

k-means-RBF 集成神经网络在工业尾气检测中的应用

刘 萍, 龚雪飞, 简家文*, 张 帆, 陈志芸

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要: 神经网络是工业尾气检测系统的一个重要组成部分. 为提高神经网络的预测精度和收敛速度, 建立 k-means-RBF 集成神经网络模型. 首先, 通过选取不同的径向基函数神经网络参数, 得到一组 RBF 神经网络; 然后, 利用 k-means 算法对生成的 RBF 神经网络进行聚类, 并筛选出各类中精度较高的神经网络; 最后, 通过简单平均法对筛选出的神经网络进行集成, 得到高性能的 k-means-RBF 集成神经网络模型. 为验证模型有效性, 搭建基于 k-means-RBF 集成神经网络模型的工业尾气检测系统进行验证. 结果表明, 与粒子群算法优化后的 Back Propagation (PSO-BP) 神经网络模型相比, k-means-RBF 集成神经网络模型的平均预测精度提高 78.27%, 收敛时间节省 99.65%.

关键词: 工业尾气检测; 径向基函数神经网络; k-means 算法

中图分类号: TP183

文献标志码: A

文章编号: 1001-5132 (2017) 01-0116-05

随着工业的快速发展, 大量工业尾气排放到大气中, 其主要成分(NO_x 、 CO 、 CO_2 等)严重污染了环境, 对人类的健康造成了威胁. 有效检测这些有害气体, 对治理污染具有重要意义. 在工业尾气检测中, 单一气体传感器之间大都存在交叉干扰问题. 因此, 人们利用气体传感器阵列与神经网络的融合技术来识别混合气体的百分含量, 从而提高混合气体质量分数检测的准确率^[1]. 目前应用于多元气体检测的神经网络模型主要利用优化算法(粒子群算法、遗传算法等)对单一神经网络进行优化来建立, 虽然在预测精度上取得了很大的改善, 但其预测精度仍然较低, 且收敛速度慢. 因此, 需要寻求一种高效的网络模型来提高预测精度和收敛速度.

径向基函数(Radial Basis Function, RBF)神经网络具有非线性能力强、收敛速度快、结构简单等优点, 现已被广泛应用^[2]. 但是单一的神经网络稳定性较差, 泛化能力较弱^[3-4]. 因此, 本文利用 k-means 算法对单一 RBF 神经网络进行聚类并筛选

出各类中精度较高的网络, 然后利用简单平均法对筛选出的网络进行集成^[5], 建立 k-means-RBF 集成神经网络模型, 将其应用于工业尾气检测.

1 工业尾气检测系统

工业尾气检测系统是模拟人的嗅觉系统对被测气体进行感知、分析和识别, 由工业尾气配气系统、气敏传感器阵列、传感器信号预处理、阵列信号预处理和模式识别五部分组成. 气体检测系统的原理如图 1 所示^[6].

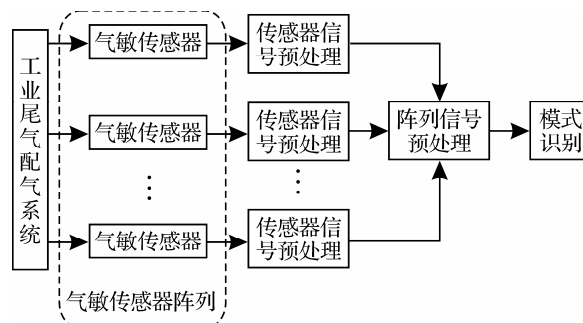


图 1 检测系统原理

收稿日期: 2015-05-21.

宁波大学学报(理工版) 网址: <http://journalg.nbu.edu.cn/>

基金项目: 国家自然科学基金(61471210); 宁波市自然科学基金(2013A610002).

第一作者: 刘萍(1989-), 女, 河北保定人, 在读硕士研究生, 主要研究方向: 新型气体传感器及其应用. E-mail: 710610290@qq.com

*通信作者: 简家文(1967-), 男, 河南淮滨人, 博导/教授, 主要研究方向: 敏感材料与传感器制备. E-mail: jianjiawen@nbu.edu.cn

1.1 数据采集

本文针对工业尾气中的4种气体—— O_2 、 CO_2 、 CO 、 NO_2 进行检测, 采用与此相关的4种标准气体, 通过4台精密流量控制器, 利用动态配气方法, 经过流量调节、配比、混合, 然后流入封闭的测试装置来模拟工业尾气环境。选用 A_2-O_2 氧气传感器、TGS4160 二氧化碳传感器、TGS2201 一氧化碳传感器、 NO_2-AE 二氧化氮传感器以及 HTG3515 温湿度传感器组成传感器阵列对应模拟尾气环境。传感器阵列的模拟输出信号由美国 NI 公司的 PCI 6221 采集板卡采集。

1.2 信号预处理电路

本文针对不同类型的传感器设计了不同的信号预处理电路。图2是电化学传感器(NO_2-AE 二氧化氮传感器、 A_2-O_2 氧气传感器)的信号预处理电路, 图3是半导体传感器(TGS4160 二氧化碳传感器、TGS2201 一氧化碳传感器)的信号预处理电路。

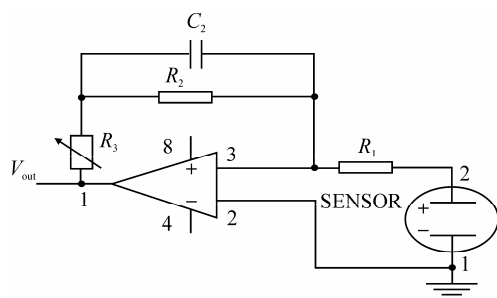


图2 电化学传感器预处理电路

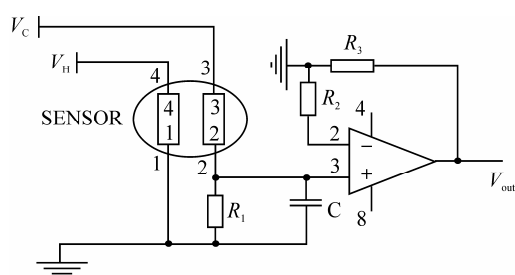


图3 半导体传感预处理电路

图2中, 传感器输出电流信号为 I_{in} , R_2 为定值反馈电阻, R_3 表示可调电阻, R_2 、 C_2 构成低通滤波器, R_2 、 R_3 及运算放大器构成反向放大电路并进行流压转换, 转换公式如下:

$$V_{out} = -I_{in} \times (R_2 + R_3). \quad (1)$$

图3中, V_H 是传感器的加热电压, V_C 表示整个回路电压, 传感器内阻 R_s 与 R_1 构成分压电路, R_2 、 R_3 及运算放大器构成电压放大电路, 输出电压求

解公式如下:

$$V_{out} = V_C \times R_1 \times (R_2 + R_3) / (R_1 + R_s) \times R_2. \quad (2)$$

根据信号预处理电路原理图, 设计并制作了PCB板, 并将传感器置于PCB板上, 然后将阵列置于密闭的测试腔中, 整个测试装置的设计图如图4所示。

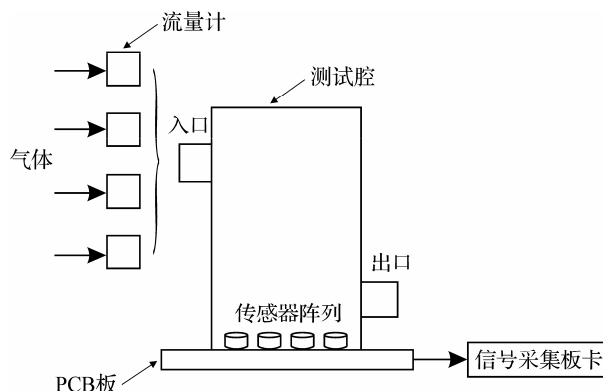


图4 测试装置设计图

2 k-means-RBF 集成神经网络模型

2.1 RBF 神经网络和 k-means 聚类算法

RBF 神经网络是前馈神经网络^[7], 网络结构分为3层——输入层、隐层和输出层, 其结构如图5所示。输入向量和输出向量的函数关系通过非线性映射表现出来。非线性映射函数一般采用高斯函数^[8], 在采用高斯函数的情况下, 输入向量和输出向量之间的函数关系如下:

$$y_l = \sum_{j=1} w_{lj} \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right), \quad (3)$$

式中, y_l 表示输出层第 l 个输出; w_{lj} 表示第 j 个隐层神经元对应第 l 个输出的权值; x 表示输入向量; c_j 和 σ_j 分别表示第 j 个隐层神经元中心和宽度。

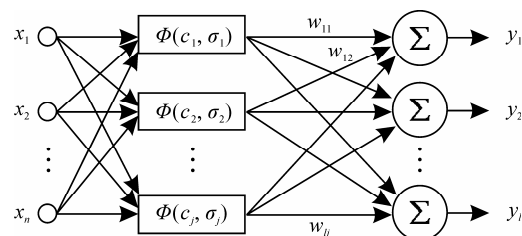


图5 RBF 神经网络结构

k-means 算法是一种常用的, 且简单、快速的无监督聚类算法^[9-10], 它将一个给定的数据集分为指定的 k 个聚簇。其主要思想是将各个聚类子集中

的样本均值作为各簇中心的代表点,再根据这些代表点进行聚类,得到新的代表点,不断进行迭代,直到代表点不发生变化,并作为最终的聚类中心,这些中心对应最终的聚类子集.

2.2 k-means-RBF 集成神经网络模型的建立

集成学习方法是合并多个弱学习器得到一个强学习器的方法,形式如下:

$$f = \sum_{i=1}^n a_i f_i, \quad (4)$$

式中, f 表示强学习器; f_i 表示弱学习器; a_i 表示各个弱学习器所对应的权重.

本文以 MATLAB 作为仿真环境,首先采用 mapminmax() 函数将训练样本归一化处理到 [0,1] 之间,然后利用 $net = newrbf(P, T, R_{SPREAD})$ 函数生成单一的 RBF 神经网络(即弱学习器),其中, net 表示单个 RBF 神经网络, $newrbf()$ 表示 RBF 神经网络生成函数, P 为训练样本的输入向量, T 为训练样本的输出向量, R_{SPREAD} 表示径向基函数的分布密度,即径向基函数的宽度.通过对 R_{SPREAD} 等间隔取值可以得到一组 RBF 神经网络.由于 R_{SPREAD} 的取值范围过大或过小都会影响网络的性能,因此经过多次实验比较,设定 R_{SPREAD} 取值范围为 3.00~13.35,并每间隔 0.15 取一个固定值进行 RBF 神经网络的建立.

研究表明,当被集成的弱学习器存在多样性和差异性时,集成学习算法可得到更好的效果^[11].因此在得到一组 RBF 神经网络后,需要对 RBF 神经网络进行筛选,以保证 RBF 神经网络的精度和单一 RBF 神经网络之间的差异性.首先,根据训练样本的平均相对误差来剔除效果不好的网络,具体操作如下:选择一个合适的阈值 t ,剔除训练样本平均相对误差大于 t 的 RBF 神经网络;其次,采用 k-means 聚类方法对其余神经网络的训练样本相对误差向量进行聚类.

具体聚类过程如下:

(1)将 RBF 神经网络训练样本的误差向量作为聚类的原始数据样本;

(2)随机选取 k 个原始数据样本作为初始的聚类中心;

(3)分别计算各样本与各聚类中心之间的距离,并按照最小距离原则将样本赋给最邻近聚类中心

所对应的类;

(4)将每类的样本均值作为新的聚类中心;

(5)重复步骤(3)和(4),直到聚类中心不再发生变化,得到最终的分类.

最后,把每个聚类子集中最小平均相对误差所对应的 RBF 神经网络作为选择后的 RBF 神经网络.经过多次试验比较设定阈值 t 为 0.4, k 为 15.利用简单平均法,对上述选择后的 k 个 RBF 神经网络赋予相同的权重,并利用式(4)进行集成得到 k-means-RBF 集成神经网络(即强学习器).

3 实验验证与分析

3.1 实验数据的获取

本文模拟工业尾气的环境,将 O_2 [18%~22%], CO_2 [$0 \sim 5\,000 \times 10^{-6}$], CO [$0 \sim 1\,000 \times 10^{-6}$], NO_2 [$0 \sim 250 \times 10^{-6}$] 气体以任意比例混合,配置成 510 组混合气体.混合气体通过装有传感器阵列和信号预处理电路的测试装置,经信号采集板卡采集得到测试数据.这些数据作为建立网络模型及检测其性能的原始数据,随机选取 450 组作为训练样本,其余 60 组作为预测样本.

3.2 实验数据的处理与分析

利用 450 组训练样本以及上文所提方法建立 k-means-RBF 集成神经网络模型,并用其余 60 组预测样本进行网络性能的测试,其测试结果如图 6 所示.通过图 6 可以看出, k-means-RBF 集成神经网络模型的预测结果与真实值基本吻合, k-means-RBF 集成神经网络模型可以精确地预测混合气体质量分数.为了更好地分析网络的性能,利用相同的训练样本分别建立 k-means-RBF 集成神经网络模型与文献[6]所提的 PSO-BP 神经网络模型,并利用相同的预测样本来分析比较两种网络模型的性能. PSO-BP 神经网络模型主要是利用粒子群优化算法的全局搜索能力对单一的 BP 神经网络权值和阈值进行寻优,得到性能较好的 PSO-BP 神经网络模型.2 种模型比较结果见表 1.从表 1 中可以看出, k-means-RBF 集成神经网络模型预测的平均相对误差百分比均小于 0.5%,耗时仅用 14.282 s,而 PSO-BP 神经网络模型预测的平均相对误差百分比均大于 0.5%,耗时 4 131.54 s.此外,分析表 1 中数据还可知,与 PSO-BP 神经网络方法相比,本文所

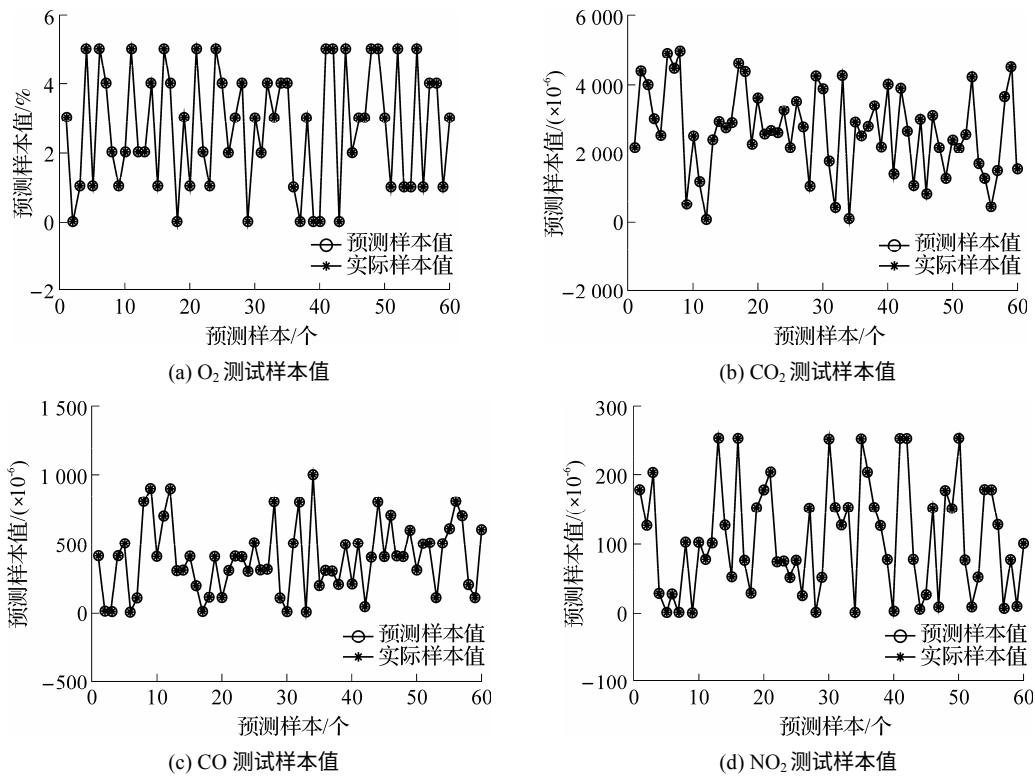


图 6 k-means-RBF 集成神经网络模型的预测性能分析

表 1 PSO-BP 与 k-means-RBF 集成神经网络性能比较

项目	PSO-BP 神经网络 ^[6]				k-means-RBF 集成神经网络			
	O ₂	CO ₂	CO	NO ₂	O ₂	CO ₂	CO	NO ₂
最大误差	0.049 2	65.127 2	15.867 6	10.197 5	0.019 2	12.52 0	8.410 7	2.082 5
最小误差	0.000 1	0.012 1	0.013 1	0.055 1	0.000 3	0.013 3	0.037 8	0.001 2
平均误差	0.015 2	16.893 9	4.586 0	2.278 0	0.004 4	2.838 9	1.100 0	0.372 8
最大相对误差	3.016 2	5.699 8	8.774 9	9.267 9	0.959 9	4.728 3	1.904 4	2.776 7
最小相对误差	0.001 3	0.000 6	0.001 6	0.028 5	0.010 2	0.000 7	0.007 6	0.000 5
平均相对误差	0.621 6	1.122 8	1.467 3	2.086 3	0.204 0	0.322 9	0.294 8	0.329 7

注: 1)误差=|网络实际输出-期望输出|; 2)相对误差=(|网络实际输出-期望输出|)/期望输出×100%; PSO-BP 神经网络训练时间为 4 131.54 s; k-means-RBF 集成神经网络训练时间为 14.282 s.

提方法预测的平均精度提高了 78.27%, 收敛速度提高 99.65%. 平均精度和收敛速度提高的百分比由(5)式和(6)式求得:

$$R_{\text{Improve},P} = \frac{(P_{\text{PSO}} - P_{\text{RBF}})}{P_{\text{PSO}}} \times 100\%, \quad (5)$$

$$R_{\text{Save},T} = \frac{(T_{\text{PSO}} - T_{\text{RBF}})}{T_{\text{PSO}}} \times 100\%, \quad (6)$$

其中, P_{PSO} 和 P_{RBF} 分别表示 PSO-BP 神经网络模型和 k-means-RBF 集成神经网络模型预测的 4 种气体平均相对误差; T_{PSO} 和 T_{RBF} 分别表示建立 PSO-BP 神经网络模型和 k-means-RBF 集成神经网络模型

的时间; $R_{\text{Improve},P}$ 和 $R_{\text{Save},T}$ 分别表示平均精度和收敛速度提高的百分比.

4 结论

为提高神经网络的预测精度和收敛速度, 本文建立 k-means-RBF 集成神经网络模型, 并将此模型应用于工业尾气检测, 搭建了一套基于传感器阵列和选择性集成 RBF 神经网络相结合的工业尾气检测系统.

与 PSO-BP 神经网络模型相比, 本文所提方法预测的平均精度提高了 78.27%, 收敛速度提高了

99.65%.

参考文献:

- [1] 郑杰. 混合有毒气体检测算法研究[J]. 计算机仿真, 2011, 28(8): 413-416.
- [2] 张志宇, 赵丹国, 侯晓宇. PSO-RBF 神经网络在城市需水量预测中的应用[J]. 水电能源科学, 2013, 31(6): 55-57.
- [3] ZHOU Z H, WU J, TANG W. Ensembling neural networks: many could be better than all[J]. Artificial Intelligence, 2002, 137(1):239-263.
- [4] 李凯, 黄厚宽. 一种基于聚类技术的选择性神经网络集成方法[J]. 计算机研究与发展, 2005, 42(4):594-598.
- [5] 王正群, 陈世福, 陈兆乾. 优化分类型神经网络线性集成[J]. 软件学报, 2005, 16(11):1902-1908.
- [6] 龚雪飞, 徐景, 孙寿通, 等. PSO-BP 神经网络在多元有害气体检测中的应用[J]. 传感器与微系统, 2015, 34(2):154-160.
- [7] 韩红桂, 乔俊飞, 薄迎春. 基于信息强度的 RBF 神经网络结构设计研究[J]. 自动化学报, 2012, 38(7):1083-1090.
- [8] 庞振, 徐蔚鸿. 一种基于改进 k-means 的 RBF 神经网络学习方法[J]. 计算机工程与应用, 2012, 48(11):161-163.
- [9] 谢娟英, 王艳娥. 最小方差优化初始聚类中心的 K-means 算法[J]. 计算机工程, 2014, 40(8):205-211.
- [10] 张靖, 段富. 优化初始聚类中心的改进 k-means 算法[J]. 计算机工程与设计, 2013, 34(5):1691-1694.
- [11] MINKU L L, WHITE A P, YAO X. The impact of diversity on online ensemble learning in the presence of concept drift[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(5):730-742.

k-means-RBF integrated neural network and its application in industrial exhaust emission detection

LIU Ping, GONG Xue-fei, JIAN Jia-wen*, ZHANG Fan, CHEN Zhi-yun

(Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China)

Abstract: Neural network is very important in detection of industrial exhaust emission. In order to improve stability and prediction accuracy of neural network and reduce the training time of neural network, k-means-RBF integrated neural network model is established based on k-means algorithm. Firstly, a set of RBF neural networks are obtained by choosing different RBF network function parameters. Secondly, the neural network with higher precision is selected from each type. Finally, high performance integration RBF neural network is obtained through integrating RBF neural network using the simple average method. In order to verify the validity of the model, the developed neural network model is used on the industrial exhaust emission detection system. The results show that, compared with the Back Propagation (BP) neural network method optimized by particle swarm optimization (PSO), the average precision predicted by the proposed neural network increases by 78.27%, and the processing time is saved by 99.65%.

Key words: industrial exhaust emission detection; RBF neural network; k-means algorithm

(责任编辑 章践立)