

# 基于深度信念网络的多品种玉米单倍体定性鉴别方法研究

于云华<sup>1,2</sup>, 李浩光<sup>1,2</sup>, 沈学锋<sup>1,2</sup>, 逢燕<sup>1</sup>

1. 中国石油大学胜利学院, 山东 东营 257061

2. 中国石油大学(华东)信息与控制工程学院, 山东 东营 257061

**摘要** 单倍体育种技术是玉米育种新方法,该方法可有效缩短产生纯合系的周期,提高育种效率。该技术需首先挑选足量单倍体籽粒,而玉米在未加人工干预时,单倍体在混合籽粒中仅占0.05%~0.1%,即使采用生物诱导技术,单倍体籽粒数一般也不到籽粒总数的10%。高速、精准地从大量混合籽粒中挑选得到占比少于10%的单倍体籽粒,才能够满足工程化育种需要,而实际育种工作中挑选单倍体时常用的分子生物学、田间形态学辨别等方法存在耗时长、成本高、破坏样本等缺点,难以高效精准地得到玉米单倍体籽粒。相关研究已经证明高油玉米的单倍体与二倍体之间具有明显含油率差异,目前低场核磁共振技术可用于检测玉米单籽粒的含油率,并根据含油率对单倍体进行鉴别,但核磁共振仅存在价格贵、维护难、速度慢、效率低等弱点,现有设备完成单籽粒分选需用时4 s,无法满足工程化育种中大量筛选的速度需求。使用VIA-VI微型近红外光谱仪能够达到0.25 s每颗的检测速度,相比核磁共振技术速度快,仪器价格较低,维护方便。使用近红外光谱仪分析技术对单倍体与二倍体籽粒进行鉴别,可以取代核磁共振鉴别单倍体的方法。采用近红外光谱定性鉴别单倍体籽粒虽然取得了一定效果,但目前研究中所采集玉米品种相对较少,研究只针对某一品种单倍体建立模型,对该品种单倍体进行分类;国内外尚无多品种混合单倍体鉴别相关研究,而工程化育种亟需一种能够识别多个品种玉米单倍体的鉴别方法。为此,本文提出一种基于深度信念网络的多品种混合玉米籽粒单倍体鉴别方法,DBN是一种多层深度神经网络,每层由受限玻尔兹曼机构成,采用逐层训练策略,可解决传统神经网络训练方法不适用于多层网络训练的问题。对比实验结果表明使用DBN方法建立多品种单倍体鉴别模型具有较高分类性能,能够满足玉米工程化育种精度要求。

**关键词** 深度信念网络;单倍体;近红外光谱;定性鉴别

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)03-0905-05

## 引言

针对单倍体玉米籽粒快速分选问题<sup>[1-7]</sup>,采用近红外定性分析鉴别玉米单倍体籽粒虽然取得了一定效果,但目前研究中所采集玉米品种相对较少,研究只针对某一个品种的单倍体二倍体建立模型,对该品种的单倍体二倍体进行分类<sup>[8]</sup>。

覃鸿等采用了微型近红外光谱仪对玉米单倍体近红外方法定性鉴别时,漫反射与漫透射两种采集方式对识别效果影响进行了比较研究,实验中分别在两种方式下采集了单一品种的玉米籽粒的近红外光谱,并使用支持向量机(support vector machina, SVM)的二分类方法鉴别单倍体,漫透射方式下分类器能够获得高于90%的分类准确率,而漫反射方式

的识别率显著低于漫透射方式,已不能满足单倍体鉴别的精度要求,研究认为漫透射方式比漫反射方式更适合玉米单倍体籽粒的鉴别<sup>[9]</sup>;覃鸿等研究了在使用MicroNIR1700微型近红外光谱仪漫透射方式下采集光谱时,光源强度以及光阑孔径对单倍体识别效果的影响,并通过实验确定了单倍体识别率最优时的光源功率及光阑孔径<sup>[10]</sup>。上述研究为采用近红外光谱漫透射方式鉴别单倍体打下了重要基础。

针对单个品种单倍体鉴别模型存在通用性不足的问题,本研究提出一种基于深度信念网络的多品种混合的玉米籽粒单倍体鉴别方法,首先采用多层受限玻尔兹曼机无监督学习及BP神经网络训练得到多品种单倍体籽粒光谱的深度信念网络模型,然后对各个品种的单倍体二倍体籽粒进行分类识别,实验结果表明与其他几种常规分类方法相比,使用深度

收稿日期: 2018-01-02, 修订日期: 2018-06-05

基金项目: 山东省教育厅科技计划项目(J18KA329), 中国石油大学(华东)信息与控制工程学院科技计划项目(KY2017006), 中国石油大学胜利学院科技计划项目(KY2015011)资助

作者简介: 于云华, 1969年生, 中国石油大学信息与控制工程学院教授 e-mail: yuyunhua@upc.edu.cn

信念网络方法建立的多品种单倍体鉴别模型具有较高的分类性能, 实验中所用 10 个品种的单倍体识别率均能达到 90% 以上, 为实现玉米工程化育种提供了新的方法。

## 1 基于深度信念网络的近红外光谱定性模型

### 1.1 原理分析

Geoffrey Hinton 教授在 2006 年提出了一种基于概率生成模型的深度信念网络 (deep believe net, DBN)。DBN 是一种多层深度神经网络结构, 每层由受限玻尔兹曼机 (restricted Boltzmann machine, RBM) 构成, DBN 采用逐层训练策略, 可解决传统神经网络训练方法不适用于多层网络训练的问题, 整个 DBN 训练过程分为预训练与调优两个阶段<sup>[13-14]</sup>。

对于近红外光谱定性分析问题, 深度信念网络以获取原始输入信息深层特征为最终目标, 通过逐层无监督训练策略, 避免传统多层神经网络在使用独立梯度下降法时陷入局部极小值的问题。

DBN 网络结构如图 1 所示, 最下层为输入层, 最上层为输出层。该网络中包含两层 RBM, RBM 是一种两层随机神经网络, 本质是使得学习所得模型能够产生符合条件的样本, 并使得该样本的概率最大。RBM 包含 1 个可见层和 1 个隐藏层, 其特点是同层节点之间并无权值连接, 因此同层中每个节点之间概率独立分布。

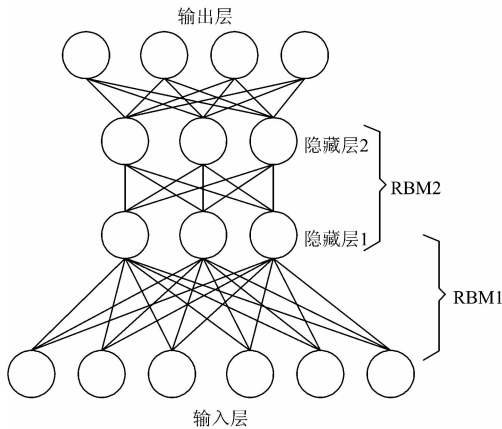


图 1 深度信念网络结构示意图

Fig. 1 Diagram of deep belief network structure

对于一组给定的状态  $(v, h)$ , 可定义如下能量函数

$$E_{\theta}(v, h) = - \sum_{j=1}^{n_v} a_j v_j - \sum_{i=1}^{n_h} b_i h_i - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} h_i w_{ij} v_j \quad (1)$$

式(1)中,  $v_j$  是可见层,  $h_i$  是隐藏层,  $w_{ij}$  是前层与后层的连接权重,  $a_j$  是可见层的偏置,  $b_i$  是隐藏层偏置。

RBM 中的调整参数是  $\theta = (w, a, b)$ , 给定训练样本后, 训练 RBM 意味着调整参数  $\theta$ , 并拟合给定的训练样本。

状态  $(v, h)$  联合概率分布为式(2)

$$P_{\theta}(v, h) = \frac{e^{-E_{\theta}(v, h)}}{Z_{\theta}} \quad (2)$$

配分函数  $Z_{\theta} = \sum_{v, h} e^{-E_{\theta}(v, h)}$  用来作为归一化因子。

通过梯度下降算法计算网络参数, 其中,  $\ln(P(v|\theta))$  关于参数  $\theta$  的偏导数如式(3)所示

$$\frac{\partial \ln(P(v|\theta))}{\partial \theta} = \sum_{i=1}^T \left[ \left\langle \frac{\partial (-E(v^{(i)}, h|\theta))}{\partial \theta} \right\rangle_{P(h|v^{(i)}, \theta)} - \left\langle \frac{\partial (-E(v, h|\theta))}{\partial \theta} \right\rangle_{P(h|v, \theta)} \right] \quad (3)$$

式中,  $\langle \cdot \rangle$  表示分布  $P$  的数学期望。  $P(h|v^{(i)}, \theta)$  表示条件为  $v^{(i)}, \theta$  的概率分布。  $P(v, h|\theta)$  表示可见单元与隐单元的联合分布。

RBM 可由如下几个步骤的对比散度方法并逐层训练得到:

(1) 设定模型的参数初值  $\theta_0 = (w_0, a_0, b_0)$ , 并设定程序结束的条件。

(2) 使  $v^0$  等于输入的向量。

(3) 计算  $h^0, v^1, h^1$ , 其激活概率计算如式(4)和式(5)

$$P_1(h_j = 1 | v, \theta) = \sigma(c_j + \sum_i v_i w_{ij}) \quad (4)$$

$$P_2(v_i = 1 | h, \theta) = \sigma(b_i + \sum_j w_{ij} h_j) \quad (5)$$

其中, 激活函数  $\sigma(x) = \frac{1}{1 + e^{-x}}$ 。

$$\Delta w_{ij} = \alpha (\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}}) \quad (6)$$

$$\Delta a_i = \alpha (\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}}) \quad (7)$$

$$\Delta b_j = \alpha (\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{recon}}) \quad (8)$$

式(6)~式(8)中,  $\alpha$  为学习率,  $\langle \cdot \rangle_{\text{data}}$  为训练样本集所定义的分布上的数学期望。  $\langle \cdot \rangle_{\text{recon}}$  为重构后的模型所定义的分布上的数学期望。

(4) 对模型参数进行更新。

(5) 若满足程序结束条件, 则训练结束, 否则将以  $\theta_{t-1}$  赋值, 并转步骤(3)。

### 1.2 算法设计

算法设计以玉米单倍体 (Haploid, H)、二倍体 (Diploid, D)<sup>[1-7]</sup> 近红外光谱为例, 建模集中每个品种玉米光谱都分为单倍体与二倍体两类, 对于本节玉米单倍体二倍体籽粒的分类任务, 深度信念网络输入为 125 维近红外光谱, 而输出为近红外光谱所对应的单倍体二倍体类别标签。

基于 DBN 的近红外光谱定性分析模型如图 2 所示, 首先选择足量近红外光谱作为建模集, 其后使用多层受限玻尔兹曼机与 BP 算法训练得到深度信念网络定性分析模型, 最后将待鉴别近红外光谱代入模型即可判定其类别。

基于 DBN 的近红外光谱定性分析模型网络首先采用逐层贪心算法训练模型参数, 其后在上一步参数初值基础上, 使用 BP 神经网络<sup>[14]</sup> 对整个网络所有节点进行参数优化, 并进行全局训练, 最终得到基于 DBN 的近红外光谱定性分析模型。

## 2 实验与结果分析

实验使用国家玉米改良中心提供的 10 个经过高油诱导的玉米品种, 二倍体籽粒胚部带有明显紫色记号, 而单倍体胚部没有紫色标记。单倍体占比约为 3%~6%, 实验所用的

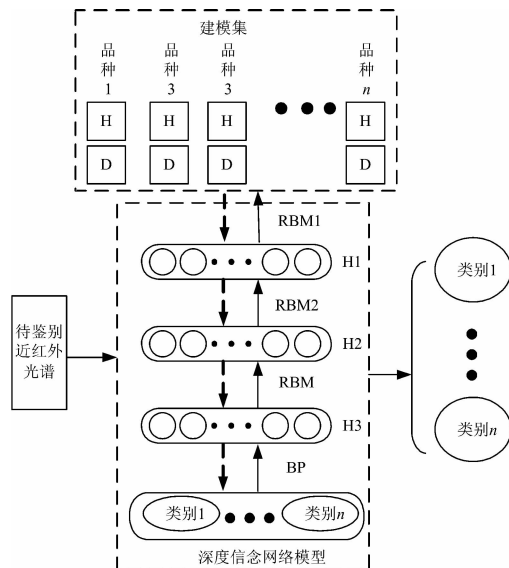


图 2 基于 DBN 的近红外光谱定性分析模型

Fig. 2 Near infrared spectrum qualitative analysis model based on DBN

玉米籽粒样本于 2015 年在北京某郊区实验基地种植并收获得到，每个品种在田间预选足量的备选籽粒，再通过有经验的技工根据 R1-nj 颜色标记从各个品种中挑选得到单倍体与二倍体籽粒<sup>[11]</sup>。其中，单倍体与二倍体籽粒分别为 100 粒，10 个品种的单倍体籽粒以及二倍体籽粒各 1 000 粒作为实验研究对象。为保证实验结果可靠性与准确性，在人工挑选阶段，只选择颜色标记非常清晰的个体，舍弃颜色标记模糊、属性难以分辨、籽粒形状畸变及胚部发育不良的籽粒。

为叙述方便，10 个品种的单倍体、二倍体混合籽粒的近

红外光谱数据分别命名为 K1, K2, ..., K10 数据集，经过光谱仪自带软件预处理后得到 125 维光谱向量，后期数据分析处理使用 Matlab2016a。

采用支持向量机描述光谱质量判定方法剔除异常光谱，预处理方法如平滑、求导等操作去除光谱噪声、提升光谱信噪比，为下一步实验做好数据准备。

针对上述 10 个品种的 2000 粒玉米单倍体二倍体混合籽粒，使用 SVM, BPR 和 DBN 三种方法分别实验。

对比实验方法设置如下：

SVM：特征提取环节采用偏最小二乘 (partial least squares, PLS) 与正交线性判别分析 (orthogonal linear discriminant analysis, OLDA) 方法，原始光谱降维后为 4 维<sup>[14]</sup>，使用 LIBSVM 工具箱，设置 SVM 分类器类型为二分类类型，以最优识别率为标准，在高斯核参数  $\sigma$  及正则化参数  $C$  指数增长的过程中，通过多重交叉验证以网格的方式搜索最优高斯核参数  $\sigma$  及正则化参数  $C$ ，其中高斯核参数  $\sigma=3.2$ ，正则化参数  $C=0.56$ 。

BPR：特征提取环节采用 PLS 与 OLDA 方法，原始光谱降维后为 4 维，使用样本选择算法 (kennard-stone, KS) 方法选择构网样本点，超香肠作为基本覆盖单元，通过交叉验证确定超香肠神经元的半径<sup>[14]</sup>。

实验 1

三种定性分析方法对比实验。在各个品种的数据集独立实验时，随机抽取各个品种数据的二分之一作为训练集，共 100 粒 (训练集中单倍体二倍体数目相等各 50 粒)，剩余二分之一作为测试集测试得到识别率，共重复 10 次取平均值。

表 1—表 3 是分别使用 SVM, 仿生模式识别 (biomimetic pattern recognition, BPR), DBN 三种方法对各个品种分别建立模型所得的正确识别率与正确拒识率。

表 1 SVM 方法各品种单倍体鉴别准确率 (%) (各品种独立建模)

Table 2 SVM; Recognition rate (%) (Independent modeling)

品种	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
CAR	80.2	91.4	82.4	91.0	78.0	93.4	96.4	74.2	92.5	97.4
CRR	82.4	91.6	81.4	92.1	79.2	95.2	96.6	76.8	90.5	94.8
CDR	81.3	92.5	83.4	91.5	78.6	94.3	96.5	75.6	91.5	95.6

表 2 BPR 方法各品种单倍体鉴别准确率 (%) (各品种独立建模)

Table 2 BPR; Recognition rate (%) (Independent modeling)

品种	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
CAR	80.6	94.2	81.3	94.5	77.3	92.1	94.7	77.4	91.5	94.3
CRR	84.2	92.8	85.7	90.9	79.1	98.1	98.5	75.2	93.1	98.1
CDR	82.4	93.5	83.5	92.7	78.2	95.1	96.6	76.3	92.3	96.2

表 3 DBN 方法各品种单倍体鉴别准确率 (%) (各品种独立建模)

Table 3 DBN; Recognition rate (%) (Independent modeling)

品种	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
CAR	85.1	92.1	82.4	91.4	75.0	98.1	95.3	77.0	91.7	94.3
CRR	81.6	96.1	85.8	94.0	74.5	92.2	99.1	78.0	93.9	98.5
CDR	83.3	94.1	84.6	93.4	79.5	95.2	97.2	77.5	92.8	96.4

由表 1—表 3 可知, 三种方法对于 K1, K3, K5 和 K8 等四个品种鉴别准确率均出现了低于 85% 的情况。其中, K5 和 K8 均低于 80%。说明单独使用这四个品种的样本集, 进行建模训练, 测试本品种数据时, 所得识别效果低于其他六个品种, 该四个品种单倍体与二倍体之间差异信息小于其他六个品种差异信息。以上现象说明在近红外光谱单倍体鉴别中, 同样的机器学习方法对于不同品种建立模型识别效果并不一致。

#### 实验 2

三种定性分析方法在各个品种的数据集上独立识别与多

个品种混合建模识别对比实验。

品种独立实验: 随机抽取各个品种数据的二分之一作为训练集, 共 100 粒(训练集中单倍体二倍体数目相等各 50 粒), 剩余二分之一作为测试集测试, 得到鉴别准确率, 重复 20 次取平均。

多品种联合实验: 2 000 粒玉米籽粒的光谱数据构成混合品种实验数据集, 从混合品种实验数据集中随机抽取各个品种数据的二分之一作为训练集(训练集中单倍体二倍体数目相等各占一半), 剩余二分之一作为测试集测试, 得到鉴别准确率, 重复 20 次取平均。

表 4 SVM 方法各品种单倍体鉴别准确率(%)(混合建模)

Table 4 SVM; Recognition rate (%) (Mixed modeling)

品种	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
单独	81.3	93.65	83.4	91.5	78.6	94.3	96.5	75.6	91.5	95.6
混合	83.1	93.8	84.8	92.3	80.3	95.1	96.8	78.3	92.1	96.3

表 5 BPR 方法各品种单倍体鉴别准确率(%)(混合建模)

Table 5 BPR method; Recognition rate (%) (Mixed modeling)

品种	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
单独	82.4	93.5	83.5	92.7	78.2	95.1	96.6	76.3	92.3	96.2
混合	84.3	94.2	85.6	93.8	81.3	96.2	97.1	79.5	93.2	97.3

表 6 DBN 方法各品种单倍体鉴别准确率(%)(混合建模)

Table 6 DBN method; Recognition rate (%) (Mixed modeling)

品种	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
单独	83.3	94.1	85.6	93	79.8	95.2	97.2	79.0	92.8	96.9
混合	91.2	95.2	90.5	94.1	91.3	96.5	97.8	91.3	92.0	96.0

表 4—表 6 为使用 SVM, BPR, DBN 三种建模方法建立定性分析模型所得鉴别准确率。

各表第一行为各品种样本分别单独建模并识别各自品种测试集样本的预测结果。各表第二行为共 10 个品种的玉米籽粒光谱混合来建立模型对单倍体二倍体进行分类, 并单独统计各个品种识别率所得识别结果。

从表 4—表 6 可见, 使用 K1—K10 共 10 个品种玉米籽粒联合建模后, 统计得到各品种识别率相对于对各品种单独建模单独测试各个品种识别率均有所提升。

以鉴别准确率较低的几个品种为例, 三种定性分析模型性能提升有如下差异:

(1) 使用 SVM 方法时, K1 品种鉴别准确率提升了 1.8%, K3 品种鉴别准确率提升了 1.4%, K5 品种鉴别准确率提升了 3.1%, K8 品种提升了 3.2%。

(2) 使用 BPR 方法时, K1 品种提升了 1.9%, K3 品种提升了 2.1%, K5 品种提升了 3.1%, K8 品种提升了 3.2%。

(3) 使用 DBN 方法时, K1 品种提升了 6.9%, K3 品种提升了 4.9%, K5 品种提升了 10.5%, K8 品种提升了 12.3%。

由表 4—表 6 可得: 使用品种混合后, 单个品种独立建模识别效果较差的品种的识别率都具有上升趋势, 三种方法

中以 DBN 方法识别率上升最为明显, 识别率上升也与对比实验中品种混合的数据集相对于单个品种独立建模时的数据量增大相关。在建模集相同的情况下, 在多个品种混合实验中, 在 DBN 方法获得了最优的分类效果, BPR 获得了次优的分类效果。

上述结果说明 DBN 方法在使用多个品种混合籽粒进行训练时, 能够充分发掘不同品种单倍体与二倍体之间共同差异信息, 同时进行多个品种学习时, 能够提升单个品种单倍体二倍体的分类性能, 实现类似于多任务学习中不同分类任务之间的信息共享。

### 3 结 论

针对玉米单倍体籽粒快速分选问题, 提出了基于深度信念网络多品种的单倍体二倍体混合籽粒定性鉴别方法, 对单个品种数据集独立建模以及多个品种混合建模, 并进行对比实验。实验结果表明采用同样方法对每个品种单独建立模型时, 识别效果有优有劣, 性能并不一致; 而使用品种混合后进行实验, 10 个品种能够达到 90% 以上的识别率, 在实验所用的 3 种机器学习方法中, 以 DBN 方法性能最优, 能够达到类似于多任务学习的目标, 实现品种之间信息共享, 并提高多品种单倍体识别效果。

## References

- [1] CHEN Shao-jiang, LI Liang, LI Hao-chuan(陈绍江, 黎亮, 李浩川). *Maize Haploid Breeding Technology*(玉米单倍体育种技术). Beijing: China Agricultural University Press(北京: 中国农业大学出版社), 2012.
- [2] Li Haochuan, Yang Jiwei, Qu Yanzhi, et al. *Journal of Maize Sciences*, 2015(1): 73.
- [3] Melchinger A E, Schipprack W, Friedrich Utz H, et al. *Crop Sci.*, 2014, 54(4): 1497.
- [4] Albrecht E Melchinger, Markus Winter, Mi Xuefei, et al. *Crop Science*, 2015, 55(5~6): 1.
- [5] Albrecht E Melchinger, Wolfgang Schipprack, Mi Xuefei, et al. *Crop Sci.*, 2015, 55(1~2): 188.
- [6] Dong X, Xu X, Li L, et al. *Molecular Breeding*, 2014, 34(3): 1147.
- [7] Dong X, Xu X, Miao J, et al. *Theoretical and Applied Genetics*, 2013, 126(7): 1713.
- [8] Liu Jin, Guo Tingting, Yang Peiqiang, et al. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, 28 (Suppl. 2): 233.
- [9] QIN Hong, MA Jing-yi, CHEN Shao-jiang, et al(覃鸿, 马竞一, 陈绍江, 等). *Spectroscopy and Spectral Analysis*(光谱学与光谱分析), 2015, 25(11): 1807.
- [10] QIN Hong, MA Jing-yi, CHEN Shao-jiang, et al(覃鸿, 马竞一, 陈绍江, 等). *Infrared Technology*(红外技术), 2015, 1(37): 78.
- [11] LI Hao-guang, LI Wei-jun, QIN Hong, et al(李浩光, 李卫军, 覃鸿, 等). *Transactions of The Chinese Society of Agricultural Machinery*(农业机械学报), 2016, 47(6): 259.
- [12] Hinton G, Deng L, Yu D, et al. *IEEE Signal Processing Magazine*, 2012, 29(6): 82.
- [13] Stellan Ohlsson. *Deep Learning*. Cambridge University Press, 2011.
- [14] Shoujue W, Jiangliang L. *Neurocomputing*, 2005, 67: 9.

## Study on Multiple Varieties of Maize Haploid Qualitative Identification Based on Deep Belief Network

YU Yun-hua<sup>1,2</sup>, LI Hao-guang<sup>1,2</sup>, SHEN Xue-feng<sup>1,2</sup>, PANG Yan<sup>1</sup>

1. Shengli College, China University of Petroleum, Dongying 257061, China

2. College of Information and Control Engineering, China University of Petroleum, Dongying 257061, China

**Abstract** Haploid breeding technology is a new method for maize breeding, which can effectively shorten the cycle of homozygous lines and improve the breeding efficiency. The technology needs to select enough haploid grains first, and the haploid grains only account for 0.05%~0.1% of the mixed grains without artificial intervention. Even with the biological induction technology, the number of haploid grains is generally less than 10%. High-speed and accurate identification of haploid grains can meet the needs of engineering breeding. However, molecular biology and morphological identification methods commonly used in practical work are time-consuming, costly and destroying samples. It is difficult to select Maize Haploid grains efficiently and accurately. Relevant studies have proved that there are obvious oil content differences between haploid and diploid of high-oil maize. At present, low-field nuclear magnetic resonance technology can be used to detect oil content of maize and identify haploid according to its oil content. However, nuclear magnetic resonance (NMR) instrument has some weaknesses, such as high price, difficult maintenance, slow speed and low efficiency. It takes 4 seconds for each single-grain sorting. It cannot meet the needs of large number identification for engineering breeding. Using VIAVI near infrared spectrometer (NIRS) can achieve the detection speed of 0.25 seconds for each maize. The NIR technology is faster, cheaper and easier to maintain. The NIR identification method can replace the method of NMR. Qualitative identification of haploid by NIRS has achieved some results, but currently there are relatively few maize varieties collected in the study. The study only establishes models for haploid of one variety, and classifies haploid of that variety. There are no studies on identification of multiple hybrid haploids at home and abroad, but engineering breeding urgently needs a method to identify multiple varieties of maize haploids. In this paper, a method for identifying haploids based on deep belief network is proposed. DBN is a multi-layer deep neural network. Each layer is composed of a restricted Boltzmann mechanism. By using layer-by-layer training strategy, the problem that traditional neural network training methods are not suitable for multi-layer training can be solved. The comparative experimental results show that the identification model of multiple varieties haploid established by DBN method has high classification performance and can meet the requirements of maize engineering breeding accuracy.

**Keywords** Deep belief net; Haploid; Near infrared spectroscopy; Qualitative identification

(Received Jan. 2, 2018; accepted Jun. 5, 2018)