

粗糙集-遗传神经网络在挖掘机故障诊断中的应用研究

胡 平, 彭纪奎

(南京工业大学 电子与信息工程学院, 江苏 南京 210009)

摘 要: 针对当前单一的故障诊断方法不能满足实际需求的问题, 提出了一种粗糙集-遗传神经网络分类器模型, 实现对挖掘机故障分类。该模型首先利用粗糙集理论对神经网络的输入进行属性约简, 以减少神经网络的工作量; 利用遗传算法优化 BP 神经网络, 解决神经网络易陷入局部极小和收敛速度慢的问题; 最后利用约简结果和优化的 BP 网络进行网络训练。实验结果验证了该方法用于故障诊断的有效性。

关键词: 粗糙集; 遗传算法; 神经网络; 故障诊断

中图分类号: TP391.5

文献标识码: A

文章编号: 1000-7180(2011)03-0055-04

Application Research of Rough Set-Genetic Algorithm-Neural Network Algorithm in Excavator Fault Diagnosis

HU Ping, PENG Ji-kui

(College of of Electronics and Information Engineering, Nanjing University of Technology, Nanjing 210009, China)

Abstract: Aiming at a single fault diagnosis method can't meet the actual need, a classification model were proposed, which based on rough set-genetic algorithm-neural network algorithm, to come true excavator fault diagnosis classification. Firstly, the attributes were reduced using rough set theory to choose neural network's input parameters, which reduced the work and calculation time. Then, in order to solve the shortcoming in the back propagation algorithm, such as trapping to the local minimum and slowness in training speed, genetic algorithm was integrated to optimizing the BP network parameters. Finally, the model carried on the training by the reduction results and the optimized BP network parameters. The experimental result shows the effectiveness of the new proposed model.

Key words: rough set; genetic algorithm; neural network; fault diagnosis

1 引言

近年来兴起的人工神经网络^[1], 特别是使用 BP 算法(backpropagation)的神经网络多层感知器模型以其独到的并行分布式处理、联想记忆、良好的自组织及自学习能力, 以及良好的容错性、鲁棒性、极强的非线性映射能力, 在大型机械故障诊断中得到了广泛应用。但是 BP 网络采用的是沿梯度下降的搜索求解算法, 存在着网络学习收敛速度慢, 容易陷入局部极小的问题。遗传算法^[2] (genetic algorithms, GA) 是一种借鉴生物界自然选择和自然遗传机制的随机搜索算法, 具有良好的全局搜索性能, 减少了局部最优解的风险, 鲁棒性强, 适用于并行处

理, 搜索不依赖于梯度信息。粗糙集理论^[3]作为一种新兴的人工智能方法, 不需要预先给定统计学中的概率分布等假设条件, 直接从给定问题的描述集合出发, 对不完整、不精确数据进行分析处理, 发现关键条件, 找出该问题中的内在规律。

针对以上不足^[4-6], 本文提出了一种基于粗糙集(RS)和遗传神经网络(GANN)的挖掘机故障诊断模型来解决以上诸多问题, 有效提高了挖掘机故障诊断的精度。

2 粗糙集-遗传神经网络分类模型

2.1 模型构成基本思路

首先利用粗糙集理论的属性约简方法对遗传神

神经网络系统的输入样本数据进行属性约简,然后选择有代表性的决策属性进行遗传神经网络算法学习与预测,并最终获得分类结果。

2.2 模型构成原理

BP 神经网络具有自学习、非线性模式识别、联想记忆及推测的能力、良好的容错性和扩展性,因此在数据分类中得到广泛应用。分类问题的实质是一种映射,神经网络的输入节点对应着样本特征属性,输出节点对应着样本类型。数据的分类就是根据给定的一组属性特征,实现属性特征到数据分类集之间的非线性映射的过程。但神经网络不能确定哪些输入特征是冗余的,哪些是有用的,不能将输入信息空间维数简化,因而当输入信息空间维数较大时,训练时间过长成为制约神经网络发展的因素之一。而粗糙集理论是基于不可分辨关系的思想,具有很强的定性分析能力,不需要预先给定某些特征或属性的数量描述,能在保留关键信息的前提下对数据进行约简,去除冗余属性和冗余样本,压缩信息空间维数,精简知识系统。因此将 BP 神经网络方法和粗糙集方法相融合,用粗糙集约简信息,去掉冗余信息,使 BP 神经网络训练集简化,以减小神经网络结构的复杂性,从而能减少训练时间、提高分类精度。另外, BP 神经网络算法自身存在的收敛速度慢、容易陷入局部极小等缺点,本文利用遗传算法对神经网络外部参数进行优化。遗传算法的全局搜索能力,可以使神经网络系统扩大搜索空间,提高计算效率。

2.3 算法实现流程

粗糙集—遗传神经网络系统运行过程包括两个主要阶段:

- ① 利用粗糙集理论进行样本特征属性选取,获得约简的数据样本;
- ② 利用遗传算法优化 BP 神经网络,得到遗传神经网络模型,并进行对测试样本的决策。算法流程如图 1 所示。

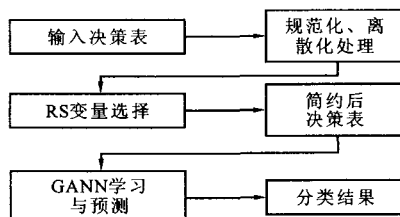


图 1 粗糙集—遗传神经网络分类器模型算法流程

2.3.1 基于粗糙集理论的样本特征属性选取

(1) 连续属性值的离散化

粗糙集理论只能处理离散型数据,但现实问题

的属性值中常常存在连续量数据,因此需要考虑数据的离散化处理。

定义 1 设非空有限集合 U, A , 其中 U 为论域, A 为属性集, 对于每一属性 $a \in A$, 存在属性值集合 $V_a = \{a(x) \mid \forall x \in U\}$, 称 $S = (U, A)$ 为信息系统。若 $A = C \cup D$, 且 $C \cap D = \emptyset$, 其中 C 为条件属性集, D 为决策属性集(一般 $D = \{d\}$), 则称 $S = (U, C \cup D)$ 为决策系统。

数据离散化过程本质上就是采用一定的断点集合对决策系统的属性空间进行划分。为了提高系统的聚类能力, 增强系统对数据噪音的鲁棒性, 应该采用尽可能少的断点来完成划分过程。从这一角度来说, 在保证系统分辨关系的前提下, 用最小的结果断点集合对系统进行的离散化就是基于粗集理论的最优离散化。本文利用阈值离散方法进行数据离散化处理, 可以大大降低计算的空间和时间复杂度。

(2) 粗糙集属性简约

定义 2 对于信息系统 $S = (U, A)$, 设 $C \subseteq A$, 称二元等价关系 $IND(C) = \bigcap_{a \in C} IND(a) = \{(x, y) \in U \times U \mid a(x) = a(y), \forall a \in C\}$ 为在论域 U 上的 C 不可分辨关系。在粗糙集理论中, 不可分辨关系是定义其他概念的基础。

定义 3 对于一个给定的决策系统 $S = (U, C \cup \{d\})$, 条件属性集合 C 的约简是 C 的一个非空子集 C' , 它满足

- ① $IND(C', \{d\}) = IND(C, \{d\})$;
- ② 不存在 $C'' \subset C'$, 使 $IND(C'', \{d\}) = IND(C, \{d\})$;

C 的所有约简的集合记作 $RED(C)$ 。 C 的所有的约简的交集叫做核, 记作 $CORE(C)$, $CORE(C) = \bigcap RED(C)$ 。

一般情况下, 样本集所构成的决策表中的知识重要程度并不相同, 一些条件属性是冗余的信息。属性约简的目的就是消除冗余属性, 保留重要的属性。

2.3.2 遗传神经网络模型的构建

遗传算法(GA)具有很强的全局优化能力。基于梯度下降的 BP 神经网络学习算法具有较高的精确性, 但存在容易陷入局部极小的缺点, 而 GA 具有很强的全局搜索能力, 可以避免局部极小。因此, 将两者结合, 利用 GA 和 NN 之间的互补, 提高优化的性能。

在传统的遗传算法优化神经网络的设计中, 大都是针对网络的初始连接权和网络结构进行优化的, 而忽视了初始阈值及学习率和动量因子。事实上, 在进行神经网络设计时, 初始连接权和阈值一般

是用随机方法产生的,而网络结构及学习率和动量因子则基本上依赖于人的经验来确定.网络外部参数(网络结构、初始连接权、初始阈值、学习率、动量因子)的选取直接影响神经网络的训练速度和处理能力,只有对它们进行全面优化,才能真正提高网络的整体性能.为此,本文利用遗传算法对神经网络外部参数进行进化设计,然后从进化神经网络中用训练样本对网络内部参数(连接权和阈值)再次寻优,形成了一种改进的基于遗传算法的神经网络学习算法.算法流程如图 2 所示.

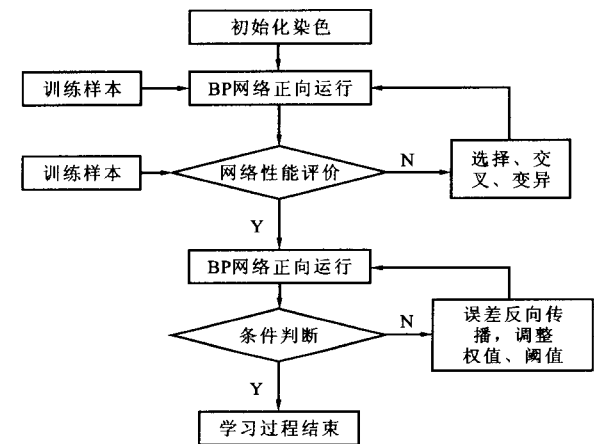


图 2 基于遗传算法的 BP 神经网络学习算法流程

3 挖掘机故障诊断实验结果与分析

以某公司挖掘机为例,运用本文提出的分类模型对挖掘机进行故障诊断.根据液压挖掘机的结构原理,分析故障信息和属性,本文从这些故障信息中选取 20 个能对故障决策属性可能有影响的条件属性($C_1, C_2, \dots, C_{19}, C_{20}$)构成决策表,进行挖掘机故障智能诊断的研究.由条件属性和决策属性所构成的挖掘机故障诊断的决策信息表(略)出发,应用基于遗传算法的粗糙集属性简化方法,得到训练样本的最小简约为: $\{C_3, C_7, C_8, C_{11}, C_{13}, C_{14}, C_{17},$

$C_{20}\}$. 本文以简约 $\{C_3, C_7, C_8, C_{11}, C_{13}, C_{14}, C_{17}, C_{20}\}$ 作为依据,构成遗传神经网络输入属性,设计遗传 BP 神经网络分类器模型.

在进行 GANN 训练时,首先设定网络的拓扑结构.网络的输入层节点为故障特征属性的个数,即简约后的属性个数为 8;隐含层节点数目设为 13;输出层节点数设为 3,分别对应挖掘机故障种类的 7 种故障类型,构成 8—13—3 的网络. GA 参数设置如下:染色体由网络的权值和阈值组成,一个染色体包含的基因数为 $L=8\times13+13\times3+13+3=159$;初始群体中的染色体数选取 $N=160$;依据适应度高的染色体被选作父染色体的概率大的原则,选 $K=50$ 对染色体作为父染色体,实际输出与期望输出误差为 0.05,遗传算法交叉概率取 0.6,变异概率取 0.03.以该挖掘机的常见故障电池阀故障、泵吸进空气、发动机故障、先导阀和换向阀故障、线路故障、溢流阀故障、柴油机转速不稳定为网络的输出,利用挖掘故障诊断中的 20 个属性值作为特征量形成训练样本.我们采集挖掘机在工作过程中 9 个有异常的样本信息作为测试数据,将其进行离散化处理得到表 1 的样本数据,将其作为网络的输入,经 312 次遗传优化训练和 698 次 BP 算法预训练,系统学习误差达到 0.001;同时利用普通的 BP 神经网络进行优化,经 3 475 次 BP 算法训练,系统学习误差达到 0.004.测试样本经过 BP 网络和粗糙遗传神经网络两种网络诊断后的输出结果和期望输出之间的比较如表 2 所示.

从诊断结果可以得知,粗糙神经网络诊断正确率达到 88.9%,而 BP 网络诊断正确率为 55.6%.在训练时间和测试时间上,粗糙遗传神经网络比 BP 网络大大减少.由此可以看出,应用该算法进行故障识别时,在收敛速度、避免局部最小点和网络性能等方面比传统 BP 算法具有明显的优势.

表 1 测试样本

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	D
1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1
2	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	2
4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	7
5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
6	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3
7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	4
8	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	5
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	6

表 2 诊断结果比较

测试	BP 网络输出			粗糙遗传神经网络输出			期望输出		
X ₁	0.000 5	0.169 6	0.979 3	0.005 9	0.011 8	0.991 6	0	0	1
X ₂	0.000 1	0.003 4	0.814 1	0.005 3	0.002 8	0.995 3	0	0	1
X ₃	0.003 2	0.990 7	0.010 0	0.002 5	0.965 1	0.018 1	0	1	0
X ₄	0.780 1	0.998 6	0.844 0	0.990 6	0.991 4	0.996 1	1	1	1
X ₅	0.001 52	0.999 9	0.916 8	0.030 7	0.995 8	0.998 0	0	1	1
X ₆	0.000 7	0.997 7	0.999 6	0.014 9	0.999 6	0.998 5	0	1	1
X ₇	0.997 2	0.004 4	0.002 6	0.997 4	0.002 9	0.000 1	1	0	0
X ₈	0.990 0	0.010 2	0.995 5	0.997 8	0.005 4	0.998 4	1	0	1
X ₉	0.995 1	0.611 6	0.607 4	0.997 1	0.893 6	0.000 3	1	1	0

4 结束语

粗糙集、遗传算法和 BP 神经网络相结合的粗糙遗传神经网络,利用粗糙集对输入 BP 网络中的特征属性约简,去除冗余信息,应用遗传算法对 BP 网络中的权值和阈值进行优化.实践证明这种方法能够优化神经网络结构,还能对挖掘机等机械故障进行准确的识别,在工程机械故障诊断领域具有较好的应用前景.

参考文献:

[1] 刘才强,时晨,于伦正. 遗传算法在故障诊断中的应用研究[J]. 微电子学与计算机,2008,25(2):97-99.
[2] Zbigniew Michalewicz. A modified genetic algorithm for optimal control problems [J]. Computers Math Applica-

tion,1992,23(12): 83-94.

[3] Rafice J, Arvani F. Intelligent condition monitoring of a gearbox using artificial neural network[J]. Mechanical System and Signal Processing,2007,21(4): 1746-1754.
[4] 郝丽娜,王伟,吴光宇,等. 粗糙集-神经网络故障诊断方法研究[J]. 东北大学学报,2003,24(3):252-255.
[5] 王国胤. Rough 集理论与知识获取[M]. 西安:西安交通大学出版社,2001.
[6] 宋立新,王玉华,李玲远. 基于遗传算法和神经网络优化的故障诊断专家系统[J]. 华中师范大学学报:自然科学版,2005,39(3):332-335.

作者简介:

胡 平 男,(1962-),副教授.研究方向为计算智能.
彭纪奎 男,(1982-),硕士.研究方向为数据挖掘.

(上接第 54 页)

参考文献:

[1] Koeser J, Haerti H. Low-latency hard real-time communication over switched Ethernet [C]// Proc ECRTS (EUROMICRO Conf Real-Time Systems). Washington: IEEE,2004.
[2] Kweon SK, Shin KG. Achieving real-time communication over Ethernet with adaptive traffic smoothing[C]// Proceedings of the IEEE Real-Time Technology and Applications Symp. Washington: IEEE Computer Society Press, 2000: 90-100.
[3] Hoang H, Jonsson M, Hagstrom U, et al. Switched real-time ethernet with earliest deadline first scheduling protocols and traffic handling[C]// Proceedings of the International Parallel and Distributed Processing Symposium. Wisconsin: IEEE, 2002: 94-99.
[4] Devi M. An improved schedulability test for uniprocessor periodic task systems[C]// Proceedings of the 15th

Euromicro Conference on Real-Time Systems. Porto, Portugal: IEEE, 2003.

[5] 沈卓伟,汪芸. 基于 EDF 调度策略的端到端实时系统可调度性分析算法[J]. 计算机研究与发展,2006,43(5): 813-820.
[6] BAKER T P. Stack-based scheduling of real-time processes[J]. The Journal of Real-Time Systems, 1991:67-99.

作者简介:

檀 明 男,(1974-),博士,讲师.研究方向为数据库与多媒体技术、工业控制网络.
魏 臻 男,(1965-),博士,教授,博士生导师.研究方向为分布式控制、计算机控制等.
韩江洪 男,(1954-),教授,博士生导师.研究方向为分布式控制、计算机控制、计算机网络等.
钟伯成 男,(1964-),博士,教授.研究方向为计算机网络.