

基于滑动窗口预测的水文时间序列异常检测

余宇峰^{1*}, 朱跃龙¹, 万定生¹, 关兴中²

(1. 河海大学 计算机与信息学院, 南京 210098; 2. 江西省水文局, 南昌 330018)

(*通信作者电子邮箱 yfyu@hhu.edu.cn)

摘要:针对水文时间序列分析与决策中存在的数质量数据问题,提出了基于滑动窗口预测的水文时间序列异常检测算法。首先基于滑动窗口对时间序列进行子序列分割,再以子序列为基础建立预测模型对未来值进行预测,并将预测值和实测值间差异范围大于预设阈值的序列点判定为异常。探讨了算法中的滑动窗口和参数设置,并以实例数据对算法进行了验证。实验结果表明,所提算法不仅能够有效挖掘出水文时间序列中的异常点,而且将异常检测的灵敏度和特异度分别提高到80%和98%以上。

关键词:时间序列;异常检测;预测模型;置信区间;滑动窗口

中图分类号: TP393.08; TP18 **文献标志码:** A

Time series outlier detection based on sliding window prediction

YU Yufeng^{1*}, ZHU Yue-long¹, WAN Dingsheng¹, GUAN Xingzhong²

(1. College of Computer and Information, Hohai University, Nanjing Jiangsu 210098, China;

2. Hydrology Bureau of Jiangxi Province, Nanchang Jiangxi 330018, China)

Abstract: To solve data quality problems for hydrological time series analysis and decision-making, a new prediction-based outlier detection algorithm was proposed. The method first split given hydrological time series into subsequences so as to build a forecasting model to predict future values, and then outliers were assumed to take place if the difference between predicted and observed values was above a certain threshold. The setup of sliding window and parameters in the detection algorithm were analyzed, and the corresponding result was validated with the real data. The experimental results show that the proposed algorithm can effectively detect the outliers in time series and improves the sensitivity and specificity to at least 80 percent and 98 percent respectively.

Key words: time series; outlier detection; forecasting model; confidence interval; sliding window

0 引言

时间序列挖掘作为数据挖掘领域的十大挑战性研究问题之一^[1],已被广泛应用于水文时间序列相似性搜索、序列模式挖掘和周期分析等领域。大部分水文时间序列挖掘的目的是为了发现频繁出现的模式,期望发现某种规律,异常数据通常被作为噪声而忽略。水文时间序列中的异常是与水文现象的数据模型或一般规律显著不同或不一致的数据对象,可能潜藏着更有价值和意义的水文信息和知识。因此,发现水文时间序列中的异常点并减少其对数据分析的影响是一项很有意义的工作。

本文采用基于滑动窗口的时间序列预测方法挖掘水文时间序列中的异常,算法通过基于滑动窗口方法预测未来数据值,并通过设定预测值和实测之间的以判定异常。实验结果表明,该方法能有效检测出水文时间序列中存在的异常,从而达到提高水文时间序列分析结果的目的。

1 相关研究

异常检测(Outlier Detection)也称为异常挖掘或异常检测,是从大量数据中提取隐含在其中的人们事先不知道的但又是潜在有用的信息和知识的过程^[2]。异常检测可以形式

化地描述如下:给定一个包含 n 个数据点或对象的集合及预期的异常点数目 k ,发现与剩余的数据相比是显著异常的、孤立的或不一致的前 k 个对象的过程。异常检测需要解决两个主要问题:在给定的数据集中定义什么样的数据是异常的;找到一个有效的方法来检测这样的异常数据。

按异常在时间序列中的不同表现形式,时间序列异常可以分为3种:序列异常、点异常和模式异常。目前,时间序列异常检测方法主要包括如下。

1) 基于窗口的方法^[3-4]。将时间序列划分成若干个固定大小的子序列(窗口),在各子序列中定位异常点,该方法的基础是时间序列中的异常点可能是其一个或多个子序列中的异常点导致。

2) 基于距离的方法^[5-6]。该方法通过特征点表征序列,利用二次回归模型实现对序列的不等长划分,以动态时间弯曲(Dynamic Time Warping, DTW)距离为基础计算子序列的异常分数,选取异常分数最大的前 k 个值来确定异常。

3) 基于密度的方法^[7-8]。该方法不是简单用Yes或No来断定一个点是否是离群点,而是用一个权值来评估它的离群度。它是局部的,意思是该程度依赖于对象相对于其邻域的孤立情况。这种方法可以同时检测出全局离群点和局部离群点。

收稿日期:2014-04-07;修回日期:2014-05-07。 基金项目:国家自然科学基金资助项目(51079040);水利部948项目(201016)。

作者简介:余宇峰(1979-),男,湖北黄冈人,高级实验师,博士研究生,主要研究方向:数据挖掘、数据质量、水利信息化;朱跃龙(1959-),男,江苏建湖人,教授,博士,主要研究方向:智能信息处理、数据挖掘;万定生(1963-),男,江苏溧阳人,教授,主要研究方向:水利信息化、数据挖掘。

4) 基于支持向量机 (Support Vector Machine, SVM) [3,9-10] 的方法。该方法采用支持向量回归 (Support Vector Regression, SVR) 技术, 对历史时间序列建立回归模型, 判断新到来的序列点与模型的匹配程度。此外, 一类支持向量机 (One-Class SVM) 技术 [3,10] 也在异常检测领域取得广泛应用。

5) 基于聚类的方法 [11]。该方法先将数据集分成若干簇, 不属于任何簇的数据点就是异常点。在异常检测领域, 聚类技术被用于无监督检测和半监督检测。但异常检测通常是聚类算法的副产品, 因此对异常点的挖掘效率较低。

综上所述, 基于窗口的方法性能依赖于窗口宽度大小, 窗口宽度过大或过小都可能影响检测结果的精度; 基于距离的方法时间复杂度较高, 效率不能保证; 基于聚类的方法依赖于所有簇的个数和数据中离群点的存在性。文献 [12-13] 认为基于预测的时间序列异常检测是最简洁直观的异常检测方法, 但该方法依赖于预测模型的预测能力, 并且难以确定合理的阈值。

本文借鉴基于窗口方法中子序列分割的思想对时间序列进行子序列分割, 再以子序列为基础建立预测模型对未来值进行预测, 并根据分割的窗口大小及用户期望的置信度动态计算阈值, 不仅能够保证准确全面挖掘出水文时间序列中的异常点, 而且提高异常检测精度和可靠性。

2 基于滑动窗口预测的时间序列异常检测

2.1 相关定义

水文现象是时变现象, 这一变化过程称为水文过程。水文数据是对水文过程的离散记录, 水文数据按其描述的物理量分为流量、水位、降雨量和蒸发量等水文时间序列。

定义 1 水文时间序列 (Hydrological Time Series)。它是由记录值和记录时间组成的元素的有序集合, 记为 $D^n = \langle d_1 = (v_1, t_1), d_2 = (v_2, t_2), \dots, d_n = (v_n, t_n) \rangle$, 其中点 $d_i = (v_i, t_i)$ 代表时刻 t_i 对应的观测值 v_i , 记录时间 t_i 是严格增加的。

水文时间序列异常检测首先要解决的一个问题, 是定义在给定的数据集中什么样的数据是异常; 异常的定义决定了异常挖掘的目标。在水文时间序列中, 不同的物理量所构成的不同时间序列间存在很大的差别, 因此, 很难给出统一的异常定义。

定义 2 水文时间序列异常。给定水文时间序列 $D^n = \langle d_1 = (v_1, t_1), d_2 = (v_2, t_2), \dots, d_n = (v_n, t_n) \rangle$, 其中点 $d_i = (v_i, t_i)$ 代表时刻 t_i 对应的观测值 v_i 。用 $\eta_i^{(k)} = \{d_{i-2k}, d_{i-2k+1}, \dots, d_{i-1}\}$ 表示点 d_i 的 k -最近邻居窗口, 其观察值集合记为 $\{v_{i-2k}, v_{i-2k+1}, \dots, v_{i-1}\}$, 若点 d_i 的实际观测值和依据其 k -最近邻居窗口模型预测值之间的差值超过某一特定阈值 τ , 则判定该点为异常点。

由上述定义可知, 最近邻居窗口 k 和阈值 τ 成为判断 d_i 是否为异常的依据。因此, 不同用户可以针对不同的水文要素, 动态调整 k 和 τ 的, 达到最优检测效果。

2.2 算法描述

本文提出的基于滑动窗口的时间序列异常检测方法的核心思想是: 根据不同的应用需求, 定义点 d_i 的最近邻居窗口

$\eta_i^{(k)}$; 建立 one-step-ahead 预测模型并以 $\eta_i^{(k)}$ 的观测值集合作为输入参数预测点 d_i 观测值 v_i' ; 计算点 d_i 对应的预测值 v_i' 的置信区间 (confidence interval), 即 $v_i' \pm \tau$; 其中阈值 τ 可由窗口宽度 k 和预测的置信度 p 计算得到; 当获取 d_i 的实际观测值 v_i 后, 比较 v_i 与预测值 v_i' 之间的关系, 若 v_i 在 $v_i' \pm \tau$ 区间之外, 则判断 d_i 为异常点; 否则为正常点。将滑动窗口后移一步, 以节点 d_i 替代 d_{i-2k} 并更新 $\eta_i^{(k)}$, 进行下一节点的判定, 直至所有节点都完成检测。算法的总体流程如图 1 所示。

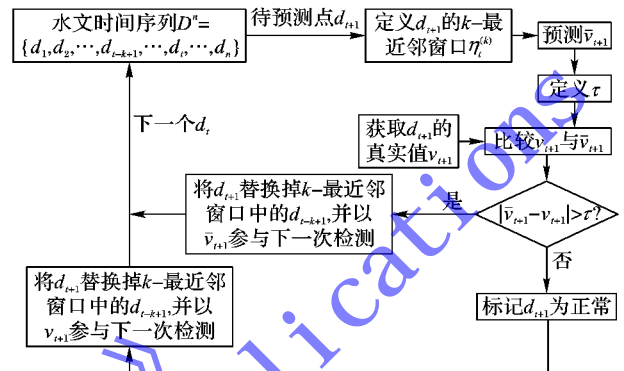


图1 基于滑动窗口的异常检测算法流程

2.2.1 滑动窗口定义

异常检测算法的第 1 步, 定义时间序列 D^n 中待检测点 d_i 的滑动邻居节点窗口 $\eta_i^{(k)}$ 。为降低算法的计算复杂性, 只需采用待检测点 d_i 的 k -最近邻节点作为预测模型的输入参数进行计算。因此, 定义合适的 k -最近邻居窗口 $\eta_i^{(k)}$ 尤其重要。

邻居节点窗口可分为单边和双边两种类型。双边邻居节点窗口适合待检测节点的前驱和后继窗口数据都已知的情况; 单边邻居节点窗口则只包含节点的前驱数据。

基于预测的异常检测算法, 只需要选择待预测节点的左邻居窗口 (已对其中可能存在的异常点进行了检测和处理的干净数据) 作为算法输入参数。单边 k -近邻窗口 $\eta_i^{(k)}$ 定义如下:

$$\eta_i^{(k)} = \{d_{i-2k}, d_{i-2k+1}, \dots, d_{i-1}\} \quad (1)$$

其中 $2k$ 是待预测节点 d_i 的滑动窗口长度, 从 $i-2k$ 到 $i-1$ 。

2.2.2 预测模型

基于预测的异常检测算法核心是建立一个 one-step-ahead 预测模型, 将滑动窗口 $\eta_i^{(k)}$ 作为输入参数预测后继节点的观测值。one-step-ahead 预测模型通过一组特定的输入值——滑动窗口 $\eta_i^{(k)}$ 来预测时间序列中未来某观测点的观测值。one-step-ahead 预测模型能减少由于采集设备故障导致数据不足而不能预测的数量。

采用滑动窗口 $\eta_i^{(k)} = \{d_1, d_2, \dots, d_i\}$ 作为 one-step-ahead 时间序列预测模型的输入参数以预测 d_{i+1} 的观测值, 其预测算法可形式化地表示为:

$$d_{i+1} = M(\eta_i^{(k)}) \quad (2)$$

其中 $M()$ 为预测模型。基于预测的异常检测算法的前提是假设 $k+1$ 时刻的观测值可以由 k 个有限前驱测量值描述, 即其隐式假定时间序列是一个 k 阶马尔可夫过程。文献 [12] 通过比较朴素预测模型、最近邻预测模型、单层线性网络预测模型和多层感知预测模型在不同数据集上异常检测效率, 实验结果表明单层线性网络预测模型能获取最理想的检测效率。在该文的工作基础上, 本文采用单层线性网络作为预测模型, 并

假设 t 时刻的观测值是其前驱邻接窗口的线性组合:

$$\bar{v}_i = \left(\sum_{i=1}^{2k} (w_{t-i} v_{t-i}) \right) / \left(\sum_{i=1}^{2k} w_{t-i} \right) \quad (3)$$

其中 $w_{t-2k}, w_{t-2k+1}, \dots, w_{t-1}$ 表示邻接窗口节点的权重向量, 节点间距离越近, 权重越大。为简化计算, 一般将权重向量 $\langle w_{t-2k}, w_{t-2k+1}, \dots, w_{t-1} \rangle$ 赋值为 $\langle 1, 2, \dots, 2k \rangle$ 。

2.2.3 异常点判定

以待测试点的邻接窗口作为输入参数, 根据预测模型计算出待测试点的预测值及其基于预测取值置信区间。预测取值置信区间给出了待检测观测值可能的取值区间, 置信度系数 $p = 100(1 - \alpha)$ 表明实际测量值在取值置信区间范围内的预期频率。若假定模型残差具有 0 均值高斯分布, 则 $p\%$ 置信度的阈值 τ 及置信区间 (Prediction Confidence Interval, PCI) 可计算如下:

$$\tau = t_{\alpha/2, 2k-1} \times s \sqrt{1 + 1/(2k)} \\ PCI = \bar{v}_{t+1} \pm \tau \quad (4)$$

其中: \bar{v}_{t+1} 是预测模型依据待测试点滑动邻接窗口计算出的预测值; $t_{\alpha/2, 2k-1}$ 是服从自由度为 $2k-1$ 的学生氏 t -分布的百分位数; s 是模型残差的标准差; k 为滑动窗口大小。若实际观测值落在该预测取值置信区间内, 则标记测试点为正常点; 否则标记为异常点并对该点进行进一步分析与处理。PCI 作为判定一个节点是否为异常点的阈值, 能有效调节窗口宽度, 从而避免因阈值选择而导致的误检概率事件。

2.3 参数选择

为了检测在时间序列中的异常值, 基于滑动近邻窗口的异常点预测方法需要计算出测试点的合理的阈值 τ 。为提高异常检测方法的效率并改进算法性能, 合理选择算法中涉及到的 2 个参数: k 和 p , 成为提高异常检测算法的关键。因此, 采用下述实验方法选择适当的参数值:

1) 窗口的宽度 k 。 k 决定了参与预测的滑动邻接点数量。 k 值越大, 参与计算的滑动邻接点越多, 计算复杂度相应增加。为选择最优滑动窗口宽度, 令 k 值范围为 3 ~ 15, 增量为 1, 即 $k = \{3, 4, \dots, 15\}$ 。

2) 置信系数 p 。 p 定义了测量值在取值置信区间范围内的预期频率, 置信系数越大, 预测取值置信区间范围越大, 令 p 值范围为 80% ~ 100%, 增量为 2%, 即 $p = \{80\%, 82\%, \dots, 100\%\}$ 。

异常检测的目标是尽可能多地检测出时间序列中存在的异常点并减少误检测的概率, 因此, 参数对 (k, p) 的取值准则是使得检测正确率和误检率之间的比值最大化。

3 实验分析

为验证本文提出的水文时间序列异常检测方法的有效性, 选择国家水文数据库中某测站的日均水位和流量数据进行实验, 并对算法的检测结果进行分析和讨论。

3.1 数据准备

选择某水文站 (测站编码 40105150, 数据来源于水利部水文局) 日均水位 (单位为 m) 和日均流量 (单位为 m^3/s) 两种不同的水文要素进行算法验证, 该水文站是黄河流域上重要的防洪控制站, 在水文数据采集和预报、防洪调度、河流控制实验和水资源开发等领域对黄河下游重要的作用。采用该测站 2006-01-01—2007-12-31 的实测数据, 图 2 为该数据对

应的曲线。从图 2 可知, 给定的时间序列数据集带有一定的周期性, 曲线总体趋势平滑; 但也存在一些明显可疑的“异常”点信息。

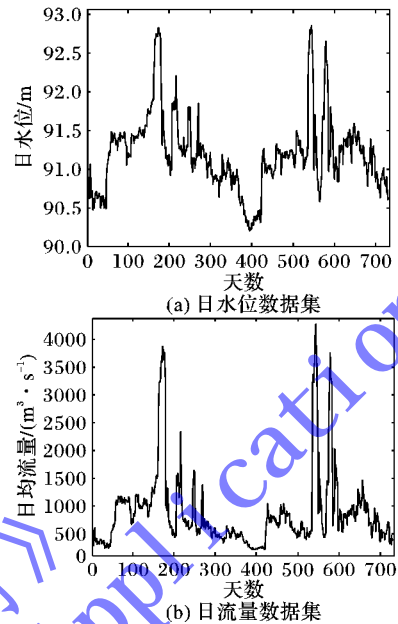


图 2 40105150 测站原始水位、流量数据集

3.2 结果评估

本实验中采用的数据是经过严格的质量控制程序后归档到国家水文数据库, 其数据质量应该相对能满足用户需求。但是, 从实验结果很容易看到一些数据点明显偏离其邻居节点, 即“异常点”。图 3 给出了检测结果。

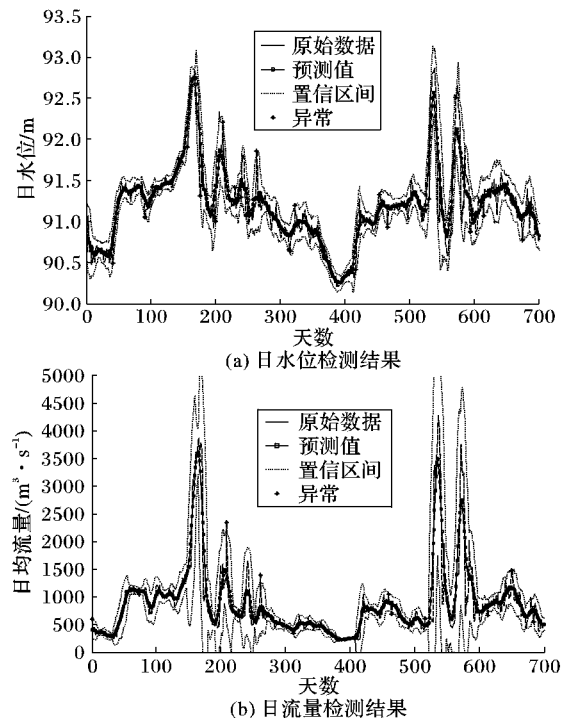


图 3 $(k, p) = \{6, 95\}$ 条件下算法异常检测结果

图 3 给出了滑动窗口宽度 $k = 6$, 置信度 $p = 95\%$ 时异常检测算法在给定时间序列数据集上的实测值、预测值及服从置信度为 95% 的置信区间。由图 3 可知, 大部分时间点的实际值都非常接近各自的预测值, 但也有部分点的实测值在其置

信区间之外,根据本文方法,这些点将被判定为异常。

检测结果表明,基于滑动窗口的水文时间序列异常检测能从水文时间序列数据集上成功检测出一定数目的异常点,具有较广阔的应用前景;但也注意到,有一些正常的数据点也被算法检测为离群值。为正确评估检测法的有效性,可以将实验结果分为4类,如表1所示。

表1 算法检测结果分类

真实值	异常	正常检测值
异常	A类:TP(True Positive), 表示异常被正确检测	B类:FN(False Negative), 表示异常被算法误检为正常
正常	C类:FP(False Positive), 表示正常被误检为异常	D类:TN(True Negative), 表示正常被正确检测

表1中的4个类别对应窗口宽度(k)和置信度(p)情况下采用算法进行异常检测的可能输出结果。A、D类相当于一个理想的情况,即一个正常或异常点能正确被检测出来;B类和C都是不希望出现的,因为这两类情况无法通过检测算法区分正常和异常数据。

基于上述检测结果,定义算法的灵敏度(Sensitivity)、特异度(Specificity)、阳性预测值(Positive Prediction Value, PPV)、阴性预测值(Negative Prediction Value, NPV)这4种评估指标,分别表征正确检测出的异常样本比例、正常样本比例、检测出的真正异常在检测出的所有异常中所占的比例及检测出的正常在检测出的所有正常中所占的比例。不同参数下算法的检测结果如表2所示。

由表2可知,当参数 $k=6, p=95\%$ 时,算法在水位数据集上正确地检测到41条异常值,即 $TP=41$;有6条正常记录被错误判定为异常值, $FP=6$;此外,还有4条异常记录未被正确检测出来, $FN=4$;最后,有678条正常记录最终被判定为正常, $TN=678(41+6+4+678=729)$;其余结果与之类似。

表2 不同参数对情况检测结果

数据集	(k, p)	TP	TN	FP	FN	Sensitivity/%	Specificity/%	PPV/%	NPV/%
水位	(5, 0.95)	38	668	14	9	78.00	98.97	84.78	98.39
	(6, 0.95)	41	678	6	4	84.00	98.97	85.71	98.82
	(7, 0.95)	35	672	12	10	86.00	97.94	75.44	98.96
流量	(5, 0.95)	24	680	13	8	83.33	98.12	69.77	99.13
	(6, 0.95)	29	688	8	4	80.56	99.28	85.29	98.99
	(7, 0.95)	26	687	10	6	66.67	99.42	86.67	98.14

评估结果表明,本文算法能有效检测出水文时间序列中存在的异常点;特别说明的是,特异度指标基本上达到了98%。这些数据值意味着,本文算法检测的异常具有很高的可靠性。除此之外,算法的灵敏度、阳性预测值、阴性预测值也都维持着较高的比率,表明算法的整体检测效果的可靠性。

通过对参数对(k, p)的计算性实验还表明,降低置信度 p 的值会导致更多的数据点被判定为异常值;减小窗口的宽度值会导致 FN 增加而 FP 减小。

3.3 算法分析

将本文方法与其他方法,如 k 近邻方法^[14]、基于混沌方法^[15]和基于中位数的自动检测算法^[13]在同一数据集上进行比较,并将比较结果展示在接收者操作特征曲线(Receiver Operating characteristic Curve, ROC)上^[16]。ROC的横轴是误

报率(False Positive Rate, FPR),纵轴是检测率TPR(True Positive Rate),直观地展示了二者的对应关系。在异常检测的实际应用中,人们希望获得高的 TPR 和低的 FPR ; TPR 表示检测算法的“灵敏度”,而 $1-FPR$ 表示检测算法的“特异性”。理想的ROC是纵轴然后转为 $y=1$ 的直线,越是接近坐标左上区域的ROC表示分类算法越精确。本文算法和其他算法在给定数据集上检测效果的ROC对比如图4所示。

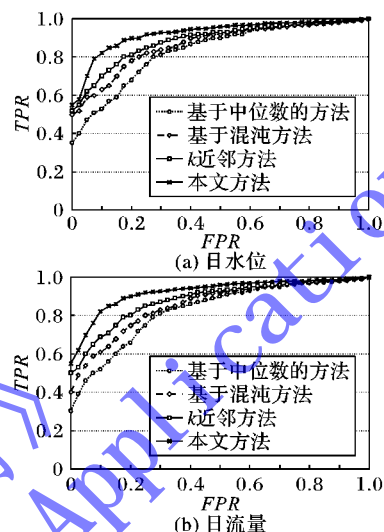


图4 不同方法在本文数据集下的检测结果ROC对比

由图4可知,本文方法的检测结果稳定且令人满意,表明本文方法能适应时间序列异常检测的要求。

4 结语

针对水文时间序列异常挖掘,本文在基于预测的异常检测算法基础上,提出了基于滑动窗口的预测算法以检测水位时间序列中异常;并采用动态参数选择方法选择最优参数模型以提高检测算法的效率。本文采用黄河流域上某测站真实水文数据进行实验,验证了该方法的可行性和有效性,并与其他经典算法进行了比较,结果表明该方法能够较准确、及时地发现时间序列的异常点,满足水文数据质量控制的目标和要求。

参考文献:

- [1] YANG Q, WU X. 10 challenging problems in data mining research [J]. International Journal of Information Technology and Decision Making, 2006, 5(4): 597-604.
- [2] HAN J, KAMBER M, PEI J. Data mining: concepts and techniques [M]. Waltham: Morgan Kaufmann Publishers, 2006.
- [3] GAO B, MA H Y, YANG Y H. HMMs (Hidden Markov Models) based on anomaly intrusion detection method [C]// Proceedings of the 2002 International Conference on Machine Learning and Cybernetics. Piscataway: IEEE Press, 2002: 381-385.
- [4] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: a survey [J]. ACM Computing Surveys, 2009, 41(3): 15.
- [5] RAMASWAMY S, RASTOGI R, SHIM K. Efficient algorithms for mining outliers from large data sets [C]// ACM SIGMOD Record, 2000, 29(2): 427-438.
- [6] YU H, WANG B, XIAO G, et al. Distance-based outlier detection on uncertain data [J]. Journal of Computer Research and Development, 2010, 47(3): 474-484. (于浩, 王斌, 肖刚, 等. 基于距离的不确定离群点检测[J]. 计算机研究与发展, 2010, 47(3): 474-484.)

(下转第2226页)

工艺入手分析帘线搭接倾角范围,进而利用从先验倾角上的灰度累积投影曲线提取的峰值能量构造投影曲线的能量特征向量,最终利用SVM分类器、位置反演计算实现缺陷图像的判别和裂痕缺陷的定位。该方法的创新点在于充分利用缺陷成像区域的灰度和纹理相异于其邻域信息的特性,摆脱了纹理基元分析在胎冠帘线搭接缺陷检测中的局限性,使算法具有普适性。但该算法仍然依赖于胎冠帘线工艺决定的搭接倾角,因此在如何去除裂痕角度先验知识以及降低算法复杂度等方面均需要作进一步的研究。

致谢 感谢山东玲珑轮胎股份有限公司对课题的资助,尤其感谢张彩明教授、郭强博士在课题研究中给予的支持和有益的讨论。

参考文献:

- [1] ABOU-ALI M G, KHAMIS M. TIREDDX: an integrated intelligent defects diagnostic system for tire production and service [J]. *Expert Systems with Applications*, 2003, 24(3): 247-259.
- [2] REN L, AN D, SHEN Y. Causes for defects of carcass ply cord in TBR tire and their countermeasures [J]. *Tire Industry*, 2004, 24(9): 559-561. (任利利, 安登峰, 申玉德. 全钢载重子午线轮胎胎体帘布质量缺陷原因分析及解决措施[J]. 轮胎工业, 2004, 24(9): 559-561.)
- [3] CHU C-H, SONG M-C, LUO V C S. Computer aided parametric design for 3D tire mold production [J]. *Computers in Industry*, 2006, 57(1): 11-25.
- [4] TÖNÜK E, ÜNLÜSOY Y S. Prediction of automobile tire cornering force characteristics by finite element modeling and analysis [J]. *Computers and Structures*, 2001, 79(13): 1219-1232.
- [5] HE Q, JI X, ZHAO S, *et al.* X-ray inspection line for all-steel radial truck tire [J]. *Tire Industry*, 2003, 23(8): 478-482. (何清滨, 姬新生, 赵松善, 等. 全钢载重子午线轮胎X光检验线[J]. 轮胎工业, 2003, 23(8): 478-481.)
- [6] ZHANG Y, LI T, LI Q. Defect detection for tire laser shearography image using curvelet transform based edge detector [J]. *Optics and Laser Technology*, 2013, 47: 64-71.
- [7] HUANG Z, LIU Z, ZHU M, *et al.* Defects on-line detection of tire textures based on statistical features [J]. *Optical Technique*, 2009, 35(1): 60-62. (黄战华, 刘正, 朱蒙, 等. 基于统计特征的轮胎纹理缺陷在线检测[J]. 光学技术, 2009, 35(1): 60-62.)
- [8] FENG X, HAO Z, FENG Y, *et al.* Edge detection in the tire X-ray detection [J]. *Computerized Tomography Theory and Applications*, 2010, 19(3): 61-66. (冯霞, 郝振平, 冯燕, 等. X射线在轮胎边缘检测中的应用[J]. CT理论与应用研究, 2010, 19(3): 61-66.)
- [9] ZHU Y, LIU W, QIU X. Development of all-steel radial tire inspection and defect recognition system [J]. *Opto-Electronic Engineering*, 2009, 36(5): 129-133. (朱越, 刘文耀, 裴稀石. 全钢子午线轮胎缺陷识别系统的研制[J]. 光电工程, 2009, 36(5): 129-133.)
- [10] ZHU Z, LI Z, HE H, *et al.* Application of finite element method in structure analysis of passenger car radial tire [J]. *Tire Industry*, 2012, 32(1): 15-21. (朱振华, 李治国, 何豪明, 等. 有限元法在轿车子午线轮胎结构分析中的应用[J]. 轮胎工业, 2012, 32(1): 15-21.)
- [11] HAN Y, YIN J, LI J. Research of an improved LBP algorithm in texture classification based on rotation invariance and statistical phase distribution [J]. *Chinese Journal of Computers*, 2011, 34(3): 583-592. (韩延彬, 尹建芹, 李金屏. 基于图元旋转不变性和相位统计信息的LBP算法在纹理分类中的研究[J]. 计算机学报, 2011, 34(3): 583-592.)
- [12] LI J, LI Q. Detection of camouflaged moving objects based on laws energy and varying frequency [J]. *Chinese Journal of Stereology and Image Analysis*, 2011, 16(1): 18-23. (李金屏, 李倩. 基于Laws能量和变化频次的伪装色移动目标检测[J]. 中国体视学与图像分析, 2011, 16(1): 18-23.)
- [13] CRISTIANINI N, SHAW-TAYLOR J. An introduction to support vector machines and other kernel-based learning methods [M]. LI Z, WANG M, ZENG H, *et al.*, translated. Beijing: Publishing House of Electronics Industry, 2004. (CRISTIANINI N, SHAW-TAYLOR J. 支持向量机导论[M]. 李正国, 王猛, 曾华军, 等译. 北京: 电子工业出版社, 2004.)
- [7] BREUNIG M M, KRIEDEL H P, NG R T, *et al.* LOF: identifying density-based local outliers [J]. *ACM SIGMOD Record*, 2000, 29(2): 93-104.
- [8] XUE A, JU S, HE W, *et al.* Study on algorithms for local outlier detection [J]. *Chinese Journal of Computers*, 2007, 30(8): 1455-1463. (薛安荣, 鞠时光, 何伟华, 等. 局部离群点挖掘算法研究[J]. 计算机学报, 2007, 30(8): 1455-1463.)
- [9] MOUNCE S R, MOUNCE R B, BOXALL J B. Novelty detection for time series data analysis in water distribution systems using support vector machines [J]. *Journal of Hydroinformatics*, 2011, 13(4): 672-686.
- [10] TIAN J, GU H. Outlier one class support vector machines [J]. *Journal of Electronics and Information Technology*, 2010, 32(6): 1284-1288. (田江, 顾宏. 孤立点一类支持向量机算法研究[J]. 电子与信息学报, 2010, 32(6): 1284-1288.)
- [11] BUDALAKOTI S, SRIVASTAVA A, AKELLA R, *et al.* Anomaly detection in large sets of high-dimensional symbol sequences, NASA TM-2006-214553 [R]. Moffett Field: NASA Ames Research Center, 2006.
- [12] HILL D J, MINSKER B S. Anomaly detection in streaming environmental sensor data: a data-driven modeling approach [J]. *Environmental Modelling and Software*, 2010, 25(9): 1014-1022.
- [13] BASU S, MECKESHEIMER M. Automatic outlier detection for time series: an application to sensor data [J]. *Knowledge and Information Systems—Special Issue on Mining Low-Quality Data*, 2007, 11(2): 137-154.
- [14] POKRAJAC D, LAZAREVIC A, LATECKI L J. Incremental local outlier detection for data streams [C]// CIDM 2007: Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining. Piscataway: IEEE Press, 2007: 504-515.
- [15] NG W W, PANU U S, LENNOX W C. Chaos based analytical techniques for daily extreme hydrological observations [J]. *Journal of Hydrology*, 2007, 342(1): 17-41.
- [16] FAWCETT T. An introduction to ROC analysis [J]. *Pattern Recognition Letters*, 2006, 27(8): 861-874.

(上接第2220页)