

文章编号: 1003-8701(2015)01-0100-04

一种改进的基于回归分析的人参智能烘干算法

李 健¹, 李晓阳¹, 李俊杰¹, 陈桂芬^{1*}, 廖梦兰², 金淑燕³

(1. 吉林农业大学信息技术学院, 长春 130118; 2. 吉林大学数学研究所, 长春 130012;
3. 白山林村中药开发有限公司, 吉林 靖宇 135200)

摘 要: 回归分析算法在工程学、农学等研究领域有着较为广泛的应用。本文基于回归分析算法, 在已有的人参烘干算法的基础上对人参烘干进行相关实验, 并对已有算法加以改进, 提出了多因素混合模型。通过与已有算法比较, 本文所提出的多因素混合模型烘干实验效果更优、优质品率更高。该模型算法及实验结果为人参烘干算法的研究做出了有益的尝试, 同时也为人参烘干工艺的研究积累了一定经验。

关键词: 回归分析; 多因素混合模型; 烘干算法

中图分类号: S11

文献标识码: A

DOI:10.16423/j.cnki.1003-8701.2015.01.023

A Improved Intelligent Ginseng Drying Model Based on Regression Analysis

LI Jian¹, LI Xiao-yang¹, LI Jun-jie¹, CHEN Gui-fen^{1*}, LIAO Meng-lan², JIN Shu-yan³

(1. College of Information Technology, Jilin Agricultural University, Changchun 130118;

2. Institute of Mathematics, Jilin University, Changchun, 130012;

3. Baishan Lin Village Medicine Development Co, Ltd. Jingyu 135200, China)

Abstract: Regression analysis algorithm is widely used in engineering, botany and other research fields. Based on regression analysis algorithm, a multivariate mixed model was proposed in the paper on the basis of the existing algorithm and some improvement based on related experiment. Comparing with existing algorithms, new multivariate mixed model presented in the paper bring better drying test results and higher quality rate. The model algorithm and experimental results made a good attempt for the study of ginseng drying algorithm, and it accumulated some experience for the study of ginseng drying process.

Keywords: Regression analysis; Multivariate mixed model; Drying algorithm

回归分析(regression analysis)^[1-4]起源于19世纪末期, 是确定两种或两种以上变量之间相互依赖的定量关系的一种最常用的统计及分析方法。近年来, 回归分析算法已经在经济学、工程学、植物学等研究领域得到了广泛应用。至目前为止, 学者们已经提出了一些关于人参烘干的基本模型。

1988年, 李云飞提出了西洋参烘干速率的数

学模型^[5], 建立了烘干速率与干球温度和主根直径的数学模型。1991年, 王修兰等通过对人参、西洋参的干燥特性的研究, 给出了在烘干人参时的温湿度关系模型^[6]。2004年, 纪庆晓通过对西洋参的快速干燥技术的研究, 提出了三次方程模型^[7]。但是这些模型中, 对温度、湿度、风速只进行单独考虑, 没有考虑其混合影响。因此上述模型会造成实验结果误差较大, 表达不准确。

本文为减少上述模型带来的较大误差, 在纪庆晓提出的单因素三次方程模型的基础上, 分析了其模型中存在的对烘干过程中温度、湿度以及风速只进行单独考虑的问题, 提出了基于回归分析算法的温湿度、风速以及时间的混合因素模型。在混合因素模型中, 对已有的单因素三次方程模型进行改进, 采用了回归分析算法中的多元非线性模型, 在多元线性回归的基础上加入混合

收稿日期: 2014-08-20

基金项目: 吉林省科技发展计划项目(重大科技攻关项目)(20140204045NY); 吉林省世界银行贷款项目(2011-Z20); 吉林省科技发展计划项目(20140204045NY); 吉林省教育厅“十二五”科学技术研究项目(2014第468号)

作者简介: 李 健(1981-), 男, 讲师, 博士, 研究方向: 物联网、空间推理。

通讯作者: 陈桂芬, 女, 教授, 博士生导师, E-mail: guifchen@163.com

项,来弥补已有模型中的不足。通过模型的比较得出结论:多因素混合模型的算法是切实可行的,并且该模型比之前处理人参烘干问题已有算法更优。

1 单因素三次方程模型

近年来,关于人参烘干应用最为广泛的模型就是单因素三次方程模型。该模型算法是通过控制烘干过程中的一个因素(温度、风速、湿度中的一个),对烘干过程进行模拟控制。由于人参含水量在这些因素的影响下随时间变化呈现出一元三次多项式的形式,所以考虑用一元三次多项式进行拟合。其一元三次多项式的一般形式如下。

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 + b \quad (i=1,2,3) \cdots (1)$$

其中 x_i 为烘干时的各个环境参数, y_i 是烘干过程中的含水量, a_0 、 a_1 、 a_2 、 a_3 为 x^0 、 x^1 、 x^2 、 x^3 项的参数, b 为误差参数。

纪庆晓通过大量的实验研究,得到的模型结果如下。

$$Y = 77.112 - 0.352X + 5.891 \times 10^{-4} X^2 - 3.561 \times 10^{-7} X^3 \cdots (2)$$

单因素三次方程模型,是通过回归分析,分别考虑温度、风速以及湿度提出的,较前人已经有了很大的提升。但是在该模型算法中对温度、湿度、风速只进行单独考虑,没有考虑其混合影响。针对上述模型的不足,下面将建立适用于复杂环境下的混合因素模型。

2 混合因素模型

2.1 混合因素模型的建立

通过对烘干工艺的研究^[8],了解到在烘干工艺中,最主要的就是对环境参数的控制。在烘干室中,环境参数包括温度、风速以及湿度。又由于烘干过程是一个含水量随时间变化的过程,因此要想建立一个较为精确的烘干算法就需要考虑时间、温度、风速以及湿度。为了方便说明,首先要定义一些符号。

定义1:记烘干过程中人参含水量的变化量为 ΔM ; 人参烘干时间为 h ; 人参烘干过程中的温度参数值、风速参数值、湿度参数值分别为 x_1 、 x_2 、 x_3 。

控制 ΔM 是整个烘干过程的关键。如果在某段时间内 ΔM 过高,代表原材料内部水分挥发过快,可能会引起碳化问题;如果某段时间 ΔM 较低,就说明原材料内部水分没有尽快挥发,就会

引起人参霉变。因此控制 ΔM 至关重要。

根据上述分析,就可以将人参烘干的实际问题抽象为在温度、湿度、风速的作用下使 ΔM 趋于平稳的数学问题。尽管实际问题中温度、湿度、风速之间存在着一定的客观联系,但是对其具体的内在联系还不甚了解,因此考虑采用只基于数据的回归分析模型对 $\Delta M = f(x_1, x_2, x_3)$ 进行求解,则 ΔM 为:

$$\Delta M = a_{11}x_1 + a_{22}x_2 + a_{33}x_3 + a_{12}x_1x_2 + \cdots + a_{13}x_1x_3 + a_{23}x_2x_3 + b \quad (3)$$

其中 a_{11} 、 a_{12} 、 a_{13} 、 a_{22} 、 a_{33} 、 a_{23} 为回归系数, b 为误差参数。

2.2 混合因素模型的求解

在本文中,所采用的模型是一个多元线性回归分析模型。其一般形式为:

$$y = b_0 + \sum_{i=1}^k b_i x_i + \varepsilon; \quad \varepsilon: N(0, \sigma^2)$$

其中 y 是因变量, x_i 为 k 个自变量, b_0 、 b_1 、 L 、 b_k 为 $k+1$ 个待估参数,定义随机误差 ε 满足正态分布。当因变量 y 与单个自变量 x_i 不呈现依赖关系时,由多元线性回归分析的一般形式衍生出多元相互作用项回归模型,其一般形式为:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + L + \varepsilon$$

即本文中的多因素混合模型。此模型中,最主要的工作就是求解出自变量 x_i 前的待估参数。

在多元线性回归分析模型中求解待估参数的方法多种多样,由于本文中实验数据量较小,现选择最小二乘法来求解待估参数。

最小二乘法中的求解形式为:

$$Q = \sum (y - y_c)^2 \cdots (4)$$

其中 y 为模型所得结果, y_c 为实验的实际结果。当 Q 取得最小值时,说明模型结果 y 与实际结果 y_c 最为接近,模型与实际相符。

通过模型的建立,已知:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + L + \varepsilon \cdots (5)$$

所以 Q 又可以表示成为如下公式:

$$Q = \sum (b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + L + \varepsilon - y_c)^2 \cdots (6)$$

在公式(6)中,要求使得 Q 最小。因此可以对待估参数 b_i 分别求偏导数,当 $\frac{\partial Q}{\partial b_i} = 0$ 时,此时 Q 可以取得最小值。因为在此算法中 x_i 和 y_c 都是已知,所以可以建立关于 b_i 的方程组,通过解方程组的方法来求解待估参数 b_i 。

求解出 b_i 后,就可以得出多元线性回归分析

模型 y 。但是模型 y 与实际值 y_c 不是一定存在相关关系,故需要通过求解相关系数 R 来验证其相关性。

下面具体给出 R 的求解方法,首先给出如下定义:

定义 2: 记相关系数为 $R = \frac{\sigma^2_{xy}}{\sigma_x \cdot \sigma_y}$, 其中 $\sigma^2_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$, 是 x, y 的协方差; $\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$, 是 x 的标准差; $\sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}$, 是 y 的标准差。

通过协方差和标准差公式,可将 R 化简成如下形式:

$$R = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \cdot \sqrt{\sum (y - \bar{y})^2}} \quad \dots\dots\dots (7)$$

当 R 值越大时,则说明模型 y 与实际值 y_c 的相关关系越强。

通过将实际模型带入上述方法中,就可以求解出 a_{11} 、 a_{12} 、 a_{13} 、 a_{22} 、 a_{33} 、 a_{23} , 此时可以得到如下公式:

$$Q = \sum (a_{11}x_1 + a_{22}x_2 + a_{33}x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + b - y_c)^2 \quad \dots\dots\dots (8)$$

再通过方程组:

$$\frac{\partial Q}{\partial a_{11}} = \frac{\partial Q}{\partial a_{22}} = \frac{\partial Q}{\partial a_{33}} = \frac{\partial Q}{\partial a_{12}} = \frac{\partial Q}{\partial a_{13}} = \frac{\partial Q}{\partial a_{23}} = 0 \quad \dots\dots\dots (9)$$

可以求解出 $f(x_1, x_2, x_3)$ 的模型结果如下:

$$\Delta M = -0.0125x_1 - 0.8268x_2 + 4.4345x_3 - 0.0008x_1x_2 + 0.0006x_1x_3 + 0.0005x_2x_3 - 0.0708 \quad \dots\dots\dots (10)$$

根据以上结果,并通过实验数据,计算出 $R = 0.641$, 可说明此模型与实际结果显著相关。

3 模型的比较

3.1 模型算法的比较

根据上述模型的建立过程,编写程序可得到多因素混合模型的结果图,如图 1 所示。图中横轴代表时间,纵轴为 ΔM 的变化量。在整个烘干过程中 ΔM 的变化区间为 0.5335 ~ 0.5625, 变化量为 0.029。

根据单因素三次方模型,可编写程序得到单因素三次方模型的结果图,如图 2 所示。图中横轴为时间,纵轴为 ΔM 的变化量。在整个烘干过程中 ΔM 的变化区间为 0.025 ~ 0.35, 变化量为 0.325。

通过程序可以得到两个模型结果的比较图,如图 3 所示。虚线为多因素混合模型的结果,实

线为单因素三次方程模型的结果。

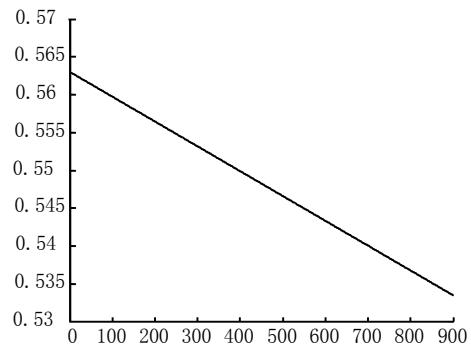


图 1 多因素混合模型结果图

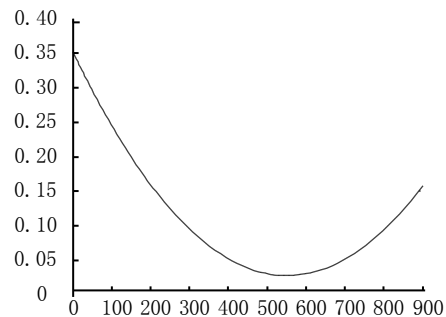


图 2 单因素三次方模型的结果图

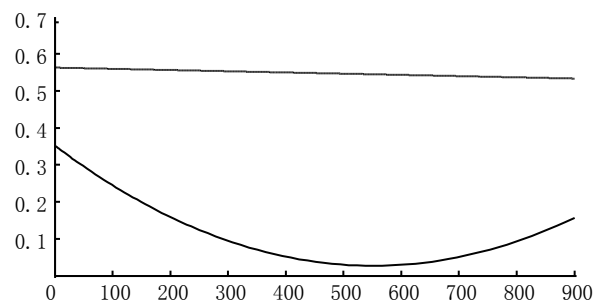


图 3 两个模型结果的比较图

从图 3 中可以直观看出,多因素混合模型中 ΔM 值较大,而且更趋于平稳。此结果说明在相同的时间下,采用多因素混合模型,水分蒸发量更大,烘干效率较高;在蒸发过程中 ΔM 变化范围小,有效地防止了水分蒸发不均匀造成的人参碳化和霉变问题的出现。

3.2 模型的实验比较

根据建立的烘干算法,进行具体的烘干实验,来验证改进后算法的准确性。

具体步骤如下:

(1)测定人参干重;(2)选定实验工况环境,使温度、湿度、风速达到稳定状态;(3)对其他实验材料进行称重,以减小误差;(4)将人参置于烘干箱中,并将温度传感器固定在实验人参的三个不同位置,测定人参烘干时的温度;(5)实验开

始:计时、称重、测定温度、湿度,每隔 10 min 测量 1 次。在称重时,需关闭风扇、保持在静态下对人参进行称重,这样可以避免因气流浮力造成的误差。当人参重量不再降低时,即达到平衡水平,结束实验。

在此实验中,具体运用两种不同算法分别对人参进行烘干实验。通过实验可以得到具体实验结果如表 1。

表 1 实验结果

实验组	含水量(%)	时间(min)
1	100	0
	80	30
	60	60
	20	600
	15	900
2	98	0
	75	30
	61	60
	18	600
	14	900
3	99	0
	85	30
	78	60
	25	600
	17	900
4	98	0
	84	30
	70	60
	30	600
	19	900

表中 1、2 组实验采用了多因素混合模型,3、4

组实验采用了单因素三次方程模型。从实际烘干实验中可以看出,实验 1、2 水分蒸发更快可以防止霉变,中间过程较平稳,可以防止碳化,烘干结束后含水量在 14%~15%,烘干效果更优。通过计算,人参的优质品率提高 2%~3%。

4 结 论

本文通过对单因素三次方程模型进行改进,得到了可以对人参进行更好地烘干的多因素混合模型。通过模型的比较,可以得到在相同的时间下,采用多因素混合模型,水分蒸发量更大,烘干效率较高,并且多因素混合模型能有效地防止水分蒸发不均匀造成的人参碳化和霉变问题的出现。根据具体实验结果,进一步说明该模型比之前处理人参烘干问题的算法效果更优。

参考文献:

[1] 王惠广. 回归分析与相关分析[J]. 经济管理学报, 1997 (3):31-32 .

[2] 姜海波,李艳茹. 通用回归分析模型[J]. 数学的实践与认识, 2006, 36(1): 174-180 .

[3] 杨文泉. 模糊回归分析研究综述[D]. 长春: 吉林大学, 2009 .

[4] 李排昌. 矩阵的子矩阵与线性模型的假设检验[J]. 吉林大学学报(理学版), 2002, 40(3): 252-254 .

[5] 李云飞. 西洋参烘干速率的数学模型[J]. 吉林农业大学学报, 1988, 10(1): 73-78 .

[6] 王修兰,徐师华. 人参、西洋参干燥特性的研究[J]. 农业工程学报, 1991, 7(1): 62-69 .

[7] 纪庆晓. 西洋参快速干燥技术研究[D]. 长春: 吉林大学, 2006 .

[8] 王茂学,陈侍良,王绍林. 人参微波干燥试验[J]. 北京农业工程大学学报, 1993, 13(1): 44-51 .



(上接第 99 页)

[14] Bouazzaoui K, LaPointe G. Use of antisense RNA to modulate glycosyltransferase gene expression and exopolysaccharide molecular mass in *Lactobacillus rhamnosus*[J]. Journal of Microbiological Methods, 2006, 65(2): 216-225 .

[15] 赵玉娟,李盛钰,牛春华,等. 鼠李糖乳杆菌 JAAS8 胞外多糖聚合和转运相关基因的克隆及生物信息学分析[J]. 基因组学与应用生物学, 2010, 29(3): 447-452 .

[16] 冀林立. 传统乳制品中产 γ -氨基丁酸乳酸菌的筛选、鉴定及发酵条件优化[D]. 呼和浩特: 内蒙古农业大学, 2008 .

[17] Mazur R, Kovalovská K, Hudec J. Changes In Selectivity Of Gamma-Aminobutyric Acid Formation Effected By Fermentation Conditions And Microorganisms Resources[J]. Journal of Microbiology, Biotechnology and Food Science, 2011, 1(2): 164-171 .

[18] Park K B, Oh S H. Production of yogurt with enhanced levels of

gamma- aminobutyric acid and valuable nutrients using lactic acid bacteria and germinated soybean extract[J]. Bioresource Technology, 2007, 98(8): 1675-1679 .

[19] Nejati F, Rizzello C G, Di Cagno R, et al. Manufacture of a functional fermented milk enriched of Angiotensin-I Converting Enzyme(ACE)-inhibitory peptides and γ -amino butyric acid (GABA)[J]. LWT - Food Science and Technology, 2013, 51(1): 183-189 .

[20] Nomura M, Kimoto H, Someya Y, et al. Production of gamma-aminobutyric acid by cheese starters during cheese ripening[J]. Journal of Dairy Science, 1998, 81(6): 1486-1491 .

[21] Wang H, Dong C, Chen Y, et al. A new probiotic cheddar cheese with high ACE-inhibitory activity and γ -aminobutyric acid content produced with koumiss- derived *Lactobacillus casei* Zhang [J]. Food Technology and Biotechnology, 2010, 48(1): 62-70 .