

李 健,王海明,李 勇,等. 基于多元线性回归与 BP 神经网络分析南疆巴旦木结构物理特性[J]. 江苏农业科学,2019,47(12):260-264.  
doi:10.15889/j.issn.1002-1302.2019.12.060

# 基于多元线性回归与 BP 神经网络分析 南疆巴旦木结构物理特性

李 健,王海明,李 勇,周 岭

(塔里木大学机械电气化工程学院,新疆阿拉尔 843300)

**摘要:**运用多元线性回归与 BP 神经网络等多种数理统计手段对南疆巴旦木结构物理特性进行研究。基于多元线性回归及方差分析方法,对南疆巴旦木结构外形进行定义和物理参数测定,运用 MATLAB 软件得出,巴旦木长、宽、缝三者之间不是简单线性关系,其外形近似为关于横、竖、纵 3 椭圆截面对称的扁壳体;在测定对应的巴旦木果仁质量后,运用 3 层 BP 神经网络结构寻找巴旦木结构外形与果仁质量间的关系,建立相应质量预测模型。可在不破坏巴旦木外壳的情况下,通过长、宽、缝来预测巴旦木果仁的质量,从而为巴旦木无损品质分级及破壳装置设计提供理论基础。

**关键词:**MATLAB;多元线性回归;BP 神经网络;巴旦木;结构物理特性

**中图分类号:**O4-34;TB303 **文献标志码:**A **文章编号:**1002-1302(2019)12-0260-05

巴旦木,蔷薇科,其果仁富含蛋白质、脂肪、淀粉、单糖、维生素等,营养价值极高,为世界著名干果,国际市场销量很大<sup>[1-2]</sup>。我国巴旦木的主栽区在南疆喀什地区,此外,和田、阿克苏等地也有栽培<sup>[2-3]</sup>。目前,我国巴旦木年产量偏少,每年仍需从美国大量进口。消费缺口的存在促进了国内巴旦木种植和加工业的发展,潜在经济效益较为显著<sup>[3-4]</sup>。

我国巴旦木加工一直以人工为主,效率低且品质难以保证。食品安全以及人们对巴旦木果仁产品多样的要求,正强力推动着巴旦木深加工的机械化进程,然而目前国际上尚无成熟的巴旦木破壳、分级装备应用于生产<sup>[5-7]</sup>。目前国内坚果破壳、分级设备的设计主要依靠经验与试验来进行,缺乏对结构物理特性及破壳力学机制的系统性研究<sup>[8-9]</sup>。因此,分析研究巴旦木的结构物理特性,为巴旦木分级、破壳设备的优化设计提供理论依据与关键参数,就显得尤为重要与迫切。

本研究运用 MATLAB 软件与数理统计原理对南疆巴旦木壳体的结构外形及物理参数进行深入细致的分析。在采集大样本的基础上,对巴旦木的结构物理特性进行测量,研究巴

旦木长、缝、宽彼此间存在的非线性关系,构建壳体几何模型,并探寻巴旦木果仁质量与巴旦木壳体外形间的关系,运用 BP 神经网络建立相应质量预测模型,以期在巴旦木加工产业化过程中,为相关机械结构的参数设计与校正提供参考。

## 1 材料与方法

### 1.1 材料与仪器

南疆巴旦木,由于土壤、气候、混种等原因,其品种非常复杂,主要品种有鹰嘴、纸皮、克西、双果、米森等,大小、形状、壳厚不一(图 1),无法通过植物学分类进行甄别,也无任何标准可查。因其主产地在南疆喀什,这里选取喀什地区市场销巴旦木作为研究对象。



图1 大小、形状、壳厚不一的南疆巴旦木

主要仪器设备包括游标卡尺:0~150 mm,测量精度为 0.01 mm,上海安亭科学仪器厂生产;面包体积测定仪:JMTY 型,测量范围为 0~1 000 mL,重复性误差≤1.0%,重现性≤0.8%,河北省虹宇仪器设备有限公司生产;万能材料试验机:WD-D3-7 型,上海卓技仪器设备有限公司生产;电子

收稿日期:2018-03-05

基金项目:国家自然科学基金(编号:11762020);塔里木大学校长基金(编号:TDZKQN201609)。

作者简介:李 健(1985—),男,山西大同人,硕士,讲师,研究方向为农业机械工程力学。E-mail:hmwang16@163.com。

通信作者:王海明,硕士,讲师,研究方向为机械设计及理论。E-mail:wanghaiming23@163.com。

中的应用[J]. 华北农学报,2017,32(增刊1):239-244.

[17]谭方利,邱 坤,杨 鹏,等. 新型能源在烟叶烘烤中应用前景和效果分析[J]. 天津农业科学,2018,24(1):59-63.

[18]王国庆,聂荣邦,周孚美. 不同热源密集烤房烟叶烘烤效果研究[J]. 作物研究,2016,30(7):782-785.

[19]王建安,刘国顺. 生物质燃烧锅炉热水集中供热烤烟设备的研制及效果分析[J]. 中国烟草学报,2012,18(6):32-37.

[20]李世军. 烤烟用生物质颗粒燃烧机及其控制方法研究[J]. 中国农机化学报,2018,39(5):43-47.

[21]倪克平,甄焕菊. 生物质压块燃料在烟叶烘烤中的应用效果[J]. 农业开发与装备,2015(11):63,80.

[22]谭方利,樊士军,董艳辉,等. 生物质压块燃料及煤炭燃料在烟叶烘烤中的应用效果对比研究[J]. 现代农业科技,2014(10):201-201,209.

天平:FA1104 型,测量精度为 0.001 g,上海安亭科学仪器厂。

1.2 试验方法

1.2.1 巴旦木结构特征测量 依据大数定律和中心极限定理,当样本数据足够大时,其分布形式必为正态分布。这里随机选取 2 000 个巴旦木进行筛选,剔除有病害、畸形、已破壳的巴旦木,将剩余的 1 412 个作为试验样品。使用游标卡尺按图 2 所示分别测量各个试验样品的横向、缝向、纵向尺寸,同时使用面包体积测定仪测量其体积。

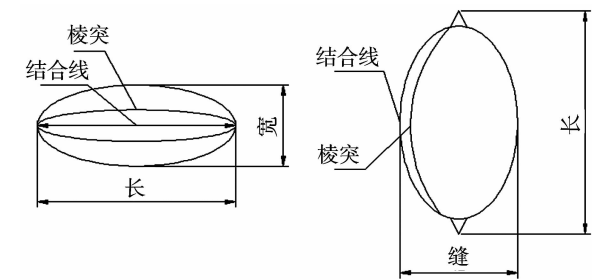


图2 南疆巴旦木结构示意图

1.2.2 巴旦木果仁质量测量 利用万能材料试验机对巴旦木破壳取仁,试验机加载速度为 240 mm/min,预加载量 30 mm/min,之后使用电子天平称取巴旦木果仁质量。

1.2.3 数据处理 试验数据处理使用 Excel 与 MATLAB2015b 软件。

2 基于 MATLAB 多元线性回归与方差分析构建南疆巴旦木几何模型

2.1 巴旦木结构物理特性的回归分析

这里选取南疆巴旦木的几何外形为研究对象,长、缝、宽各自的样本容量分别为 1 412 个。调用 MATLAB 中的  $H = kstest(X)$  函数分别对长、缝、宽各自的数据(表 1)进行正态分布检验,测试水平  $\alpha$  为 5%,其中  $H$  为正态分布检验值。将长、缝、宽各自的数据分别代入  $X$  中,在 MATLAB 中的运行  $kstest(X)$  函数得  $H = 0$ ,因此认为长、缝、宽各自均服从正态分布。

对所研究对象的三维尺寸数据进行比较发现,巴旦木彼此间的长、缝、宽数据存在相互交错的现象,但对单个巴旦木而言,依然存在着长 > 缝 > 宽的规律。为进一步明确巴旦

表 1 南疆巴旦木样本数据

序号	长 (mm)	缝 (mm)	宽 (mm)	仁质量 (g)	实测体积 (mm <sup>3</sup> )	预测体积 (mm <sup>3</sup> )
1	23.52	16.76	12.48	1.081	28 650.91	20 606.99
2	25.08	15.98	11.58	1.032	29 576.61	19 440.23
3	22.98	15.54	12.96	1.038	26 011.46	19 386.29
4	24.24	15.84	11.86	1.094	23 710.25	19 074.85
5	21.50	15.5	12.32	1.058	28 186.62	17 197.66
6	23.60	15.72	11.38	1.127	31 368.40	17 684.61
7	25.70	17.02	12.86	1.228	29 097.36	23 562.55
8	26.68	17.66	12.34	1.043	24 840.65	24 354.56
9	23.20	16.44	12.10	1.294	32 519.99	19 331.42
10	21.48	16.42	11.78	1.160	28 068.20	17 403.69
11	21.22	15.72	12.12	1.181	33 447.26	16 935.15
12	26.20	15.44	12.28	1.036	22 626.18	20 808.25
13	19.46	15.78	12.52	1.054	39 851.57	16 104.33
14	23.52	15.84	12.34	1.020	24 095.44	19 257.34
15	26.02	16.02	11.84	1.081	28 650.91	20 606.99
16	26.02	16.02	11.84	1.097	28 451.63	20 673.31
17	22.06	15.84	11.92	1.253	25 214.75	17 447.19
18	22.24	16.42	11.82	1.310	38 741.18	18 080.65
19	22.66	15.70	11.80	1.015	28 714.81	17 584.51
20	21.38	16.02	11.08	1.031	33 883.78	15 896.39
21	20.72	15.46	12.98	1.224	26 543.03	17 416.57
22	21.34	13.56	12.42	1.190	39 396.02	15 054.43
23	20.22	14.22	11.78	1.086	21 429.37	14 187.79
24	20.78	15.02	11.96	1.296	37 458.21	15 636.35
25	22.32	15.96	11.78	1.158	37 026.03	17 577.66
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1 398	27.64	22.12	17.44	1.160	31 786.82	44 664.07
1 399	24.72	17.52	16.06	1.095	35 352.10	29 135.11
1 400	25.04	19.62	18.02	1.226	25 372.33	37 083.16
1 401	24.22	20.28	17.54	1.050	53 230.05	36 087.79
1 402	28.44	21.36	17.88	1.091	21 890.79	45 497.44
1 403	25.22	20.72	18.32	1.199	30 131.17	40 100.42
1 404	29.62	21.38	15.58	1.122	44 364.41	41 328.42
1 405	25.02	18.24	16.38	1.177	33 322.04	31 312.28
1 406	26.72	19.24	15.02	1.117	26 612.45	32 344.47

续表 1

序号	长 (mm)	缝 (mm)	宽 (mm)	仁质量 (g)	实测体积 (mm <sup>3</sup> )	预测体积 (mm <sup>3</sup> )
1 407	23.88	19.96	18.04	1.018	27 759.95	36 018.03
1 408	25.12	19.44	15.98	1.139	26 133.62	32 687.47
1 409	23.74	18.86	15.96	1.171	46 601.16	29 932.56
1 410	24.78	19.52	17.76	1.051	36 834.78	35 984.27
1 411	23.74	18.84	16.52	1.206	26 507.76	30 949.97
1 412	25.82	19.52	18.02	1.318	28 313.89	38 043.41

木三维尺寸间的关系,对长、缝、宽各数据间的相关性进行深入研究,并作回归分析。首先对长、缝、宽数据两两间的关系进行回归分析,从图 3 可以看出,长、缝、宽三者彼此间存在一定的正相关关系,其相关系数分别为  $r_1=0.994\ 8$  (长-缝)、 $r_2=0.971\ 2$  (长-宽)、 $r_3=0.962\ 2$  (缝-宽),相关系数均大于 0.95,表明长、缝、宽数据两两间的线性关系相当高。这一点与标准椭球体长、缝、宽间的分布关系极为相符,因此可认为南疆巴旦木外形与标准椭球壳有相似之处。

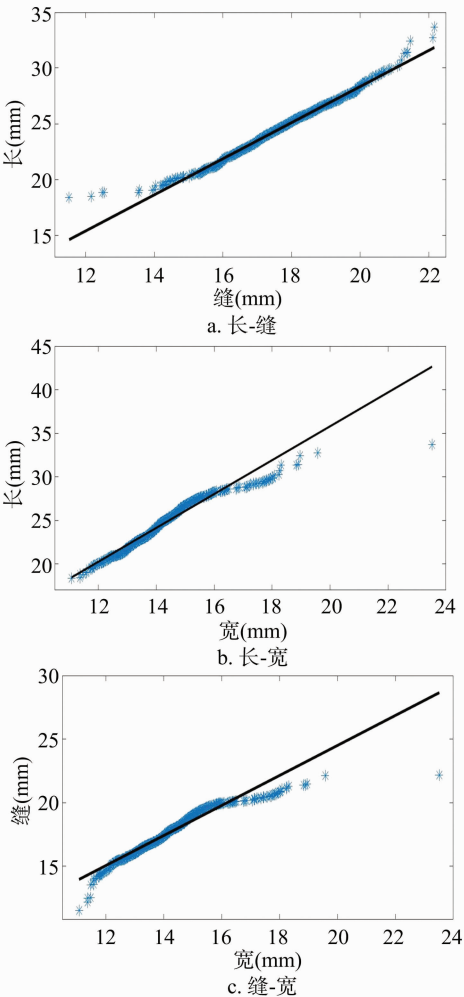


图3 巴旦木长、缝、宽数据两两间的线性回归关系

然后,进一步考虑长与缝、宽间,缝与长、宽间的关系,宽与长、缝间。对上述三者间的关系进行回归分析,求解出线性回归方程,分析结果见表 2,长与缝、宽间,缝与长、宽间,宽与长、缝间 3 个回归方程各自的拟合优度  $R$  都约等于 1,3 个回

归方程各自的概率  $P$  都远小于显著性水平 0.05,说明预测的结论很准确,符合预期,长、缝、宽数据三者间的关系是高度线性化的。长与缝、宽间,缝与长、宽间,宽与长、缝间,各自的检验值  $F$  都不相同、逐步递减,且均方差 ( $S$ ) 为 0,表明巴旦木壳体的不同截面间存在着差异,与标准椭球体长、缝、宽三者间的分布关系不相符,因此可认为南疆巴旦木的外形结构与标准椭球壳相似,但不能如标准椭球壳般视为回转壳体处理。

2.2 巴旦木结构物理特性的方差分析

表 1 数据中,第 5 列为南疆巴旦木实测体积,由 JMTY 型面包体积测定仪测得,第 6 列为由长、缝、宽数据依据标准椭球壳计算出的体积,此 2 列样本各自独立,由  $kstest(X)$  函数检验,都为正态分布。用 MATLAB 软件计算第 5 列、第 6 列数据的方差,分别为  $4.251\ 3 \times 10^7$ 、 $4.059\ 7 \times 10^7$ ,二者差距小于 5%,因此可认为二者方差近似相等。对第 5 列、第 6 列数据进行方差分析,即分析南疆巴旦木外形结构与标准椭球壳间的差别,结果见表 3。从表 3 可以看出, $P=0<0.05$ ,因此可认为南疆巴旦木实测体积与按照标准椭球壳计算出的体积间差距明显;二者间的差别即为  $F$  统计量 770.534 6。

2.3 巴旦木几何模型的建立

综上可知,南疆巴旦木虽为壳体但并非如椭球壳一样是轴对称回转壳体,其外形可近似如图 4 所示。巴旦木外形可近似为关于横、竖、纵 3 椭圆截面对称的壳体,设长、缝、宽分别为  $a$ 、 $b$ 、 $c$ 。从图 4 可以看出,XOZ 截面称为横椭圆截面,截面椭圆离心率为  $e_1$ ;XOY 截面称为竖椭圆截面,截面椭圆离心率为  $e_2$ ;YOZ 截面称为纵椭圆截面,截面椭圆离心率为  $e_3$ 。分别计算所研究的 1 412 个巴旦木样本的  $e_1$ 、 $e_2$ 、 $e_3$ ,之后进行数据统计处理,结果见表 4。

从表 4 可以看出,巴旦木 3 截面离心率的变异系数均小于 13%,说明利用离心率均值来探讨巴旦木截面形状具有很高的可靠性;且  $0<e_3<e_2<e_1<1$ ,因此巴旦木 3 截面的外形可作为椭圆面看待,其中横截面椭圆最扁,与实际情况一致。

3 基于 BP 神经网络构建巴旦木果仁质量预测模型

调用 MATLAB 中的  $H=kstest(X,[X,p],alpha)$  函数对表 1 中巴旦木果仁质量的数据进行正态分布检验,测试水平  $alpha$  为 5%。其中, $p$  代表各种分布形态的累计分布函数,这里引入的不同种分布类型有 normcdf 正态分布、gamedf 伽马分布、poisscdf 泊松分布、expcdf 指数分布、raylcdf 瑞利分布。将各种不同分布形态下的累计分布函数分别代入  $kstest(X,[X,p],alpha)$  函数中,在 MATLAB 中运行,最后得  $H=1$  (图 5),即巴旦木果仁质量的数据不服从以上任何类型的分布形态。此时,若要寻找巴旦木果仁质量与巴旦木壳体外形间的

表 2 巴旦木长、缝、宽数据三者间的多元线性回归方程

指标	回归系数	回归系数估计值	回归系数置信区间
长-缝、宽	$\beta_0$	-4.228	$[-4.361\ 3, -4.094\ 7]$
	$\beta_1$	1.3243	$[1.296\ 7, 1.351\ 8]$
	$\beta_2$	0.3794	$[0.345\ 5, 0.413\ 2]$
		$R^2 \approx 1, F = 9.000\ 3, P < 0.000\ 01, S^2 = 0$	
回归方程		$y = -4.228 + 1.324\ 3x_1 + 0.379\ 4x_2$	
缝-长、宽	$\beta_0$	2.8748	$[2.773\ 9, 2.975\ 8]$
	$\beta_1$	0.6517	$[0.638\ 2, 0.665\ 3]$
	$\beta_2$	-0.0855	$[-0.112\ 6, -0.058\ 3]$
		$R^2 \approx 1, F = 6.865\ 6, P < 0.000\ 01, S^2 = 0$	
回归方程		$y = 2.874\ 8 + 0.651\ 7x_1 - 0.085\ 5x_2$	
宽-长、缝	$\beta_0$	3.122 1	$[2.819\ 5, 3.424\ 7]$
	$\beta_1$	0.673 5	$[0.613\ 4, 0.733\ 6]$
	$\beta_2$	-0.308 4	$[-0.406\ 3, -0.210\ 4]$
		$R^2 \approx 1, F = 1.205\ 3, P < 0.000\ 01, S^2 = 0$	
回归方程		$y = 3.122\ 1 + 0.673\ 5x_1 - 0.308\ 4x_2$	

表 3 巴旦木实测体积与估算结果间的方差分析

来源	平方和	自由度	均方	F 值	P 值
组间方差	$1.145\ 0 \times 10^{11}$	1 411	$8.115\ 1 \times 10^7$	770.534 6	0
误差	$1.487\ 1 \times 10^8$	1 412	$1.053\ 2 \times 10^5$		
总和	$1.146\ 5 \times 10^{11}$	2 823			

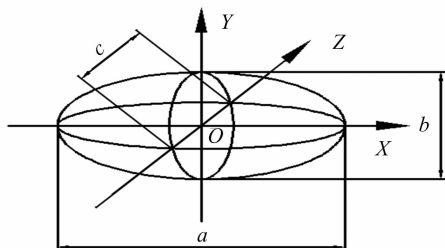


图 4 南疆巴旦木外形近似示意

表 4 巴旦木 3 截面的离心率数据统计结果

截面椭圆离心率	均值	标准差	变异系数
$e_1$	0.810 2	0.041 9	0.051 7
$e_2$	0.686 7	0.059 2	0.086 2
$e_3$	0.587 3	0.072 5	0.123 4

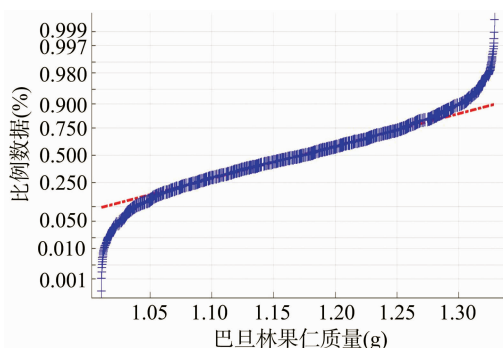


图 5 巴旦木果仁质量的正态分布检验

内在联系,由于长、缝、宽服从正态分布,而果仁质量不是任一种常见的分布形态,就不能简单利用多元线性回归来进行分析。

本研究采用 3 层 BP 神经网络结构来进行分析。在 1 412 组数据中,随机抽取 1 332 组数据作为 BP 神经网络的训练样本,其余 80 组数据为 BP 神经网络的测试样本。经

MATLAB 多次试验,确定输入层结点数目为 3 个,隐含层节点数目为 8 个,输出层节点数目为 1 个,训练函数为 trainlm,传递函数为 tansig,输出函数选用 purelin,最大训练次数为 300 次,学习速率为 0.001,训练目标最小误差为 0.000 02,最大确认失败次数为 10 次。

从图 6 可知,训练样本数据、验证样本数据与测试样本数据之间在训练过 1 次之后基本没有差异,接近于设定的训练目标误差值,为 0.213 6。而在训练过 3 次以后,训练样本、验证样本、测试样本三者形成的曲线完全重合,且三者均与目标值曲线重叠,达到训练预期。

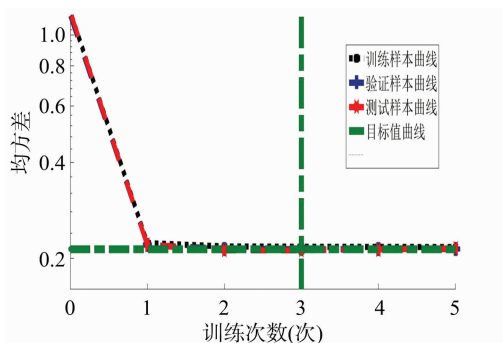


图 6 训练效果及误差

在 BP 神经网络的运行过程中,训练状态的优劣取决于几个关键参数的变化(图 7)。由于设定的最大确认失败次数为 10 次,从图 7 可以看出,确认失败次数达最大值时,训练次数为 13 次,在此时训练误差已不再减小,训练效果最佳,将停止训练。此时,训练过程中的误差梯度由 1 减小到 0.003 2,不再变化;整体训练精度在训练过 1 次后趋于稳定,而后在第 8 次下降,在第 9 次达最小值,之后逐渐恢复并不再变化,保持为  $10^{-4}$ 。

用训练好的 BP 神经网络进行预测函数输出,预测结果如图 8 所示。由图 5 与图 8 可知,期望输出即巴旦木果仁的



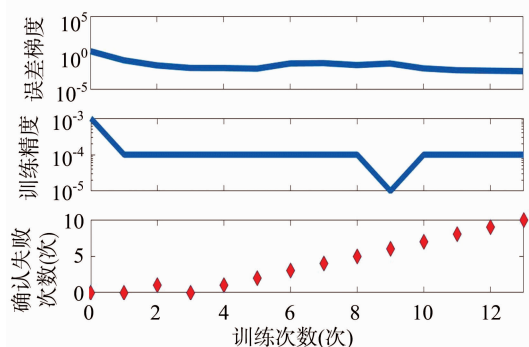


图7 训练状态

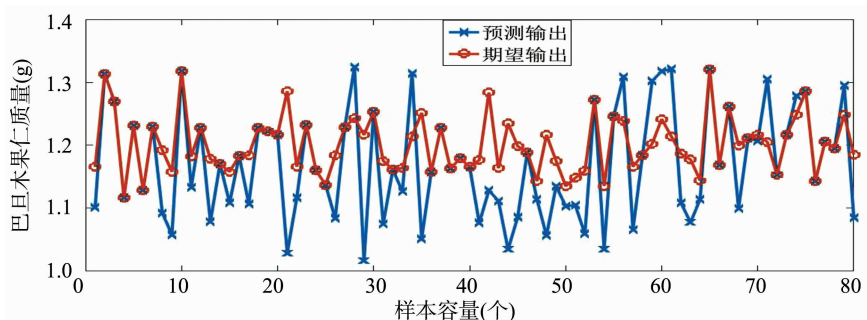


图8 BP神经网络预测输出

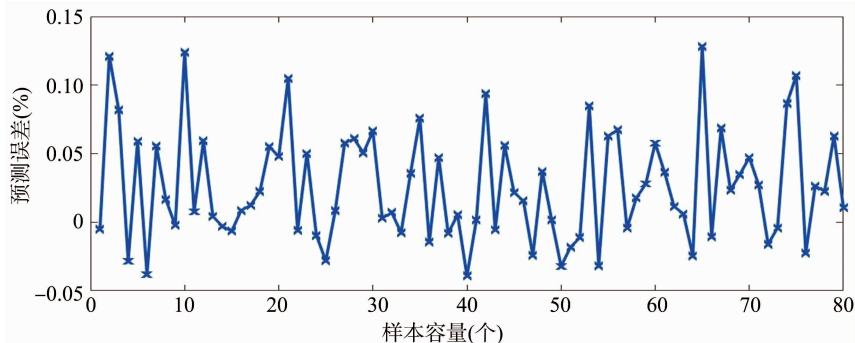


图9 BP神经网络预测误差百分比

实测质量分布在 1.00 ~ 1.35 g 之间,但分布杂乱,规律性较差;用 BP 神经网络预测的输出值主要分布在 1.1 ~ 1.3 g 之间,分布规律与实测值较为吻合,但彼此间仍有一定的误差存在,误差在 0 ~ 0.15 g 之间。

BP 神经网络预测输出和期望输出的误差如图 9 所示。从图 8、图 9 可以看出,虽然 BP 神经网络具有较高的拟合能力,但是 BP 神经网络预测结果仍有一定误差,预测误差在 13% 以内,某些样本点的预测误差较大,整体准确率在 87% 以上。在实际生产中,机械化作业效率高,1 次处理的量很大,87% 以上的准确率足以满足生产与经济需求。

#### 4 结论

本研究基于多元线性回归 BP 神经网络对南疆巴旦木结构物理特性进行研究,结论如下:

(1) 对南疆巴旦木长、缝、宽数据进行回归分析与方差分析发现,长、缝、宽数据两两间高度线性相关,相关系数均大于 0.95,与标准椭球体长、缝、宽间的分布关系相符;长与缝、宽间,缝与长、宽间,宽与长、缝间的回归方程也都存在高度线性关系,但都含有常数项,因此巴旦木外形不是标准椭球壳。(2) 通过构建南疆巴旦木几何模型,巴旦木的外形可近似为一异形椭球薄壳,且从横截面到纵截面椭圆离心率逐步减小,横截面椭圆最扁。(3) 建立基于南疆巴旦木长、缝、宽数据和巴旦木果仁质量测定参数的 BP 神经网络模型,预测准确率在 87% 以上。

#### 参考文献:

- [1] 殷继英,张强,田嘉,等. 新疆地区扁桃坚果品质分析[J]. 经济林研究,2015,33(4):58-64.
- [2] 邵丽珊,李鹏,朱秋萍,等. 新疆野扁桃生物学特性研究[J].

- 新疆农业科学,2016,53(7):1237-1244.
- [3] Fadhil A B, Aziz A M, Altamer M H. Potassium acetate supported on activated carbon for transesterification of new non-edible oil, bitter almond oil[J]. Fuel, 2016, 170: 130-140.
- [4] Kostic M D, Velickovic A V, Jokovic N M, et al. Optimization and kinetic modeling of esterification of the oil obtained from waste plum stones as a pretreatment step in biodiesel production[J]. Waste Management, 2016, 48: 619-629.
- [5] 吐鲁洪·吐尔迪,刘小龙,刘旋峰,等. 新疆巴旦木加工机械现状及解决的技术问题[J]. 农机化研究,2015(1):254-257.
- [6] Sharifian F, Derafshi M H. Mechanical behavior of walnut under cracking conditions[J]. Journal of Applied Sciences, 2008, 8(5): 886-890.
- [7] Shahbzi F. Effects of moisture content, impact direction and impact energy on the cracking characteristics of apricot pit[J]. World Applied Sciences Journal, 2012, 20(11): 1520-1528.
- [8] 吐鲁洪·吐尔迪,阿依木妮莎·拜克热. 巴旦木破壳机的试验研究[J]. 农机化研究,2011,33(3):166-169.
- [9] 范修文,曾勇,兰海鹏,等. 圆锥式巴旦木破壳分离机的设计[J]. 塔里木大学学报,2015,27(2):86-89.