Static Node Classification Algorithm Based on Improved Rough Set Theory in Wireless Sensor Networks*

WANG Xiaoyan[1*], ZHANG Nan[2]

(1.College of Computer Engineering, Shanxi Vocational University of Engineering Science and Technology, Jinzhong Shanxi 030619, China;
2.School of Mechanical and Electrical Engineering, Shanxi Datong University, Datong Shanxi 037003, China)

Abstract: Because the static nodes in wireless sensor networks contain some uncertain information, the decision level fusion technology is difficult to effectively describe the data, resulting in poor classification of static nodes. A static node classification algorithm for wireless sensor networks based on improved rough set theory is proposed. The multiple sampling method is used to weight the test results, the weighted moving average algorithm is used to test the neighbor nodes, and the hybrid algorithm based on node density is used to denoise the data. The improved rough set model is constructed by using the variable precision rough set in the error classification rate. Based on the improved rough set theory, the static node characteristics of wireless sensor network are extracted. The static node classification model of wireless sensor network is built by using support vector machine to classify nodes. The simulation results show that the classification accuracy of the proposed method is always higher than 96.852%, the node recall rate is more than 97.321%, the F1 value is more than 97%, and the energy consumption is between 12.52 J and 90.20 J, which proves that the proposed algorithm has good network static node classification performance.

Key words: wireless sensor; network static node classification; improve rough set theory; weighted moving average algorithm; variable precision rough set

EEACC: 6150P **doi**: 10.3969/j.issn.1004-1699.2023.08.019

改进粗糙集理论的无线传感器网络静态节点分类算法*

王晓艳[1*], 张 楠[2]

(1.山西工程科技职业大学计算机工程学院, 山西 晋中 030619; 2.山西大同大学机电工程学院, 山西 大同 037003)

摘 要: 因无线传感器网络中静态节点包含一定的不确定信息, 决策级融合技术很难对数据进行有效刻画, 导致网络静态节点分类效果差。提出一种基于改进粗糙集理论的无线传感器网络静态节点分类算法。采用多次采样法对测试结果加权处理, 通过加权移动平均算法对邻居节点进行测试, 基于节点密度的混合式算法对数据去噪处理。利用错误分类率中的变精度粗糙集构建改进的粗糙集模型, 基于改进粗糙集理论提取无线传感器网络静态节点特征, 通过支持向量机构建无线传感器网络静态节点分类模型对节点分类处理。仿真结果表明, 所提方法的分类正确率始终高于 96.852%、节点召回率在 97.321%以上、F1 值多于 97%、能量消耗在 12.52 J~90.20 J 之间, 证明所提算法具有良好的网络静态节点分类性能。

关键词: 无线传感器; 网络静态节点分类; 改进粗糙集理论; 加权移动平均算法; 变精度粗糙集

中图分类号: TP393.04 **文献标识码**: A **文章编号**: 1004-1699(2023)08-1310-06

无线传感器网络是一种分布式网络, 利用传感器可以获取更加精准的数据变化趋势, 同时还可以采集数据特征, 将获取的全部信息直接发送给用户。由于受到周围环境和自身等因素的影响, 在无线传感器网络数据中存在大量的噪声, 导致无法准确区分不同类型的无线传感器网络静态节点[1-2]。

为此, 国内外相关专家针对无线传感器网络静态节点分类方面的内容展开了大量研究, 樊成等[3]将图卷积网络、多任务学习以及自适应加权策略三者有效结合, 并且在模型训练过程中加入对应的节点相关性, 以此为依据展开权重任务划分操作, 最终实现网络节点分类处理。谢润山等[4]主要基于函数链神经网络构建深度分类器, 通过分类器完成节点分类处理。Uppal 等[5]利用下降曲线对传感器节点进行预测, 实现静态节点分类。但因无线传感器网络中静态节点包含一定的不确定信息, 通常认为不确定信息主要来源于传感器对目标的不充分观测所获取的不精确数据, 以及对目标的局部观测所得

项目来源: 2022 年全国煤炭行业教育研究课题项目(ZMZA20220007)
收稿日期: 2022-12-01 **修改日期**: 2023-01-03

到的不完整数据。不精确数据和不完整数据都会使静态节点分类问题变得十分复杂,且极易造成错误的分类结果。传统的基于粗糙集理论的决策级融合技术很难对数据中的不确定信息进行有效的刻画。

在以上几种网络节点分类算法的基础上,提出一种改进粗糙集理论的无线传感器网络静态节点分类算法。采用多次采样法对数据进行加权处理,利用 DHA 算法对含有噪声的无线传感器网络静态节点数据进行滤波处理,通过错误分类率中的变精度粗糙集对粗糙集理论进行改进,提取静态节点特征,基于支持向量机构建无线传感器网络静态节点分类模型,对节点进行分类处理。仿真结果表明,所提算法可以得到精准的节点分类结果。

1　无线传感器网络静态节点分类方法设计

根据多次采样法采集节点数据后,对数据进行加权处理,通过 DHA 算法去除节点数据中的噪声,利用错误分类率中的变精度粗糙集和错误分类器,改进粗糙集模型,提取静态节点特征,结合支持向量机完成节点分类处理。

1.1　无线传感器网络静态节点数据滤波

采集无线传感器网络静态节点数据过程中,如果节点数据中的随机噪声 $R_{i,t}$ 服从期望趋近为 0,则对应的噪声均值可以表示为:

$$U_{i,t}=\frac{1}{r_{i,t}}\sum_{t=1}R_{i,t} \tag{1}$$

式中:$U_{i,t}$ 代表噪声均值;$r_{i,t}$ 代表在 t 时间段内第 i 个带有噪声的节点数据。

在给定噪声均值在允许误差范围内以及节点数量稳定的情况下,假设 n 次采样的平均随机噪声落在允许误差范围内,则采用其代替真实数据的置信度,对应的计算式为:

$$d_{i,t}=P(x,y)*\{\tilde{E}+U_{i,t}\} \tag{2}$$

式中:$d_{i,t}$ 代表无线传感器网络静态节点数据对应的置信度值;$P(x,y)$ 代表真实节点数量的置信度;$\tilde{E}$ 代表平均随机噪声对应的误差范围。

假设给定无线传感器网络静态节点数据的置信度之后,需要估计节点数量,对应的计算式为:

$$\alpha_{i,t}=\frac{U_{i,t}}{P(x,y)-1}\times d_{i,t}(\tilde{E}+1) \tag{3}$$

式中:$\alpha_{i,t}$ 代表无线传感器网络静态节点的数量估计值。同时,通过对上述公式分析可知,当无线传感器网络静态节点数量开始增加,对应的节点置信度也随之增加。

为了有效滤除无线传感器网络静态节点数据中的噪声,主要引入 Kalman 线性滤波或者线性回归给出一个预测区域 $[a_{it},b_{it}]$,假设节点落在预测区域内,则说明节点满足置信度需求,不需要去噪处理;假设节点落在预测区域外,主要通过两种不同的方式有效滤除节点数据中的噪声,分别为:

①本地测试:

采用多次采样的方法提高数据的置信度;在测试邻居节点时,对测试结果展开加权处理,得到去噪处理后的数据。

②邻居测试:

在邻居节点测试过程中,各个节点数据之间的置信度和邻居节点数量两者之间存在密切关联,检测到邻近样本之间的节点数量越多,对应节点的置信度取值也就越高。反之,则说明邻居节点数量比较少,对应的置信度也就越低,同时对应的节点数量也没办法达到设定的置信度需求。

在给定置信度展开邻居测试过程中,采集的节点数据达到设定的置信度时,则节点需要的最少邻居节点需要满足设定的要求。优先分析无线传感器网络静态节点的稠密情况,通过 WMA 算法对邻居节点进行测试处理,由于此时邻居节点的数量比较大,所以各个节点都需要接收大量的邻居节点数量,对应的网络通信能量也随之增加,加重网络的负担,导致网络运行不畅。为了有效避免上述情况的发生,确保数据达到设定的置信度,需要设定各个节点的能量消耗范围 d_{pt},具体的计算式如下:

$$d_{pt}=\frac{1}{[a_{it},b_{it}]}\times P(x,y)\times(d-p_{lx}) \tag{4}$$

式中:d 代表节点消耗的能量值;p_{lx} 代表节点接收一个字节所消耗的能量。

为了有效控制无线传感器网络静态节点测试的通信消耗,确保数据准确性后,节点对应的邻居节点会形成一个随机数据 r,如果 $0<r<[a_{it},b_{it}]$ 则节点需要将感知数据发送给另外一个节点;反之,则不需要将感知数据发送给另外一个节点。通过随机变量 s_i 判断节点 i 的邻居节点 j 是否需要将感知数据发送给另外一个节点,对应的计算式如下:

$$s_i=\begin{cases}1 & \text{邻居节点 } j \text{ 把感知数据发送给节点 } i\\ 0 & \text{邻居节点 } j \text{ 不把感知数据发送给节点 } i\end{cases} \tag{5}$$

受到网络带宽问题的影响,当网络中需要接收和发送的数据包数量达到一定程度,大量信息堆积在一起会产生拥堵。发生拥堵后,传感器网络中的数据无法接收和传送,同时数据受到噪声的影响,影响了网络的正常通信效果,导致网络的通信功能无法顺利展开。所以,进行邻居测试时,在节点数量较多的情况

下,有效调节邻居数量也就显得十分重要。在计算节点稠密度的过程中,采用 DHA 算法对数据进行去噪处理可以有效避免数据拥堵情况的发生,确保节点通信的正常运行。通过上述分析,给出无线传感器网络静态节点数据滤波的详细操作步骤[6-8]:

①优先对无线传感器网络静态节点展开密度测试,对决策网络进行初始化处理,每个节点对应一个节点 ID 消息包,可以同时接收周围邻居节点信息。

②分别计算无线传感器网络静态节点时间维和空间维的加权平均值 m_{it}^{*} 和 r_{it}^{*},如式(6)所示:

$$\begin{cases} m_{it}^{*}=\dfrac{\omega_{it}\times v_{it}+\cdots\beta_{it}\times\tau_{it}}{u} \\ r_{it}^{*}=\dfrac{\sum\limits_{j\in N}c_{ij}\times\omega_{it}\times v_{it}}{e_{it}\times\omega_{it}} \end{cases} \tag{6}$$

式中:ω_{it}代表无线传感器网络静态节点的方差值;v_{it}代表节点的传播速度;β_{it}代表邻居节点形成的广播消息包;τ_{it}代表邻居节点在传感器网络中的总数;u代表网络静态节点密度;N代表邻居节点总数;c_{ij}代表无线传感器网络静态节点通信过程中产生的开销;e_{it}代表邻居节点的抽样总数。

③随着时间的推进,通过样本数据可以更加精准反映网络静态节点的运行情况,在完成式(6)计算后,引入 DHA 算法对含有噪声的无线传感器网络静态节点数据进行滤波处理,进而获取去噪后的节点数据[9-10],详细的计算式如式(7)所示:

$$f(x,y)=\frac{\{u\times m_{it}^{*}\times e_{it}\times\omega_{it}\}^{2}}{u+\omega_{it}} \tag{7}$$

式中:$f(x,y)$代表经过滤波处理后的无线传感器网络静态节点滤波处理结果。

1.2 改进粗糙集理论的无线传感器网络静态节点分类

随着数据规模的不断增加以及数据类型日益复杂多样,粗糙集理论的应用场景也日益多样化[11-13]。粗糙集理论是对数据进行分析及处理的理论,通过错误分类率中的变精度粗糙集对粗糙集理论进行改进,提高数据的综合处理效果。将改进的粗糙集理论应用到无线传感器网络静态节点分类过程中。在不完备的信息系统中,设定两个随机无线传感器网络静态节点的差异度为 $S_a(x,y)$,对应的计算式如下:

$$S_a(x,y)=\frac{|g(a)-g(b)|}{|g_{\max}-g_{\min}|} \tag{8}$$

式中:$g(a)$和 $g(b)$分别代表节点数据对应的属性值;$g_{\max}$和 $g_{\min}$分别代表节点数据的最大属性值和最小属性值。

由于网络静态节点数据的不完备性,需要根据得到的节点数据差异度进行划分处理,将明确属性的节点数据划分到对应的类中[14-15]。为了避免在划分过程中出现划分结果不准确的问题,需要引入阈值展开进一步的计算和分析,设定无线传感器网络静态节点之间的综合差异度 $F_b(x,y)$为:

$$F_b(x,y)=\frac{\{g(a)-g(b)\}\times S_a(x,y)}{m} \tag{9}$$

式中:m代表无线传感器网络静态节点总数。

在不对数据造成损失的情况下,需要确保数据的综合处理效果得到明显提升,通过错误分类率中的变精度粗糙集构建改进的粗糙集模型,如式(10)所示:

$$\kappa(x,y)=\frac{|F_b(x,y)-S_a(x,y)|}{\{u\times m_{it}^{*}\times e_{it}\times\omega_{it}\}^{2}} \tag{10}$$

式中:$\kappa(x,y)$代表改进后的粗糙集模型;在改进处理后的模型中,近似区域和各个参数之间存在密切的关联,假设参数的取值发生改变,则近似区域相应地也会发生改变。在粗糙集模型中引入错误分类器,主要目的是进一步对模型进行优化处理,进而获取更大和更小的上下近似集,同时也为后续的无线传感器网络静态节点分类奠定坚实的基础。采用改进后的模型提取无线传感器网络静态节点,详细的操作流程如图 1 所示。

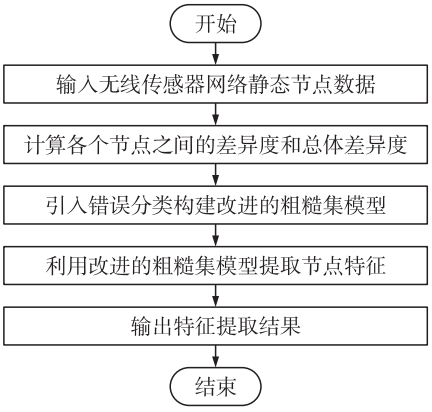

图 1 基于改进粗糙集理论的无线传感器网络静态节点特征提取方法

通过式(11)计算随机两个节点之间的入度和出度相加的和 $C(u_{xy})$:

$$C(u_{xy})=\frac{\{e_{ij}-1\}}{D(x,y)+D(x,y-1)} \tag{11}$$

式中:e_{ij}代表传感器网络静态节点的维度特征;$D(x,y)$代表传感器网络静态节点的聚集系数,通过节点的聚集系数描述节点对应的聚集程度,对应的计算式为:

$$D(x,y)=\frac{|C(u_{xy})|}{|g(a)-g(b)|} \tag{12}$$

经过上述分析,引入支持向量机[16]建立无线传感器网络静态节点分类模型完成节点分类处理[17-18],详细的操作步骤如下所示:

①为了提升分类模型的泛化能力,需要在支持向量机目标函数以及约束条件的基础上加入松弛函数以及惩罚参数展开校正处理,进而获取优化处理后的目标函数 $p(x,y,z)$:

$$p(x,y,z)=\frac{D(x,y)\times C(u_{xy})}{|g_{\max}-g_{\min}|\times\{u\times m_{it}^{*}\times e_{it}\times\omega_{it}\}^{2}} \tag{13}$$

②通过拉格朗日优化方法将相关函数的优化问题转换处理,形成对偶问题求解,详细的操作步骤如下所示:

$$S(x,y,z)=\frac{\{\tilde{E}+U_{i,t}\}+|g_{\max}-g_{\min}|}{|F_b(x,y)-S_a(x,y)|} \tag{14}$$

式中:$S(x,y,z)$代表对偶问题的求解结果。

③通过高斯核函数替换径向基函数,同时将全部无线传感器网络静态节点特征映射到高维特征空间内,进行线性分类处理 $K(x,y,z)$,其计算式为:

$$K(x,y,z)=\frac{S(x,y,z)}{\kappa(x,y)\times F_b(x,y)\times S_a(x,y)} \tag{15}$$

④利用处理后的函数在高维特征空间内建立最优分类超平面,得到分类函数,通过分类函数结合支持向量机建立无线传感器网络静态节点分类模型,将提取的静态节点特征输入到分类器中,通过上述步骤完成无线传感器网络静态节点分类处理。

2　仿真结果与分析

为了验证所提改进粗糙集理论的无线传感器网络静态节点分类算法的有效性,在 Github 数据集中选取 900 个节点作为仿真节点,将无线传感器网络静态节点数据输入到分类模型中,计算节点之间的差异度后,利用四种方法提取无线传感器网络静态节点特征,并将其输入到分类器中,对节点进行分类处理。仿真参数设置如表 1 所示。

选取四个指标评估各个算法的分类性能,分别为:

①正确率(Accuracy):正确分类节点在节点总数中所占据的比例,正确率越高说明该算法的分类性能越好;

②召回率(Recall):节点被正确划分到所属差别的比例,主要是用来衡量对节点的查找是否全面,召回率越高,查找效果越好;

③F1 值:是正确率和召回率两者的综合评估指标,F1 值越大,说明该算法的分类效果越好;

④能量消耗:是传感器静态节点分类过程中消耗的能量,消耗能量越小,分类效果越好。

表 1　仿真参数设置

变量名	含义	取值
$\alpha_{i,t}$	静态节点数量估计值	400~2 000
d_{pt}	节点能量消耗范围	[0,50]
d	节点消耗能量值	2
p_{lx}	节点接收一个字节消耗能量	1
m	静态节点总数	900

选取文献[3]算法、文献[4]算法和文献[5]算法作为对比算法,仿真结果如下:

2.1　Accuracy 仿真

首先分析四种算法的 Accuracy 变化情况,仿真结果如图 2 所示。

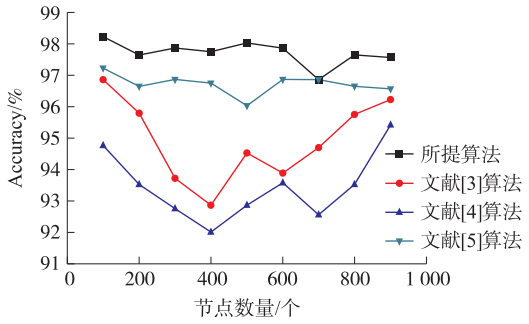

图 2　不同分类算法的 Accuracy 结果对比

分析图 2 中的数据可知,各个方法的 Accuracy 会随着节点数量的变化而变化,文献[3]算法、文献[4]算法和文献[5]算法的 Accuracy 范围分别为 92.854%~96.852%、92.002%~95.411%和 96.021%~97.215%,而所提算法的 Accuracy 值始终高于 96.852%。因为所提算法对全部无线传感器网络静态节点数据展开了滤波处理,所以后期获取的无线传感器网络静态节点正确率取值会更高一些,进一步验证了所提算法的分类性能。

2.2　Recall 仿真

进一步分析四种不同分类算法的 Recall 变化情况,仿真结果如图 3 所示。

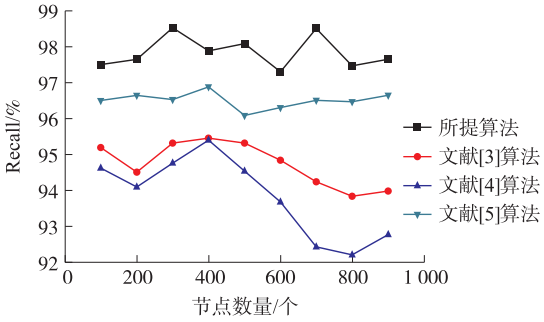

图 3　不同分类算法的 Recall 结果对比

从图 3 可以看出,各个算法的 Recall 结果随着节点数量的变化而变化,但是各个算法的变化幅度并不是很大。在四种分类算法中,文献[3]算法、文献[4]算法和文献[5]算法的 Recall 范围分别为 93.854%~95.471%、92.220%~95.415%和 96.102%~96.899%,而所提算法获取的 Recall 值始终高于 97.321%,证明所提算法的召回率更高,查找更全面。

2.3 F1 值仿真

F1 值结果如图 4 所示。

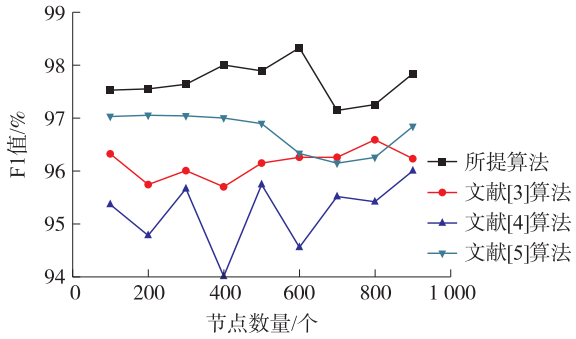

图 4 不同分类算法的 F1 值对比

由图 4 可知,在相同的节点数量下,文献[3]算法的 F1 值在 95%左右,文献[4]算法的 F1 值在 96%左右,文献[5]算法的 F1 值在 96.5%左右,而所提算法的 F1 值最高,在 97%以上,说明所提算法的分类性能较好,验证了所提算法的优越性。

2.4 能量消耗仿真

四种算法的能量消耗情况如图 5 所示。

由图 5 中的数据可知,当节点数量呈直线上升时,各个分类算法产生的能量消耗也会相应增加。经过对比分析可知,在四种算法中,文献[3]算法、文献[4]算法和文献[5]算法的能量消耗范围分别为 15.66 J~105.69 J、16.99 J~120.99 J 和 15.52 J~100.20 J,而所提算法的能量消耗在 12.52 J~90.20 J,能量消耗最低。主要是因为所提算法对节点数据展开了滤波处理,有效简化了分类环节,进而降低分类过程中产生的能量消耗,验证了所提算法可有效降低传感器静态节点分类过程中的能量消耗。

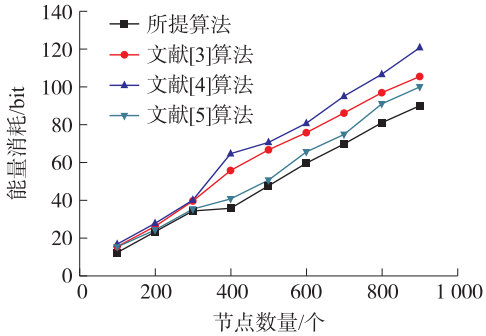

图 5 不同分类算法的能量消耗结果对比

3 结束语

为了更好区分无线传感器网络中不同类型的静态节点,提出一种改进粗糙集理论的无线传感器网络静态节点分类算法。仿真结果证明:

①所提算法的 Accuracy 值始终高于 96.852%,分类性能更好;②所提算法获取的 Recall 值始终高于 97.321%,证明该算法查找更全面、查找效果最好;③所提算法的 F1 值高于 97%,说明该算法的分类效果更好;④所提算法的能量消耗在 12.52 J~90.20 J 之间,能量消耗最低,分类效果最好。

通过上述结果证明,采用所提算法对无线传感器网络静态节点分类可以获取高精度的分类结果,同时还可以有效降低分类过程中产生的能量消耗。

参考文献:

[1] 徐逸夫,段隆振. 基于蛙跳算法的无线传感器网络节点重部署[J]. 计算机仿真,2021,38(10):328-332.

[2] 刘颖,车鑫. 基于图网络优化及标签传播的小样本图像分类算法[J]. 信号处理,2022,38(1):202-210.

[3] 樊成,王布宏,田继伟. 基于多任务学习图卷积模型的航空网络节点分类[J]. 系统工程与电子技术,2022,44(7):2341-2349.

[4] 谢润山,王士同. 基于函数链神经网络的深度分类器[J]. 南京航空航天大学学报,2020,52(5):736-745.

[5] Uppal M,Gupta D,Anand D,et al. Fault Pattern Diagnosis and Classification in Sensor Nodes Using Fall Curve[J]. Computers,Materials & Continua,2022,72(1):1799-1814.

[6] 张科,姜海旭,王靖宇. 基于滑动最速跟踪微分器的遥测数据滤波方法[J]. 西北工业大学学报,2020,38(3):515-522.

[7] 刘政奇,甘淑. 改进移动曲面最小二乘拟合的 Lidar 数据滤波优化[J]. 应用激光,2022,42(3):154-160.

[8] Javaid N,Khan K U,Khattak A M,et al. Community Search in a Multi-Attributed Graph using Collaborative Similarity Measure and Node Filtering[C]//2021 15th International Conference on Ubiquitous Information Management and Communication(IMCOM),Seoul,South Korea,IEEE,2021:1-8.

[9] 马国庆,王泽坤,李丽丽. 基于自注意力机制深度学习的重磁数据网格化和滤波方法[J]. 石油地球物理勘探,2022,57(1):34-42,5.

[10] 王南天,王雄,孙鹏,等. 连续变迎角试验数据自适应分段拟合滤波方法[J]. 北京航空航天大学学报,2021,47(6):1145-1151.

[11] 洪壮壮,黄兆华,万仲保,等. 基于 GMM 的文本规则挖掘的粗糙集方法研究[J]. 中文信息学报,2020,34(2):56-62.

[12] Zhou K,Wang Q,Gu F. Research on Cloud Service Selec-

tion Based on Rough Set Theory[J]. Journal of Physics: Conference Series,2022,2174(1):012088.
[13] 汤国林,杨文栋,刘培德. 基于区间二型模糊决策粗糙集的三支决策方法[J]. 控制与决策,2022,37(5):1347-1356.
[14] 谢小杰,梁英,王梓森,等. 基于图卷积的异质网络节点分类方法[J]. 计算机研究与发展,2022,59(7):1470-1485.
[15] 蔡瑞初,李烁,许柏炎. 基于异构动态图模型的社交网络节点分类方法[J]. 计算机应用研究,2021,38(9):2635-2639,2645.
[16] Mohapatra S, Khilar P M. Fault Diagnosis in Wireless Sensor Network Using Negative Selection Algorithm and Support Vector Machine[J]. Computational Intelligence, 2020,36(3):1374-1393.
[17] 袁自勇,高曙,曹姣,等. 基于异构图卷积网络的小样本短文本分类方法[J]. 计算机工程,2021,47(12):87-94.
[18] 王乐,韩萌,李小娟,等. 基于动态加权函数的集成分类算法[J]. 计算机应用,2022,42(4):1137-1147.

王晓艳(1981—)女,汉族,山西太谷人,硕士,山西工程科技职业大学计算机工程学院讲师,毕业于兰州大学信息科学与工程学院计算机软件与理论专业,研究方向为人工智能与数据挖掘、粗糙集理论,wangxiaoyan79@163.com;

张　楠(1981—),男,汉族,山西大同人,硕士,教授,硕士生导师,研究方向:无线传感器网络在煤矿的应用研究。