

基于 Shapley 值的大豆种子均径组合预测模型

顿国强,陈海涛,杨纪龙,李若曦,董冰哲,明向兰

(东北农业大学 工程学院,黑龙江 哈尔滨 150030)

摘要:大豆种子均径是育种专用排种盘型孔设计的重要依据,精准构建大豆种子均径预测模型具有非常重要的意义。利用线性回归模型、粒重物理模型及二次回归模型分别对大豆种子均径进行预测,并对其预测误差进行了对比分析。在此基础上,应用 Shapley 值法确定组合预测模型中各单一预测模型的权重,依此构建了大豆种子均径的组合预测模型,并对随机选取的 5 粒东农 52 大豆样本均径进行了预测。结果表明:组合预测模型预测精度高于选定的各单一预测模型,平均预测误差较小,且预测误差波动有所降低,对大豆种子均径的预测是可行且有效的。同时,为其他作物种子的相关物理参数预测提供了一种实用的新方法。

关键词:大豆种子均径;Shapley 值;组合预测

中图分类号:S223.2

文献标识码:A

DOI:10.11861/j.issn.1000-9841.2014.06.0910

Combined Forecasting Model of Soybean Seed Average Diameter Based on the Shapley Value

DUN Guo-qiang, CHEN Hai-tao, YANG Ji-long, LI Ruo-xi, DONG Bing-zhe, Ming Xiang-lan

(Engineering College, Northeast Agricultural University, Harbin 150030, China)

Abstract: Average diameter of soybean seed, as an important reference of breeding seed plate designing, the precise construction of average diameter of soybean seed forecasting model is very important. The linear regression model, grain weight physical model, quadratic regression model were built to predict the average diameter of soybean seed, and then contrastive analyzed prediction error of every forecasting model. Then the weight of every single forecasting model was determined by using Shapley value method, the combination forecasting model was built and used to predict average diameter of 5 Dongnong-52 soybean seed samples which were selected randomly. The results indicated that combination forecasting model's prediction accuracy was higher than the selected individual forecasting models, the combined model had a smaller average error and decreased the fluctuation of error, the prediction of average diameter of soybean seed was feasible and effective. Meanwhile, the research provided a practical new method for relative physical parameters' prediction of other crop seeds.

Key words: Soybean seed average diameter; Shapley value; Combination prediction

大豆育种专用排种器对品种的适应性是当前技术研究的关键问题之一,主要解决的是排种盘型孔如何适应不同大豆品种间的尺寸及形状的差异。在实际生产实践及科学研究过程中,大豆种子的净度、发芽率、百粒重等参数通常为已知量,而作为排种盘型孔设计的重要依据的种子形状及尺寸为未知量^[1-5],因此,构建未知变量(型孔的直径及深度,大豆种子的均径等)关于已知变量(大豆粒重及品种百粒重等)的精确预测模型具有非常重要的意义。相关研究^[6-11]已经确定大豆种子均径与排种盘型孔尺寸间的尺寸关系模型。由于大豆种子单粒重与品种百粒重亦存在一定的函数关系,因此,利用一种新的方法,构建大豆种子均径与粒重间的模型,桥接种子百粒重与排种盘型孔尺寸关系,将为育种专用排种盘型孔的

设计提供重要的理论设计依据。

本文首先选用线性回归模型、粒重物理模型及二次回归模型分别对大豆种子均径进行模型预测,并运用合作博弈中的 Shapley 值法^[12],确定组合预测模型中各单一预测模型的权重,以构建大豆种子均径组合预测模型,并对随机选取的 5 粒东农 52 大豆样本均径进行组合预测,并对各单一预测模型及组合预测模型的预测精度进行对比分析,确定各模型预测精度的主次。

1 大豆物理参数测定

试验材料为东农 52 大豆种子,实测百粒重为 20.7 g,含水率 12.1%。随机选取完整种子 150 粒,利用 JM-D200 型分析天平(量程 200 g,精度 0.001 g)

收稿日期:2014-03-13

基金项目:现代农业产业技术体系建设专项(GARS-04);公益性行业(农业)科研专项(20130311)。

第一作者简介:顿国强(1986-),男,博士,主要从事设施农业机械装备研究。E-mail:dunguoqiang110@126.com。

通讯作者:陈海涛(1962-),男,教授,博导,主要从事农业机械装备及生物质材料技术研究。E-mail:htchen@neau.edu.cn。

及 MASTERPROOF-158 型电子游标卡尺(量程 160 mm,精度 0.01 mm)测量种子粒重及对应的粒形尺寸(粒长 a 、粒宽 b 及粒厚 c 见图 1),150 粒大豆测试样本种子粒重服从均值 0.202 g、标准差 0.021 g 正态分布,按粒重从小到大依次排序,每 6 粒取均值,共 25 组,并计算每组的种子的均径 d ,测量结果如表 1 所示。

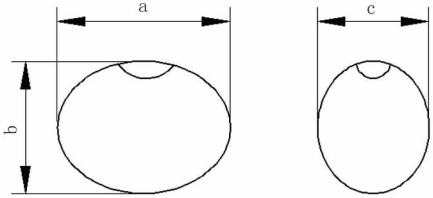


图 1 大豆外形尺寸

Fig.1 Soybean dimensions

表 1 各组大豆种子外形尺寸及其平均粒径

Table 1 Soybean dimensions and its average seeds diameter

| 组别 Groups | 粒重 Grain weight/g | 长度 Length /mm | 宽度 Width /mm | 厚度 Thickness /mm | 平均粒径 Average of diameter /mm |
|--------------|----------------------|---------------------|--------------------|------------------------|------------------------------------|
| 1 | 0.164 | 6.46 | 6.33 | 5.89 | 6.23 |
| 2 | 0.167 | 6.66 | 6.48 | 5.80 | 6.31 |
| 3 | 0.168 | 6.50 | 6.36 | 5.93 | 6.26 |
| 4 | 0.176 | 6.71 | 6.60 | 5.91 | 6.41 |
| 5 | 0.183 | 6.91 | 6.57 | 5.97 | 6.48 |
| 6 | 0.186 | 6.68 | 6.61 | 6.10 | 6.46 |
| 7 | 0.187 | 6.85 | 6.65 | 6.07 | 6.53 |
| 8 | 0.190 | 6.88 | 6.75 | 6.06 | 6.56 |
| 9 | 0.194 | 6.80 | 6.76 | 6.26 | 6.60 |
| 10 | 0.199 | 7.02 | 6.83 | 6.26 | 6.70 |
| 11 | 0.200 | 7.02 | 6.85 | 6.19 | 6.69 |
| 12 | 0.201 | 7.11 | 6.78 | 6.20 | 6.70 |
| 13 | 0.201 | 7.21 | 6.82 | 6.16 | 6.73 |
| 14 | 0.204 | 6.94 | 6.80 | 6.36 | 6.70 |
| 15 | 0.206 | 7.04 | 6.85 | 6.35 | 6.75 |
| 16 | 0.207 | 7.08 | 6.97 | 6.34 | 6.79 |
| 17 | 0.209 | 7.23 | 6.91 | 6.43 | 6.86 |
| 18 | 0.211 | 7.16 | 6.89 | 6.28 | 6.78 |
| 19 | 0.219 | 7.18 | 7.02 | 6.49 | 6.90 |
| 20 | 0.220 | 7.16 | 7.09 | 6.47 | 6.90 |
| 21 | 0.223 | 7.16 | 7.10 | 6.57 | 6.94 |
| 22 | 0.226 | 7.20 | 7.05 | 6.43 | 6.90 |
| 23 | 0.231 | 7.30 | 7.16 | 6.60 | 7.02 |
| 24 | 0.237 | 7.31 | 7.18 | 6.56 | 7.01 |
| 25 | 0.240 | 7.40 | 7.11 | 6.59 | 7.04 |

2 单项预测模型的建立

2.1 线性回归模型

由表 1 可知,大豆的平均粒径 d 随着粒重 w 的增加而不断增大,因此运用一元线性回归分析法^[12]进行预测,应用 Excel 2007 软件建立预测模型:

$$d = 10.97w + 4.473$$

(1)

式中, d -大豆种子平均直径, mm; w -大豆粒重, g。

对回归模型进行方差分析,其显著性检验尾概率小于 0.000 1,模型极显著,决定系数 $R^2 = 0.972$,拟合精度较高。

2.2 粒重物理模型

以椭球体构建大豆种子模型,如图 1 所示,则大豆种子的体积为:

$$V = \frac{4}{3} \times \pi \times \frac{a}{2} \times \frac{b}{2} \times \frac{c}{2} = \frac{abc\pi}{6} \quad (2)$$

式中, V -大豆种子体积, mm^3 ; a -大豆种子长, mm ; b -大豆种子宽, mm ; c -大豆种子厚, mm 。

大豆粒重 w 为:

$$w = \rho V = \frac{abc\rho\pi}{6} \quad (3)$$

式中, w -大豆粒重, g ; ρ -大豆种子密度, $\text{g} \cdot \text{mm}^{-3}$ 。对表 1 数据进行统计分析得出, 平均含水率 12.1% 的东农 52 大豆种子平均密度 ρ 为 $0.001\,283\,\text{g} \cdot \text{mm}^{-3}$ 。

由表 1 数据, 利用一元线性回归分析法建立种子均径 d 与种子长 a 、种子宽 b 和种子厚 c 间的线性回归模型, 为便于反求模型, 线性模型截距均设定为 0, 则模型分别如式(4)、(5)和(6)所示。

$$a = 1.0471d \quad (4)$$

$$b = 1.0204d \quad (5)$$

$$c = 0.9346d \quad (6)$$

将式(4)、(5)和(6)带入式(3), 得出大豆种子平均粒径 d 的粒重物理模型为

$$d = 11.42 \sqrt[3]{w} \quad (7)$$

2.3 二次回归模型

为提高拟合精度, 选用二次函数拟合预测模型, 采用 SAS 软件建立预测模型, 如式(8)所示。

$$d = -48.63w^2 + 30.63w + 2.509 \quad (8)$$

回归模型决定系数 $R^2 = 0.985\,2$, 拟合精度较高, 回归模型方差分析结果表明模型极显著。

随机选取 5 粒东农 52 种子样本, 用 Z_1 、 Z_2 、 Z_3 分别代表线性回归模型、粒重物理模型和二次回归模型, 对大豆种子均径进行预测, 并将预测结果与实际值进行对比, 预测相对误差见表 2。通过对上述 3 种预测模型相对误差的平均值及标准差进行对比分析可知, 模型预测结果相对误差均值及误差标准差具有相同的趋势, 其中, 二次回归模型最大, 线性回归模型次之, 粒重物理模型最小。

表 2 模型预测结果及各预测模型相对误差

Table 2 Forecast result of the forecast model and relative error of each forecast model

| 粒重 Grain weight | 实际值 Actual value | Z ₁ 线性回归模型 Linear regression model | | Z ₂ 粒重物理模型 Grain weight physical model | | Z ₃ 二次回归模型 Quadratic regression model | |
|----------------------------------|------------------------|--|--------------------------|--|--------------------------|---|--------------------------|
| | | 拟合值 Fitted value | 相对误差 Relative error/% | 拟合值 Fitted value | 相对误差 Relative error/% | 拟合值 Fitted value | 相对误差 Relative error/% |
| | | | | | | | |
| 0.173 | 6.39 | 6.371 | 0.297 | 6.363 | 0.423 | 6.353 | 0.579 |
| 0.192 | 6.59 | 6.579 | 0.167 | 6.588 | 0.03 | 6.597 | 0.106 |
| 0.202 | 6.69 | 6.689 | 0.015 | 6.701 | 0.164 | 6.712 | 0.329 |
| 0.215 | 6.87 | 6.832 | 0.553 | 6.841 | 0.422 | 6.847 | 0.335 |
| 0.233 | 7.01 | 7.029 | 0.271 | 7.027 | 0.243 | 7.006 | 0.057 |
| 绝对值平均值 Average of absolute value | | | 0.261 | | 0.256 | | 0.281 |
| 误差标准差 Error standard deviation | | | 0.197 | | 0.170 | | 0.209 |

3 组合预测模型

3.1 模型的建立

由 Bates 和 Granger 于 1969 年提出的组合预测 (combined forecasting) 方法, 是对不同预测模型的预测结果进行适当组合, 以充分提取不同预测模型中的有用信息, 具有提高预测精度, 增加预测结果可靠性和稳定性的特点^[13-15]。

设对同一预测对象, 有 N 种不同的预测模型, 则由这 N 个单一预测模型构成的组合预测模型为:

$$f_i = \sum K_2 f_{iu} \quad (9)$$

式中 f_i -组合预测模型的预测值;

f_{iu} -第 i 种预测模型的预测值, $i = 1, 2, \dots, N$;

K_2 -第 i 种模型的权重, $i = 1, 2, \dots, N$, $\sum K_1 = 1$ 。

3.2 Shapley 值法及组合预测权系数的确定

Shapley 于 1953 年提出的 Shapley 值法, 是一种解决多人合作对策问题的数学方法^[16]。其主要应用于合作收益在各合作方之间的分配问题, Shapley 值实现各个合作成员对该合作联盟的平均贡献, 突出反映了每个成员在合作中的重要性。此方法具有原理和结果易于被各合作方视为公平且结果易于被各方接受的特点。因此, 本文应用 Shapley 值法来确定各预测模型的权系数。

组合预测模型中各预测模型之间共同作用产生组合预测总误差。可认为各预测模型间为了共同目的形成“合作关系”。把组合预测总误差看成合作的“总收益”。将这个总收益向各预测模型进行分配, 利用分配结果来确定各预测模型在组合预测模型中的权系数。

用 N 种预测模型来进行组合预测,则记为 $I = \{1, 2, \dots, N\}$, 对于集合 I 的任一子集 u, v (表示为 N 种方法中的任一组合), $E(u), E(v)$ 表示各自组合的误差。定义如下:

(1) 集合 I 的任一子集 u, v , 有 $E(u) + E(v) \geq E(u \cup v)$ 。

(2) N 种预测模型参与组合预测产生的总误差 E , 将在 N 种预测方法间进行分配, 即 $E = \sum_{i \in N} E'_i$, 其中 E'_i 为模型 i 所分误差, 即 Shapley 值, $i = 1, 2, \dots, N$ 。

设第 i 种预测模型误差预测的绝对值平均值为 $E'_i, E'_i \leq E_i$, 且有

$$E = \frac{1}{N} \sum_{i=1}^N E_i$$

(10)

Shapley 值法的权系数分配式为

$$E'_i = \sum_{u_i \in u} W(|u_i|) [E(u_i) - E(u_i - \{i\})]$$

(11)

$$W(|u_i|) = \frac{(N - |u_i|)! (|u_i| - 1)!}{N!}$$

(12)

式中 $w(|u_i|)$ 一加权因子, 表示组合中模型 i 应承担的组合边际贡献;

u -包含模型 i 的所有子集;

$u_i - \{i\}$ -去除模型 i 的组合;

$|u_i|$ -组合预测中预测模型个数;

则组合预测模型中单一预测模型的权系数计算式为:

$$w_i = \frac{E - E'}{E(N - 1)} (i = 1, 2, \dots, N)$$

(13)

依据表 2, 按公式 (10) 计算组合预测模型总误差为: $E = (0.261 + 0.256 + 0.281) / 3 = 0.266$ 。

以 1, 2, 3 分别代表 3 种单一预测模型, 根据 Shapley 值的概念, 参总误差分摊的“合作关系”成员为: $I = \{1, 2, 3\}$, 其所有组合的子集为: $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$ 。对应组合的误差见表 3。

表 3 组合子集相对误差
Table 3 Combined subset relative error

| $E\{1\}$ | $E\{2\}$ | $E\{3\}$ | $E\{1, 2\}$ | $E\{1, 3\}$ | $E\{2, 3\}$ | $E\{1, 2, 3\}$ |
|----------|----------|----------|-------------|-------------|-------------|----------------|
| 0.261 | 0.256 | 0.281 | 0.2585 | 0.271 | 0.2685 | 0.266 |

根据公式 (11) 计算模型 Z_1, Z_2, Z_3 的 Shapley 值

$$\begin{aligned} E'_1 &= \frac{2! \times 0!}{3!} [E\{1\} - E(\{1\} - \{1\})] + \\ &\frac{1! \times 1!}{3!} [E\{1, 2\} - E(\{1, 2\} - \{1\})] + \\ &+ \frac{1! \times 1!}{3!} [E\{1, 3\} - E(\{1, 3\} - \{1\})] + \\ &\frac{0! \times 2!}{3!} [E\{1, 2, 3\} - E(\{1, 2\} - \{1\})] \\ &= 0.0849 \end{aligned}$$

同样, 计算模型 Z_1, Z_2, Z_3 的分担误差为: $E'_2 =$

0.0812, $E'_3 = 0.0999$, 3 个分担误差之和为 0.266, 等于总误差 E , 说明上述计算正确。

根据公式 (13) 计算各单一预测模型的权系数:

$w_1 = 0.3404, w_2 = 0.3474, w_3 = 0.3122$ 。

则组合预测模型如式 (14) 所示。

$$\begin{aligned} Z_4 &= 0.3404Z_1 + 0.3474Z_2 + 0.3122Z_3 \\ &= 3.967\sqrt[3]{w} - 15.182w^2 + 13.297w + 2.306 \end{aligned}$$

(14)

同样, 利用组合预测模型 Z_4 对上述随机选取的 5 粒东农 52 大豆样本粒径均值进行预测, 组合预测模型的预测值及对应的相对误差见表 4。

表 4 组合模型预测结果及相对误差

Table 4 Forecast result of the combined forecast model and relative error

| 粒重 Grain weight | 实际值 Actual value | 组合预测模型 Combined model | |
|----------------------------------|---------------------|-----------------------|-----------------------|
| | | 拟合值 Fitted value | 相对误差 Relative error/% |
| 0.173 | 6.39 | 6.362 | 0.438 |
| 0.192 | 6.59 | 6.588 | 0.030 |
| 0.202 | 6.69 | 6.700 | 0.149 |
| 0.215 | 6.87 | 6.84 | 0.437 |
| 0.233 | 7.01 | 7.021 | 0.157 |
| 绝对值平均值 Average of absolute value | | | 0.242 |
| 误差标准差 Error standard deviation | | | 0.185 |

由表4可知,组合预测模型的最大拟合误差小于0.438%,预测平均误差0.242%,模型预测精度较强,且将组合预测模型对大豆种子均径的预测结果及其预测结果的相对误差与表2中个单一预测模型进行比较,可知,其预测相对误差均值小于各单一预测模型,提升了模型预测的精度;同时,由表2及表4中,各单一预测模型及组合预测模型的预测结果误差标准差可知,按模型预测误差波动状况由大到小排序为二次回归模型、线性回归模型、组合预测模型及粒重物理模型,组合预测模型误差标准差为0.185,小于单一预测结果误差标准差均值0.192,从整体趋势上,降低了模型预测结果的波动;因此,模型可用于大豆种子均径的预测。

4 结 论

采用线性回归模型、粒重物理模型和二次回归模型3种预测模型对大豆种子均径进行预测,其预测精度从低到高依次为:粒重物理模型、线性回归模型、二次回归模型。

应用Shapley值法确定各单一模型预测的权系数,依此构建了大豆种子均径的组合预测模型,其预测精度高于任一单一预测模型,模型具有较小的预测误差,并且降低了模型预测结果的波动。

参考文献

- [1] 牛远,徐宇,李广军,等.大豆种子大小和粒形的驯化研究[J].大豆科学,2012,31(4):522-528. (Niu Y, Xu Y, Li G J, et al. Domestication of Seed size and shape traits in soybean[J]. Soybean Science, 2012, 31(4): 522-528.)
- [2] 汪霞,徐宇,李广军,等.大豆百粒重QTL定位[J].作物学报,2010,36(10):1674-1682. (Wang X, Xu Y, Li G J, et al. Mapping quantitative trait loci for 100-seed weight in soybean[J]. Acta Agronomica Sinica, 2010, 36(10): 1674-1682.)
- [3] 韩秉进,潘相文,金剑,等.大豆植株性状相关性与产量回归分析[J].中国生态农业学报,2008,16(6):1429-1433. (Han B J, Pan X W, Jin J, et al. Correlation and regression analysis of trait and yield of soybean[J]. Chinese Journal of Eco-Agriculture, 2008, 16(6): 1429-1433.)
- [4] 王福林,尚家杰,刘宏新. EDEM颗粒体仿真技术在排种机构研究上的应用[J].东北农业大学学报,2013,44(2):110-114. (Wang F L, Shang J J, Liu H X. Application of EDEM particles simulation technology on seed-metering device research[J]. Journal of Northeast Agricultural University, 2013, 44(2): 110-114.)
- [5] 谢宇峰,许剑平,陈宝昌,等.播种盘型孔参数的理论计算与试验分析[J].农机化研究,2006,3(3):127-130. (Xie Y F, Xu J P, Chen B C, et al. Theoretic analysis and research of cell of high-speed pneumatic feed wheel for soybean[J]. Journal of Agricultural Mechanization Research, 2006, 3(3): 127-130.)
- [6] 张宇文.机械式多功能精密排种器的设计[J].农业机械学报,2005,36(3):51-54. (Zhang Y W. Research and design for making a new type of mechanized and multiple functions of precision seed-drilled appliance[J]. Transactions of the Chinese Society for Agricultural Machinery, 2005, 36(3): 51-54.)
- [7] 于建群,王刚,心男,等.型孔轮式排种器工作过程与性能仿真[J].农业机械学报,2012,42(12):83-87. (Yu J Q, Wang G, Xin N, et al. Simulation analysis of working process and performance of cell wheel metering device[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 42(12): 83-87.)
- [8] 张庆峰,耿端阳,李少寸.倾斜圆台式精量排种器的设计研究[J].农机化研究,2012(12):54-57. (Zhang Q F, Geng D Y, Li S C. Study on the declined cone frustum type precision seed-metering[J]. Journal of Agricultural Mechanization Research, 2012(12): 54-57.)
- [9] 宋井玲,杨自栋,杨善东.一种新型内充种式精密排种器[J].农机化研究,2013(6):90-93. (Song J L, Yang Z D, Yang S D. A novel inside-filling seed metering device[J]. Journal of Agricultural Mechanization Research, 2013(6): 90-93.)
- [10] 赵佳乐,贾洪雷,姜鑫铭,等.大豆播种机偏置双圆盘气吸式排种器[J].农业机械学报,2013,44(8):78-83. (Zhao J L, Jia H L, Jiang X M, et al. Suction type offset double disc seed metering device of soybean seeder[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(8): 78-83.)
- [11] 赵宇,陈海涛,王业成,等.垄上双行插装式排种器性能试验研究初报[J].大豆科学,2013,32(1):98-101. (Zhao Y, Chen H T, Wang Y C, et al. Preliminary study on the performance of insert-metering sub-district seeder[J]. Soybean Science, 2013, 32(1): 98-101.)
- [12] 陈启明,陈华友.一类组合预测模型的权系数确定的Shapley值方法[J].安徽大学学报,2012,36(2):29-34. (Chen Q M, Chen H Y. Shapley value approach to determining the weights of a kind of combined forecasting model[J]. Journal of Anhui University, 2012, 36(2): 29-34.)
- [13] 熊光红.基于Shapley值的西安市生活需水量组合预测研究[J].水资源与水工程学报,2013,24(1):10-13. (Xiong G H. Research on combination forecast method in predication of domestic water demand in Xi'an based on Shapley value[J]. Journal of Water Resources & Water Engineering, 2013, 24(1): 10-13.)
- [14] 傅培华,鲍福光,李进.基于组合预测模型的航空货运吞吐量预测研究[J].上海管理科学,2012,34(2):48-52. (Fu P H, Bao F G, Li J. Research on the air cargo forecast based on combination forecast model[J]. Shanghai Management Science, 2012, 34(2): 48-52.)
- [15] 张淑娟,冯岫,介邓飞.基于Shapley值的农机装备水平组合预测[J].农业工程学报,2008,24(6):160-164. (Zhang S J, Feng S, Jie D F. Combination prediction of agricultural equipment level based on Shapley value[J]. Transactions of the Chinese Society of Agricultural Engineering, 2008, 24(6): 160-164.)
- [16] 黄宜,赵光洲,王艳伟.基于Shapley值的云南省能源消费综合预测方法研究[J].能源研究与信息,2013,29(1):57-61. (Huang Y, Zhao G Z, Wang Y W. Research on the comprehensive prediction methods for energy consumption of Yunnan based on the Shapley value[J]. Energy Research and Information, 2013, 29(1): 57-61.)