

胶料中炭黑分散性的 Fisher 判别法

王明兰

(华南师范大学 数学系, 广东 广州 510631)

摘要: 提出运用数理统计方法中的 Fisher 判别法对胶料中炭黑分散性进行判别并定出分散度等级的新方法。该方法通过计算机对胶料的彩色照片进行灰度化及黑白化处理, 使其形成鲜明的黑白图形文件, 然后计算不同大小炭黑颗粒的面积及个数, 再代入 Fisher 数学模型进行炭黑分散度等级判别。该方法具有准确、快速和客观性强的优点。

关键词: 胶料; 炭黑分散性; Fisher 判别法; 假设检验; 样本判别函数

中图分类号: T Q330. 38⁺1 文献标识码: B 文章编号: 1000-890X(2000)09-0551-04

橡胶厂建立准确、快速的胶料中炭黑分散性的检测体系, 有利于提高橡胶制品质量和优化生产工艺。

目前, 国内普遍采用显微照像法检测胶料中炭黑的分散性, 但该法是靠人的眼睛进行比较定级, 不可避免地受到人的主观因素的影响。本工作运用计算机技术和数理统计法中的 Fisher 判别法对胶料中炭黑分散性进行统计并定级。

1 方法简述

Fisher 判别法是将被鉴定胶料的彩色照片经扫描仪按 1:1 的比例扫描成 256 色 8 位彩色的 bmp 图形文件, 存入计算机内, 由设计的软件“橡胶分散性的模式识别”对图形进行处理; 将存入电脑中的胶料切面的彩色图像进行灰度化处理, 转换成 256 色灰度图像; 输入适当的灰度值, 转换成黑白图像, 在黑白图像中, “白块粒子”的面积和数量分别与炭黑点的面积和数量密切相关, 用计算机图形学中的种子填充算法可计算出“粒子”面积值落在某一范围内的粒子个数与这些粒子的面积之和, 计算机将给出这些重要的数量指标; 最后运用 Fisher 判别法判

定胶料中炭黑的分散度等级。

在实际生产中, 绝大部分胶料中炭黑分散度等级处于三~六级, 在质检中, 分散度等级为三和四级的是不合格产品, 五和六级是合格产品, 故本工作也在此范围内考虑。按照国家标准, 炭黑分散度共分十级, 遇到分散度等级为一、二、七、八、九和十级的可以采用同样程序处理。

2 炭黑分散性的判别指标^[1]

衡量炭黑分散度大小的重要指标是未分散炭黑颗粒的大小和数量, 对不同炭黑分散度级别的胶料照片计算其炭黑团块面积(以像素点为单位), 着重考虑以下两类共 8 个指标的分析数据:

y_1 ——图片中面积介于 10~60 个像素点的炭黑团块的面积之和占整张图片面积的比例;

y_2 ——图片中面积介于 60~100 个像素点的炭黑团块的面积之和占整张图片面积的比例;

y_3 ——图片中面积介于 100~150 个像素点的炭黑团块的面积之和占整张图片面积的比例;

y_4 ——图片中面积大于 150 个像素点的炭黑团块的面积之和占整张图片面积的比例;

y_5 ——图片中面积介于 10~60 个像素点

基金项目: 广东省自然科学基金资助项目(99044)

作者简介: 王明兰(1962-)女, 河南南阳人, 华南师范大学讲师, 硕士, 从事数理统计与经济信息管理方面的教学工作。

的炭黑团块的个数占整张图片中面积大于9个像素点的炭黑团块个数的比例;

y_6 ——图片中面积介于60~100个像素点的炭黑团块的个数占整张图片中面积大于9个像素点的炭黑团块个数的比例;

y_7 ——图片中面积介于100~150个像素点的炭黑团块的个数占整张图片中面积大于9个像素点的炭黑团块个数的比例;

y_8 ——图片中面积大于150个像素点的炭黑团块的个数占整张图片中面积大于9个像素点的炭黑团块个数的比例。

为了方便,在进行数据处理时,将以上8个指标值同时乘以1000,使数据间的差异更加显著。本工作的数据处理均在SAS(Statistical

Analysis System)下完成。

3 Fisher判别法

待定级的彩色胶料照片,扫描成彩色图形文件后,经过以下5步,就可定出其炭黑分散度等级。

3.1 图形处理(灰度化,黑白化)^[2]

分散性判定的软件“橡胶分散性的模式识别”是在Windows 95操作系统下,用Microsoft的Visual Basic 4.0 for enterprise作为工具设计完成的,制成可执行文件后,在32位486以上机Windows 95环境下运行。

3.1.1 应用程序的处理流程

应用程序处理流程见图1。

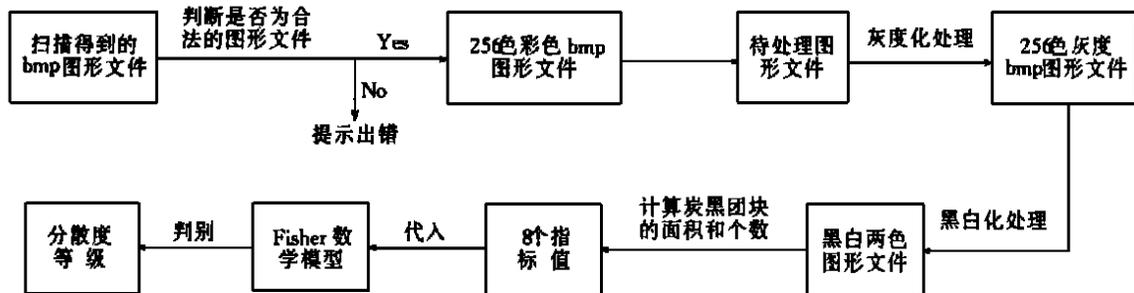


图1 应用程序处理流程图

3.1.2 图形处理的功能^[3]

(1) 判断文件格式

判断扫描后的文件是否为256色彩色图形文件,并确定其文件类型。这一功能可防止扫描选取了不符合要求的文件格式,避免文件无法在软件中处理的情况。

(2) 彩色文件的灰度化处理

将256色8位彩色的bmp图形文件转换成256色8位灰度bmp图形文件。

(3) 灰度文件的黑白化处理

在灰度图形文件中,灰度较浅的连通区域代表了原胶料中未分散的炭黑团块,为了易于计算机的识别,将代表炭黑团块的部分置为白色,其余部分置为黑色,即将256色灰度图形文件转换成黑白两色的图形文件。

(4) 分界值的确定

在灰度文件中,灰度值(反映该点灰色的深

浅程度)的范围为0~255。经反复尝试得出,当分界值定为180时,黑白化后图形中的白块可真实地反映原图形中炭黑团块的大小和数量。

3.2 计算8个指标值

要得到8个指标值,必须统计出被测图形中所有炭黑团块的面积和个数,这一步主要应用计算机图形学中的种子填充算法。种子填充算法是一种区域填充算法,用一种颜色或图案来填充一个二维区域。利用种子填充算法计算炭黑团块的面积和个数只需在填色的同时进行计数即可。

获得8个指标值后,需对表征橡胶分散性的数据进行标准化。由多个总体的Fisher判别法的理论可知,若4个级别的样本协方差矩阵相差甚远,则判定的效果很差;反之,若4个级别的样本协方差矩阵相等,则判别效果很好。

但实际上 4 个级别原始数据的样本协方差矩阵相差甚远。要做到严格的各样本协方差矩阵相等比较困难, 但能使它们尽量接近。为了做到这一点, 可对 $y_i (i=1 \sim 8)$ 进行标准化, 标准化后的 8 个指标值变为 x_i 。

3.3 代入 Fisher 线性判别函数^[4,5]

(1) 检验假设

设炭黑分散度等级分别为三级、四级、五级和六级的总体分别为 π_1, π_2, π_3 和 π_4 ; 样本协方差矩阵分别为 S_1, S_2, S_3 和 S_4 ; 样本个数 n_1, n_2, n_3 和 n_4 分别为 18, 87, 66 和 16。当不知道总体分布时, 可根据中心极限定理, 利用极限分布, 对总体参数作近似检验。

检验有两种假设, $H_0: S_1 = S_2 = S_3 = S_4$;
 $H_1: S_1, S_2, S_3, S_4$ 不全相等。设

$$\bar{x}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^{(k)} \quad (k = 1, 2, 3, 4)$$

$$A_k = \sum_{i=1}^{n_k} [x_i^{(k)} - \bar{x}^{(k)}][x_i^{(k)} - \bar{x}^{(k)}]'$$

$$A = A_1 + A_2 + A_3 + A_4$$

$$n = n_1 + n_2 + n_3 + n_4 = 187$$

于是检验假设的似然比统计量 λ 为:

$$\lambda = \frac{\left| \frac{A}{187} \right|^{-93.5}}{\prod_{k=1}^4 \left| \frac{A_k}{n_k} \right|^{-\frac{n_k}{2}}} = \frac{187^{748} \prod_{k=1}^4 |A_k|^{\frac{n_k}{2}}}{|A|^{93.5} \prod_{k=1}^4 n_k^{4n_k}}$$

以 λ 为基础的似然比检验不是无偏的, 但如果将 λ 中的 n_k 修改为 $n_k - 1$, 从而使 187 变成 183, 经过修改后的 λ 记为 λ^0 , 即

$$\lambda^0 = \frac{183^{732} \prod_{k=1}^4 |A_k|^{\frac{n_k-1}{2}}}{|A|^{91.5} \prod_{k=1}^4 (n_k - 1)^{4(n_k-1)}}$$

以 λ^0 为基础的检验是无偏的。由显著水平 $\alpha = 0.05$ 的 x^2 分布表可得, 不能拒绝 H_0 , 即这 4 个总体 (三级、四级、五级、六级) 的协方差矩阵没有显著差异。

(2) Fisher 判别法的思想方法

设总体 π_1 的 18 个样本为 x_{1j} , 总体 π_2 的 87 个样本为 x_{2j} , 总体 π_3 的 66 个样本为 x_{3j} ,

总体 π_4 的 16 个样本为 $x_{4j} (j=1, 2, \dots, n_k)$, 以 x_{kj} 作为 x_k 的列, 即 $x_{kj} = (x_{k1}, x_{k2}, \dots, x_{kn_k})$, 那么, 对于第 k 个总体, 样本均值 \bar{x}_k 和样本协方差矩阵 S_k 分别为:

$$\bar{x}_k = \sum_{j=1}^{n_k} x_{kj} / n_k$$

$$S_k = \sum_{j=1}^{n_k} (x_{kj} - \bar{x}_k)(x_{kj} - \bar{x}_k)' / (n_k - 1)$$

总的均值向量 \bar{x} 为

$$\bar{x} = \sum_{k=1}^4 n_k \bar{x}_k / \left(\sum_{k=1}^4 n_k \right)$$

样本组间差矩阵 B_0 为

$$B_0 = \sum_{k=1}^4 (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})'$$

样本组内差矩阵 W 为

$$W = \sum_{k=1}^4 (n_k - 1) S_k$$

Fisher 的思想是: 变换多元观测 x 到二元或三元观测 x' , 使得从总体 π_k 和 π_j 产生的 x 尽可能地分离开; 为了便于数学上处理, 应用上简便, 把 x' 取做 x 的线性组合。

事实上, $S_p = W / (n_1 + n_2 + n_3 + n_4 - 4)$ 为 S 的无偏估计, 因 W 是 S_p 的一个倍数,

$\max \frac{l'B_0 l}{l'S_p l}$ 与 $\max \frac{l'B_0 l}{l'W l}$ 就有相同的解 l , 并且若 $W^{-1} B_0 \hat{e} = \lambda \hat{e}$, 就有:

$$S_p^{-1} B_0 \hat{e} = (n_1 + n_2 + n_3 + n_4 - 4) \lambda \hat{e}$$

因此, 我们可用 $W^{-1} B_0$ 的特征向量 \hat{e}_k 来表示极值问题的解。

进而可获得 $W^{-1} B_0$ 的特征值与特征向量; 因为 $S = 3$, 故取其特征向量为:

$$\hat{e}_1 = (0.009\ 842\ 09, -0.082\ 086, 0.053\ 650\ 2, \\ -0.021\ 413\ 9, 0.795\ 88, 0.376\ 194, \\ 0.233\ 469, 0.400\ 467)$$

$$\hat{e}_2 = (0.090\ 588\ 7, -0.447\ 224, 0.705\ 953, \\ 0.000\ 537\ 334, -0.136\ 002,$$

$$0.006\ 772\ 86, -0.521\ 76, -0.051\ 373\ 3)$$

$$\hat{e}_3 = (0.117\ 625, 0.621\ 598, -0.549\ 696,$$

$$0.017\ 743\ 8, 0.309\ 533, -0.326\ 285,$$

$$0.225\ 774, 0.209\ 887)$$

则: $l_1 = \hat{e}_1$, 使

$$\frac{I' B_0 I}{I' W I} = \frac{I' \left[\sum_{k=1}^4 (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})' \right] I}{I' \left[\sum_{k=1}^4 \sum_{j=1}^{n_k} (x_{kj} - \bar{x}_k)(x_{kj} - \bar{x}_k)' \right] I}$$

达到最大;线性组合 $I_1' x = \hat{e}_1' x$ 叫做第一样本判别函数,线性组合 $I_2' x = \hat{e}_2' x$ 叫做第二样本判别函数,线性组合 $I_3' x = \hat{e}_3' x$ 叫做第三样本判别函数,故得分散性判定的 Fisher 样本判别函数为:

$$g_1 = 8.53 \times (0.009\ 842\ 09x_1 - 0.082\ 086x_2 + 0.053\ 650\ 2x_3 - 0.021\ 413\ 9x_4 + 0.795\ 88x_5 + 0.376\ 194x_6 + 0.233\ 469x_7 + 0.400\ 467x_8)$$

$$g_2 = 3.6 \times (0.090\ 588\ 7x_1 - 0.447\ 224x_2 + 0.705\ 953x_3 + 0.000\ 537\ 334x_4 - 0.136\ 002x_5 + 0.006\ 772\ 86x_6 - 0.521\ 76x_7 - 0.051\ 373\ 3x_8)$$

$$g_3 = 2.81 \times (0.117\ 625x_1 + 0.621\ 598x_2 - 0.549\ 696x_3 + 0.017\ 743\ 8x_4 + 0.309\ 533x_5 - 0.326\ 285x_6 + 0.225\ 774x_7 + 0.209\ 887x_8)$$

g_1, g_2 和 g_3 三个函数中的 $x_1 \sim x_8$ 是原始指标值 $y_1 \sim y_8$ 经标准化后所得数据。

3.4 代入 Fisher 判别法则

Fisher 判别法则:

$$F_3 = \text{SQRT}[(g_1 + 0.004\ 566\ 8)^2 + (g_2 + 0.003\ 064\ 6)^2 + (g_3 - 0.000\ 507\ 426)^2]$$

$$F_4 = \text{SQRT}[(g_1 - 0.011\ 074\ 9)^2 +$$

$$(g_2 + 0.000\ 968\ 732)^2 + (g_3 - 0.000\ 914\ 210)^2]$$

$$F_5 = \text{SQRT}[(g_1 + 0.002\ 152\ 1)^2 + (g_2 - 0.001\ 341\ 5)^2 + (g_3 - 0.000\ 792\ 879)^2]$$

$$F_6 = \text{SQRT}[(g_1 - 0.003\ 518\ 2)^2 + (g_2 + 0.001\ 133\ 4)^2 + (g_3 + 0.000\ 939\ 147)^2]$$

3.5 定级

比较 F_3, F_4, F_5 和 F_6 的大小,取 $\min\{F_3, F_4, F_5, F_6\}$ 。若 $\min\{F_3, F_4, F_5, F_6\} = F_m$ ($m = 3, 4, 5, 6$), 则判定该张图片为 m 级。

4 结语

利用 Fisher 判别法进行胶料中炭黑分散度等级的确定,准确、快速,并杜绝了显微照相法中人为因素的影响,客观性强,是胶料中炭黑分散性判定的有效方法。

参考文献:

- [1] 姜启源. 数学模型[M]. 北京: 高等教育出版社, 1993. 10-120.
- [2] 赵胜利. Visual Basic 3.0 for Windows 编程实用手册[M]. 北京: 人民邮电出版社, 1997. 64-137.
- [3] 孙家广. 计算机图形学[M]. 北京: 清华大学出版社, 1994. 92-280.
- [4] 王学仁, 王松桂. 实用多元统计分析[M]. 上海: 上海科学技术出版社, 1996. 346-390.
- [5] 孙文爽. 多元统计分析[M]. 北京: 高等教育出版社, 1994. 153-181.

收稿日期: 2000-04-01

2001年《橡胶参考资料》征订启事

《橡胶参考资料》创刊于1971年,由西北橡胶塑料研究设计院主办。主要报道世界发达国家和国内橡胶材料及制品的配合技术、加工工艺、物理测试、化学分析、机械设备及理论研究。内容涵盖轮胎、胶管胶带及工业橡胶制品等。

2001年将重点报道新材料开发与应用、新工艺、功能橡胶及工程橡胶研究及其专利技术。

《橡胶参考资料》为月刊,大16K,全年12期共收费100元。

地址: 陕西省咸阳市12号信箱; 邮编: 712023; 电话: (0910) 3319370; 传真: (0910) 3319360; E-mail: xbyfsd@mail.sn.cninfo.net; 帐号: 20624914885; 开户行: 陕西省咸阳市工人西办; 收款单位: 西北橡胶塑料研究设计院《橡胶参考资料》编辑部; 联系人: 高桂芝。