

基于 B 树的否定选择算法及其应用

许小润¹, 吴贵芳¹, 张庆华²

(1 河南科技大学 电子信息工程学院, 河南 洛阳 471003; 2 洛阳圣瑞机电技术有限公司, 河南 洛阳 471003)

摘 要: 针对传统的方法在带钢表面缺陷图像检测中存在的问题, 提出了基于 B 树的否定选择算法. 根据影响成熟检测器生成的因素, 通过低频和高频变异生成初始检测器集合, 改变检测器的表示方式, 以数值型的检测数代替二进制字符串形式的检测器, 提取自体数值、检测数值和待检数值, 使用 B 树建立索引并构建自体信息树和检测信息树, 来实现其质量检测. 实验表明, 该方法明显优于传统的图像检测算法, 提高了检测效率.

关键词: 带钢; 缺陷检测; B 树; 否定选择算法

中图分类号: TP31

文献标识码: A

文章编号: 1000-7180(2011)06-0105-04

Negative Selection Algorithm Based on B Tree and Applications

XU Xiao-run¹, WU Gui-fang¹, ZHANG Qing-hua²

(1 Electronic Information Engineering College, Henan Uni. of Sci. & Tech., Luoyang 471003, China;

2 Luoyan Sunray Technology Co. Ltd., Luoyang 471003, China)

Abstract: As the traditional algorithms of image detection are applied to the inspection of strip surface, the treatment effect is not very ideal. According to this problem, a new strip surface defects detection technology is proposed negative selection algorithm based on B tree. According to factors of generation of mature detectors, the candidate detector are produced variation through low frequency and high frequency, representation of detector is changed to value-based detector instead of binary-based detector, and 3 properties such as self-number, detector-number and some antigen-number are extracted to construct indexes using B tree based negative selection algorithm to achieve quality of steel strip surface defect detection. Experiments show that the method is significantly superior to the traditional image detection algorithms in detection efficient.

Key words: steel strips; defect detection; B tree; negative selection algorithm

1 引言

表面质量的优劣是带钢的一项重要指标, 由于传统的人工检测方法无法获得令人满意的检测结果, 因此研究带钢表面缺陷检测技术具有重要的现实意义.

人工免疫系统^[1-2] (Artificial Immune System, AIS) 是在模拟生物免疫系统组织和工作机制的基础上, 快速发展而来的新兴智能技术, 具有快速准确的识别系统中异常数据信息^[3] 的特性, 与带钢表面缺陷检测问题^[4] 具有本质的相似性. New Mexico

大学 Forrest 教授等人在 1994 年提出了否定选择算法^[5-6], 是人工免疫系统研究的重要理论, 通过模拟 T 细胞的成熟机制选择成熟检测器^[7-8], 然后判断检测器与抗原是否匹配, 与检测器匹配的抗原为非法自体, 从而达到检测异常信息的目的. 否定选择算法可使用空间向量^[9] 或二进制字符串两种形式来表示抗原和检测器, 并衍生了一系列改进的算法, 由于计算机系统最终均以二进制字符串方式来存储检测器和抗原, 以二进制字符串形式表示检测器的方法, 使得生成检测器和检查访问请求时存在时间和空间开销大、冗余开销较多和难以分析检测器识别

收稿日期: 2010-08-16; 修回日期: 2010-11-01

基金项目: 国家自然科学基金项目(50074010)

异常信息的能力等问题. 本文针对二进制字符串表示法, 提出了基于 B 树的否定选择算法, 从而实现了带钢表面缺陷的质量检测.

2 元素的定义

下面对相关元素进行定义.

定义 1 论域 $V = \{0, 1\}^l$ 表示所有长度为 l 的二进制字符串, 包含所有的抗原和检测器, 属于人工免疫算法讨论范围.

定义 2 带钢表面用抗原 k 表示: $k = k_1 k_2 \cdots k_l (k_i \in \{0, 1\})$, 为长度为 l 的二进制字符串, 是否定选择算法的一个检测对象, 抗原有自体和非自体组成.

定义 3 带钢表面非缺陷用自体 S 表示: $S \subseteq V$, 是合法抗原的集合.

定义 4 带钢表面缺陷用非自体 NS 表示: $NS \subseteq V$, 是非法抗原的集合.

根据上面的定义, 可以得出推论一.

推论一: $S \cap NS = \emptyset$ 且 $S \cup NS = V$

定义 5 检测器 $t: t = (t_1, t_2, \cdots, t_l) (t_i \in \{0, 1\})$, 是长度为 l 的二进制字符串, 用于检测抗原的合法性, 便于分析带钢表面缺陷和非缺陷.

定义 6 匹配阈值用 r 表示: 是判断抗原 k 与检查器 t 是否匹配的标准.

定义 7 特征子串用 q 表示: 是检查器用于检测抗原的长度超过匹配阈值的子串.

3 检测数值的转换算法

检测器用二进制字符串表示法是影响生成成熟检测器所需空间和时间开销的主要原因, 使得现有的人工免疫算法在选择成熟检测器时, 存在着反复遍历所有的自体子串、重复提取特征子串、逐位比较字符串、浪费大量系统开销等问题, 此外还存在难以分析选择成熟检测器对非自体的覆盖率.

根据检测器在检测抗原时, 只使用其中的两个因素: 特征子串和对应的起始位置, 本文用二元组形式来表示检测元素, 每个检测器中有 $l - r + 1$ 个检测元素. 检测元素的概念如定义 8 所示.

定义 8 检测元素 (q, p) , q 表示检测器中某一个特征子串, p 表示 q 在检测器中相对于最左端的偏移量.

本文根据匹配阈值 r , 提取检测器中的检测元素, 然后将检测元素转换为检测数值 t_m , 删除重复检测元素, 用数值形式比较来代替逐位的二进制字符

串比较, 提高了检测效率. 检测数值转化如式 (1) 所示.

$$t_m = l_q * 2^l + \sum_{i=0}^{l_q} q_i * 2^{l-p-i} \quad (1)$$

式中, l_q 表示特征子串 q 的长度, q_i 表示特征子串中任一位的值 (1 或 0), p 表示与之对应的特征子串相对于最左端的偏移量. 当计算检测数 t_m 时, 首先根据检测器 t 中特征子串 q 的位置将其右边的位置用 0 补齐, 然后转换成十进制的子串数值, 特征子串的长度各不相同 ($l_q \in [r, l]$), 用 l_q 与 2^l 相乘获得检测元素的长度值, 最后用子串数值与检测元素的长度值相加得到唯一的检测数值 t_m .

将检测器中的检测元素转换为检测数值后, 本文算法仅保存从检测元素转换得到的检测数值, 不再保存长度为 l 的检测器, 删除了相同的检测元素, 一次性转换并得到了全部的检测数值, 不需要反复提取特征子串, 避免了检测抗原所带来的重复开销, 为建立索引提供了基础, 并提高了检测效率.

4 基于 B 树的否定选择算法

传统的否定选择算法, 采用的检测方法是逐位比较抗原与特征子串, 而检测器中用于检测抗原的特征子串的长度各不相同, 无法排序, 因此检测抗原时需要遍历全部的检测器, 检测效率比较低. B 树是一种多路平衡查找树, 具有较高的搜索性能, 能有效地提高检索数据的效率.

初始检测器集合的优劣直接影响到合格检测器生成的效率和质量, 本文不再使用随机方式生成, 而是通过低频变异和高频变异^[10]来生成初始检测器集合, 从而提高了检测器生成的效率, 减少了计算量. 具体流程如图 1 所示.

图 1 中先对各个数据集合进行初始化, 读取自体的数据集, 然后判断其所属集合, 如果属于自体集, 进行高频变异, 否则进一步判断该数据值是否属于合格检测器集, 如果是则进行低频位变异, 加入初始检测器集, 否则就删除该数据, 然后获取新的模式数据, 直到达到系统设置阈值停止.

为了提高检测抗原的效率, 本文引入了 B 树, 提出了基于 B 树的否定选择算法. 先将自体、检测器和抗原中用于检测的元素转换为数值形式, 用数值形式比较代替传统算法中检测抗原时逐位比较二进制字符串的方法, 然后使用 B 树建立检测数值的索引, 代替遍历所有检测器的方法, 提高了检测效率. 基于 B 树的否定选择算法的关键是: 构建检测

树和构建检测抗原. 具体的流程如图2所示.

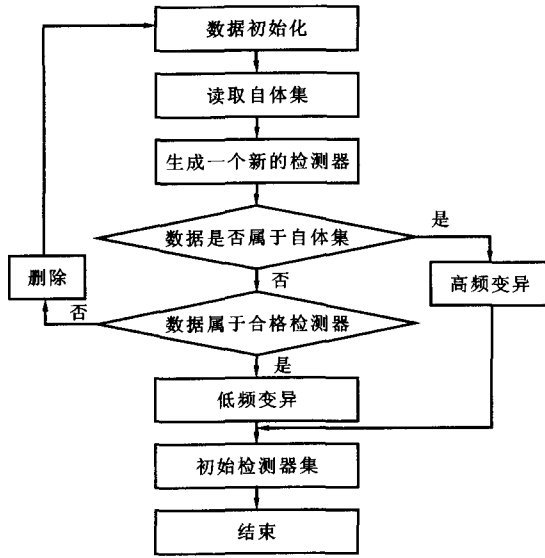


图1 初始检测器生成流程

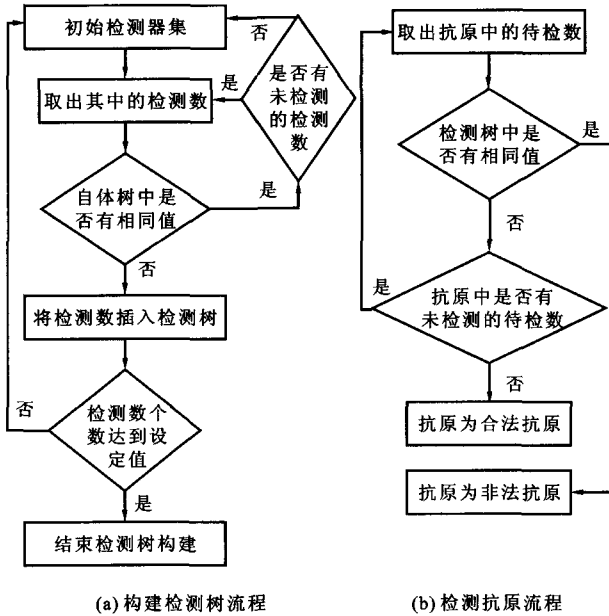


图2 基于B树的否定选择算法

图2(a)构建检测树流程目的是选择检测器, 包括两个主要过程: 构建自体信息树和构建检测信息树. 自体信息树的构建: 先从所有的自体中提取自体数 s' , 如式(2)所示, 然后用二元组形式的自体元素 $sn(s', f)$ 来表示, 其中 f 表示从全部自体中转换得到的相同自体数的次数, 最后使用B树建立索引, 构成自体信息树.

$$s' = p * 2^{l_s} + \sum_{i=p}^{p+l_s} s_i * 2^{l-i} \quad (2)$$

式中, p 是子串在自体中相对于最左端的偏移量, s_i 是子串中某一位的数值, l_s 是提取自体中的子串长度, i 为某一位相对子串左端的偏移量.

图3给出了当匹配阈值为7时, 从一个长度为8的检测器中提取自体数的过程, 以及自体数的计算结果.

子串, 相对左端的偏移量使用式(2)计算				自体数
自体1	11100000	1100000,1	$96+1*2^7$	224
		1110000,0	$112+0*2^7$	112
自体2	11100001	1100001,1	$97+1*2^7$	225
		1110000,0	$112+0*2^7$	112
				重复

图3 从自体中提取和转换自体数的示意图

定义9 检测数值集合 $T: T = \{(t_m, r, \text{type})\}$, $t_m \in N, r \in N, 1 < r < l, \text{type} \in \{0, 1\}$, 其中 t_m 是一个检测数值, r 是其对应的匹配阈值, type 表示检测器的类型为1或0.

构建检测树的过程: 首先将提取生成检测器中包含的检测数作为关键字, 查找用B树建立的自体信息数中是否包含相同值, 选择其中未包含的数值为检测数 t_m , 并保存到检测数值集合 T 中, 且检测器类型的值默认为0, 当挑选出的检测数的个数达到系统设置阈值时, 完成了检测树构建工作, 否则重新生成一个新的检测器, 继续选择符合条件的检测数, 补充到检测信息树中.

图2(b)检查抗原流程目的是判断抗原的合法性, 首先提取抗原中的待检数 k' (计算方法如式(3)所示), 再将提取的待检数作为关键字, 查找检测树中是否存在相同值, 如果存在则判断该抗原为非法, 如果抗原中所有的待检数均不存在检测树中, 则判断该抗原为合法.

$$k' = p * 2^{l_k} + \sum_{i=p}^{p+l_k} k_i * 2^{l-i} \quad (3)$$

式中, l_k 是抗原中子串的长度 $l_k \in [r, l]$, k_i 是子串中的某一位的数值, p 是子串在抗原中相对于最左端的偏移量, i 为某一位相对子串左端的偏移量.

基于B树的否定选择算法, 将自体信息、检测器和抗原中与之匹配、比较相关的元素, 一次性转化为数值形式, 不需要反复地提取子串, 降低了开销, 使用B树建立自体数和检测数的索引, 改变了遍历检测器的方法, 提高了检测效率, 在自体信息树、检测信息树及待检数集合中, 删除了重复值, 避免了重复比较, 提高了检查效率.

5 实验与结果分析

设计了一个对比实验, 将本文算法的检测效果与形态滤波检测算法进行对比, 以验证基于B树的否定选择算法在带钢表面缺陷检测技术中的效果.

从现场采集了白斑缺陷样本图像, CCD 采集的图片大小为 129×129 像素。

图 4(a)是样本图像中的一个例图,可以看出,白斑缺陷图像(a)受光照和噪声干扰,视觉效果不明显,经过本文算法检测后,得到了白斑缺陷效果定位图像(b)。

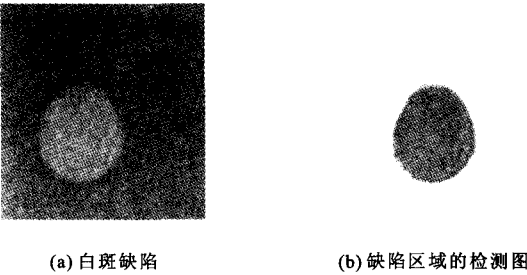


图 4 白斑缺陷图像及处理效果图

表 1 是白斑缺陷的特征参数表,表 2 是本文方法和形态滤波检测算法在 CPU Core2 T5550 1.9 GHz 以及 1 GB 内存的计算机中对采集的 20 幅带钢白斑缺陷样本的特征参数进行检测的结果。可以看出,利用本文算法进行缺陷检测的时候,平均时耗 0.10 ms,缺陷的平均检测速度提高了 10.1 倍,识别准确率可以达到 95.63%,比形态滤波检测算法提高了 10% 以上,很大程度上提高了缺陷检测的效率。

表 1 白斑缺陷的主要特征参数表

指标	缺陷区面积	缺陷区周长	缺陷区长宽比	缺陷区圆弧度
白斑	1 962.9	157.8	25.8:33.6	29.323

表 2 不同缺陷检测算法的识别结果

对比指标	形态滤波检测算法	本文方法
识别率/%	85.46	95.63
计算时间/ms	0.79~1.19	0.03~0.17

本文为了验证其抗噪性,采用的性能指标为信噪比(SNR),并作为对比的标准,如式(4)所示:

$$SNR = 20 \times \lg \frac{255}{\sqrt{1/[AB \cdot \sum_{a=0}^A \sum_{b=0}^B (X_{ab} - X'_{ab})^2]}}$$

(4)

式中, X_{ab} 和 X'_{ab} 分别表示为原始图像和处理后的图像灰度值, A 和 B 分别表示图像长度和宽度的像素个数。

从表 3 中可以看出,基于 B 树的否定选择算法处理带钢表面缺陷的图像,其信噪比明显高于原图像和经直方图均衡化和同态滤波算法处理后的图像。

表 3 不同算法的图像信噪比(SNR)对比

	原来图像	形态滤波检测算法	本文方法
指数	68.67	73.89	86.89

6 结束语

在检测带钢表面缺陷时,针对否定选择算法在检测抗原时存在重复检测、检测器查找效率较低及逐位比较开销较大等问题,本文首先通过低频和高频变异生成初始检测器集合,再提取检测器中用于检测的元素,转化为唯一的数值形式,为提高检测算法的效率提供了的基础;然后设计了基于 B 树的否定选择算法,分别提取自体信息中的自体数和抗原信息中的待检数,使用 B 树建立自体信息树和检测信息树,由于是一次性地提取了自体信息、检测器和抗原中用于匹配比较所需要的元素,因此有效地避免了反复提取;最后在生成检测器和检测抗原时,分别以检测数和待测数作为关键字,查找自体信息树和检测信息树中是否包含相同值,判断是否匹配,提高了带钢表面缺陷检测的效率。通过设计相关实验,验证了基于 B 树的否定选择算法相比传统的缺陷检测算法具有更高的检测效率。

虽然 B 树在带钢表面缺陷检测时取得了较好的检测效果,但 B 树并不能对高维数据进行有效的索引,所以下一步需要研究特殊的能适应多维特性的空间索引方式,进一步提高否定选择算法的检测效率。

参考文献:

[1] 肖人彬,王磊. 人工免疫系统:原理、模型、分析及展望[J]. 计算机学报,2002,25(12):1281-1293.

[2] 何申,罗文坚,王煦法. 一种检测器长度可变的非选择算法[J]. 软件学报,2007,18(6):1361-1368.

[3] Jorge L Mamaral, Jose F M. An immune fault detection system for analog circuits with automatic detector generation[J]. IEEE Congress on Evolutionary Computation, 2006,16(21):2966-2971.

[4] 张洪涛,段发阶,丁声华. 钢板表面缺陷在线视觉检测系统研究[J]. 传感技术学报,2007,20(8):58-61.

[5] Tan minsheng, Yu honfxiao, Zhao zhiguo. An artificial immunity-based proactive defense system[J]. IEEE international Conference on Robotics and Biomimetics, 2007,15(18):2239-2243.

[6] 王茜,傅思思,葛亮. 基于人工免疫的新型检测器生成模型[J]. 计算机应用,2006,26(11):2618-2621.

(下转第 112 页)

种情况,由图示可知,该情况下比较器失调处理后残余的等效输入失调电压为 $33.2 \mu\text{V}$,满足精度要求。

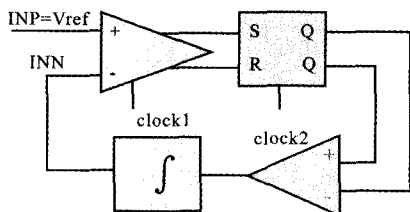


图4 钟控比较器动态失调测试电路

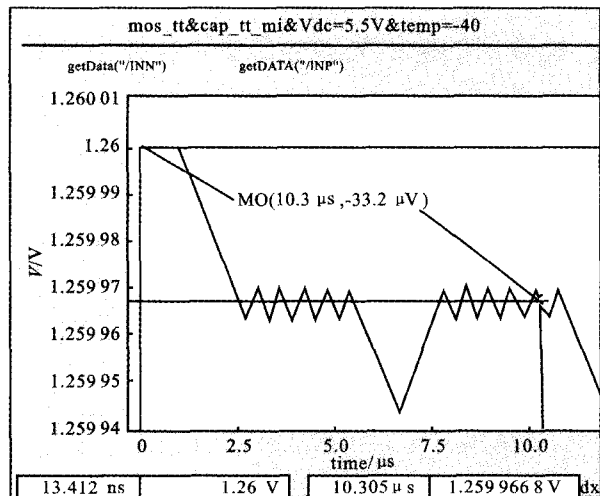


图5 比较器输入失调电压仿真结果

5 结束语

通过采用三级前置运算放大器和 latch 再生电路的比较器结构以及输入失调储存和输出失调储存级联的消失调技术,实现了高精度要求。带负反馈网络的动态失调电压测试电路有效提高了设计和仿真

效率。仿真表明,比较器能够分辨的最小信号为 $33.2 \mu\text{V}$,满足 14 bit SAR ADC 对比较器的精度要求,但是比较器的速度只有 30 MHz,可以进一步优化设计,实现高速高精度,扩大比较器的应用范围。

参考文献:

- [1] 裴晓敏. 8 通道 10 b 的 R-C 混合式 SAR ADC 的设计[J]. 现代电子技术, 2008(9): 83-85.
- [2] 陈娟娟, 钟德刚, 徐静平. 用于便携式设备的 12 位低功耗 SAR A/D 转换器[J]. 微电子学, 2008, 38(3): 401.
- [3] Long Shanli, Wu Jianhui, Xia Xiaojuan, et al. A 1.8-V 3.1 mW successive approximation ADC in system-on-chip[J]. Analog Integrated Circuits and Signal Processing, 2008, 56(3): 205-2011.
- [4] 冯楚华, 陶建中, 于宗光. 一种用于数字功放的低功耗宽输入电压比较器[J]. 微电子学与计算机, 2008(11): 109-112.
- [5] 殷湛, 郭立, 杨吉庆. 一种用于流水线 ADC 的高速电压比较器[J]. 微电子学与计算机, 2006(2): 182-184.
- [6] 孙彤, 李冬梅. 一种 0.2-mV 20-MHz 600-μW 比较器[J]. 微电子学, 2007, 37(2): 270-272.
- [7] 李亮, 臧佳峰, 徐振, 等. 高速低功耗钟控比较器的设计[J]. 半导体技术, 2008, 33(1): 11-13.
- [8] 林武平, 郭良权, 于宗光. 新型高速低功耗动态比较器[J]. 半导体技术, 2008, 33(12): 1119-1120.

作者简介:

陈幼青 女, (1983-), 硕士研究生. 研究方向为模拟集成电路设计.

何明华 男, (1971-), 博士, 教授. 研究方向为嵌入式系统与系统级芯片设计.

(上接第 108 页)

- [7] 鲁云平, 宋军, 姚雪梅. 基于免疫原理的网络入侵检测算法改进[J]. 计算机科学, 2008, 35 (9): 116-119.
- [8] 蔡涛, 鞠时光, 仲巍, 等. 面向存储安全系统的新型人工免疫算法[J]. 计算机科学. 2008, 35 (8): 60-63.
- [9] 陈云芳, 王汝传. 基于免疫算法的分类器设计[J]. 计算机科学, 2008, 35 (12): 133-135.
- [10] 陈君波, 叶庆卫, 周宇, 等. 一种新的混合变异粒子群算法[J]. 计算机工程与应用, 2007, 43 (7): 59-60.

作者简介:

许小润 女, (1982-), 硕士. 研究方向为数字图像处理、模式识别.

吴贵芳 男, (1978-), 博士, 副教授. 研究方向为数字图像处理、模式识别.

张庆华 男, (1964-), 硕士, 高级工程师. 研究方向为机器视觉.