

遗传神经网络在稻米垩白度检测中的应用研究

黄星奕¹, 吴守一¹, 方如明¹, 罗玉坤²

(1. 江苏大学生物与环境工程学院, 江苏 镇江 212013; 2. 中国水稻研究所, 杭州 310006)

摘要: 新的优质稻谷国家标准中, 垩白度是 4 个定级指标之一, 被用来代表稻谷的商品品质。垩白度的检测目前仍由人工目测完成。为使检测结果更具客观性、一致性, 建立了遗传神经网络对垩白像素和胚乳其它像素进行了识别, 从而实现了垩白度的自动无损检测。对两种市售粳米进行了检测, 计算机视觉的检测结果与人工检测结果的误差小于 0.05。试验结果表明所建立的新方法是可行的, 它为开发垩白度在线检测系统提供了科学依据。

关键词: 稻米; 垩白度; 人工神经网络; 遗传算法; 品质检测

中图分类号: TP389.1; S37

文献标识码: A

文章编号: 1002-6819(2003)03-0137-03

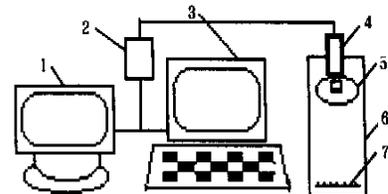
1 引言

中国于 1999 年 9 月颁布了优质稻谷国家标准^[1], 并于 2000 年 4 月起在全国执行。新国标的 4 个定级指标中以垩白度代表商品品质, 可见垩白度检测的重要性。存在垩白的米粒结构疏松、硬度小, 加工时易破碎, 严重影响出米率, 也影响米的外观品质, 降低其市场价值, 并影响食味^[2]。目前江苏粳米比东北大米差的主要原因就是垩白度太高, 而其他指标并不逊色^[3]。垩白度也是影响我国大米在国际市场上竞争力的一个重要因素。我国是稻米生产大国, 但由于品质原因及缺乏高水平的检测技术, 稻米的出口创汇情况很不理想^[4]。虽然垩白是稻谷在田间生产过程中形成的, 加工行业无法控制其产生, 但通过检测, 剔除垩白粒, 可以提高大米的品质等级, 提高其市场价值和增强国际竞争能力。这对于入世后的我国农业如何应对农产品所面临的巨大挑战, 是具有十分积极的意义的。随着社会经济的发展和人民生活水平的提高, 人们对稻米品质的要求越来越高, 但是目前垩白度的检测仍由人工目测完成。人工判断结果通常随着年龄、性别、识别能力等的不同而存在相当大的个体差异。即使同一人员也随其身体状况和情绪的变化产生不同的结果, 难以坚持统一、客观的标准。侯彩云等^[5]用 3 维图像处理系统对心白米粒垩白度进行了测定, 提出了可行的途径。该方法利用染料对样品进行了预处理, 所用系统对原料的观测是破坏性的, 不适合于解决优质稻谷国家标准中垩白度等指标的快速检测。孙明等^[6]在 MATLAB 环境中用计算机视觉技术对大米垩白进行了检测。所用固定阈值方法当采样环境或待测米样品种两者中任一方发生变化时都无法实现自动检测。本文建立了一个遗传神经网络用于垩白像素和胚乳其它像素的识别, 从而实现垩白度的无损自动检测, 为开发垩白度自动在线检测系统打下良好的基础。为适应我国粮食流通体制的改革, 为商品稻谷收购及市场流通

过程中按质论价提供了科学依据。

2 计算机视觉与图像采集

所用的计算机视觉系统(如图 1)由 Panasonic WL-CL500CCD 彩色摄像头, CA540 真彩色图像采集卡, Pentium III 计算机, 监视器及光源箱组成。光源箱设计成可调式四棱柱体暗箱, 内壁配以无反差白色表面, 以形成均匀漫反射。在 20W 环形荧光灯照射下采集单层米样的反射图像(如图 2)。



1 监视器 2 图像采集卡 3 计算机
4 摄像头 5 光源 6 光源箱 7 样本

图 1 计算机视觉系统示意图

Fig. 1 Computer vision system



图 2 部分垩白米样

Fig. 2 Sample of chalky kernels

垩白度的检测涉及到垩白面积和米粒面积的精确计算, 在对图像分割时, 分割阈值的选取直接影响面积计算的精度。因此分割前先对图像进行平滑, 以减小图像的各种寄生效应对分割阈值的影响^[7]。

3 遗传神经网络的建立

对平滑滤波后的图像用灰度阈值分割法把研究对

收稿日期: 2002-04-08

基金项目: 教育部留学回国人员启动基金项目

作者简介: 黄星奕, 博士, 副教授, 江苏镇江 江苏大学生物与环境工程学院, 212013



象(米粒)与背景分离出来。去除背景,再用最大方差阈值法求得歪白与胚乳其他部分分离的阈值 T_c ,用 T_c 将米粒部分分割成歪白区域与胚乳无歪白区域,理论上可以求得有关歪白的各个指标值。然而歪白区域与胚乳其他区域的交界处虽然存在灰度的差异,但却并不形成十分陡峭的阶跃性突变,而这部分交界区域内的像素的识别对歪白指标的计算十分重要,少量像素的误识别会使面积计算的误差成倍增加。尽管已采用最大方差阈值法求取分割阈值 T_c ,但由于 T_c 附近的像素多而且重要, T_c 稍有偏差都会造成大量像素的误分类。因此采用遗传算法建立一个神经网络识别系统对这部分交界区域内的像素进行识别^[8]。

3.1 遗传神经网络的构筑

遗传算法是一种基于达尔文生物进化论思想的新的优化算法。它是对生物进化过程中自然选择规律的模拟^[9]。针对具体问题,首先选择一组可能解作为原始种群,通过选择、杂交、变异等使这组可能解逐渐改变,并在物竞天择、优胜劣汰原则下使种群在若干代后的素质得以提高,这些子代中素质最好的即问题的最优解得以生存。

遗传算法对优化问题基本上没有限制。对目标函数及约束既不要求连续也不要求可微。对种群进行的优化操作实现了全局并行操作其中隐含着一种并行计算机制。其搜索空间遍及解空间,因而易于得到全局最优解,而不至陷入局部最小,说明其有较强的鲁棒性^[10]。遗传算法能获得较高的效率,加之它的通用编码技术和简单有效的遗传操作,使其应用领域正越来越广泛。

基于遗传算法的上述优点,将其运用到神经网络的设计中,以使得所构建的神经网络具有尽可能好的函数估计和分类功能。

3.2 遗传算法的实现

遗传算法模拟了自然选择和自然遗传过程中的繁殖、交配和变异现象,它将每个可能的解看作是群体(所有可能解的集合)中的一个个体,并将每个个体编码成字符串的形式,根据预定的目标函数对每个个体进行评价,给出一个适配值。因此,首先对所求的所有网络参数按照一定的顺序进行编码,形成一个字符串(一个个体),一个字符串代表了神经网络参数的一种可行解,然后根据编码规则产生初始种群(由若干个个体组成)。

根据歪白检测的需要,将网络的输入结点数取为 3,用于输入米粒的 RGB 值(归一化为 rgb)。采用一个隐层,隐层结点为 4,输出层结点为 2,对应歪白和非歪白像素。其中输入层到隐层的权阵取为 W_1 ,隐层到输出层的权阵为 W_0 ,采用先 W_1 后 W_0 的顺序编排。同时对每个权重矩阵按行的顺序先后对各权重参数进行编码,每个权重参数用 8 位二进制码表示。一个串的长度等于网络权重参数总数乘以 8。

随机产生 50 个串作为初始种群,得到相应的 50 组网络权重参数。这样便完成了遗传算法的准备工作,下面进行遗传算法的 3 个基本操作步骤:

1) 复制

复制过程是个体串按照它们的适配值进行复制,本研究中个体的适配值按公式(1)、(2)进行计算。

$$f = 1/E \quad (1)$$

$$E = \frac{1}{2} \sum_{k=1}^N \sum_{m=1}^{N_0} (D_{km} - O_{km})^2 \quad (2)$$

式中 f —— 种群中个体的适配值; E —— 网络评价函数; D_{km} —— 第 k 个输入模式时,第 m 个输出层神经元期望输出值; O_{km} —— 第 k 个输入模式时,第 m 个输出层神经元实际输出值; N —— 样本数; N_0 —— 输出层结点数。

此时,适配值相当于自然界中的一个生物为了生存所具备的各项能力的大小,它决定了该串是被复制还是被淘汰。概括地说,复制操作对模式的影响是使高于平均适配值的模式数量增加,低于平均值的模式数量减少。

2) 交叉

交叉操作可以分为如下两个步骤。第 1 步是将新复制产生的匹配池中成员随机两两匹配,第 2 步是进行交叉繁殖。

设串的长度为 l ,则串的 l 个数字位之间的空隙标记为 $1, 2, \dots, l-1$ 。随机地从 $[1, l-1]$ 中选取一个整数位置 k ,则将两个父母串中从位置 k 到串末尾的子串互相交换,从而形成两个新串。

交叉概率的选择决定了交叉操作的频率。频率越高,可以更快地收敛到最有希望的最优解区域,但是太高的频率也可能导致收敛于一个解。本研究选择交叉概率为 0.8。

交叉过程是串之间的有组织的而又又是随机的信息交换,它在创建新结构的同时,最低限度地破坏复制过程所选择的高适配值模式。

3) 变异

变异是以很小的概率随机地改变一个串位的值。对于二进制串,是将随机选取的串位由 1 变为 0 或由 0 变为 1。变异的概率通常是很小的,本研究中取为 5%。这个操作的作用相对于复制和交叉而言是相对次要的,其目的是为了防止丢失一些有用的遗传因子。特别是当种群中的个体,经遗传运算可能使某些串位的值失去多样性,从而可能失去检验有用遗传因子的机会时,变异操作可以起到恢复串位多样性的作用。

根据模式理论,随着复制、交叉、变异一代一代地进行,那些短的、低次数、高适配值的模式将越来越多,最后得到的串即这些模式的组合,因而期望性能越来越得到改善,并最终趋向全局的最优点。

4 网络训练与歪白识别

训练集由多个训练模式对构成。为了使训练集中的模式对涵盖歪白像素和非歪白像素的各种情形,训练模式对的选取采用在一定的组织中随机选择的方法,即先将歪白像素分成歪白中心区域和边缘区域像素,然后在两区域中随机选择一定量的样本,非歪白像素的选择也照此办理。训练集包括了 80 个样本,训练 100 00 次时,

训练集的回判正确率为 98.75%, 测试集的正确率为 96%。

如果米粒上每个像素都用该网络来测试, 则工作量太大, 也没有这个必要。因此, 实际运用中首先采用前面已平滑过的图像, 求得垩白与胚乳其余部分区分的阈值 T_c 。将灰度小于 $T_c - 15$ 的像素归入胚乳区域, 灰度大于 $T_c + 15$ 的像素归入垩白区域, 而将灰度在 $[T_c - 15, T_c + 15]$ 区间内的像素用网络进行分类。

按照新国标中垩白度检测的操作规程, 由人工和计算机视觉分别对两种市售粳米进行垩白度的检测, 结果如表 1 所示。

比较检测结果可知, 计算机视觉与人工目测方法的误差小于 0.05。人工目测方法中尺度的掌握主要凭经验, 而用计算机视觉技术则可以排除检测人员的个体差异和其它主观因素的影响, 提高检测结果的客观性和一致性。今后还应将专家检测经验引入新方法中, 用以指导程序中相应参数的选择和标定, 进一步提高新方法的权威性。

表 1 垩白度检测结果

Table 1 Results of chalk degree inspection %		
品种	人工目测	计算机视觉检测
1	12.8	13.4
2	15.2	15.9

5 结论

测试表明: 遗传算法作为对进化论思想的计算机模拟, 这一非数学型自适应优化搜索算法能够有效地解决

网络的构筑及结合权值的确定等问题。所建立的遗传神经网络能有效地识别垩白像素和胚乳其它像素, 提高了垩白检测的客观性和一致性, 为实现垩白度的无损自动在线检测打下了良好的基础。另外, 只要提供合适的训练集, 所建立的网络亦可用于其它农产品的特征识别和品质检测。

[参考文献]

- [1] 中华人民共和国国家标准 优质稻谷 GB/T17891-1999 [S]. 北京: 中国标准出版社, 1999.
- [2] 李智念, 徐海波, 王光明. 温度与稻米垩白形成的研究综述 [J]. 耕作与栽培, 2001(6): 58~60.
- [3] 缪炳良, 黄宝才, 张长斌. 对照新国标看我省目前的稻米品质 [J]. 江苏农业科学, 2001(1): 8~12.
- [4] 农业部国外农业调研组. 世界农业及发展趋势 [J]. 世界农业, 1998(3): 3~6.
- [5] 侯彩云, 等. 3 维图像处理系统在稻米品质检测中的应用研究 [J]. 农业工程学报, 2001, 17(3): 92~95.
- [6] 孙明, 凌云, 王一鸣. 在 MATLAB 环境中基于计算机视觉技术的大米垩白检测 [J]. 农业工程学报, 2002, 18(4): 146~149.
- [7] Abdullah M Z. Color vision system for ripeness inspection of oil palm [J]. Journal of Food Processing and Preservation, 2002(3): 86~98.
- [8] 黄星奕. 基于计算机视觉的稻米外观质量检测系统研究 [D]. 镇江: 江苏理工大学, 1999.
- [9] 程国胜. 连续时间遗传算法模型及其收敛性分析 [J]. 应用数学, 2002, 15(4): 23~28.
- [10] Srinivas M. Genetic Algorithms: A Survey [J]. Computer, 1994(6): 17~26.

Inspection of chalk degree of rice using genetic neural network

Huang Xingyi¹, Wu Shouyi¹, Fang Ruming¹, Luo Yukun²

(1. College of Biological and Environmental Engineering, Jiangsu University, Zhenjiang

212013, China; 2. China National Rice Research Institute, Hangzhou 310006, China)

Abstract: Chalk degree is one of the four important criteria for judgment of rice quality according to China National Standard of Rice. It has been determined by human inspection exclusively so far. A new method was developed to identify chalk and to grade chalk degree of rice using genetic algorithm and neural network in conjunction with computer vision. Genetic neural network was trained to identify chalk pixels and other pixels of endosperm and subsequently to evaluate chalk degree of rice. Two different kinds of rice bought on market were tested to evaluate system performance. Compared experiment results of new method using genetic neural network with that of human inspection, the error rate was less than 0.05. This method is proved to be robust and consistent. It paves the way for on-line automated judgment of chalk degree of rice.

Key words: rice; chalk degree; artificial neural network; genetic algorithm; quality inspection