曲线拟合

通过农药残毒检测系统得到的融合数据必须定量

的表示成微处理机 (MSP430) 所能够识别的数学表达式形式, 因此要对融合后的数据进行曲线拟合, 便于微处理机对检测的数据进行判决。通常先验数据如果能够反映检测的真实性, 则拟合数据构成的曲线能够呈现或部分的呈现线性特征, 基于此, 论文采用线性迭代建模方法对融合后的数据进行曲线拟合。

不失一般性, 论文假设融合数据的一元线性回归关系式为:

$$y' = ax + b \quad (2)$$

线性迭代建模中对模型参数进行优化估计的最普遍的方法是小二乘法, 对式(2)中的模型参数  $a$  和  $b$ , 假设要找的  $a$  和  $b$  的最佳估计值为  $\tilde{a} \neq \tilde{b}$ , 则

$$\tilde{a} = \frac{S_{xx}S_y - S_{xy}S_x}{\gamma S_{xx} - S_x^2} \quad (3)$$

$$\tilde{b} = \frac{\gamma S_{xy} - S_x S_y}{\gamma S_{xx} - S_x^2} \quad (4)$$

其中  $\gamma$  为样品浓度与光强关系数组的点数,  $S_x = \sum_{i=1}^{\gamma} x_i$ ,

$$S_{xx} = \sum_{i=1}^{\gamma} x_i^2, \quad S_y = \sum_{i=1}^{\gamma} y_i, \quad S_{xy} = \sum_{i=1}^{\gamma} x_i y_i。$$

每个实际检测值到拟合后的相对误差为:

$$e_i = y_i - y'_i \quad (5)$$

则相对误差平方和为

$$J(a, b) = \sum_{i=1}^{\gamma} e_i^2 = \sum_{i=1}^{\gamma} (y_i - b - ax_i)^2 \quad (6)$$

有必要说明的是, 个别情况下得到的融合数据线性特征不明显, 可利用增加先验数据提高数据的线性特征, 进而进行线性拟合; 或者, 直接采用多元回归法等非线性方法进行直接拟合。

## 3 应用案例

利用上述数据融合方法, 对农药残毒检测系统中单传感器 (光电倍增管) 采集到的数据进行信息融合。此时所取参数  $\varepsilon_1 = 0, \varepsilon_2 = 0.30, \varepsilon_3 = 0.15$ 。以表 1 中 0mg/L 的样品浓度为例, 第一组数据 [1.17, 1.19] 所得支持矩阵

$$W = \begin{bmatrix} 1.0000 & 0.9333 \\ 0.9333 & 1.0000 \end{bmatrix}$$

最大模特征值  $\lambda = 1.9333$ , 其对应的正的特征向量  $\alpha = (0.7071, 0.7071)$ , 从而得到第一组数据的综合支持度  $w = (0.50, 0.50)$ 。利用式 1 可得第一组数据的融合数据为 1.1800V。同样可以得到同一样品浓度水平下的其他四组测量数据的融合数据分别为: 1.1993V,

1.1865V, 1.1894V, 1.1897V。五个融合后的数据组成新的一组再次融合, 其对应的支持矩阵

$$W = \begin{bmatrix} 1.0000 & 0.9357 & 0.9783 & 0.9687 & 0.9677 \\ 0.9357 & 1.0000 & 0.9573 & 0.9670 & 0.9680 \\ 0.9783 & 0.9573 & 1.0000 & 0.9903 & 0.9893 \\ 0.9687 & 0.9670 & 0.9903 & 1.0000 & 0.9990 \\ 0.9677 & 0.9680 & 0.9893 & 0.9990 & 1.0000 \end{bmatrix}$$

最大模特征值  $\lambda = 4.8889$ , 其对应的特征向量  $\alpha = (0.4437, 0.4416, 0.4497, 0.4506, 0.4505)$ , 从而得到综合支持度  $w = (0.1984, 0.1975, 0.2011, 0.2015, 0.2017)$ 。计算可得在 0mg/L 样品浓度水平下的光强 (电压信号) 融合数据为 1.1896V。同样可以计算出其他样品浓度水平下的光强数据的融合信息分别为: 1.1916V, 1.1767V, 1.0200V, 0.8145V, 0.5139V, 0.2767V, 0.0546V。

根据得到的融合数据, 利用最小二乘法进行处理, 其关系趋势图如图 2 所示。由图可以看出, 曲线在 0.07mg/L 处的电压信号有明显变化, 故最低检出率为 0.07mg/L; 曲线在 0.09mg/L-0.35mg/L 处呈线性关系。根据 2.4 节, 可得上述融合数据的拟合曲线关系为  $y = -3.456x + 1.324$ ,  $J(-3.456, 1.324) = 0.0315$ 。

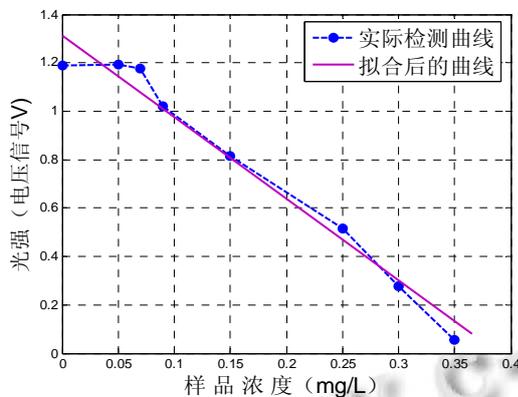


图 2 样品浓度—光强 (电压信号) 趋势图

通过上述融合结果可以发现本文的数据融合方法比传统数字滤波方法具有更高的可操作性, 该方法将数据间的相互支持程度模糊化, 更能反映农药残毒检测快速检测仪所采集数据环境的真实性; 同时该方法

可以有效解决农药残毒数据组中某些数据偏差大的问题, 比传统数字滤波方法具有更强的抗干扰性。

#### 4 总结

本系统通过综合运用化学发光技术、光电转换技术和单片机技术, 不仅可以建立快速检测农药残毒的新方法, 同时将为研制微型化仪器提供理论依据。论文将一种适合于单传感器重复采样的数据融合方法应用于农药残毒检测数据的处理中, 利用统计学理论对数据分组, 同时将支持概率模糊化, 对支持区间的确定采用分段讨论, 使融合数据比以往利用绝对化支持概率和支持区间所得结果更加稳定, 进而保证数据融合具有更好的抗干扰性。

#### 参考文献

- 1 杜习光. 药残留光电快速检测系统研究. 重庆: 西南大学, 2010.
- 2 Liu Y, Song QJ, Wang L. Development and characterization of an amperometric sensor for triclosan detection based on electropolymerized molecularly imprinted polymer. *Microchemical Journal*, 2009, 91(3): 222-226.
- 3 Chen ZL, Song QJ, Cao GQ, Chen YF. Photolytic degradation of triclosan in the presence of surfactants. *Chemical Papers* 2008, 62(6): 608-615.
- 4 刘萍, 张进忠. 有机磷农药残留检测技术研究进展. *环境污染与防治*, 2006, 1(1): 55-66.
- 5 浦昭邦. 光电测试技术. 北京: 机械工业出版社, 2005. 19-21.
- 6 郭培源, 付扬. 光电检测技术与应用. 北京: 北京航空航天大学出版社, 2006. 10-11.
- 7 李小昱, 雷廷武, 王为, 张军. 基于统计理论数据融合的水流泥沙含量测量仪. *农业工程学报*, 2004, 20(5): 110-113.
- 8 涂国平, 叶素萍. 一种传感器数据的融合算法. *传感器技术*, 2003, 22(3): 30-32.
- 9 付华, 刘亚宁. 基于 ARM 和数据融合技术的网络化矿井安全监测系统. *计算机系统应用*, 2007, 16(5): 60-62.

# 混合蚁群算法在城市交通路径规划中的应用<sup>①</sup>

陈家琪, 刘 跃

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘 要:** 为诱导车辆在城市交通路网中规划较优出行路线以提高出行质量, 研究并建立了城市道路权值模型。此外, 分析比较粒子群算法和蚁群算法各自优缺点, 将两个算法分别做了改进, 提出一种混合粒子群蚁群算法, 并利用道路权值模型对三种算法在路径寻优效果上做了比较和分析。实验结果表明该改进的混合算法在路径规划准确度和算法计算效率上均优于粒子群和蚁群算法, 取得了较好的效果。

**关键词:** 粒子群算法; 蚁群算法; 道路权值模型

## Application of Hybrid Ant Colony Optimization to Urban Transportation Path Planning

CHEN Jia-Qi, LIU Yue

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** In order to guide vehicles to plan better travel routes for improving the quality of travel in urban transportation network, this paper studies and establishes the urban road weights model. Besides, it analyzes and compares the advantages and disadvantages of particle swarm optimization and ant colony optimization. This two algorithms are improved respectively and integrated into one algorithm. The urban road weights model is used in analyzing and comparing the effects of roads planning by three algorithms. The experimental results show that the improved hybrid algorithm is superior to other two algorithms in path planning accuracy and computation efficiency.

**Key words:** particle swarm optimization; ant colony optimization; road weights model

随着城市规模和相配套的城市交通变得日益庞大和复杂, 与此同时, 以车辆为载体的人们对出行质量的要求也越来越高。在这现实矛盾下, 采用有效手段实时诱导车辆规划合理行车路线, 降低出行成本就显得尤为重要。在路径规划研究中, 核心是路径规划算法, 一个算法的好坏直接影响着路径规划质量的优劣。目前, 作为一种群智能优化算法的粒子群算法(Particle Swarm Optimization, PSO)<sup>[1]</sup>和一种模拟进化算法的蚁群算法(Ant Colony Optimization, ACO)<sup>[2]</sup>在许多领域都有很好的应用, 尤其在路径搜索和网络路由策略等领域应用广泛。前者是 1995 年由 Kenndy 和 Eberhart 等人提出, 此算法具备设置参数少、实现简单等优点。后者由 Marco Dorigo 等人在 20 世纪 90 年代提出, 该算法具有较强的鲁棒性、正反馈、并行性

和分布式协作等特点, 并且易于与其它算法结合, 在求解二次分配问题<sup>[3]</sup>、旅行商问题<sup>[4]</sup>等方面相比遗传算法、模拟退火算法等具有很大的优势。同样, 粒子群算法和蚁群算法也有不少缺点, 如粒子群算法收敛速度慢, 蚁群算法搜索时间长、易陷入局部优先和出现停滞现象等问题。本文针对上述算法不足和实际交通路径规划对算法计算时间和精准度上的高要求, 分别对粒子群和蚁群算法做出改进, 提出改进型混合算法, 充分发挥两个算法在路径搜索上的优势。

### 1 城市道路权值模型

传统最优路径指的是车辆在起点和终点间选择一条距离最短的路径。这时道路权值仅和行车距离有关。目前很多最优路径搜索研究是基于图论以简单的距离

① 收稿时间:2011-05-07;收到修改稿时间:2011-07-08

长短来衡量路径是否最优,市场上绝大多数导航仪也是基于这个准则生产的。然而在实际背景下,道路权值需要考虑如道路等级、拥挤程度、路段限速、突发交通事故、道路临时交通管制、过路费、燃油费等诸多因素,这些因素制约着驾驶员对路段的选取,从而影响最后最优路径的结果。为更加贴近实际情况,将道路权值定为出行时间和出行费用的线性组合,提出如下权值模型:

$$w_{ij} = \begin{cases} \lambda_1 t_{ij} + \lambda_2 f_{ij}, & \text{允许通行} \\ \infty, & \text{禁止通行} \end{cases} \quad (1)$$

式(1)中禁止通行条件为以*i*、*j*为起止节点的*ij*路段临时交通管制或该路段在交叉口选择中为禁止转向的路段; $w_{ij}$ 为*ij*路段的权值; $t_{ij}$ 为车辆在*ij*路段的行驶时间和路口*j*停车等待时间, min;  $f_{ij}$ 为车辆在*ij*路段的行驶费用, yuan;  $\lambda_1$ 、 $\lambda_2$ 为时间费用权衡系数,分别为/min和/yuan。根据文献[5]研究,  $\lambda_1=1$ ,  $\lambda_2=2$ 。

介于城市内部交通网中收费站较少,与整个出行时间相比微乎其微,暂不考虑车辆在收费站延误时间。此时对于  $t_{ij}$  有:

$$t_{ij} = t'_{ij} + \Delta t_j \quad (2)$$

式(2)中  $t'_{ij}$  为车辆在 *ij* 路段实际行驶时间, min;  $\Delta t_j$  为车辆在 *j* 节点交叉口停车延误时间, sec。其中对于  $t'_{ij}$  的描述为:

$$t'_{ij} = \frac{L_{ij} [1 + (V_{ij} / C_{ij})^{\alpha_{2ij} + \alpha_{3ij}} (V_{ij} / C_{ij})^3]}{\alpha_{1ij} U_{sij}} \quad (3)$$

式(3)中,  $L_{ij}$  为 *ij* 路段的长度, km;  $V_{ij}$  为 *ij* 路段的实际交通量, veh/h;  $C_{ij}$  为 *ij* 路段通行能力, veh/h;  $U_{sij}$  为 *ij* 路段设计车速, km/h;  $\alpha_{1ij}$ 、 $\alpha_{2ij}$ 、 $\alpha_{3ij}$  为 *ij* 路段回归参数。各等级道路的  $C_{ij}$ 、 $\alpha_{1ij}$ 、 $\alpha_{2ij}$ 、 $\alpha_{3ij}$  的取值见文献[6]。交叉口停车延误国内研究不是很多,对有信号灯控制交叉口通常采用 Webster 延误模型,对无信号灯控制交叉口延误模型参见文献[7]。介于城市交通交叉口通常有信号灯控制这一情况,采用 Webster 延误模型,其公式为:

$$\Delta t_j = \frac{P_j(1-g_j)^2}{2(1-g_jx)} + \frac{x^2}{2V_{ij}(1-x)} - 0.65(P_j/V_{ij}^2)^{1/3} x^{(2+5g_j)} \quad (4)$$

式(4)中  $P_j$  为 *j* 节点交叉口信号灯周期长度, sec;

$g_j$  为车道绿信比;  $x_j$  为饱和度,  $V_{ij}/C_{ij}$ ;  $V_{ij}$  为实际交通流量, veh/h。对于出行费用  $f_{ij}$  来说, 有

$$f_{ij} = L_{ij}r + M_{ij} \quad (5)$$

式(5)中  $r$  为单位路程油耗费用, yuan/km;  $M_{ij}$  为 *j* 节点过路费, yuan。此时最优路径定义为起止点时间费用综合最小的路径,并且在以下算法改进研究中均将道路权值由原有单一距离改为时间和费用的线性组合。

## 2 粒子群蚁群混合算法及改进

实时路径规划,这对核心搜索算法除搜索准确的同时也对算法计算效率提出了更高的要求。蚁群算法具有出色的鲁棒性、运用正反馈、并行性和分布式协作等特点,是一种新的进化算法,运用到路径搜索方面具有很大的优势。但是该算法初期信息素相同,蚁群在路径上形成差异性的信息素需要很长一段时间。针对这一缺点,引入一种基于群集智能优化的粒子群算法,并对两个算法做了改进和融合。在路径搜索初期采用具有较强全局搜索能力、参数少、实现容易的粒子群算法进行粗搜索,并增加最优路径上的初始信息素浓度,后期则改成具有较强鲁棒性、正反馈的蚁群算法进行细搜索。

### 2.1 粒子群算法描述及改进

在粒子群算法中,将每个优化问题的可能解定义为粒子(Particle)。粒子在搜索时根据自身飞行经验和同伴飞行经验动态调整飞行速度。数目为  $S$  的粒子都有一个被优化的函数决定的适应值(fitness value),并且知道自身到目前为止发现的最好位置(particle best, Pbest)和当前位置,这是粒子自身的飞行经验。此外,每个粒子都知道到目前为止群体中所有粒子发现的最好位置(global best, Gbest),这是粒子同伴的飞行经验。粒子在每一次迭代中根据自身位置、当前速度、当前位置与 Pbest 间的距离、当前位置与 Gbest 间的距离来改变自身的搜索位置。最后经过多次迭代找到问题的最优解。Shi 和 Eberhart<sup>[8]</sup>于 1998 年提出标准粒子群算法,该算法的数学描述为:

$$V_i^{k+1} = \omega V_i^k + c_1 r_1 (P_{best}^k [i] - x_i^k) + c_2 r_2 (G_{best}^k [g] - x_i^k) \quad (6)$$

$$x_i^{k+1} = x_i^k + V_i^{k+1} \quad (7)$$

式(6)(7)中,  $\omega$ 为惯性权重;  $c_1$ 、 $c_2$  为非负常数的学习因子;  $r_1$ 、 $r_2$  为[0,1]间均匀分布的随机数;  $V_i^k$  和  $x_i^k$  分别为第  $k$  次迭代粒子  $i$  当前速度和位置;  $P_{best}^k$  和  $G_{best}^k$  分别为第  $k$  次迭代粒子  $i$  当前找到的最好位置和第  $k$  次迭代整个种群当前找到的最好位置; 此外,  $V_i^k \in [V_{min}, V_{max}]$ , 若  $V_{ik}$  超过  $V_{max}$  时将  $V_i^k$  限制为  $V_{max}$ 。

虽然标准 PSO 加入惯性权重改善了原始算法的收敛性能, 但 PSO 在实际搜索过程中是非线性且高度复杂的, 致使惯性权重线性递减的策略不能反映实际的搜索过程。文献[9]表明, 在初始权值和最终权值相同的情况下,  $\omega$ 凹函数递减策略优于线性策略, 而线性策略优于凸函数策略, 凹函数能在不影响算法收敛精度的情况下较大幅度提高算法的收敛速度。将惯性权重改为如式(8)所示,

$$\omega(g) = \omega_s - (\omega_s - \omega_e) \sqrt{\frac{g}{g_{1max}}} \quad (8)$$

式(8)中,  $\omega_s$  和  $\omega_e$  分别为初始迭代惯性权值和终止迭代惯性权值;  $g$  为算法当前迭代次数;  $g_{1max}$  为粒子群算法最大迭代次数; 经过 Shi 等大量计算,  $\omega_s$  和  $\omega_e$  分别取 0.95 和 0.4 时会大大提高算法的收敛速度和求解精度。

### 2.2 蚁群算法描述及改进

在求解不同类型的问题, 如旅行商问题 (TSP)、车辆路径问题 (VRP)、二次分配问题 (QAP) 等时蚁群算法相应的模型描述也不尽相同, 但是基本原理都是一样的。首先初始化一组一定数量的蚂蚁, 每只蚂蚁在初始位置出发, 根据当前路径上信息素浓度和启发信息前行, 并通过一定的状态转移规则选择下一条合适的路径。同时, 蚂蚁在爬行过程中自身也分泌一种具有启发作用的信息素 (路径上留下的信息素也会随着时间有一定量的挥发), 以此一直爬行到终点。这时每只蚂蚁所走过的路线为一个可行解, 该组所有蚂蚁完成一次搜索为一次迭代, 算法经多次迭代后找到理想的全局最优解。为描述方便, 本文以  $n$  个城市的 TSP 问题为例介绍基本蚁群算法。旅行商从某一城市出发, 要访问  $n$  个城市中的每一个城市有且只能访问一次, 结果要求以最小旅行距离回到起点。蚂蚁  $k$  在城市节点  $i$  时, 选

择城市  $j$  的状态转移概率为:

$$P_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{s \in allowed} [\tau_{is}(t)]^\alpha [\eta_{is}]^\beta}, & j \in allowed \\ 0, & others \end{cases} \quad (9)$$

式(9)中  $allowed$  为该蚂蚁可访问的城市节点集;  $\tau_{ij}(t)$  为  $t$  时刻  $ij$  路段上信息素浓度;  $\eta_{ij}$  为启发函数, 反映蚂蚁由节点  $i$  转移到  $j$  的期望程度, 一般为  $1/L_{ij}$ ;  $\alpha$  和  $\beta$  分别为信息素启发因子和期望启发因子。蚂蚁  $k$  爬过的  $ij$  路径上留下相应的信息素增量

$$\Delta \tau_{ij}^k(t) = Q / L_{ij} \quad (10)$$

式(10)中  $Q$  为常量, 表示信息素浓度;  $L_{ij}$  为  $ij$  路径长度。若蚂蚁  $k$  未经过  $ij$  路径, 则  $\Delta \tau_{ij}^k$  为 0。当该蚂蚁搜索完  $n$  个城市后, 信息素局部更新规则为:

$$\tau_{ij}^k(t+1) = (1-\rho)\tau_{ij}^k(t) + \rho\tau_0 \quad (11)$$

式(11)中,  $\rho$  为信息素挥发系数;  $\tau_0$  为各条路径上信息素的初始值。当  $m$  只蚂蚁都完成搜索时  $ij$  路径上增加的信息素总量为:

$$\Delta \tau_{ij}(t) = \sum_{k=1}^m \Delta \tau_{ij}^k(t) \quad (12)$$

当所有蚂蚁完成一次对  $n$  个城市遍历搜索时, 将  $m$  条路径中最优的一条进行信息素更新, 全局更新规则为:

$$\tau_{ij}(t+n) = (1-\rho)\tau_{ij}(t) + \rho\Delta \tau_{ij}(t) \quad (13)$$

为进一步提高蚁群算法在路径搜索的性能, 对其做了部分改进, 具体改进如下:

#### 1) 最大最小信息素

为在提高算法收敛速度的同时兼顾蚁群扩大全局搜索的能力, 引入最大最小信息素概念<sup>[12]</sup>。对某一时刻  $ij$  路径上的信息素  $\tau_{ij}(t)$  进行适当限制, 将  $\tau_{ij}(t)$  限制在  $[\tau_{min}, \tau_{max}]$  内, 大小及关系见<sup>[10]</sup>。若某一时刻  $\tau_{ij}(t) > \tau_{max}$ , 令  $\tau_{ij}(t) = \tau_{max}$ 。目的是抑制某条过热路径的信息素浓度, 避免后续蚁群跟风导致算法过早陷入局部最优; 若某一时刻  $\tau_{ij}(t) < \tau_{min}$ , 令  $\tau_{ij}(t) = \tau_{min}$ 。目的是增加某些过冷路径的信息素, 迫使蚁群扩大搜索范围, 寻找全局最优解, 避免算法出现局部优先解和停滞现象。

#### 2) 改进 $\tau_0$ 、 $\Delta \tau_{ij}^k(t)$ 、 $\eta_{ij}$ 和状态转移条件

随着迭代次数的增加, 蚂蚁搜索的路径解越来越优化。为发挥当前搜索最优路径解对后续蚁群的指导

作用，将原有静态启发函数改为动态启发函数，同时对  $\tau_0$  和  $\Delta\tau_{ij}^k(t)$  作出必要修改，见(14)(15)。

$$\tau_{0ij} = \begin{cases} \tau_0 + \Delta\tau_{ps0} & , ij \in L_{psobest} \\ \tau_0 & , others \end{cases} \quad (14)$$

$$\Delta\tau_{ij}^k(t) = Q / \omega_{ij} \quad (15)$$

$$\eta_{ij}(g) = \frac{1 + \mu(g/g_{2max})^2}{\omega_{ij}}, ij \in L_{best}(g) \quad (16)$$

式(16)中， $\eta_{ij}(g)$  为第  $g$  次迭代的启发函数； $g_{2max}$  为蚁群算法最大迭代次数； $L_{best}(g)$  为前( $g-1$ )次迭代中的最优路径； $\mu$  为迭代指导因子；状态转移概率中 **allowed** 改为当前节点蚂蚁在交叉口可以选择的下一节点集合，而非所有未经过的节点集合。

### 3) 改进信息素全局更新规则

为减少不必要的蚁群资源浪费在较差的路径上，本文将每次迭代中最差的一段搜索路径且  $\Delta\tau_{ij}(t)$  最小的那段路径上的信息素进行适当抑制，使后续蚁群朝着较优的路径行进，从而提高算法搜索最优解的效率。改进后的信息素全局更新规则如下：

$$\tau_{ij}(t+n) = \begin{cases} \Delta\tau_{ij}(t), ij \in L_{worst} \ \& \ \Delta\tau_{ij}(t)_{min} \\ (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t), others \end{cases} \quad (17)$$

## 3 示例仿真

本文以模拟城市局部交通网模型为测试对象，基于改进的粒子群蚁群混合算法，实验平台为 Visual studio2005，编程语言为 C++。为研究方便，将城市交通网简化为点线模型，假设车辆在行驶中能接收到交通网中各条路段路况信息(包括节点  $j$  属性( $P_j, g_j, x_j, V_{ij}, M_j$ ))，路段  $ij$  属性 ( $RL_{ij}, U_{sij}, C_{ij}, L_{ij}, V_{ij}, \alpha 1ij, \alpha 2ij, \alpha 3ij$ )，其中  $RL_{ij}$  为公路等级)，以便车辆在行驶中能够实时调整规划路径。规定在交通网模型中画线如 4-11 的为 1 级公路，10-11 为 2 级公路，3-4 为 3 级公路，2-9 为 4 级公路，如图 1 所示；收费站如交叉点 9，通过一次收费 10 元；处于不同等级公路的交叉点在计算  $P^i, g^j, x^j$  时均取其中最高等级路段的最大值；车辆行驶的起点为 1、终点为 49，路段 30-31 为交通管制路段，节点 38 禁止

转向 45。路径 1-2 上的 8.9 和 410 分别代表  $L_{ij}$  和  $V_{ij}$ 。在仿真前初始化道路参数和算法参数，见表 1。实验结果如表 2 和图 2 所示。

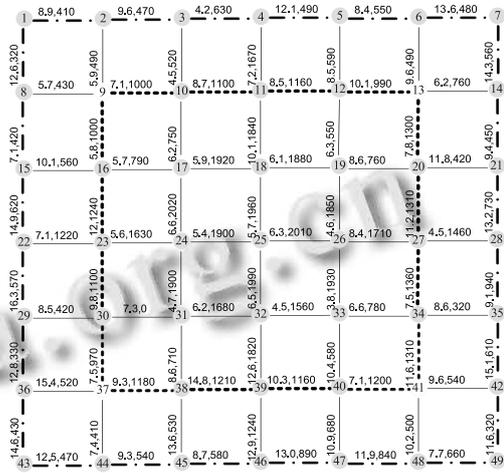


图 1 城市局部路网模型

表 1 道路模型参数及算法参数设置

$\lambda_1$	$\lambda_2$	$r$	$\omega$	$\omega_s$	$\omega_c$	$c_1$
1	2	0.6	1	0.95	0.4	2.0
$c_2$	$g_{1max}$	$V_{min}$	$V_{max}$	$S$	$m$	$g_{2max}$
2.0	150	0	$U_s$	50	50	250
$\mu$	$\rho$	$Q$	$\alpha$	$\beta$	$\tau_0$	$\Delta\tau_{ps0}$
2	0.2	100	1	5	0	0.2
RL	$U_s$	$P$	$g$			
1	100	80	0.6			
2	80	70	0.6			
3	60	60	0.6			
4	40	50	0.6			

表 2 三种算法求解性能比较

算法	最优路径	权值	时间	费用
PSO	1-2-3-4-			
	11-12-19-26-	242.4	128.4	57
	33-34-			
ACO	35-42-49			
	1-2-9-16-			
	17-24-31-	261.0	132.8	64.1
	32-33-34-			
	35-42-49			

	1-2-3-4-			
I-PSO-	11-18-25-	233.2	112.4	60.4
ACO	26-27-28-35-			
	42-49			

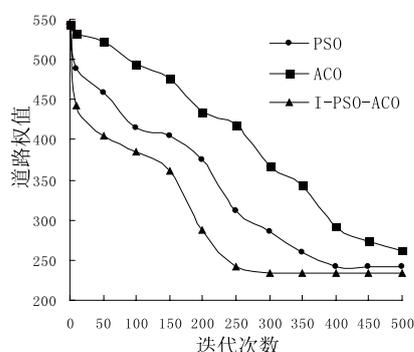


图2 三种算法收敛与求解情况

改进后的混合算法 (I-PSO-ACO) 虽然在行驶距离上略大于粒子群 (PSO) 和蚁群算法 (ACO), 但却有效避开流量大的路段而获得较少行驶时间, 该算法最优路径权值为 233.2, 相比于粒子群算法的 242.4 和蚁群算法的 261.0 更接近理论最优路径权值 228.8, 见表 2。因此改进后的算法在路径求解性能上优于另外两个算法。另外, 改进后的混合算法在迭代 150 后开始迅速收敛, 算法迭代到 250 次后开始出现最优解。而粒子群算法和蚁群算法分别在 400 次和 500 次迭代后才开始出现最优解, 见图 2。因此在时间效率上, 改进后的算法优于粒子群算法, 粒子群算法优于蚁群算法。

#### 4 结语

以城市局部路网模型为测试对象, 针对粒子群算法收敛速度慢、蚁群算法易陷入局部优先和收敛速度慢等缺点, 对两个算法做了相应改进。并结合粒子群算法较强的全局搜索能力和蚁群算法正反馈等优点, 将两个算法混合运用于最优路径搜索, 实验表明该改进后的混合算法相比粒子群算法和蚁群算法在交通路径规划求解准

确度和算法计算效率上都有显著的提高。

文中提到车辆在行驶过程中能接收到交通网中各条路段路况车况信息这一假设是基于高度信息化的城市交通网络之上, 限于当前交通信息化水平, 车辆在行驶中获取实时交通信息这一技术问题还有待解决。另外, 混合算法中参数众多, 大多数参数是根据以往研究经验设定, 对于不同问题合理搭配参数以达到最佳效果将是下一步具有重要研究意义的课题。

#### 参考文献

- 1 Kennedy J, Eberhart RC. Particle swarm optimization. Proc. of IEEE International Conference on Neural Networks. 1995: 1942-1948.
- 2 Dorigo M, Maniezzo V, Colomi A. Ant system: optimization by a colony of cooperating agents. IEEE Trans. on Systems, 1996, 26(1): 28-41.
- 3 Maniezzo V, Colomi A. The ant system applied to the quadratic assignment problem. IEEE Trans. on Knowledge Data Engineering, 1999, 11(5): 769-778.
- 4 Dorigo M, Gambardella LM. Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Trans. on Evolutionary Computation, 1997, 1(1): 53-66.
- 5 田璠, 基于道路拥挤收费的出行时间价值研究. 大连理工大学, 2009.
- 6 王炜, 邓卫, 杨琪, 等. 公路网络规划建设与管理方法. 北京: 科学出版社, 2006. 90-108.
- 7 王炜, 高海龙, 李文全, 等. 公路交叉口通行能力分析. 北京: 科学出版社, 2001. 12-30.
- 8 Shi Y, Eberhart R. A modified particle swarm optimizer. IEEE World Congress on Computational Intelligence, 1998: 69-73.
- 9 陈贵敏, 贾建援, 韩琪. 粒子群优化算法的惯性权值递减策略研究. 西安交通大学学报, 2006, 40(1): 2-4.
- 10 Thomas S, Holger HH. MAX-MIN ant system. Future Generation Computer Systems, 2000, 16(8): 889-914.

# ACOA-RBF 网络模型在短期负荷预测中的应用<sup>①</sup>

王 涛, 王洋洋, 郭长娜

(辽宁工程技术大学 电气与控制工程学院, 葫芦岛 125000)

**摘 要:** 为了进一步提高 RBF 神经网络的性能, 实现准确、快速预测短期电力负荷的目的, 将蚁群优化算法 (ACOA) 作为 RBF 神经网络的学习算法, 建立了一种新的蚁群优化算法的 RBF (ACOA-RBF) 网络预测模型, 利用山西某地区电网的历史数据进行短期负荷预测。仿真表明, 这一算法与传统的 RBF 神经网络预测方法相比, 能达到更好的预测效果。该优化算法改善了径向基神经网络的泛化能力, 提高了山西电网短期负荷预测的精度, 可有效用于电力系统的短期负荷预测。

**关键词:** 蚁群优化算法; 径向基神经网络; 短期负荷预测; 预测精度; 隐含层

## Application of Ant Colony Optimization Algorithm and Radial Basis Function Network Model to Short-Term Load Forecasting

WANG Tao, WANG Yang-Yang, GUO Chang-Na

(Department of Electrical Engineering, Liaoning Technical University, Huludao 125000, China)

**Abstract:** To improve the capacity of RBF neural network and make short-term load forecasting more accurate and faster, a neural network ant colony optimization algorithm and Radial Basis Function neural network forecasting model is established by using the ant colony optimization algorithm to train the RBF neural network. Using the method and history load data of shanxi power system, the short-term load forecasting was carried out. The simulation results show that the forecasting results by the proposed method are better than those by RBF neural network method. The optimization algorithm improves the RBF neural network generation capacity, and the short-term load forecasting accuracy is improved in Shanxi power system. So it can be effectively used in short-term load forecasting of power system.

**Key words:** ant colony optimization algorithm; radial basis function neural network; short-term load forecasting; forecasting accuracy; hidden layer

电力负荷的准确预测对于电力生产和电网安全运行以及国民经济都有重要意义。多年来, 负荷预测一直是国内外研究的热点, 前人在此领域已取得了一定的成果<sup>[1]</sup>。传统的预测方法, 如时间序列法、回归分析法和模式识别法等都存在缺陷。

径向基函数(Radial Basis Function 简称 RBF) 神经网络是一种新型的神经网络, 与 BP 网络相比, 不仅具有生物学基础和数学基础, 而且结构简单, 学习速度快, 隐节点具有局部特性, 逼近能力更强<sup>[2]</sup>。目前

已被广泛运用于图像处理、负荷预测<sup>[3,4]</sup>、语音识别<sup>[5]</sup>、交通<sup>[6]</sup>等领域。

蚁群优化算法 (Ant colony optimization algorithm, ACOA) 是 20 世纪 90 年代初, 意大利学者 M.Dorigo 等人提出的一种新的模拟进化算法。该算法是受到自然界的蚁群行为的启发而提出的, 是一种随机搜索算法, 与遗传算法、模拟退火算法等模拟进化算法一样, 是通过候选解组成的群体在进化过程中寻求最优解。该方法采用了正反馈并行自催化机制, 具有较强的鲁

① 收稿时间:2011-05-08;收到修改稿时间:2011-06-11

棒性、优良的分布式计算机制、易于与其他算法相结合等特点，在复杂优化以及智能搜索方面已经展现其优异的性能和巨大的发展潜力<sup>[7]</sup>，已成功应用于一些实际问题，如 TSP 问题<sup>[8]</sup>、分配问题，并取得一系列较好的试验结果。本文采用蚁群优化算法选取合适的基函数中心、宽度及隐含层与输出层之间的权值，建立了蚁群优化算法的径向基 (Ant colony optimization algorithm and Radial Basis Function, ACOA-RBF) 网络预测模型。该模型在山西某地区应用情况表明，该方法的预测精度和稳定性较传统的 RBF 神经网络预测模型均有较大的提高。

## 1 基于 ACOA 的 RBF 网络模型构造

### 1.1 RBF 神经网络的结构及原理

RBF 神经网络是一种 3 层前馈神经网络，包括输入层、隐含层和输出层，其结构如图 1 所示。其中输入层只把输入数据简单映射到隐含层，并不对输入数据进行处理；隐含层中的函数是径向对称的径向基函数，对输入数据实现非线性映射处理，为输出提供数据，通常由高斯函数构成；隐含层与输出层之间是一种线性映射关系。

设一个 RBF 网络的输入层有  $M$  个神经元，隐含层有  $N$  个神经元， $G_i$  为基函数，它是第  $i$  个隐单元的激励输出，输出层有  $J$  个神经元，隐含层与输出层权

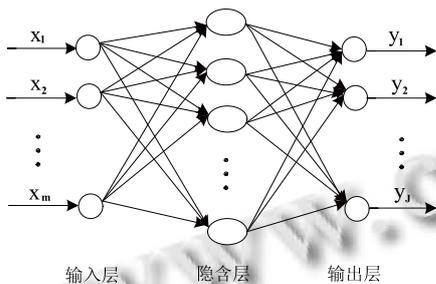


图 1 RBF 网络结构

值为  $w_{ij}$  ( $i=1, 2, \dots, N; j=1, 2, \dots, J$ )。

假定输入样本为  $X=[x_1, x_2, \dots, x_m, x_M]$ ，对应的网络输出为  $Y=[y_1, y_2, \dots, y_j, \dots, y_J]$ ，隐单元的高斯函数为  $G(X, C_i) = \exp[-\|x - c_i\|^2 / 2\sigma_i^2]$ ；则网络第  $j$  个神经元的实际输出

$$y_j(X) = \sum_{i=1}^N w_{ij} G(X, C_i), (j=1, 2, \dots, J) \quad (1)$$

式中， $C_i$  为高斯函数的中心， $\sigma_i$  为高斯函数的宽度。

在 RBF 神经网络中，需要处理的参数包括隐含层中高斯函数的中心、高斯函数中心所对应的宽度以及隐含层与输出层之间的权值。

### 1.2 ACOA 的基本原理

蚁群优化算法 (ACO) 是一种模拟蚂蚁群体觅食行为的随即搜索优化算法。蚂蚁个体之间通过信息素相互通信，即蚂蚁在运动过程中在它所经过的路径上撒播该种物质，且在运动中能感知信息素的存在和强度，并朝着信息素强的方向移动，直到找到最优路线来寻找食物。

ACO 具有以下特性：1) 正反馈作用有助于快速发现问题较好的解；2) 分布式的计算可避免迭代过程中的早熟现象；3) 启发性收敛则使搜索过程中更早地发现可接受解成为可能。故从其原理分析可用作 RBF-NN 的学习算法，可有效克服传统梯度下降法的不足，且可提高模型的泛化能力，从而提高精度。这些特性使得 ACOA 成为有发展前景的优化方法，ACO A 在求解复杂优化和离散优化问题有明显的优势，在电力系统研究方面也已经得到了初步应用。

### 1.3 基于 ACOA 的 RBF 神经网络模型训练

假定网络中有  $m$  个参数，包括 RBF 神经网络的隐含层中高斯函数的中心、高斯函数中心所对应的宽度以及隐含层与输出层之间的权值。 $m$  个参数用  $p_i (1 \leq i \leq m)$  表示，取其  $N$  个可能的随机非零值，组成集合  $I_{p_i}$ ， $\tau_j(I_{p_i})(t)$  表示  $t$  时刻集合  $I_{p_i} (1 \leq i \leq m)$  中第  $j$  个元素的信息素。

假定开始时有  $h$  只蚂蚁的蚁群从蚁巢出发去寻找食物，每只蚂蚁从集合  $I_{p_i}$  中随机地选择一个元素，它表示赋予该蚂蚁的信息素。当蚁群中的蚂蚁在所有集合中完成选择元素后，即到达食物源。此后，按一定的规则调节集合中元素的信息素，这一过程反复进行，直到进化趋势不明显或是达到指定迭代次数时搜索结束。

1) 初始化过程。令时间  $t=0$ ，迭代次数  $NC=0$ ，设置最大迭代次数  $NC_{max}$ ，初始化所有集合  $I_{p_i} (1 \leq i \leq m)$ ，令每个集合中的每个元素在初始时刻的信息素  $\tau_j(I_{p_i})(0) = 0$ ，且  $\Delta \tau_j(I_{p_i}) = 0$ ，将所有蚂蚁置于蚁巢。

2) 启动所有蚂蚁。对蚂蚁  $k (k=1 \dots h)$  根据下述路径选择概率规则选择第  $j$  个元素  $p_j(I_{p_i})$ 。

$$\text{Prob}(\tau_j^k(I_{p_i})) = \tau_j(I_{p_i}) / \sum_{g=1}^N \tau_g(I_{p_i}) \quad (2)$$

3) 重复步骤 2), 直到蚁群中的所有蚂蚁全部到达食物源。

4) 置  $t:t+m$ ,  $NC:NC+1$ , 依据各蚂蚁选择的权值计算网络的输出和误差, 记录当前的最优解。根据下述的信息素调节规则更新集合  $I_{p_i}$  中的每个元素的信息素。

经历了  $m$  个时间单位后, 所有蚂蚁都从蚁巢运动到达食物源, 各路径上的信息素要按下式进行调整:

$$\tau_j(I_{p_i})(t+m) = \rho\tau_j(I_{p_i})(t) + \Delta\tau_j(I_{p_i}) \quad (3)$$

其中:

$$\Delta\tau_j(I_{p_i}) = \sum_{k=1}^h \Delta\tau_j^k(I_{p_i}) \quad (4)$$

式中: 参数  $\rho \in (0,1)$ , 表示信息素的残留程度,  $\Delta\tau_j^k(I_{p_i})$  表示第  $k$  只蚂蚁本次循环中在集合  $I_{p_i}$  的第  $j$  个元素上留下的信息素, 可用下式计算:

$$\Delta\tau_j^k(I_{p_i}) = \begin{cases} Q/e^k, & \text{第 } k \text{ 只蚂蚁在本次} \\ & \text{循环中选择 } p_j(I_{p_i}) \\ 0, & \text{其它} \end{cases} \quad (5)$$

式中  $Q$  是常数, 用作调节信息素的挥发速度,  $e^k$  是将第  $k$  只蚂蚁选择的元素作为神经网络参数时全网的平均误差平方和。如下式所示:

$$e^k = \frac{1}{2} \sum_{i=1}^a \sum_{k=1}^j (y_{i,k} - \overline{y_{i,k}})^2 \quad (6)$$

式中:  $a$  表示训练样本集的样本个数;  $j$  表示输出矢量维数;  $y_{i,k}$  和  $\overline{y_{i,k}}$  分别表示实际输出和期望输出。

5) 当进化趋势不明显或迭代次数  $NC \geq NC_{max}$ , 则迭代结束, 输出最优解, 网络训练完成; 否则转到步骤 2 重新执行。

在 ACO 算法中, 蚂蚁的数目  $h$  会影响蚁群算法的全局搜索能力及算法的稳定性, 当蚂蚁数目远远大于问题规模的时候, 会使大量曾被搜索过路径上的信息量变化趋于平均, 信息正反馈的作用不明显, 搜索的随机性虽然得到了加强, 但是算法的收敛速度相应减慢, 因此增大蚂蚁数目对算法的影响不大, 即过大的蚂蚁数量对算法的寻优性能提高不多, 只是徒增了算法的时间复杂度。而如果蚁群数目过少, 搜索的随机性减弱, 虽然收敛速度加快, 但会使算法的全局收敛性能降低, 容易出现过早停滞现象。在蚁群算法中蚂蚁数量的选择一般可取在  $[\sqrt{n}, n/2]$  区间 (其中  $n$  为问题的规模)。

$Q$  和  $\rho$  也是重要的参数, 参数  $Q$  的大小决定路径上信息量的更新程度, 变量  $\rho \times \tau$  的物理含义为残留的

信息量, 需要忘记一部分过去积累的信息, 可更有效利用最新的信息。这两个参数的选取对计算结果影响较大。但该参数的如何选取目前还难以用解析法来确定其最佳组合。因此参数  $Q$  和  $\rho$  是通过反复匹配和调整后再凭经验得到的。考虑到  $Q$  的取值范围较大, 而参数  $\rho$  的取值范围相对固定 ( $(0,1)$  之间), 因此可先随机设定  $\rho$  的值后再来调整  $Q$ , 得到较理想的解, 在基本确定  $Q$  值后, 再反过来调整  $\rho$ , 寻找更优的解, 如此反复, 最终逼近较为理想的参数组合。

## 2 历史负荷数据的预处理

历史负荷数据中往往包含有非真实的数据, 通常称之为“不良数据”或“坏数据”, 如果利用这些伪数据进行负荷建模, 必然会给负荷预测带来较大的误差, 甚至导致预测算法的发散。因此, 在利用这些数据之前, 应先对其进行预处理。本文的预处理主要是根据当月某一时刻的平均值来对样本数据是否异常进行判断。若样本值与平均值相差过大, 则将该数据作为异常值清除。

为了避免出现神经元饱和这一问题, 在经过上述的处理后, 要将输入负荷数据归一化处理。以恰当的方式对负荷数据进行归一化处理可以加速 RBF 神经网络的收敛。

$$y_i = \frac{x_i - \frac{1}{2}(\max(x) - \min(x))}{\frac{1}{2}(\max(x) - \min(x))} \quad (7)$$

在利用 RBF 神经网络预测结束后, 需用反归一化来计算出负荷值:

$$x_i = \frac{1}{2}(\max(x) - \min(x))y_i + \frac{1}{2}(\max(x) + \min(x)) \quad (8)$$

式中,  $x_i$ ,  $y_i$  分别表示输入样本归一化前后的值,  $\max(x)$ ,  $\min(x)$  分别表示样本的最大负荷和最小负荷值。

## 3 仿真计算

### 1) 预测模型的网络结构

一般日负荷有如下关系: 某一天的负荷曲线同其前一天的负荷曲线比较相似, 同其一星期前的同一天的负荷曲线也比较相似; 一天中某一点的负荷同前几天同一点的负荷相差不大。即可认为一天中的某一点负荷与前一天同一点附近的负荷, 与前几天同一点负

荷,与一星期前同日同一点附近的负荷相关性比较大。本文充分考虑了以下 4 种环境因素的影响:日期类型、天气情况、最高温度和最低温度。为降低问题的求解规模,对一天 24 h 每一小时负荷的预测分别建立其单独的预测模型。

根据以上 ACOA-RBF 神经网络模型,采用山西某地区电网 2010 年 8 月至 9 月的负荷数据,以 MATLAB 为工具对 9 月 2 号 24h 的负荷进行预测。

2) 采用 3 层 RBF 神经网络模型:12 个输入节点、10 个隐含层节点、1 个输出节点。设置  $\rho = 0.7$ ,  $Q = 30$ ,  $h = 100$ 。

3) 预测的误差指标采用平均绝对百分误差  $E_{MAPE}$ 、均方根相对误差和最大相对误差  $E_{MAX}$ , 其中,

$$E_{MAPE} = \frac{1}{N} \sum \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\% \quad (9)$$

$$E_{MSE} = \sqrt{\frac{1}{N} \sum \left( \frac{Y_i - \hat{Y}_i}{Y_i} \right)^2} \times 100\% \quad (10)$$

式中:  $Y_i$  为实际值,  $\hat{Y}_i$  为预测值,  $N$  为负荷数据点组成, 这里为 24。

对该地区电网 9 月 2 日的预测结果对比如表 1 所示, ACOA-RBF 网络预测模型的平均绝对百分误差  $E_{MAPE}$ 、均方根相对误差  $E_{MSE}$  和最大相对误差  $E_{MAX}$  分别为 1.19%、1.41% 和 2.43%; 而 RBF 网络预测模型的  $E_{MAPE}$ 、 $E_{MSE}$  和  $E_{MAX}$  分别为 2.62%、2.80% 和 4.13%。可见, ACOA-RBF 网络预测模型的各项误差指标均各小于 RBF 网络预测模型的误差指标。

实际负荷和预测负荷曲线如图 2 所示, 由此可见, ACOA-RBF 网络预测模型的负荷预测值更接近实际值, 预测精确度更高。

图 3 为预测日 24h 两种预测模型的负荷预测误差曲线, 可见采用 ACOA-RBF 网络预测模型得到的百分误差变化更加均匀, 稳定性更好。

表 1 9 月 2 日各小时负荷预测结果比较

时间	实际值 /MW	RBF		ACOA-RBF	
		预测值 /MW	相对误差/%	预测值 /MW	相对误差/%
0:00	6199.51	5971.99	3.67	6332.80	2.15

1:00	5697.53	5557.94	2.45	5632.58	1.14
2:00	5615.01	5410.06	3.65	5526.85	1.57
3:00	5535.51	5378.86	2.83	5523.89	0.21
4:00	5486.77	5429.16	1.05	5474.15	0.23
5:00	5527.56	5419.22	1.96	5480.58	0.85
6:00	5711.32	5480.01	4.05	5663.34	0.84
7:00	6062.87	5934.94	2.11	5915.54	2.43
8:00	6478.55	6708.54	3.55	6325.01	2.37
9:00	7211.12	7025.79	2.57	7087.81	1.71
10:00	7335.87	7071.05	3.61	7186.22	2.04
11:00	7297.34	7206.85	1.24	7184.23	1.55
12:00	7042.15	6811.17	3.28	7000.60	0.59
13:00	6656.89	6480.48	2.65	6549.71	1.61
14:00	6711.36	6490.56	3.29	6632.17	1.18
15:00	6647.34	6554.94	1.39	6646.01	0.02
16:00	6723.64	6798.94	1.12	6733.05	0.14
17:00	7121.08	7415.18	4.13	7130.34	0.13
18:00	7893.56	8165.10	3.44	7857.25	0.46
19:00	8249.76	8100.44	1.81	8235.74	0.17
20:00	8169.43	8280.53	1.36	8027.28	1.74
21:00	7923.65	8216.03	3.69	7757.25	2.10
22:00	7199.87	7047.95	2.11	7074.59	1.74
23:00	6481.05	6363.09	1.82	6385.78	1.47
最大相对误差			4.13	2.43	
平均绝对误差			2.62	1.19	
均方根相对误差			2.80	1.41	

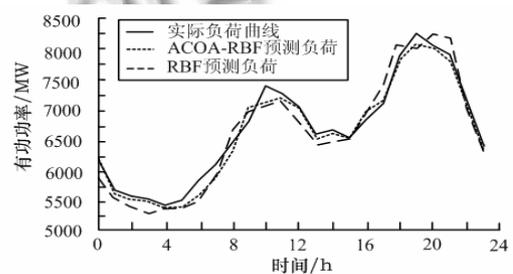


图 2 9 月 2 日 24h 实际和预测的电力负荷

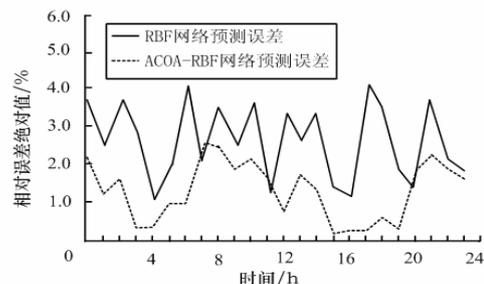


图 3 9 月 2 日电力负荷预测的百分误差

#### 4 结 语

短期负荷预测对保证安全供电及电力运输和电力系统安全分析有重要的意义。本文提出了一种基于 ACOA 训练 RBF 神经网络的预测模型,该算法充分利用蚁群算法的并行、全局寻优等特点来优化径向基函数的中心、宽度及隐含层与输出层之间的权值。仿真表明,基于蚁群优化算法的 RBF 神经网络预测模型比传统的 RBF 网络模型预测精度更高,稳定性更好,是一种行之有效的方法。

#### 参 考 文 献

- 1 雷绍兰,孙才新,周淑,张晓星,程其云.基于径向基神经网络和自适应神经模糊系统的电力短期负荷预测方法.中国电机工程学报,2005,25(22):78-82.
- 2 高隽.人工神经网络原理及仿真实例.北京:机械工业出版社,2007.
- 3 李如琦,褚金胜,谢林峰,王宗耀.IAFSA-RBF 神经网络在短期负荷预测中的应用.电力系统及其自动化学报,2011,(23)2:142-146.
- 4 吴建龙,郭滨钊.PCA\_RBF 网络在电力负荷预测中的应用研究.计算机仿真,2010,27(11):270-273.
- 5 Oglesby J, Mason J S. Radial Basis Function networks for Speaker Recognition. Proc. of International Conference on Acoustics, Speech, and Signal Processing. Toronto, Canada: Causal Production Pty Ltd,1991,393-396.
- 6 袁立行,王玲芝.基于 RBF 网络的道路交通能力预测.西安邮电学院学报,2011,(16)2:42-45.
- 7 李士勇,陈永强,李研,等.蚁群算法及其应用.哈尔滨:哈尔滨工业大学出版社,2004.
- 8 Dorigo M, Maniezzo V, Colomi A. Ant System Optimization by a Colony of Cooperating Agent. IEEE Trans. on Systems, Man and Cybernetics, 1996,26(1):29-41.

(上接第 117 页)

但图中缺陷部分出现了断裂,需要在进行形态学方面的处理。

#### 参 考 文 献

- 1 Nian B, Fu ZZ, Wang L, Cao XX. Automatic detection of defects in solar modules. 2010 6th International Conference on Wireless Communications, Networking and Mobile Computing, 2010.
- 2 张舞杰,李迪,叶峰.硅太阳能电池视觉检测方法研究.计算机应用,2010,30(1):249-252.
- 3 丁叶飞,李红波,张滢清,薛永胜,刘小宇,陈鸣波.硅太阳能电池组件工艺缺陷的红外检测技术.太阳能,2010,(2):36-38.
- 4 Gonzalez RC, et al.阮秋琦,等译.数字图像处理.第 2 版.北京:电子工业出版社,2007.215-217.
- 5 Hunt LP. Total Energy Use in the Production of Silicon Solar Cells from the Raw Material to Finished Product. 12th IEEE Photovoltaics Specialists Conference. Baton Rouge, 1976: 347-352.
- 6 齐丽娜,张博,王战凯.最大类间方差法在图像处理中的应用.无线电工程,2006:25-26.
- 7 郝颖明,朱枫.二维 Otsu 自适应阈值的快速算法.中国图象图形学报,2005,4.
- 8 Marchadier J, Arquès D, Michelin S. Thinning Grayscale Well-Composed Images: A New Approach for Topological Coherent Image Segmentation. DGCI 2002, LNCS 2301, 360-371.
- 9 Pujol FA, Suau2 P, Pujol M, Rizo R, Pujol MJ. Selection of an Automated Morphological Gradient Threshold for Image Segmentation Application to Vision-Based Path Planning. Iberamia 2004, LNAI 3315.667-676.

# 数据融合技术在无线传感器网络中的应用<sup>①</sup>

徐世武, 王 平

(福建师范大学 医学光电科学与技术教育部重点实验室, 福州 350007)

**摘 要:** 在大规模的无线传感器网络中, 传输数据量巨大, 必然存在着数据传输可靠性、拥塞以及能耗等问题, 高效的数据融合技术能够有效的解决这些问题。本文结合分簇路由算法的特征, 采用两层融合技术, 首先簇内节点与簇首节点的融合, 簇内节点根据阈值来判断是否需要发送数据, 簇首节点根据接收到的数据, 进行数据一致性检验, 剔除异常数据, 第二层采用 BP 神经网络算法对簇首节点与基站的融合, 得到所需要的结果。实验表明, 进行融合后的数据可靠性高, 较大减少了数据的传输量与冗余度、降低了能量的消耗, 从而提高了整个网络的性能。

**关键词:** 无线传感器网络; 数据融合; BP 神经网络; 智能温室; 数据可靠性

## Data Confusion Algorithm in Wireless Sensor Networks

XU Shi-Wu, WANG Ping

(Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, Fujian Normal University, Fuzhou 350007, China)

**Abstract:** In the large-scale wireless sensor networks, which need to transfer large amounts of data, there must be some problems, such as, transmission reliability, congestion and energy consumption. High-efficiency data fusion technology can effectively solve these problems. In this paper, we combined the characteristics of clustering routing algorithm, and adopted two-layer data fusion technology. The first layer is about the data fusion between cluster node and cluster-head node. Cluster node determines whether to send data according to threshold. In terms of received data, cluster-head conducts consistency check to the data, excluding abnormal data. At second layer, BP neural network algorithm is adopted to fuse head-cluster node and base station. Experiment results show that data reliability is high after fusion, which greatly decreases transmission quantity and redundancy of data, reduce energy loss, and thereby improves the performance of entire network.

**Key words:** wireless sensor networks; data fusion; BP neural network; intelligent greenhouse; data reliability;

无线传感器网络集成了传感器技术、计算机技术和通信技术, 实用与战场通信、抢险救灾和公共集会等突发性、临时性场合。由于节点存储与功耗的限制, 因此功耗的高低是衡量一个无线传感器网络性能的主要标志之一。在较大规模的无线传感器网络中, 如工业自动化与检测以及环境检测等, 传输数据量大, 冗余度高, 会引发数据传输碰撞, 存在功耗、数据传输的可靠性等问题, 射频收发一个比特的数据消耗的能量是执行一条指令消耗能量的数百倍, 使用数据融合

可以显著减少射频收发的次数。因此对节点传输数据进行处理非常重要, 利用节点的计算与处理能力, 我们可以去除掉冗余信息, 提高整个网络的性能<sup>[1,2]</sup>。

数据融合技术涉及到检测技术、模式识别、决策论、不确定性理论、估计理论、最优化理论等众多学科领域。目前关于无线传感器网络数据融合技术的发展情况, 大体可以分为有损融合、无损融合、依赖于应用的数据融合, 如应用层开发面向应用的数据融合接口, 在网络层开发与路由相结合的数据融合技术。

① 基金项目:福建省自然科学基金(2008J0211);福建省教育厅资助项目(JB09071;JB09076)

收稿时间:2011-05-09;收到修改稿时间:2011-05-31

独立于应用的数据融合、基于分布式数据库的数据融合、基于中心的数据融合等<sup>[3]</sup>。本文主要针对在网络层与路由相结合的数据融合技术的研究,目前关于这方面比较典型的方法有 LEACH 分簇算法<sup>[4]</sup>,使用分簇的方法使得数据融合技术在算法中起到了非常重要的作用。然而在该算法中偏重路由技术,在簇首节点融合数据中,并没有提到具体的融合技术。

## 1 数据融合算法的设计

整个无线网络可以利用分簇路由机制,如 LEACH 算法分成不同的簇,则整个网络中大体分成了三类节点,基站,簇首节点以及簇内节点。本文主要根据簇形结构,设计两层的数据融合机制,第一层为簇内节点与簇首节点的融合,簇内节点根据阈值来判断是否需要发送数据,簇首节点根据接收到的数据,进行数据一致性检验,剔除异常数据,第二层采用 BP 算法对簇首节点与基站的融合,得到所需要的报警结果。

### 1.1 簇内节点与簇首节点之间的第一层数据融合

在无线传感器网络中,簇内节点基本作为传感节点,用来采集数据的,假设簇内节点 A 当前采集到的数据值为  $D_i$ ,则上一次采集到的数据值就为  $D_{i-1}$ ,簇内节点的阈值为  $P$ ,则簇内节点就可以根据所设定好的阈值来决定是否发送当前的数据值,即当  $|D_i - D_{i-1}| > P$  时,簇内节点发送数据,簇内节点更改存储器的值,替换上一次采集到的数值,否则不发送数据,存储器的值保持上一次的不变。通过这样的方式,在采集到的数据值没有发生较大变化时,变化量大小取决于阈值  $P$  的大小,簇内节点可以减少较多的数据发送次数,可以防止向其簇头节点发送大量重复和高度相似的数据。节省了簇成员节点的能量消耗<sup>[5]</sup>。

假设簇首节点拥有 10 个簇内节点,且每个簇内节点都有各自不同的编号,分别从 1 到 10 编号,簇首节点还没有接收到簇内节点的数据时,存储的是上一次簇内节点采集过来的数据,则当簇首节点接收到簇内节点的数据时,替换对应节点的数据值,存储本次采集过来的数据,否则,说明该簇内节点数据值并未发生较大变化,则直接调用对应簇内节点上次采集过来的数据值代表本次的数据。以这样的方式簇首节点就可以知道对应 10 个簇内节点当前采集过来的数据值,以温度为例,假设采集过来的 10 个数据值按照从小到大排列后的数值为  $T_1, T_2, T_3, \dots, T_{10}$ ,因为节点可

能受到外界的干扰,以及自身硬件的故障等影响,簇首节点首先对采集到的数据值进行一致性检验,采用的是数据探测技术中的分布图法剔除掉疏误差的数据<sup>[6]</sup>,假设中位数用  $TM$  表示,上四分位数用  $FU$  表示,下四分位数用  $FL$  表示以及四分位数离散度用  $dF$  来表示。则根据以上采集到的数据可以得到,  $TM = (T_5 + T_6) / 2$ ,下四分位数  $FL$  的区间为  $[T_1, TM]$ ,上四分位数  $FU$  的区间为  $[TM, T_{10}]$ ,四分位数离散度用  $dF = FU - FL$ 。则我们就认定与中位数的距离大于  $adF$  的数据为离异数据,应该剔除掉。即无效数据的判别区间为  $[Ti - TM] > adF$ 。式中  $a$  为常数,可以根据实际需要来设定大小。通过这样的方式,簇首节点就可以简单有效的剔除掉异常的数据。

### 1.2 簇首节点与基站之间的第二层数据融合

虽然经过了第一层的数据融合后,簇内节点可以减少大量不必要的数据传送,然而簇首节点仍然有较多的数据需要转发到基站,对于一些报警的场合,用户只需要知道报警程度,并不需要知道每个节点具体采集到的数值情况,这时我们可以采用 BP 神经网络算法实现簇首节点与基站的第二层融合,直接得到用户所需要的报警结果,从而再次减少了簇首节点大量的数据转发到基站。我们采用的是三层的神经网络,分别为输入层,中间层(隐层)以及输出层。输入层与隐层节点的数量可以根据实际情况设置,如果需要的报警结果为 1 维表示,如 0 正常,1 异常,则输出层采用一个节点就够了。

具体的算法流程如下<sup>[7-8]</sup>:

步骤 1、对各变量初始化,假设输入层与中间层的权值矩阵为  $V$ ,阈值为  $\theta$ ,中间层与输出层的权值矩阵为  $W$ ,阈值为  $T$ 。

步骤 2、输入第  $P$  对训练样本  $x_i^p$ ,计算各层输出。

$$\text{隐层的净输入} \quad net_j = \sum_{i=1}^n v_{ij} x_i^p - \theta_j \quad (1)$$

$$\text{隐层的输出} \quad y_j^p = f(net_j), \quad j=1, 2, \dots, m, \quad (2)$$

$$\text{输出层的净输入} \quad net_k = \sum_{j=1}^m w_{jk} y_j^p - T_k \quad (3)$$

$$\text{输出层输出} \quad o_k^p = f(net_k), \quad k=1, 2, \dots, L, \quad (4)$$

步骤 3、计算输出误差

设共有  $P$  对训练样本,网络对于不同的样本具有不同

的误差。则误差 
$$E^p = \sqrt{\sum_{k=1}^L (d_k^p - o_k^p)^2} \quad (5)$$

步骤 4、检查是否对所有样本完成一次轮训。

步骤 5、计算各层误差信号

隐层误差信号 
$$\delta_k^{p,o} = (d_k^p - o_k^p) o_k^p (1 - o_k^p) \quad (6)$$

式(6)中,  $k=1,2,\dots,L; p=1,2,\dots,P;$

输出层误差信号, 
$$\delta_j^{p,y} = \left( \sum_{k=1}^L \delta_k^{p,o} w_{jk} \right) y_j^p (1 - y_j^p) \quad (7)$$

式(7)中,  $j=1,2,\dots,m; p=1,2,\dots,P;$

步骤 6、调整各层权值、阈值

调整输出层权值 
$$w_{ij} \leftarrow w_{ij} + \eta \sum_{p=1}^P \delta_k^{p,o} y_j^p \quad (8)$$

式(8)中,  $j=1,2,\dots,m; k=1,2,\dots,P;$

调整输出层阈值 
$$T_k \leftarrow T_k - \eta \sum_{p=1}^P \delta_k^{p,o} \quad (9)$$

式(9)中,  $k=1,2,\dots,L;$

调整隐层权值 
$$v_{ij} \leftarrow v_{ij} + \eta \sum_{p=1}^P \delta_k^{p,y} x_i^p \quad (10)$$

式(10)中,  $j=1,2,\dots,m; i=1,2,\dots,n;$

调整隐层阈值 
$$\theta_j \leftarrow \theta_j - \eta \sum_{p=1}^P \delta_k^{p,y} \quad (11)$$

式(11)中,  $j=1,2,\dots,m;$

步骤 7、检查网络总误差是否达到精度要求。

经过上述 7 个步骤结束即网络训练过程结束后, 就可以把训练好的神经网络内部的权值和阈值数据提取出来, 将其直接用于节点的神经网络的程序中, 基站收到这组数值后通过一次前向传播即可得到相应时间的温度数据的拟合值。

## 2 具体运用实例分析

我们以温室环境监控为例, 通过采集到的 100 个样本数据进行训练, 隐层节点作用函数选择对称型 S 函数(tansig), 输出节点为 S 型对数函数(logsig)。BP 网络训练过程主要分成两个阶段, 首先通过设置网络结构和迭代的权值和阈值, 从网络的第一层向后计算各神经元的输出; 其次则对权值和阈值进行修改, 计算对总误差的影响, 2 个阶段反复交替, 直到收敛为

止。训练的结果如图 1 所示, 横轴 Epoch 表示训练步数, 纵轴表示误差, 从图 1 可以看出经过 20 次训练后性能达到 0.00886247, 目标是 0.01。我们随机选出 10 个样本数据采用交叉验证的方法进行验证训练后网络的可靠性能, 0 表示正常情况, 即该区域温度处于正常范围, 1 表示异常, 即该区域温度处于异常范围。将 10 个样本数据分成 5 组, 轮流将其中 4 组做训练 1 组做测试, 结果如表 1 所示。从表 1 中我们可以看出经过训练后的 BP 网络, 误差非常小, 几乎不会影响到报警效果。则用户就可以得到所需要的报警结果, 当温度出现 1 的情况, 用户就可以对相应的区域进行处理, 让该区域温度恢复到正常情况。如图 2 所示, 图 2 中, 0 节点表示基站, 1、8、15、22 四个节点为每个区域的簇首节点, 我们以 D 区域为例, 其中 16、17、18、19、20 五个节点为传感节点, 采集温度, 假设五个节点上次采集到的温度值分别为 30.20, 31.20, 31.60, 30.60, 31.80。本次采集到的温度值分别为 30.80, 30.60, 31.30, 31.30, 35.80。假设传感节点的阈值为 0.5, 则节点 18 在本次中不再发送数据到簇首节点, 簇首节点则直接调用节点 18 上次采集温度值作为本次的值, 簇首节点 15 在根据会合的数据值进行一致性检验, 剔除掉节点 20 的数值, 其余的数值经过训练好的神经网络, 得到结果为 1, 表明该区域温度值超过 300, 出现异常, 应该对 D 区进行维护。

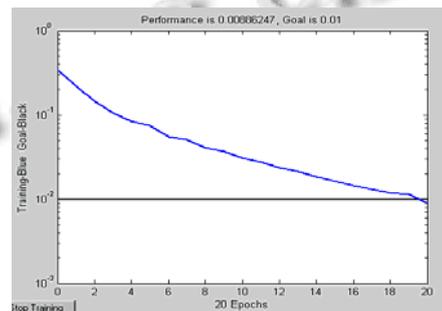


图 1 网络训练结果

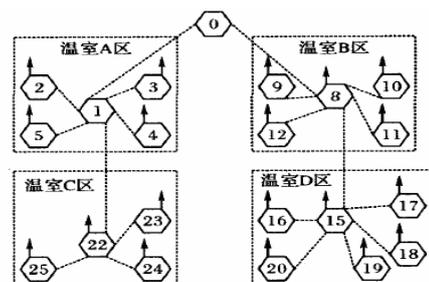


图 2 温室环境监控网络结构图

表 1 交叉验证结果

训练数据	测试数据	网络输出	实际要求输出	误差	均值
除样本1、2	样本1、2	0.0152 0.0120	0 0	0.0152 0.0120	0.0136
除样本3、4	样本3、4	0.0140 0.0641	0 0	0.0140 0.0641	0.03905
除样本5、6	样本5、6	0.9985 0.9984	1 1	-0.0015 -0.0016	-0.0016
除样本7、8	样本7、8	0.9966 0.9748	1 1	-0.0034 -0.0252	-0.0143
除样本9、10	样本9、10	0.0070 0.0068	0 0	0.0070 0.0068	0.0069

### 3 结语

本文采用两层的数据融合算法，在第一层簇内节点与簇首节点的数据处理中，可靠有效的减少了簇内节点数据的转发次数。第二层采用 BP 算法进行数据融合后，直接可以得到用户所需要的结果，无需传输大量的原始采集数据。无线网络中信息量的减少对网络的好处是显而易见的，可以减少节点信息重发的次数，节省网络有限的能量，提高信息准确度，延长网络的生存期。但是有一些问题也需要注意，本文在簇内节点设置阈值，对一些精度要求非常高的场合，数

据变化量较敏感的场所，如果设置的阈值过大，就会较大的影响测量结果，还有数据融合算法过于复杂，会给网络带来一些性能损失，如网络时延，实时性恶化，路由节点计算压力过大而导致网络效率降低等。因此在实际的运用中，应该在效率和复杂度上做一个权衡，由具体应用决定使用何种数据融合方式。

### 参考文献

- 1 付华,胡雅馨.一种改进的无线传感器网络信息融合技术.计算机系统应用,2010,19(7):183-185.
- 2 王平,程明传,翁宗煌.数据融合处理算法在 ZigBee 中的应用.计算机应用,2009,29(7):1897-1900.
- 3 康健,左宪章,唐力伟.无线传感器网络数据融合技术.计算机科学,2010,37(4):31-35.
- 4 Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. Proc. of the 33rd Hawaii International Conference on System Science. Hawaii, January, 2000: 3005-3014.
- 5 张西军.农田土壤信息无线传感器网络与数据融合算法研究.无锡:江南大学,2009.
- 6 滕召胜.基于多传感器数据融合的热处理炉温度测量方法.计量学报,2000,21(2):148-152.
- 7 韩力群.人工神经网络教程.北京:北京邮电大学出版社,2006.
- 8 肖晓,王国军,马征.BP 神经网络在 WSN 数据融合中的应用.企业技术开发,2009,28(9):15-17.