

# 粗糙集和神经网络在数据融合中的应用研究

卜益民<sup>1</sup>, 陈小惠<sup>2</sup>

(1. 南京邮电大学 继续教育学院, 江苏 南京 210003;

2. 南京邮电大学 自动化学院, 江苏 南京 210003)

**摘要:**传感器聚类状态的切换是多传感器数据融合的难点之一,也就是某个传感器在某一时刻应该向哪个方向融合数据的问题。文中采用粗糙集进行知识的获取,把 1 天内的 54 个传感器的可融合典型聚类分布作为数据样本空间形成决策表——“数据-融合分布”决策表;然后对一个月内的数据运用粗糙集的知识约简算法,去除冗余的属性和样本;根据神经网络聚类分析方法,形成多传感器数据的融合分布规则。仿真结果表明该模型的分类效率较好、实现传感器聚类分布的判断较快速。

**关键词:**粗糙集;神经网络;数据融合;传感器网络;Tossim

中图分类号:TP24

文献标识码:A

文章编号:1673-629X(2013)04-0221-05

doi:10.3969/j.issn.1673-629X.2013.04.054

## Application Study of Data Fusion Using Rough Set and Neural Network

BO Yi-min<sup>1</sup>, CHEN Xiao-hui<sup>2</sup>

(1. College of Continuing Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** The difficulties of fusing multi-sensor data lie in the switching of the state of sensor clusters. That is, at a given moment which direction the sensor should fuse data into. First the rough set is used for acquisition of knowledge. The typical clustering distributions of 54 sensors within one day are regarded as sample room for the decision-making table of the "data-fusion distribution". Next, based on rough set of method of simplified knowledge, for one month date, remove redundant properties and samples. Then, the neural network is used to analyze clustering. And finally the patterns of multi-sensor data fusion distribution are formed. The model is proved experimentally to be efficient in classification and rapid in sensor clustering distribution decision.

**Key words:** rough set; neural network; data fusion; sensor network; Tossim

## 0 引言

多传感器数据融合是无线传感器测量网络技术领域中的一个重要研究内容,它的有效作用之一是减少测量网络的通信信息量,同时是延长传感器测量网络生存时间和提高信息处理精度的一个有效方法。目前,多传感器数据融合已成为无线传感器测量网络中的一个研究热点,已经越来越引起人们的重视。利用数据融合对事件的状态和身份进行多方位、多层次的综合优化组合处理的能力,获得对事件更精确和完整

的不同威胁和态势程度的正确估计和评估<sup>[1]</sup>。

不确定性和决策理论、优化和估计理论及神经网络等<sup>[2,3]</sup>是进行数据融合研究时常用的基本理论与方法。目前,许多学者和研究人员已开展基于神经网络的数据融合方法的研究<sup>[4]</sup>,但已有的多传感器神经网络数据融合系统和方法存在一个共同的缺点:当多传感器数据融合网络系统结构庞大或数据融合网络系统中的传感器数据量很大时,会使得数据融合过程中的神经网络训练时间过长、神经网络的训练负担加大,较大地降低多传感器数据融合网络系统的性能。

在大数据量、冗余数据以及不确定数据等的处理方面,粗糙集理论有很多的优点,许多学者和研究人员开始越来越多的注意到,应该把神经网络和粗糙集结合起来<sup>[5,6]</sup>,利用粗糙集的优点进行检测数据的前期处理,这样可以使训练神经网络的数据样本得以简化,加快训练神经网络的速度。在工业<sup>[7-9]</sup>、农业<sup>[10]</sup>、金

收稿日期:2012-06-29;修回日期:2012-10-09

基金项目:国家自然科学基金资助项目(61104216);江苏省科技支撑计划项目(BE2011843);南京邮电大学人才引进项目

作者简介:卜益民(1970-),男,讲师,硕士,主要开展通信技术、信号处理与信息融合等方面的研究;陈小惠,教授,博士,主要开展网络化测控技术、传感器网络与信息融合等方面的研究。

融<sup>[11]</sup>等方面,研究人员和学者已开展把神经网络和粗糙集结合进行应用方法的研究和推广。

基于上述分析,文中的研究特点在于:

(1)在具有大量传感器的多传感器系统中,将粗糙集与神经网络集合起来进行数据融合的研究;

(2)多传感器系统的实验环境是一个由 54 个传感器组成的网络,传感器网络系统中的数据检测量、数据复杂程度都比前述论文要高;

(3)在多传感器网络系统中,选用实际的修改版 TinyOS 网络协议栈进行实验,这样可以非常方便地把文中所研究的 OS 数据融合层移植到实际系统中。

## 1 粗糙集理论

波兰学者 Z. Pawlak 在 20 世纪 80 年代提出了粗糙集论,粗糙集论是一个用于数据分析的有效的数学方法和工具,由于它在分析与处理不完备、不一致和不精确的信息事件等方面拥有优良的性能,故作为在知识获取、数据分析与处理等方面的一种有效的和极具潜力的数学方法和工具,已越来越受到许多学者和研究工作者的极大关注。

在运用粗糙集论处理信息事件时,其特点是可以采用一张数据表的形式来描述一个信息系统,而描述信息系统的数据库中的每个元素值都是由对象的各个属性值决定的。其中,数据库的列描述为对象的属性,数据库的行描述为研究对象。

用数学形式来描述一个信息系统,即:  $S = (U, A)$ , 其中,  $A$  为对象属性,  $U$  为对象论域,  $A$  和  $U$  为非空有限集合;对象属性元素定义为:  $a: U \rightarrow V_a$ , 其中,  $V_a$  为对象属性  $a$  的值域;对于二元关系  $I(B)$  中的不可区分关系,定义如下:  $(x, y) \in I(B)$  当且仅当  $a(x) = a(y)$ , 且所有  $a \in A$ , 其中每个  $x$  元素在  $a$  上的取值由  $a(x)$  表示。一个族等价类别  $I(B)$ , 即为一个  $B$  的分类,由  $U/I(B)$  来表示,简单的用  $U/B$  来表示;一个等价类别  $I(B)$ , 等价于区块分割的  $U/B$ 。

进一步把属性分为两类:决策属性和条件属性,同时由决策表来描述信息系统。则定义:  $S = (U, C, D)$ , 这里  $D$  为决策属性集,  $C$  为条件属性集,并且  $C \cup D = A$ 。

一个信息系统中往往存在许多属性是重复的和冗余的,需对信息系统中的属性表述进行约简,约简的目的在于使信息表述中不包含多余的属性,同时确保最小属性集分类的正确性,即约简后的信息系统属性表述既最简洁又正确。采用数学描述为:等价关系  $B$  属于  $A$ , 当存在属性  $b \in B$ , 可使  $U/B = U/(B - b)$  时,称  $b$  为  $B$  中可省略的,否则称  $b$  是  $B$  中不可省略的。当  $B$  中不存在可省略的属性时,称  $B$  为独立的,即为  $B$  的一

个约简,记作为  $\text{red}(B)$ 。

## 2 自组织神经网络

自组织神经网络是一种能对任意复杂环境的输入模式完成自组织识别和自稳定的网络系统<sup>[12]</sup>,可以在未知数据类别的情况下进行自组织聚类。该网络使用自稳机制原理和竞争学习方法进行稳定的无监督分类,可以实现实时的学习,并对已学习过的模式实现自动识别和快速的响应。

自组织神经网络结构有注意子系统和取向子系统两部分。注意子系统由特征表示场  $F_1$  和类别表示场  $F_2$  以及两组它们之间的长期记忆系数 LTM 组成,  $F_1$  为网络输入比较层,是整个系统的核心,  $F_2$  是网络识别层,完成网络中各神经元的竞争学习。取向子系统则由重置机构组成。对于每个神经元可能有两种输入,一种输入为兴奋激励,另一种为抑制激励。 $S$  为输入特征向量,当第  $k$  个特征向量输入后,  $F_1$  层、 $F_2$  层按式(1)~式(7)进行变化。

$$U^{(k)} = \frac{V^{(k)}}{\|V^{(k)}\| + e} \quad (1)$$

$$W^{(k)} = S^{(k)} + aU^{(k)} \quad (2)$$

$$Q^{(k)} = \frac{P^{(k)}}{\|P^{(k)}\| + e} \quad (3)$$

$$X^{(k)} = \frac{W^{(k)}}{\|W^{(k)}\| + e} \quad (4)$$

$$P^{(k)} = F(X^{(k)}) + bF(S^{(k)}) \quad (5)$$

$$V^{(k)} = F(X^{(k)}) + bF(S^{(k)}) \quad (6)$$

$$Y_j = \sum_{i=1}^m b_{ij} p_i \quad (7)$$

$$\text{式中: } F = (f_1, f_2, \dots, f_m); f_i(x) = \begin{cases} x, & x \leq \theta \\ 0, & x > \theta \end{cases}; \theta \text{ 为噪声}$$

抑制系数;  $b_{ij}$  为底向上矢量;  $p_i$  为顶向下矢量。把最大值对应的神经元作为获胜神经元  $J$ , 即:

$$Y_j = \max \{Y_j, j = 1, \dots, l\} \quad (8)$$

神经元  $J$  竞争获胜后能否作为系统输出模式,必须根据警戒参数  $\rho$  和相似度  $R$  做进一步判断,按式(5)调整  $P$ ,按式(9)计算相似度:

$$R = \frac{U + cP}{\|U\| + c\|P\| + e} \quad (9)$$

式中:  $R = (r_1, r_2, \dots, r_m)$ ,  $c$  为系数。

若  $\|R\| \leq \rho$ , 将该神经元的竞争资格取消,同时在  $F_2$  剩余的神经元中重新选择、进行竞争;否则,把竞争中获胜的神经元按式(10)和式(11)实施长期的记忆系数学习。

$$t_j(k+1) = t_j(k) + ad(1-d) \left[ \frac{U_j^{(k)}}{1-d} - t_j(k) \right] \quad (10)$$

$$b_j(k+1) = b_j(k) + ad(1-d) \left[ \frac{U_j^{(k)}}{1-d} - b_j(k) \right] \quad (11)$$

这里: $U_j^{(k)}$  为在第  $k$  次获胜时神经元  $J$  对应的  $U$ ,  $a$  为学习速率;反复学习,长期记忆将收敛于一个稳定值,这样就完成了对输入模式的识别。

当所有神经元均不能匹配时,在  $F_2$  中添加一个新的神经元,创建新的类别。

3 多传感器测量网络的数据融合应用

为了将前述的数据融合算法应用于多传感器测量网络中,设相对应警戒系数  $\rho$ ,采用自组织神经网络聚类,实现数据融合。

3.1 传感器网络数据

在 TOSSIM 环境下进行模拟,实验数据选用来自 inter Berkeley Research lab 的从 2004 年 2 月 28 日到 4 月 5 日的 54 个 Mica2Dot 传感器数据。

考虑一定的数据量,采用在设定的时间范围内,每个传感器规定每隔 31 秒发送一个含有温度、湿度、电压和光照等数值的数据包,把所有传感器发送的所有数据构成一个数据文件(大小约为 150MB,条数约为 2300 万)。在这些传感器测量数据中,湿度为经过温度校正的相对湿度,范围从 0 ~ 100%,光照单位是 lux (10 万 lux 相当于日光直射的强度,400lux 大致为一个明亮的办公室的亮度,1lux 等价于月光强度),温度单位是摄氏。电压单位为伏特,范围是 2 ~ 3V。在一般情况下,传感器采用锂电池供电,锂电池供电的优点是在使用期内它的供电电压相当稳定,其电压值的高低仅受电池的工作环境温度的一定影响。

3.2 传感器网络数据的粗糙约简

在传感器网络中,来自相邻传感节点的检测数据具有一定的相似性,也即这些数据中存在一定的冗余特征。另一方面,在短时间内来自同一个传感器的检测数据也具有一定的相似性,具有冗余特征。根据 Govindan 的研究报告,假定区域  $S$  内的分布密度为  $P(x,y)$ ,则区域  $S$  内,传感器节点检测数据之间的冗余数为  $\eta = \xi Se^{\rho}$ 。

为了把粗糙集理论应用于传感器网络的数据处理中,把不同时间传感器节点检测数据:温度、湿度、电压与光照值,构成一个数据样本空间,提取传感器节点检测数据的典型值组成数据样本空间的属性;在某一给定时间范围内,把可以代表某一组群的某个传感器数据作为数据样本空间的决策属性。根据粗糙决策理论,由决策表产生分辨矩阵,分辨矩阵中浓缩了决策表中相关属性区分信息,分辨矩阵在粗糙集论中非常重要,它可用于决策表的约简和核的求解,而核的不同将直接影响到传感器检测数据的关键属性表述。

24 小时内 54 个传感器检测数据中的温度和光照度的关键属性如图 1 所示。

(a) 光照度的关键属性描述

(b) 温度值的关键属性描述

图 1 关键属性描述

3.3 神经网络聚类

约简传感器测量系统中冗余的属性,形成新的最简决策表,进入神经网络训练这一步。同时,约简传感器测量数据的条件属性,形成传感器的典型数据,并作为数据样本聚类训练自组织神经网络。

在自组织神经网络中,选择不同警戒参数  $\rho$  可以获得不同的分类粗细程度; $\rho$  值越大,分类条件越严格,分类数目越多;警戒系数  $\rho$  越小,模式分类越粗糙,分类数目越少。

在多传感器网络聚类中,如果警戒系数取得较大,就有可能产生错误的分类情况,从而使得传感器数据精度较过度的降低;如果警戒系数取得较小,就会产生分类数目过多的问题,分类越频繁,就越达不到把传输数据量减少的目的。

经过测试,选取  $\rho = 0.9720$  时分类效果最理想。一共输入 15524 个样本进行聚类分析,最后分为 27 类。图 2(a)、(b)分别表示第 1 种和第 27 种传感器数据融合分类分布图。

3.4 传感器测量网络数据融合层协议

在使用文中所述算法时,把数据融合层增加到传感器测量网络协议中,该协议数据融合层的主要功能为:

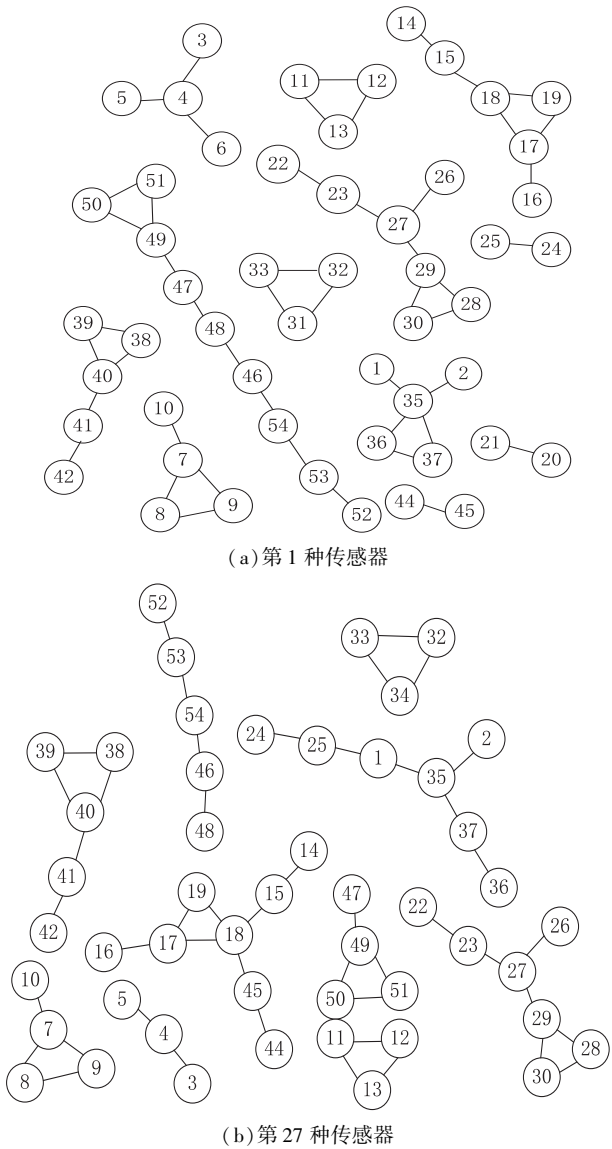


图 2 传感器融合分布情况

(1) 将传感器网络测量系统中的多个传感器的检测数据分别进行数据融合处理,即:数据关联、数据统计、数据分类及优化组合等,形成比特流更加确定、更加精细的出口报文;

(2) 把数据转发给下一跳,完成对感知对象的融合识别与融合估计等。

推荐采用 5 层结构的数据融合功能传感器网络。针对传感器测量网络的特点,加州大学伯克利分校专门为传感器测量网络研究了一个称作 TinyOS<sup>[13]</sup> 的操作系统,TinyOS 操作系统是一个基于事件的易于构造、高效和模块化的组件型应用软件,程序开发和设计人员能够非常方便地通过 nesC 语言进行程序的编制,描述组件与组件之间的事件/命令。通常,采用一种自上到下的树状结构,即:上层组件发命令给下层组件,而下层组件发布信号通知事件的发生,最底层组件通过接口直接与硬件进行交互。

添加数据融合层到网络中,在 TinyOS 协议栈中的

实现方法为:

(1) 把数据融合模块、路由模块添加到 TinyOS 协议栈中;

(2) 修改与通信相关的模块代码。

修改后的通信协议栈结构如图 3 所示:

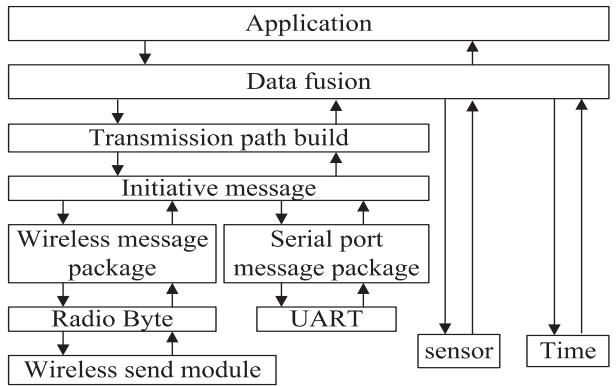


图 3 修改后的通信协议栈结构

4 仿真实验

在仿真实验中,选用 TinyOS 所附的 TOSSIM 模拟器<sup>[14,15]</sup>。由于 TOSSIM 模拟器具有与传感器网络硬件运行相同代码的特点,TOSSIM 模拟器的仿真编译器能够直接在 TinyOS 应用组件表中编译和运行仿真程序。为使 TOSSIM 硬件中断改为离散仿真事件,需更换 TinyOS 下层的部分硬件相关组件;仿真器提供事件中断驱动,TinyOS 组件、上层的应用组件等不变。

传感器工作模式的相关流程图如图 4 所示,图 4 中采用 Python 语言来描述控制。

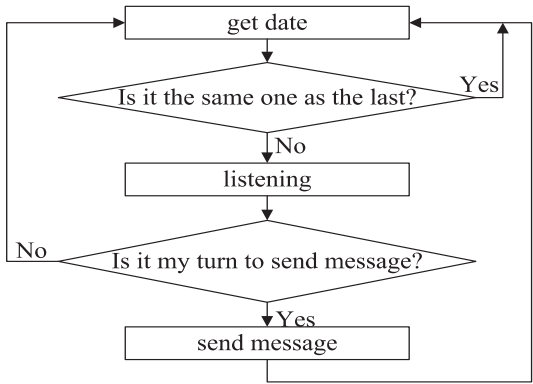


图 4 Python 控制传感器工作流程图

图 5 所示为传感器网络接收和发送数据包的数量对比。由图 5 可见,运用粗糙集与神经网络聚合后,传感器的收发信息量明显减少。因为收发数据包是传感器网络的主要能量消耗,传感器执行计算消耗较小电能,如:传感器执行上千条计算指令与传输 1 比特信息所需要的电能相当;同时,通信的能耗大约是传送距离的三次方,很明显,减少了不必要的数据传输,节约了多传感器网络的能量消耗。



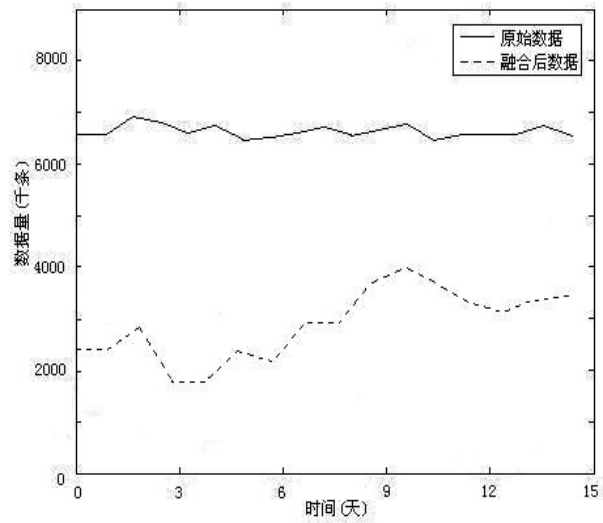


图5 传感器接收和发送信息量对比图

5 结束语

文中把 54 个传感器组成的网络作为仿真实验环境,在该较为复杂的仿真网络环境下,模拟对比了神经网络的聚类特性与原始传感器网络特性,由图 5 可见,神经聚类方法明显改善了网络性能。

利用神经网络的学习能力、粗糙集的约简性能,并将两者的优点进行有效地结合,较大地提高了多传感器测量网络数据融合性能;在传感器测量网络中的 OS 中添加网络数据融合协议层,可使传感器测量网络的数据传递量和网络能耗有较大幅度的降低;在数据融合中,较好地解决了多传感器融合方向的问题;尽管仍存在着一些缺点——如:传感器数据处理要求较高、网络融合结构样本难以总结,在不久的将来,随着传感器硬件性能的不断完善,这些问题便能够有效地得到解决。

参考文献:

[1] Waltz E, Llinas J. Multisensor Data Fusion[M]. Boston London; Artech House, 1990.

[2] D]. 济南: 济南大学, 2008.

[6] Lammel R. Google's MapReduce programming model revisited[M]. Redmon, USA; Data Programmability Team Microsoft Corp., 2007.

[7] Azzedin F, Maheswaran M. Evolving and Managing Trust in Grid Computing Systems[C]//Proceedings of the IEEE Canadian Conference on Electrical Computer Engineering. New York: IEEE Press, 2002: 1424-1429.

[8] Foster I, Kesselman C, Tuecke S. The anatomy of the grid: enabling scalable virtual organizations[J]. International Journal

[2] Krunz W, Muqattash A. Transmission Power Control in Wireless Ad Hoc Networks: Challenges, Solutions, and Open Issues [J]. IEEE Network, 2004(9/10): 115-122.

[3] Ross A, Jain A K. Information fusion in biometrics[J]. Pattern Recognition Letters, 2003, 24(13): 2115-2125.

[4] 樊四良, 常亮, 郭丙君. 基于模糊神经网络的通用模型自适应控制[J]. 自动化与仪器仪表, 2009(5): 9-11.

[5] 张腾飞, 李云. 基于粗糙-神经网络的非线性系统逆模型控制[J]. 仪器仪表学报, 2009, 30(8): 1726-1733.

[6] 刘决仕, 金晶, 陈华曦. 一种结合粗糙集与神经网络的多传感器信息融合算法[J]. 计算机应用与软件, 2009, 26(2): 14-16.

[7] 李孝全, 庄德慧, 张强. 基于粗糙径向基神经网络的电网故障诊断新模型[J]. 电力系统保护与控制, 2009, 37(18): 20-24.

[8] 邓武, 杨鑫华, 赵慧敏, 等. 粗糙集、神经网络和专家系统模型用于电力系统故障诊断[J]. 高电压技术, 2009, 35(7): 1623-1628.

[9] 李界家, 房智超, 孙璐璐. 粗糙集-神经网络在铝电解故障诊断中的应用[J]. 沈阳建筑大学学报, 2009, 25(4): 792-795.

[10] 孙健敏, 吴昊, 李书琴. 粗糙神经网络在农业病虫害诊断中的应用[J]. 农机化研究, 2009(11): 197-199.

[11] 王天娥, 叶德谦, 季春兰. 粗糙集属性约简方法在股票预测中的应用研究[J]. 计算机工程与应用, 2009, 45(30): 227-229.

[12] Wann Chin-Der, Thomopoulos S C A. A comparative study of self-organizing clustering algorithms dignet and ART2 [J]. Neural Networks, 1997, 10(4): 737-753.

[13] Hinden R. Virtual Router Redundancy Protocol (VRRP) [S]. IETF3768, 2004.

[14] Wolf F I. Commbench 2A telecommunications benchmark for network processors [C]//IEEE International Symposium on Performance Analysis of Systems and Software. Austin, Tx; [s. n.], 2000: 154-162.

[15] Karlin S, Peterson L. VERA: An extensible router architecture [J]. Computer Networks, 2002, 38(3): 277-293.

(上接第 220 页)

on Supercomputer Applications, 2001, 15(3): 200-222.

[9] Subramani V, Kettimuthu R, Srinivasan S, et al. Distributed Job Scheduling on Computational Grids Using Multiple Simultaneous Requests[C]//Proceedings of 11th IEEE International Symposium on High Performance Distributed Computing. Washington, DC, USA: IEEE Computer Society, 2002: 359-366.

[10] 逢华, 王龙, 王剑辉. 一种基于移动 Agent 的分布式并行计算模型[J]. 计算机技术与发展, 2011, 21(2): 73-76.

粗糙集和神经网络在数据融合中的应用研究

|          |   |
|----------|---|
| 作者:      | <a href="#">卜益民, 陈小惠</a>  |
| 作者单位:    | <a href="#">卜益民(南京邮电大学 继续教育学院, 江苏 南京210003), 陈小惠(南京邮电大学 自动化学院, 江苏 南京210003)</a> |
| 刊名:      | <a href="#">计算机技术与发展</a>  |
| 英文刊名:    | <a href="#">Computer Technology and Development</a>                             |
| 年, 卷(期): | 2013(4)   |

本文链接: [http://d.g.wanfangdata.com.cn/Periodical\\_wjtz201304056.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjtz201304056.aspx)