

文章编号:1007-5321(2014)增-0023-07

支持向量机回归预测在网络故障检测中的应用

孟洛明¹, 朱杰辉¹, 杨 杨¹, 孟玲莉², 张平平², 高志鹏¹

(1. 北京邮电大学 网络与交换技术国家重点实验室, 北京 100876; 2. 国网天津市电力公司, 天津 300010)

摘要: 无线传感器网络主要用于收集环境数据,然而传感器可靠性低,容易出现故障并返回异常的监测数据. 因此,如何检测传感器故障成为关乎无线传感器网络监测性能的重要问题. 提出了基于支持向量机回归预测的故障检测算法,通过充分利用历史数据预测传感器的测量值,并根据测量值与实际值的对比有效地划分信誉等级,最后基于置信数据集计算出的置信条件进行故障的检测. 仿真结果证明了算法在节约能量和故障检测率上有着优异的表现.

关键词: 无线传感器网络; 支持向量机回归; 信誉度; 故障检测

中图分类号: TP212;V241.6

文献标志码: A

A Fault Detection Algorithm for Wireless Sensor Networks Based on Support Vector Regression

MENG Luo-ming¹, ZHU Jie-hui¹, YANG Yang¹, MENG Ling-li²,
ZHANG Ping-ping², GAO Zhi-peng¹

(1. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. State Grid TianJin Electric Power Company, Tianjin 300010, China)

Abstract: The main role of wireless sensor networks is to collect environmental data. As for the sensor nodes are vulnerable and work in unpredictable environments, the sensors are possible out off work and return to unexpected response. Therefore, fault detection is important in wireless sensor networks. The authors propose a fault detection algorithm based on support vector regression, which predicts the measurements of sensor nodes by using historical data. Credit levels of sensor nodes will be determined by a contrast between predictions and actual measured values. Then the dependable data set which is constructed by high credit level measurements will be used to detect sensor faults. Simulations demonstrate that the algorithm works very well in conserving energy and raising failure detection rate.

Key words: wireless sensor network; support vector regression; credibility level; fault detection

无线传感器网络 (WSN, wireless sensor networks)是目前最流行的环境侦测技术,其具有快速部署性,自组织性,无需基础设施支持等优点,在环境监测领域有着广阔的应用前景. 传感器可靠性低、能量有限以及部署环境不可预知等缺点则是无线传感器网络技术需要克服的问题.

由于传感器的低可靠性,传感器故障广泛存在. 其中数据故障危害最大,数据故障均表现为采集的数据不正常波动或者超出合理的取值范围. 出现数据故障的传感器节点的采集数据会发生异常,因此传感器在排除数据故障前只保留路由转发功能. 及时发现数据故障一方面可以通过关闭数据采集功能

收稿日期: 2013-10-12

基金项目: 国家自然科学基金项目 (61272515); 国家自然科学基金创新研究群体科学基金项目 (61121061); 教育部博士点基金项目 (20110005110011); 国家电网公司科技项目 (SGTJXT00XXJS1300048)

作者简介: 孟洛明 (1955—), 男, 教授, 博士生导师; 朱杰辉 (1989—), 男, 硕士生, E-mail: int1425@gmail.com.

节省传感器的能耗,另一方面,错误的传感器数据会大大降低网络的服务质量.笔者提出的故障检测算法主要针对数据故障.

现存的故障检测算法普遍存在效率不高的缺点,这些算法对于计算能力以及能量相对有限的传感器节点来说实用性不高.此外,以往的故障检测算法应用范围局限在小规模,低故障率的传感器网络,实际应用效果十分有限.针对现存故障检测算法的上述不足,提出了基于支持向量机回归预测的故障检测算法.

1 相关工作

总的来说,节点故障管理算法分为两种,集中式的和分布式的.

Ramanathan N 等^[1]介绍了 Sympathy 集中式算法,算法要求所有节点周期性交换邻居列表信息,这将带来较多的能量消耗.

当信息流量不大的时候,集中式故障管理简单而有效.然而,当应用于大规模网络的时候,集中式故障检测算法将带来不可忍受的低效率和高延迟.

Ding M 等^[2]描述的算法是一个分布式本地算法,该算法中节点将自身的测量值与其邻居节点的测量值的中值进行比较,如果差值的绝对值大于阈值,则该节点会被认为是故障的.

Chen J 等^[3]介绍了一种基于分簇思想的故障检测算法,算法选取一些好节点作为基准节点,再把节点跟基准节点进行比较来判断节点是否存在故障.文献[4]中,簇头节点向其成员节点发送询问剩余能量状态,此后簇头节点掌握簇的全局信息并根据全网信息来计算子节点是否处于故障状态.Imieliński T 等^[5]对 Srisathapornphat C 等提出的方法^[4]作出了改进,每个节点只有 p 的概率会向簇头发送状态信息,这样的改进让网络消息数大大减少.

然而上述算法面对大规模故障时效果有限,此外,上述算法的运行速度也不尽理想.笔者针对故障检测算法的运行速度和面对大片故障的能力上提出了新型的分布式故障检测算法.

2 基于支持向量机回归的信誉度划分故障检测

2.1 网络模型及分簇算法

无线传感器被部署为一个分层网络结构.在分层网络结构下,无线传感器网络根据节点的地理位

置分为若干个自组织子网络,被称为簇(Cluster).其中,每个簇中有一个集中式管理节点,该节点被称为簇头(CH, cluster head),簇头掌握着整个簇的组成情况,是将所有簇内传感器节点的数据上报到基站(BS, base station)的传输中枢.除此之外的簇内节点被称为簇成员(CM, cluster member),簇成员是簇内进行数据采集的执行主体.

无线传感器节点可向覆盖范围内的节点发送或接收消息.此外,无线传感器可开启高功率发射模式将发射半径大幅扩大到最大发射半径.假设所有传感器节点具有相同的覆盖半径和最大发射半径,节点有专用的信道往基站发送数据,且节点间通信不会发生丢包现象.

传感器节点的低能储和充电困难的特点很容易成为传感器网络的性能瓶颈,为此,需要一个准确高效的能量模型来描述节点耗能,笔者采用 Heinzelman B 等^[6]提出的能量模型:

$E_T(k, d)$ 表示传感器向距离为 d 的另一个传感器发送 k 比特数据时消耗的能量

$$E_T(k, d) = \begin{cases} k(E_{\text{elec}} + \varepsilon_{\text{FS}} d^2), & d \in (0, d_0] \\ k(E_{\text{elec}} + \varepsilon_{\text{MP}} d^4), & d \in (d_0, \infty) \end{cases} \quad (1)$$

其中: $d_0 = \sqrt{\frac{\varepsilon_{\text{FS}}}{\varepsilon_{\text{MP}}}}$, E_{elec} 表示电子能量, ε_{FS} 表示发送放大器在覆盖范围内发送数据的能耗参数, ε_{MP} 表示在最大发射范围模式下发送放大器的能耗参数.

$E_R(k)$ 表示相应的接收器消耗的能量

$$E_R(k) = kE_{\text{elec}} \quad (2)$$

根据上述的能量模型,基于能量节约以及能量均衡原则提出了下面的分簇算法:

每个节点均向邻居节点广播消息包,包内包含了节点的地理位置和节点规模,节点规模是一个常数,一般认为节点规模越大,该常数越大,向其他节点发送信息的可能性就越大.

节点根据收集的邻居节点的地理位置(x_i, y_i),求出相对距离 d ,计算出把消息发送到所有邻居节点的总能量消耗占自身能量的比例 E_s ,称作承载能力.承载能力 E_s 的计算方法如下

$$E_s = \frac{\sum_i E_T(k, d)}{E_{\text{self}}} = \frac{\sum_i n_i (E_{\text{elec}} + \varepsilon_{\text{FS}} d_i^2)}{E_{\text{self}}} \quad (3)$$

其中: n_i 为收发节点的平均连通度,即自身与所有邻居节点的节点度的均值,用于衡量节点规模.而 E_{self} 为自身的剩余能量.充分交互节点状态后,若自

身的 E_s 比所有邻居的 E_s 值都要大, 则宣布自身成为簇头。

若邻居中没有簇头, 则加入 E_R 最小的邻居所属的簇。否则所有非簇头节点都将选择邻居簇头中接收耗能 E_R 最低的簇加入。其中接收耗能由下面的公式确定

$$E_R(i) = \sum_i n_i \times E_{\text{elec}} \quad (4)$$

根据上述分簇算法, 簇头和簇成员都能保证以最节能的方式组织簇结构, 同时能量高的节点更可能担任高耗能的簇头, 保证了能量节约和均衡。

2.2 基于支持向量机回归的信誉度划分算法

在 WSN 的研究中, 大部分的算法将所有节点作相同的故障检测操作, 这种操作往往是复杂的。但事实上, 对于大部分具有良好故障检测历史的节点来说, 存在故障的概率很低, 因此对这些可靠的节点来说复杂的操作是不必要的, 这种无差别的处理将会带来巨大的能量和计算能力上的浪费。

针对上述问题, 笔者提出基于支持向量机回归预测的故障检测算法 (SFDA, SVR-based fault detection algorithm)。SFDA 包括 5 个步骤:

步骤 1 训练阶段。收集长度为 L 的数据样本集合 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)\}$, 一般 L 为 200 ~ 300, 产生预测序列 $(x_1, x_2, x_3, x_4, \dots, x_L)$, 并训练出支持向量机相关参数。

步骤 2 回归预测。根据支持向量机回归预测算法, 由 T 预测下一时刻的采集数据值 $y_{i+1} = f(x_i)$ 。

步骤 3 信誉度计算。把预测值 y_i 与实际值 $d_m = x_{i+1}$ 作比较, 计算出信誉度 λ_{i+1} 。

步骤 4 信誉等级划分。根据信誉度 λ_{i+1} 划定信誉等级, 并根据各个节点的信誉等级调节上交数据到簇头进行簇头检测的频率。

步骤 5 基于置信数据集的故障检测机制。根据基于置信数据集的故障检测机制, 将故障节点信息上传至基站。同时, 根据簇头检测的结果反馈性地调整各个簇成员节点的信誉等级。

在信誉划分的过程中, 支持向量回归 (SVR, support vector regression) 预测算法是本算法的核心, 预测结果的准确与否关乎信誉划分的权威性。

Vapnik 等人根据统计学习理论 (SLT, statistical learning theory) 提出了支持向量机 (SVM, support vector machine) 学习方法。SVM 方法基于结构风险最小化原理, 在传感器数据预测上明显优于基于经验风险最小化原则的人工神经网络法。经验风险最

小仅仅能让预测值相对于训练数据的误差最小化, 而对于数据取值范围广泛, 随机性大的 WSN 数据预测来说是远远不够的。而结构风险最小则可以保证 VC 维数 (用于衡量泛化误差) 的上限最小化, 这使得 SVM 具有更好的泛化能力, 还能有效地避免“过学习”的问题。

SVM 学习方法具有无可比拟的泛化性, 非常适合应用于 WSN, 而 SVR 算法是 SVM 最重要的应用, SVR 是一种基于数据的机器学习方法。SVR 算法借助各个传感器节点前期的有效数据构造数据样本, 然后找到数据内在的相互依赖关系, 用于预测未来的数据, 可以解决区域节点大规模失效的情况下的数据预测。

下面介绍对传感器数据进行预测的 SVR 算法:

假设归一化后的传感器数据与相应预测值组成样本集合 (X, Y) , 该集合上存在某个概率分布 $F(x, y)$, 选取独立同分布的样本点进行处理。然而, 由于预测的过程不可能完全准确, 在 SVM 理论中, 往往用损失函数来衡量结构风险。假设损失函数为 $\text{Lost}(y, f(x))$, 于是期望结构风险可以由下式计算

$$R[f] = \int \text{Lost}(y, f(x)) dF(x, y) \quad (5)$$

根据结构风险最小化 (SRM, structure risk minimization) 原理, 为了达到最好的预测效果, 需要找出最优的预测函数 $f(x)$, 使得 $R[f]$ 尽可能的小。多种损失函数都广泛应用于研究之中, 其中高斯函数能够有效地克服序列内的高斯噪声, 同时具有良好的稀疏特性和泛化性, 更能排除高斯噪声干扰, 达到更好的预测效果。高斯损失函数的表达式如下

$$\text{Lost}(y, f(x)) = (y - f(x))^2 \quad (6)$$

在统计学习理论中, 结构风险定义为经验风险与置信区间之和。然而 WSN 中传感器数据随机性大、变化区间广, 在误差较小的区间内, 普遍认为这些传感器节点数据对经验风险与置信区间的影响微乎其微, 因此其所带来的结构风险往往可以忽略。在此区间内进行结构风险的衡量反而很容易因为数据的不规则波动造成误判, 进而错误估计了系统的结构风险。针对上述问题, 提出了针对 WSN 的 ε -不敏感高斯函数

$$\text{Lost}(y, f(x)) = \begin{cases} (y - f(x))^2 - \varepsilon^2, & |y - f(x)| \geq \varepsilon \\ 0, & |y - f(x)| < \varepsilon \end{cases} \quad (7)$$

其中: ε 是一个常量, 表示无线传感器网络中能容

忍的最大误差,在 $(-\varepsilon, \varepsilon)$ 内的误差均认为是不会带来结构风险的. ε -不敏感高斯函数可以在保持原有的抗高斯噪声的优点的同时弥补了原高斯函数在处理 WSN 数据时容易发生的“过敏感”问题.

损失函数所表示的结构风险在 SVM 分类问题中的数学含义是样本集中的任意点离超平面的距离. 多维空间中的超平面可表示为

$$f(x) = \langle \omega, x \rangle + b \quad (8)$$

根据结构风险最小化原理 (SRM, structure risk minimization), 求损失函数的最小值等价于调整超平面的斜率 ω 使得超平面离样本集里最靠近的点的距离尽可能大, 即使得 $\frac{\varepsilon}{\sqrt{1 + \|\omega\|^2}}$ 最大. 此时的超平面的划分具有最高的置信度, 称为最优超平面.

由于 ε 为常数, 因此求 $\frac{\varepsilon}{\sqrt{1 + \|\omega\|^2}}$ 的最大值等价于求 $\frac{1}{2} \|\omega\|^2$ 的最小值. 结合式 (7) ε -不敏感高斯函数, 达到结构风险最小化的超平面等同于下列规划问题的解

$$\min \quad \frac{1}{2} \|\omega\|^2 \quad (9)$$

$$\text{s. t.} \quad |(\langle \omega, x_i \rangle + b) - y_i| \leq \varepsilon, \quad i = 1, 2, \dots, L \quad (10)$$

条件中 L 表示样本容量. 这是一个凸二次规划问题, 引入松弛变量 ξ 和 ξ^* , 上述问题可以转化为下列规划问题

$$\min \quad Q(\omega) = \frac{1}{2} \|\omega\|^2 + \frac{C}{L} \sum_{i=1}^L (\xi_i^2 + \xi_i^{*2}) \quad (11)$$

$$\text{s. t.} \quad \begin{cases} (\omega \cdot \varphi(x_i) + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (\omega \cdot \varphi(x_i) + b) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, L \end{cases} \quad (12)$$

其中: C 为惩罚常数, 反映了 SVR 算法中函数回归模型的复杂度和样本拟合精度之间的折中, C 越大, 拟合程度越高; ω 为超平面的法向量, ξ_i 和 ξ_i^* 为随机误差, b 为偏差. 约束条件中 ε 除了表示允许的最大误差外, 还起着控制回归的支持向量个数和泛化能力的作用, ε 越大, 支持向量数越少.

为求解上述规划问题, 引入 Lagrange 乘子 $\alpha = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_L)$ 并将上面的问题转化为对偶规划问题. 定义 Lagrange 函数, 并代入后化简可得

$$L(\omega, b) = \frac{1}{2} \|\omega\|^2 + \frac{C}{L} \sum_{i=1}^L (\xi_i^2 + \xi_i^{*2}) + \sum_{i=1}^L (\alpha_i - \alpha_i^*) (\omega \varphi(x_i) + b - y_i) - \varepsilon \sum_{i=1}^L (\alpha_i + \alpha_i^*) \quad (13)$$

根据 KKT (Karush-Kuhn-Tucker) 最优化条件:

1) Lagrange 函数的偏导数等于零; 2) 等式约束等于零; 3) 不等式约束与相应 Lagrange 乘子的乘积等于零. 根据 KKT 条件第 1) 条, $L(\omega, b)$ 分别对 ω, b 取偏导并令其等于 0

$$\left. \begin{aligned} \frac{\partial L(\omega, b)}{\partial \omega} &= \omega + \sum_{i=1}^L (\alpha_i - \alpha_i^*) \varphi(x_i) = 0 \\ \frac{\partial L(\omega, b)}{\partial b} &= \sum_{i=1}^L (\alpha_i - \alpha_i^*) = 0 \end{aligned} \right\} \quad (14)$$

此外, 根据 KKT 条件第 2) 条又能得到

$$\left. \begin{aligned} \alpha_i g(x_i) &= 0 \\ \alpha_i^* h(x_i) &= 0 \end{aligned} \right\} \quad (15)$$

式 (15) 的含义在于, 只有当 α_i, α_i^* 取 0 时, 约束条件才起约束作用. 实际上, 只有若干 α_i, α_i^* 不为 0, 此时的约束条件 $g(x_i), h(x_i)$ 等于 0, 也就是说回归样本点落在了约束边界上, 这样的约束向量被称为支持向量 (SV, support vector).

把式 (14) 代入式 (13), 可以得到

$$L(\omega, b) = -\frac{1}{2} \sum_{i,j=1}^L (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) \varphi(x_i, x_j) - \varepsilon \sum_{i=1}^L (\alpha_i^* + \alpha_i) + \sum_{i=1}^L (\alpha_i^* - \alpha_i) \quad (16)$$

但是式 (16) 中的 $\varphi \langle x_i, x_j \rangle = \varphi(x_i) \cdot \varphi(x_j)$ 是一个高维度特征空间上的映射之间的内积, 具有很大的处理难度. 在 SVM 理论中, 解决这个内积计算的办法是核技巧. 即寻找一个核函数 $K(x_i, x_j)$, 使得 $K(x_i, x_j) = \langle x_i, x_j \rangle = \varphi(x_i) \cdot \varphi(x_j)$, 通过核函数替代巧妙地避免了计算高维内积所需的计算代价. 由于 SVR 算法虽然把结构风险最小化问题转化为求高维度超平面的问题, 旨在把常规维度上难以处理的非线性不可分问题转化为高维度线性可分问题, 然而算法的最后会把高维度向量 ω 约去, 因此问题最后又回到常规维度上, 换言之核技巧处理是合理的. 核技巧是整个 SVM 理论的核心. 用核函数 $K(x_i, x_j)$ 替代高维度内积, 最终转化为一个严格凸函数的对偶二次规划问题

$$\max L(\omega, b) = -\frac{1}{2} \sum_{i,j=1}^L (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j)$$

$$K(x_i, x_j) - \varepsilon \sum_{i=1}^L (\alpha_i^* + \alpha_i) + \sum_{i=1}^L y_i (\alpha_i^* - \alpha_i) \tag{17}$$

$$\text{s. t.} \begin{cases} \sum_{i=1}^L (\alpha_i - \alpha_j^*) = 0 \\ 0 \leq \alpha_i, \alpha_j \leq \frac{C}{L} \quad i, j = 1, 2, \dots, L \end{cases} \tag{18}$$

由于这个二次规划并没有关于 x 的不等式约束,因此可以简单地用线性方程求解,解上述凸二次规划问题可以求得 α_i, α_i^* .

选取研究成果最为成熟的高斯径向基核函数核函数 $K(x_i, x_j)$ 以及式 (14) 代入超平面表达式 (8) 中,则得到最优超平面,即结构风险最小的 WSN 传感器数据预测函数

$$f(x) = \sum_{i=1}^L (\alpha_i - \alpha_i^*) \exp\left(\frac{-\|x_i - x\|^2}{2\sigma^2}\right) + b \tag{19}$$

根据 SVR 预测函数的预测结果与实际值的偏差可以定义信誉系数. 然而仅由一次预测的数据来判定一个节点的信誉容易发生“长期误信”,即当一个低信誉度节点被误判为高信誉度节点后,在一个长时间内都不会得到修正的现象. 考虑到的历史数据和当前的预测结果符合逻辑回归模型,因此可结合 Logistic 回归模型来评价节点的信誉度, Logistic 回归模型的核心是 sigmoid 曲线 $P_n = \frac{1}{1 + e^{-g(n)}}$, 引入历史信誉系数的线性加权作为 P_n 的输入,最后得到动态的信誉系数的表达式如下

$$\lambda_i = \left(\frac{2}{1 + e^{-\sum_{j=1}^{i-1} \beta_j \lambda_j}} - 1 \right) \times \frac{|f(x_i) - d_i|}{d_i} \tag{20}$$

根据可以把节点分成 5 个信誉度等级,分别为确信、良好、一般、可疑、故障. 簇成员节点根据自身的信誉度等级,将选择相应的对簇头的故障数据上报频率,而处于最低信誉级别的节点将被认为是故障节点. 同时,簇头也会根据信誉度的不同来对簇成员进行不同的处理,信誉度等级划分标准及处理详情如表 1 所示.

需要指出的是,簇头节点有权限根据基于置信数据集的故障检测算法结果发送指令更改簇成员的信誉等级.

表 1 节点信誉等级及相应处理方式

信誉等级	评定标准	CM
确信 (sure)	信誉系数 5% 以内	仅进行本地检测,自身数据发送给 CH 作为故障检测的置信数据
良好 (good)	信誉系数 5% ~ 10%	以低频率向 CH 发送数据副本,若无故障则数据作为置信数据
一般 (common)	信誉系数 10% ~ 20%	以中频率向 CH 发送数据副本检测故障
可疑 (doubt)	信誉系数 20% ~ 40%	以高频率向 CH 发送数据副本检测故障
故障 (fault)	信誉系数大于 40%	向 CH 发送故障报告

2.3 基于置信数据集的故障检测算法

分簇算法把地理位置接近的节点分为一个簇,这个簇的覆盖范围有其实际意义. 因为在簇的覆盖范围内,可以认为传感器检测的物理量差别较小,在训练阶段可以得到簇头数据差别量的阈值 U_{thres} .

根据信誉级别的不同,节点会以不同的频率向簇头发送数据,而簇头根据这些数据与置信数据集 (DDS, dependable data set) 内的数据进行比较来判定节点是否出现故障,这个机制称为基于置信数据集的故障检测机制. 置信数据集由 3 个数据集通过操作得到:簇头数据以及簇成员节点中具有“确信”和“良好”信誉度的成员节点上报的数据,如下式表示

$$S_{DDS} = S_{CH} \cup S_{sure} \cup S_{good} \tag{21}$$

其中 S_{CH} 为簇头数据集 S_{sure} 和 S_{good} 分别为“确信”和“良好”簇成员节点上报的数据集.

若检测数据与 S_{DDS} 中数据的均值之差大于 U_{thres} 则判定为故障,若大于 $0.7 U_{thres}$ 则信誉等级下降 1 级,若差值小于 $0.3 U_{thres}$ 则信誉度上升 1 级. 信誉度的调整会通过信誉度调整消息传达到相应的簇成员节点. 基于置信数据集的故障检测机制运作流程图如图 1 所示.

3 仿真结果与分析

首先根据 3.1 节中的算法对节点进行分簇,然后再进行故障检测. 参照 Heinzelman B^[6] 中能量模型,仿真参数取值如表 2 所示.

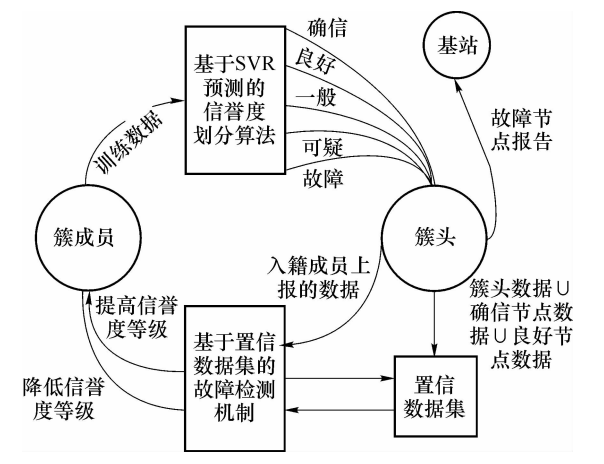


图 1 基于置信数据集的故障检测机制

表 2 仿真参数说明

参数名称	参数描述	参数值
length	仿真区域边长	500
node_quantity	传感器节点数	100, 150, 200, 250
d_0	节点覆盖半径	50
E_{elec}	电子能量	50 nJ/bit
ε_{FS}	覆盖半径内发送放大器的能耗参数	10 pJ/(bit · m ²)
ε_{MP}	最大发射范围模式下发送放大器的能耗参数	0.004 pJ/(bit · m ⁴)

3.1 基于 SVR 预测的信誉度划分算法仿真

本节验证 SVR 算法的预测效果. SVR 模型的回归预测效果由如下 3 个指标来评价:精度 A , 均方误差 MSE 和相关系数 ρ .

首先通过训练提高 SVR 模型的回归预测效果, 获取模型最优参数. 对训练数据进行归一化处理, 可以得到取值范围在 $[-1, +1]$ 之间的训练序列 T . 对 SVR 模型进行训练, 得到最优参数: ε - 不敏感高斯函数的最大允许误差 $\varepsilon = 512.0$, 高斯核函数中 $\sigma = 8.0$.

实际测量值序列如图 2 所示. 图中黑色实心圆形节点表示无故障且测量值随机误差极小的节点, 空心圆形节点表示无故障且测量值随机误差在允许范围内的节点, 红色实心方形节点表示故障节点. 若采用经典线性回归算法——最小二乘法进行预测, 并把预测值代入式(20)计算节点的信誉度, 可以得到回归划分的效果图如图 3(a)所示. 其中绿色区域表示最小二乘法划分的高信誉度区域, 蓝色区域表示中信誉度区域, 红色区域表示低信誉度区域. 图中

可以看出, 信誉度的划分出现了大量的误划分点. 经统计, 划分的正确节点数为 201/248, 正确率约为 81%. 若用 SVR 回归预测算法, 根据训练参数 $\varepsilon = 512.0$ 以及 $\sigma = 8.0$ 代入第 3 节中式(17)~式(21)对传感器节点的采集数据进行预测, 并与实际测量值进行比较计算出信誉度系数. 根据进行信誉度等级的划分, 可得到信誉等级划分如图 3(b)所示. 显而易见, 图 3(b)中误划分点显著减少. 经统计, 基于 SVR 预测的信誉度划分算法的正确划分点数为 240/248, 信誉度划分正确率高达 96.8%. 上述仿真说明, 基于 SVR 预测的信誉度划分算法具有优异的性能.

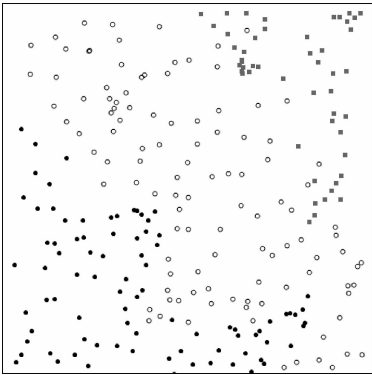
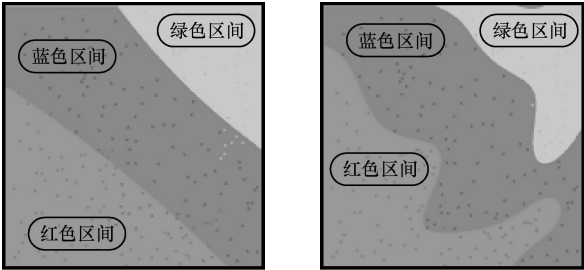


图 2 实际预测序列



(a) 最小二乘法预测效果 (b) SVR 预测效果

图 3 预测效果对比

3.2 故障检测算法仿真

为验证算法的正确性和高效性, 采用文献[2]和文献[3]中算法作对比. 文献[2]提出了 Localized Event Boundary Detection Algorithm, 这是一个本地算法, 根据节点的本地邻居数据判定节点是否存在故障; 而文献[3]中的 Tendency-based Localized Faulty Sensor Detection Algorithm 是一个基于分簇思想的算法, 簇头节点会在簇成员中选择一个可信的标准节点, 并把该标准节点的数据作为检测簇内节点故障的依据. 仿真分别测试 SFDA 以及上述两个

对比算法在不同网络规模(100、150、200、250 个节点)下的故障检测率,每个网络规模重复仿真 15 次。SFDA 的仿真结果如图 4 所示。

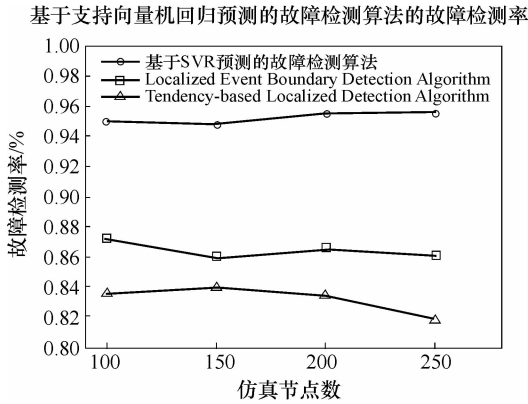


图 4 故障检测率比较

由图 4 可得, SFDA 算法在故障检测率上远远优于对比算法. 在不同的网络规模下故障检测率均保持在 94% 以上, 高于算法 [2] 的约 87% 和算法 [3] 的约 84%. 此外, 文献 [2] 和文献 [3] 中的算法随着节点规模的增大, 在故障检测率上呈现着下降的趋势. 而 SFDA 算法的故障检测率随着节点规模的增大仍能保持稳定. 由上述分析可得, SFDA 算法通过高精度的 SVR 预测算法和可靠的置信数据集故障检测, 能够达到可观的故障检测率, 且算法效果不会随着网络规模的上升而受到影响.

4 结束语

根据 WSN 的特点将支持向量机回归预测理论以及信誉度划分思想应用在 WSN 的故障检测中, 解决了现有大部分故障检测算法效率过低以及无法应对大规模传感器网络的不足. 仿真结果说明, 基于支持向量机回归预测的信誉度划分算法能对传感器数据进行准确预测并正确划分信誉等级, 其故障检测率在不同的网络规模下均能高达 94% 以上.

参考文献：

[1] Ramanathan N, Kevin Chang, Kapur R, et al. Sympathy for the sensor network debugger[C]//Proceedings of the 3rd international Conference on Embedded Networked Sensor Systems. ACM, 2005: 255-267.

[2] Ding Min, Chen Dechang, Xing Kai, et al. Localized fault-tolerant event boundary detection in sensor networks [C]//INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies. [S. l.]: IEEE, 2005: 902-913.

[3] Chen Jinran, Kher S, Somani A. Distributed fault detection of wireless sensor networks[C]//Proceedings of the 2006 Workshop on Dependability Issues in Wireless Ad hoc Networks and Sensor networks. ACM, 2006: 65-72.

[4] Srisathapornphat C, Shen Jaikaeo C. Sensor information net-working architecture[C]// 2000 International Workshops on Parallel Processing. Toronto: [s. n.], Canada), 2000: 23-30

[5] Imieliński T, Goel S. DataSpace—querying and monitoring deeply networked collections in physical space[C]// Proceedings of the 1st ACM international workshop on Data Engineering for Wireless and Mobile Access. ACM, 1999: 44-51.

[6] Heinzelman B, Chandrakasan P, Balakrishnan H, An application-specific protocol architecture for wireless microsensornetworks [C]//WirelessCommunications, 2002: 660-670.