

文章编号:1673-0062(2012)04-0014-04

基于 SVM 的核电站环境辐射监测网络中 传感器节点缺失值估计算法

王龙辉¹, 高 嵩², 屈 星²

(1. 南华大学 经济管理学院, 湖南 衡阳 421001; 2. 南华大学 电气工程学院, 湖南 衡阳 421001)

摘 要: 传感器节点监测数据缺失会影响核电站外围环境辐射监测的有效性, 需要对缺失数据进行准确估计. 提出一种基于支持向量机的监测数据缺失值估计算法, 对传感器节点缺失监测数据进行估计. 用实际监测数据对算法进行了验证, 用均方误差和相关系数评价实验结果. 与现有的基于神经网络的估计算法进行了性能比较. 实验结果表明, 本文所提出的算法具有较高的估计精度.

关键词: 核电站; 环境辐射监测; 缺失值; 估计算法; 支持向量机

中图分类号: TP183 **文献标识码:** B

SVM Based Missing Data Imputation Algorithm in Nuclear Power Plant's Environmental Radiation Monitor Sensor Network

WANG Long-hui¹, GAO Song², QU Xing²

(1. School of Economics and Management, University of South China, Hengyang, Hunan 421001, China;

2. School of Electric Engineering, University of South China, Hengyang, Hunan 421001, China)

Abstract: Monitor data mission in sensor node will influence the validity of environmental radiation monitoring around a nuclear power plant. To solve the problem, a missing data imputation algorithm based on support vector machine(SVM) is proposed to impute the missing data. This Algorithm is validated with real radiation monitoring data, evaluated by using MSE and correlation coefficient and compared with ANN based imputation algorithm. The experimental results demonstrate that the proposed algorithm can achieve higher imputation accuracy.

key words: nuclear power plant; environmental radiation monitor; missing data; imputation algorithm; SVM

核电在保障我国能源方面发挥着重要作用, 核 安全问题也随之日益受到人们的重视. 核电在运行

收稿日期: 2012-11-07

基金项目: 湖南省科技厅科研基金资助项目(2012FJ4332)

作者简介: 王龙辉(1962-), 男, 湖南长沙人, 南华大学经济管理学院讲师. 主要研究方向: 数据处理.

过程中,会向外围环境释放放射性物质,对核电站外围环境的 γ 辐射水平进行连续监测,是保障核安全的重要措施.我国在核电站周边区域都建立了区域 γ 辐射监测网络,采集核电站外围 γ 剂量率数据,监测核电站向外围辐射的放射性物质所产生的 γ 辐射水平及其变化趋势,从而确保核电站的安全运行,防止发生核污染事故.无线传感器网络(WSN)能够在无人值守的状态下及恶劣的环境中连续工作,非常适合用于核电站外围环境辐射监测,给核电站外围环境辐射连续监测系统的建设提供了新的技术途径^[1-3].但是传感器节点采集的监测数据缺失可能由于各种因素的影响而丢失,从而造成环境辐射连续监测数据的中断,破坏了监测数据的完整性,影响对核电站外围环境 γ 辐射水平评估的准确性与可靠性,甚至可能导致错误的评估结果,影响对核电站外围环境辐射水平的有效监控,因此对传感器节点缺失数据进行估计是十分重要的.

传感器监测数据缺失不可避免,缺失数据估计问题已成为无线传感器网络数据处理研究的新热点.数年来,国际上对该问题的研究重点是应用数据挖掘方法,发现现有数据间的知识,利用这些知识估计缺失值^[4].M. Halatchev 提出了估计传感器缺失数据的 WARM 算法^[5],L. Gruenwald 对 WARM 算法进行了改进,提出了 FARM 算法^[6].这些算法利用数据间的时间关联规则或空间关联规则估计缺失数据.

虽然国内外对无线传感器网络中传感器节点数据缺失值的估计问题的研究还处在起步阶段,但是我国学者对传感器感知数据缺失值估计的研究,也做出了一定的贡献.文献[7]基于传感器感知数据的空间相关性,提出了基于 K-最近邻缺失数据估计算法,文献[8]提出了一种基于感知数据时空相关性的缺失值估计算法,但须对缺失数据进行两次估计.

核电站外围环境 γ 辐射剂量率监测数据虽然在时间上具有某种规律性,但空间上具有某种相关性,但是这种关系是复杂的,非线性的.很难采用传统的模型来描述.支持向量机(SVM)具有很强的逼近非线性函数的能力,支持向量机用于回归,能较好的解决数据估计的非线性问题^[9].本文提出一种基于 SVM 的核电站环境辐射监测网络中传感器节点监测数据缺失值估计算法,以提高数据估计的精度.

1 SVM 估计模型

对给定的 N 个训练样本:

$\{(x_i, y_i) : i = 1, 2, \dots, N\}$, 其中 $x_i \in R^n$ 是输入,

$y_i \in R$ 是输出. SVM 用下式函数对缺失数据进行估计:

$$y(x) = (W \cdot X) + b \quad (1)$$

由支持向量机理论可知^[10],式(1)可通过解下面的最优化问题求解:

$$\min_{a,b} \frac{1}{2} \|\omega\|^2 \quad (2)$$

约束条件为:

$$s. t. \left. \begin{aligned} (\omega \cdot x_i) + b - y_i &\leq \varepsilon \\ y_i - ((\omega \cdot x_i) + b) &\leq \varepsilon \end{aligned} \right\} \quad (3)$$

式中, ε 是允许估计误差,引入 Lagrange 函数:

$$L(\omega, b, \alpha^{(*)}) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^m \alpha_i (\varepsilon + y_i - (\omega \cdot x_i) - b) - \sum_{i=1}^m \alpha_i^* (\varepsilon - y_i + (\omega \cdot x_i) + b) \quad (4)$$

其中 $\alpha = (\alpha_1, \dots, \alpha_m)^T$, $\alpha^* = (\alpha_1^*, \dots, \alpha_m^*)^T$ 为 Lagrange 乘子向量且 $\alpha_m, \alpha_m^* > 0$.考虑式(2)和式(3),有:

$$\min_{\alpha^{(*)} \in R^{2m}} \frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i \cdot x_j) + \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) - \sum_{i=1}^m y_i (\alpha_i^* - \alpha_i) \quad (5)$$

$$s. t. \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \quad \alpha_i^{(*)} \geq 0, i = 1, \dots, m \quad (6)$$

可解得:

$$f(x) = y = \sum_{i=1}^m (\alpha_i^* - \alpha_i) (x_i \cdot x) + b \quad (7)$$

对于非线性问题,引入核函数 $K(x_i, x_j)$ 代替样本向量的内积运算,将数据映射到高维特征空间,最后到缺失数据估计函数:

$$y = \sum_{i=1}^m (\alpha_i^* - \alpha_i) K(x_i \cdot x) + b \quad (8)$$

核函数 $K(x_i, x_j)$ 须满足 mercer 条件,本文核函数取径向基函数:

$$k(x_i)(x_1) = \exp(-\|x - x_i\|^2 / 2\sigma^2) \quad (9)$$

2 实验仿真

2.1 样本预处理

用某核电站环境 γ 辐射监测网络的实际数据对 SVM 缺失数据估计模型进行估计实验.核电站外围环境 γ 辐射水平监测,要考虑正常运行监测和事故应急监测两种情况, γ 辐射探测器的探测量程为 $0.01 \mu\text{Gy/h} \sim 10 \text{Gy/h}$.核电站正常运行时的外围环境 γ 剂量率是远远小于量程上限的,我国监测到的核电站外围环境连 γ 辐射剂量率年均值在

100 nGy/h 至 130 nGy/h 之间^[11]. 采用式(10)对样本数据进行归一化处理,将样本数据均转化为 $[0,1]$ 之间的数.

$$x'_k = \frac{x_k - x_{\min}}{x_{\max} - x_{\min}} \quad (10)$$

其中, x_k 为 γ 辐射剂量率实测值; x'_k 为归一化处理后的数值; $x_{\min} = 0.01 \mu\text{Gy/h}$,考虑瞬时值的波动,取 $x_{\max} = 0.25 \mu\text{Gy/h}$.

2.2 实验结果

选取连续300组数据作样本,其中200组数据作为训练样本,另外100组数据作为测试样本,采用均方误差MSE和相关系数R来对算法的性能进行评估.图1中a为SVM模型对传感器节点

剂量率缺失数据估计值与实测值的比较,从图中可以看出估计值非常接近实测值,图中b是估计误差.作为比较,图2中a给出了用神经BP神经网络估计算法对同样本数据的估计结果.图2中b是估计误差.

由仿真结果得到SVM模型的缺失数据估计的最大相对误差为7.9%,均方误差 $MSE = 0.000675$,相关系数 $R = 0.821735$.基于BP神经网络模型的最大相对估计误差为10.1%,均方误差 $MSE = 0.000965$,相关系数 $R = 0.796375$.SVM模型估计精度比神经网络测量模型提高了2.2%.

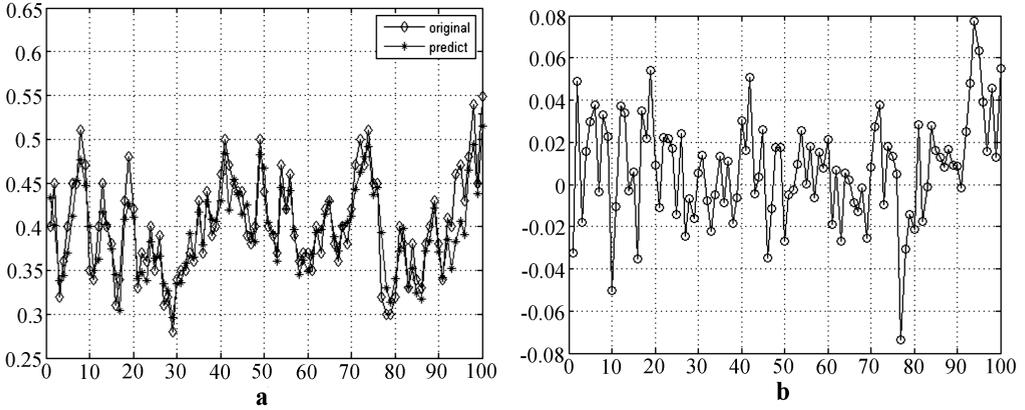


图1 SVM算法缺失值估计结果

Fig. 1 Imputation results with SVM imputation algorithm

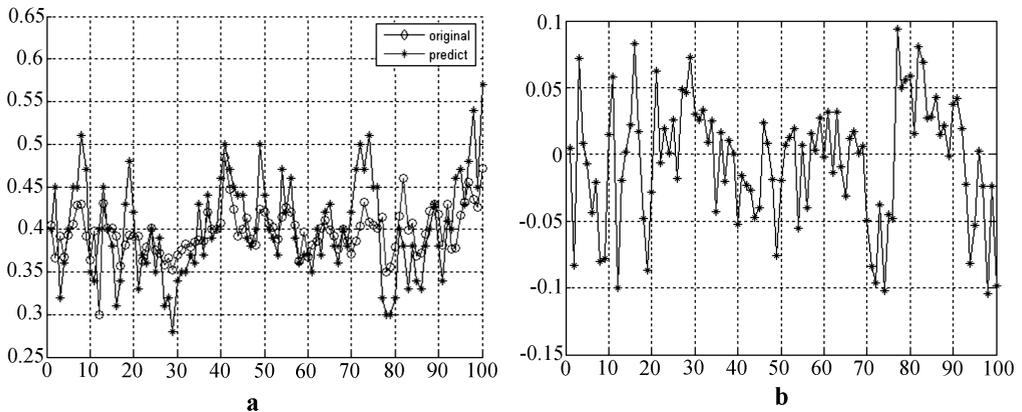


图2 BP神经网络算法缺失值估计结果

Fig. 2 Imputation results with SVM imputation algorithm

3 结 论

本文研究了一种核电厂外围环境 γ 辐射无线传感器监测网络传感器节点缺失数据估计算法.建立了基于SVM的节点缺失数据估计模型,通过实验对算法进行了验证,与现有的基于神经网络的估计算法比较,本文提出的对传感器节点缺失数据估计精度更高,为估计核电厂外围环境 γ 辐射水平无线传感器监测网络中传感器节点缺失数据提供了一种实用的方法.

参考文献:

- [1] Brennan S, Mielke A, Torney D, et al. Radiation detection with distributed sensor networks[J]. Computer, 2004, 37(8):57-59.
- [2] Barbaran J, Diaz M, Esteve I, et al. A mobile framework for radiation monitoring in nuclear power plants[J]. International Journal of Electronics, Circuits and Systems, 2007, 1(2):104-109.
- [3] Vujo Drndarevic, Miodrag Bolic. Gamma radiation monitoring with internet-based sensor network[J]. Instrumentation Science & Technology, 2008, 36(2):121-133.
- [4] Chok H. Spatio-temporal association rule mining framework for real-time sensor network applications [C]// Proc. of CIKM'09. New York, USA: ACM Press, 2009: 1761-1764.
- [5] Halatchev M, Gruenwald L. Estimating missing values in related sensor data streams[J]. Lecture Notes in Computer Science, 2007, 4443:981-987.
- [6] Gruenwald L. Using data mining to estimate missing sensor data[C]//Proc. of ICDMW'07. Washington, C. USA: IEEE Computer Society, 2007:207-212.
- [7] Liqiang Pan, Jianzhong Li. K-nearest neighbor based missing data estimation algorithm in wireless sensor networks [J]. Wireless Sensor Network, 2010, 2(2):115-122.
- [8] 潘立强, 李建中, 骆吉洲. 传感器网络中一种基于时空相关性的缺失值估计算法[J]. 计算机学报, 2010, 33(1):1-10.
- [9] Vapnik V N. Statistical learning theory[M]. New York: John Wiley, 1998.