



基于近红外光谱技术鉴别植食性金龟子的初步研究*

吕欣航^{1**} 高传部¹ 田有文² 王小奇^{1***} 方红¹

(1. 沈阳农业大学植物保护学院, 沈阳 110161; 2. 沈阳农业大学信息电气与工程学院, 沈阳 110161)

摘要 【目的】植食性金龟子是我国的重要农林害虫, 探索一种快速而准确地鉴别植食性金龟子的新方法, 为将此法推及至其他鞘翅目昆虫的识别来建立研究基础。【方法】利用近红外光谱法对金龟子进行鉴别, 提出了用支持向量机 (Support vector machine, SVM) 算法对 15 种植食性金龟子近红外光谱图 (数据) 进行分析, 经过噪声波段去除后, 用平滑求导与标准化法对的光谱进行预处理, 选取金龟子标本 150 个, 针对不同分类阶元和分类单元将 66% 样本图谱作为校正集, 用 SVM 建立鉴别模型并对模型进行自身检验, 用剩余样本图谱作为预测集对这些模型进行验证。【结果】模型的自身检验显示在金龟科 4 个亚科的鉴别模型中, 鳃金龟亚科正确识别率为 86%, 其他样本的识别准确率均大于 95%, 在亚科不同属和属下不同种的鉴别模型中, 除疏纹星花金 *Protaetia cathaica* (Bates) 外, 其他样本的识别准确率均为 100%; 模型的预测集验证结果显示, 在不同分类阶元和分类单元的鉴别模型中, 由于色斑鳃金龟 *Polyphylla laticollis* Lewis 样本较少未能正确识别, 其他样本的识别准确率均为 100%。整体试验结果较为理想, 说明模型性能较好。【结论】基于已定金龟子建立的模型能够很好地鉴别大部分样本, 采用近红外光谱扫描技术结合支持向量机得到的植食性金龟子鉴别模型具有很强的推广能力。

关键词 金龟子, 快速鉴别, 近红外光谱, 支持向量机, 模型

Identification of chafers using near-infrared spectra

LÜ Xin-Hang^{1**} GAO Chuan-Bu¹ TIAN You-Wen² WANG Xiao-Qi^{1***} FANG Hong¹

(1. College of Plant Protection, Shenyang Agricultural University, Shenyang 110161, China;

2. College of Electricity Information and Engineering Institute, Shenyang 110161, China)

Abstract [Objectives] To explore a new method for the rapid and accurate identification of chafers, which are important agricultural and forestry pests in China, and extend this to other beetles. [Methods] Near-infrared reflectance spectroscopy (NIRS) was used to analyze near-infrared spectrum data from 15 chafers with the support vector machine (SVM) algorithm. The spectra were processed using the methods described by Savitzky-Golay for smoothing and autoscaling, after removal of the noise region. Using the spectra of 150 specimens, 66% of the NIR spectra of chafer samples were used to establish SVM-based discrimination models, and the remaining 34% of NIRS spectra were used to validate these models. The models were self-validated. [Results] The self-validated results of calibration samples are shown below. In the identification model of 4 subfamilies of the Scarabaeidae family, the accuracy rate for identifying the Melolonthinae was 86% and that for other subfamilies 95%. Identification models for species and genera all had accuracy rates of 100%, except that for *Protaetia cathaica* (Bates). The self-validated results of prediction samples showed that all accuracy rates reached 100%, except that for *Polyphylla laticollis* Lewis which was because of the low number of specimens available. These results are consistent with those obtained by morphological identification. [Conclusion] This novel method has a considerable potential for application

*资助项目 Supported projects: 辽宁省教育厅重点实验室项目 (LS2010140)

**第一作者 First author, E-mail: lxh2011syau@163.com

***通讯作者 Corresponding author, E-mail: wxq1120@sina.com

收稿日期 Received: 2015-03-16, 接受日期 Accepted: 2015-06-12

in the future due to its speed, reliability and easy of use, as well as its veracity and credibility in pest control and inspection and quarantine.

Key words chafer, rapid identification, NIRS, SVM, models

金龟子是鞘翅目中一个大而独特的类群,植食性金龟子(Chafer)是我国农林业中的重要害虫,常造成严重损失,主要存在于花金龟亚科(Cetoniinae)、犀金龟亚科(Dynastinae)、臂金龟亚科(Euchirinae)、鳃金龟亚科(Melolonthinae)中。金龟子不仅种类繁多,且在形态、大小、颜色等方面存在多型现象,正确的种类鉴别是昆虫研究及相关学科的前提条件,如不能准确鉴别出昆虫种类,一切研究都将很难进行下去(Klarica *et al.*, 2011)。形态学鉴别在实际应用中存在一定难度,分子生物学及蛋白质组学的方法由于操作复杂、价格昂贵,在应用上也存在局限性。因此对金龟子进行快速有效的识别鉴别对农业害虫防治与植物检疫工作具有重要意义。

近红外光谱(Near infrared reflectance, NIR)分析技术是物质定性定量分析研究的重要手段之一,其原理是有机分子一般都包含 O-H、N-H、C-H 等含氢基团,它们的不同倍频与合频振动特征及其各种方式的组合所形成的振动信息包含了有机分子含氢基团的主要结构信息,这些信息都能够加载到近红外光谱上,使得该谱曲能够分析各种物质(严衍禄等, 2013)。近红外光谱谱区的波长范围是 770~2 500 nm,频率范围是 $13\ 000\sim 4\ 000\text{ cm}^{-1}$ 。近红外光谱分析技术具有快速、准确、无污染、无损坏、低成本和易操作的特点(Paliwal *et al.*, 2004; Munck, 2007),近年来作为一种新型检测技术被越来越多地应用于农作物质量分析与快速鉴别(田有文等, 2004; 田有文和牛妍, 2009; 赵秋剑等, 2013)、食品质量检验(Huang *et al.*, 2008)、药品定量与定性分析(滕乐生等, 2008)、医学的症状监测(Wolf and Ferrari, 2007; Zheng *et al.*, 2013)和石油化工(陆婉珍, 2007)等各个领域。

自 20 世纪 90 年代以来,近红外光谱技术在昆虫学的研究越来越多,在储藏害虫的检测研究中, Baker 等(1999)利用近红外光谱技术检测

和识别被象虫金小蜂 *Anisopteromalus calandrae* (Howard)寄生的米象 *Sitophilus oryzae* (L.), 结果显示可以准确识别出被金小蜂计生的米象的卵和蛹。Maghirang 等(2003)采用近红外光谱检测了小麦粒内的害虫,结果识别蛹的准确率为 94%, 识别大、中、小型幼虫的准确率分别为 93%、84%和 63%。Perez-Mendoza 等(2002, 2004)将近红外光谱分析技术应用于家蝇 *Musca domestica* (L.)和 3 种储粮害虫龄期的预测,结果显示此分析技术受温度、性别等因素的影响不大。国内外研究表明,昆虫表皮碳氢化合物可以用于不同阶元物种的鉴别(Carlson, 1979; 高明媛, 2001; Howard and Blomquist, 2005),利用近红外光谱技术对昆虫种类鉴别或近缘种区分有 Cole 等(2003)利用 NIR 技术对 2 种茧蜂 *Cotesia flavipes* 和 *C. sesamiae* 鉴别的准确率为 85%; Aldrich 等(2007)成功鉴别了白蚁的 3 个种和 2 个亚种,准确率分别为 95%和 80%;在近缘种的幼虫期的鉴别研究中有 Siegwart 等(2015)利用偏最小二乘法对苹果蠹蛾 *Carpocapsa pomonella* (L.)和梨小食心虫 *Grapholitha molesta* (Busck)的幼虫进行了有效区分;另外在蚊(Mayagaya *et al.*, 2009)、蚁(Antonialli *et al.*, 2008)、啮虫(Lazzari *et al.*, 2010)、蝇(Jaime *et al.*, 2011)等类群种近缘种的鉴别也都有报道。但应用近红外光谱技术对鞘翅目昆虫的鉴别和分类尚无报道,同时上述的研究所用的样本数量都比较大,对于小样本量的昆虫近似或近缘种识别的研究也比较少。

鉴于上述代谢组学研究中应用近红外光谱分析技术鉴别昆虫的可行性,本试验选取了 15 种植食性金龟子共 150 头,结合近红外光谱漫反射测量技术并建立基于支持向量机的识别模型,来讨论基于近红外光谱技术进行快速有效鉴别植食性金龟子的可行性。

1 材料与方法

选择具有代表性的金龟子进行标本采集,扫描标本样本的近红外光谱,将光谱数据进行预处理,通过数学计算建立分类鉴别模型,最后用未知标本样本对所建模型进行验证。

1.1 材料

试验所选金龟子依据形态特征进行鉴别,参考《中国北方常见金龟子彩色图鉴》(刘广瑞等,1997)等书籍,最后由沈阳农业大学张治良教授确认定种。将样本保存至 75%酒精中备用,扫描前用蒸馏水浸泡并冲洗,置于通风处晾干,金龟

子标本信息见表 1。

1.2 仪器设备与测量

采用德国 Burker 公司的 MPA 型近红外光谱仪(图 1)(附件检测器 TE-InGaAs 固体光纤探头)进行全谱测定扫描,扫描前仪器预热 0.5 h,波数范围 12 000~4 000 cm^{-1} ,间隔 8 cm^{-1} 。将光纤探头紧贴于金龟子鞘翅体壁进行扫描测量光谱的吸光度(图 2),重复扫描 32 次得出平均光谱,每头金龟子均从鞘翅的 5 个不同位置进行测量以减少散射误差。工作软件为 OPUS7.2 近红外处理软件。

表 1 金龟子标本信息
Table 1 Information of chafer dug specimen

亚科 Subfamily	属 Genus	种类 Species	采集信息 Collection information	数量(头) Number
鳃金龟亚科 Melolonthinae	<i>Polyphylla</i>	云斑鳃金龟 <i>P. laticollis</i> Lewis	2012.7.16 辽宁鞍山 Anshan, Liaoning	3
	<i>Hoplosternus</i>	灰胸突鳃金龟 <i>H. incanus</i> Motschulsky	2014.7.5 辽宁沈阳 Shenyang, Liaoning	20
	<i>Holotrichia</i>	暗黑鳃金龟 <i>H. parallela</i> Motschulsky	2014.06.29 辽宁沈阳 Shenyang, Liaoning	8
	<i>Holotrichia</i>	东北大黑鳃金龟 <i>H. diomphalia</i> (Bates)	2014.07.1 辽宁沈阳 Shenyang, Liaoning	9
	<i>Maladera</i>	阔胫玛绢金龟 <i>M. verticalis</i> Fairmaire	2012.7.16 辽宁鞍山 Anshan, Liaoning	11
	<i>Maladera</i>	黑绒金龟 <i>M. orientalis</i> (Motschulsky)	2012.5.26 辽宁葫芦岛 Huludao, Liaoning	12
犀金龟亚科 Dynastinae	<i>Pentodon</i>	阔胸禾犀金龟 <i>P. mongolicus</i> Motschulsky	2012.8.23 吉林松原 Songyuan, Jilin	11
	<i>Blitopertha</i>	苍翅黎丽金龟 <i>B. pallidipennis</i> Reitter	2011.7.29 辽宁丹东 Dandong, Liaoning	11
	<i>Anomala</i>	蒙古异丽金龟 <i>A. mongolica</i> Faldermann	2012.8.4 辽宁阜新 Fuxin, Liaoning	10
丽金龟亚科 Rutelinae	<i>Anomala</i>	多色异丽金龟 <i>A. chamaeleon</i> Ohause	2014.7.5 辽宁沈阳 Shenyang, Liaoning	9
	<i>Mimela</i>	褐足彩丽金龟 <i>M. testaceipes</i> (Motschulsky)	2012.7.16 辽宁鞍山 Anshan, Liaoning	11
	<i>Cetonia</i>	华美花金龟 <i>C. magnifica</i> Ballion	2011.7.16 辽宁抚顺 Fushun, Liaoning	12
	<i>Protaetia</i>	白星花金龟 <i>P. brevitorsis</i> (Lewis)	2012.7.16 辽宁鞍山 Anshan, Liaoning	10
	<i>Protaetia</i>	疏纹星花金龟 <i>P. cathaica</i> (Bates)	2012.7.16 辽宁鞍山 Anshan, Liaoning	4
花金龟亚科 Cetoniinae	<i>Gametis</i>	小青花金龟 <i>G. jucunda</i> (Faldermann)	2012.6.19 辽宁沈阳 Shenyang, Liaoning	9

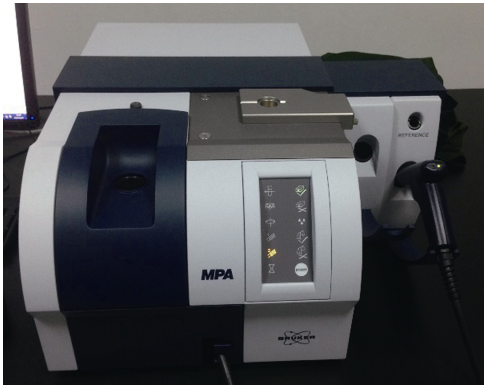


图 1 MPA 型多用途傅立叶变换近红外分析仪
Fig. 1 MPA multi purpose FT-NIR analyzer

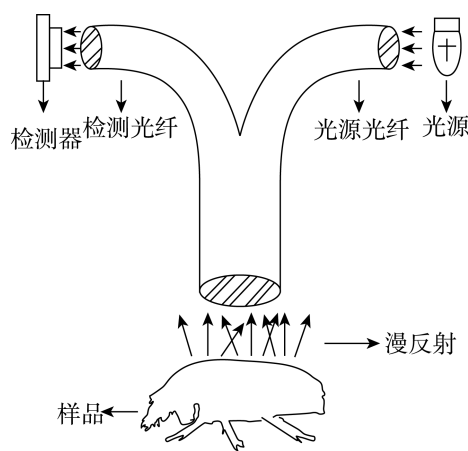


图 2 金龟子扫描图示
Fig. 2 Diagram of scanning chafers

1.3 建立模型

1.3.1 支持向量机分类原理 支持向量机 (Support vector machine, SVM) 是一种基于结构风险最小化原理 (Structural risk minimization, SRM) 处理数据的统计方法, 通过非线性映射技术把样本空间映射到高维特征空间中, 用线性化手段进行分类求出不同类样本的最优分类超平面 (奉国和, 2011)。与许多传统模式识别算法相比, 支持向量机避开了从归纳到演绎的传统过程, 建模所需的先验干预小, 实现了高效的从训练样本到预测样本的转导, 大大简化了分类问题。在解决非线性、小样本、高维数和局部极小值的模式识别上更有优势, 适用于回归分析、模式识别、预测和综合评价等问题的处理 (王莉等, 2008)。

1.3.2 基于支持向量机的模型建立方法与步骤 本文利用软件 Matlab8.4 和 PLS_TOOLBOX 对近红外光谱进行数据处理, 基于支持向量机对 15 种植食性金龟子的扫描光谱从高级分类阶元向低级阶元进行逐级识别, 然后通过不同的分类单元进行识别; 即先对光谱数据进行不同亚科的区分, 再对每一亚科的不同属进行区分, 最后对每一属下不同种进行区分, 相当于建立 1 个鉴别模型总库下按不同级别分出 8 个不同分类单元的鉴别模型子库 (图 3)。

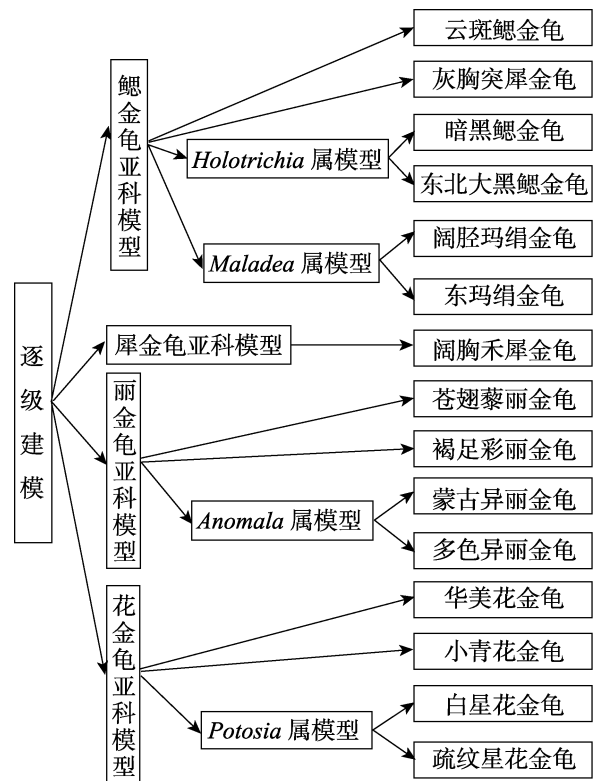


图 3 逐级建模过程图示
Fig. 3 Diagram of modeling process by different taxonomic levels

一个完整的支持向量机模型包括数据处理的方法、核函数和模型参数三部分, 因此基于支持向量机的鉴别模型包括样本数据集的建立、确定数据样本预处理方法、选择核函数与相关参数、预测评价鉴别模型 4 个步骤。

(1) 样本数据集的建立

依据不同分类单元将金龟子光谱数据分为 8 个子库, 分别对应要建立的 8 个鉴别模型, 其次采用 K-S (Kennard-Stone) 算法将每一个子库的

样本分为校正集和预测集,分类原则为采用 66% 样本作为校正集,剩余样本作为预测集,校正集用于模型的建立与自身验证,预测集用于对所建模型进行预测评价(表 2)。

(2) 光谱的预处理

为了消除来自高频随机噪声、基线漂移、样本不均匀、光散射等影响需进行光谱预处理来消

除噪声。对每个样本的 5 条光谱数据取平均数。本文采用求导法对平均光谱进行预处理(Barnes *et al.*, 1989; 褚小立等, 2004), 同时采用不同的参数(导数、求导阶数与平滑点数)进行测试、评价并改进,最终得出各个鉴别模型的最适光谱预处理方法。如图 4 所示,在全光谱 10 000~12 000 cm^{-1} 光谱的噪声也比较明显,因此剔除建模波段采用

表 2 样本数据集的分类
Table 2 Classifying of sample data sets

鉴别模型 Models	校正集数(个) Calibration samples	预测集数(个) Prediction samples
金龟科 4 个亚科 4 subfamilies of family Scarabaeidae	101	45
丽金龟亚科 3 个属 3 genera of subfamily Rutelinae	44	19
鳃金龟亚科 4 个属 4 genera of subfamily Melolonthinae	29	12
花金龟亚科 3 个属 3 genera of subfamily Cetoniinae	21	11
<i>Holotrichia</i> 属 2 个种 2 species of genus <i>Holotrichia</i>	12	5
<i>Maladera</i> 属 2 个种 2 species of genus <i>Maladera</i>	16	7
<i>Anomala</i> 属 2 个种 2 species of genus <i>Anomala</i>	13	6
<i>Protaetia</i> 属 2 个种 2 species of genus <i>Protaetia</i>	10	4

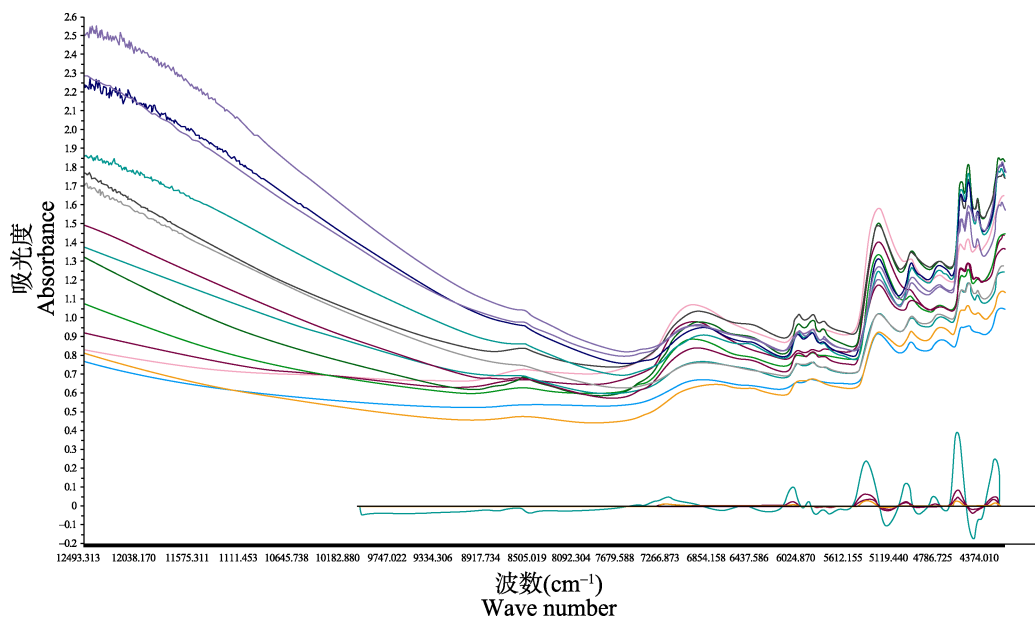


图 4 金龟子标本近红外扫描光谱

Fig. 4 Near infrared spectroscopy of chafer dug specimen

上方: 原始光谱; 下方: 预处理光谱。

Above: Original spectra; Below: Pretreated spectra.

4 000~10 000 cm⁻¹ (图 4)。

(3) 模型核函数与参数的选择

本文采用的支持向量机分类类型为 C-SVC, 采用径向基核函数作为鉴别模型的核函数(陈婷等, 2015)。惩罚参数 C 和核函数参数 g 作为影响支持向量机模型性能的关键参数, 因此需要对 C 与 G 值进行参数作为寻优变量来确定最适值(鞠鲁峰等, 2014)。本文采取网格寻优算法进行参数寻优, 同时采用交叉验证(Cross validation, CV) 法对参数进行测试检验, 即对模型自身进行检验。本文采用识别率和拒绝率对模型自身验证进行识别, 以判断所建立的支持向量机判别模型的可行性, 并最终确定模型的参数值。

识别率表示模型预测的准确性, 拒绝率表示模型预测的专属性即准确地测出待测样本的特性。公式如下。

$$\text{识别率} = \frac{\text{识别自身类样本个数}}{\text{该类样本的总个数}} \times 100\%,$$
$$\text{拒绝率} = \frac{\text{拒绝自身类样本个数}}{\text{其他类样本的总个数}} \times 100\%。$$

(4) 模型的预测评价

确定好 SVM 模型后取用验证集数据对所建

立的各个模型进行预测, 用识别率和拒绝率对模型预测结果进行评价以判断所建立的鉴别模型的推广能力。

2 结果与分析

2.1 光谱的预处理方法

8 个子库鉴别模型的光谱预处理方法见表 3。

8 个鉴别模型中都采用了 Savitzky-Golay 卷积求导法与标准化法相结合的方法对光谱进行预处理(表 3), 参数中 Savitzky-Golay 卷积平滑点数都是 15, 但 Savitzky-Golay 卷积求导法在不同分类单元下的鉴别模型所选取的多项式次数又不完全相同(Savitzky-Golay 卷积求导法、多项式次数、移动窗口宽度)。

2.2 各个模型的参数与自身验证结果

所有模型的核函数均为径向基函数, 表 4 显示所有模型最后确定的的惩罚参数 C 都是 100, 在不同分类单元的模型中核参数的确定虽有不同但都比较小, 因此所建的 SVM 模型可以适用于各个分类阶元的金龟子鉴别。

模型自身验证结果可以看出: 金龟科 4 个亚

表 3 8 个模型的光谱预处理方法
Table 3 Pretreatment methods of near-infrared spectra for 8 models

鉴别模型 Models of different categories	光谱预处理方法 Pretreatment methods of near-infrared spectra
金龟科 4 个亚科 4 subfamilies of family Scarabaeidae	一阶导数(阶数 2, 平滑点数 15); 标准化 1st derivative (2nd order polynomial, smoothing point 15); normalize
丽金龟亚科 3 个属 3 genera of subfamily Rutelinae	二阶导数(阶数 5, 平滑点数 15) 标准化 2nd derivative (5th order polynomial, smoothing point 15); normalize
鳃金龟亚科 4 个属 4 genera of subfamily Melolonthinae	二阶导数(阶数 5, 平滑点数 15) 标准化 2nd derivative (5th order polynomial, smoothing point 15); normalize
花金龟亚科 3 个属 3 genera of subfamily Cetoniinae	一阶导数(阶数 2, 平滑点数 15) 标准化 1st derivative (2nd order polynomial, smoothing point 15); normalize
Holotrichia 属 2 个种 2 species of genus Holotrichia	一阶导数(阶数 2, 平滑点数 15) 标准化 1st derivative (2nd order polynomial, smoothing point 15); normalize
Maladera 属 2 个种 2 species of genus Maladera	一阶导数(阶数 2, 平滑点数 15) 标准化 1st derivative (2nd order polynomial, smoothing point 15); normalize
Anomala 属 2 个种 2 species of genus Anomala	一阶导数(阶数 5, 平滑点数 15) 标准化 1st derivative (2nd order polynomial, smoothing point 15); normalize
Protaetia 属 2 个种 2 species of genus Protaetia	一阶导数(阶数 5, 平滑点数 15) 标准化 1st derivative (2nd order polynomial, smoothing point 15); normalize

表 4 鉴别模型的参数与交叉验证
Table 4 Discriminant models' parameters and cross validation

模型 Model	惩罚系数 Cost	核参数 Gamma	模型交叉验证 Self verification of models		
			分类单元 Taxon	识别率 (%) Recognition rate	拒绝率 (%) Rejection rate
金龟科 4 个亚科 4 subfamilies of family Scarabaeidae	100	0.001	Melolonthinae	98	98
			Dynastinae	86	100
			Rutelinae	97	96
			Cetoniinae	95	100
			<i>Polyphylla</i>	100	100
丽金龟亚科 3 个属 3 genera of subfamily Rutelinae	100	0.00000316	<i>Hoplosternus</i>	100	100
			<i>Holotrichia</i>	100	100
			<i>Maladera</i>	100	100
			<i>Blitopertha</i>	100	100
鳃金龟亚科 4 个属 4 genera of subfamily Melolonthinae	100	0.00001	<i>Mimela</i>	100	100
			<i>Anomala</i>	100	100
			<i>Cetonia</i>	100	100
			<i>Protaetia</i>	100	82
花金龟亚科 3 个属 3 genera of subfamily Cetoniinae	100	0.001	<i>Gametis</i>	33	100
			<i>H. parallela</i>	100	100
			<i>H. diomphalia</i>	100	100
<i>Maladera</i> 属 2 个种 2 species of genus <i>Maladera</i>	100	0.000001	<i>M. verticalis</i>	100	100
			<i>M. orientalis</i>	100	100
<i>Anomala</i> 属 2 个种 2 species of genus <i>Anomala</i>	100	0.000001	<i>A. mongolica</i>	100	100
			<i>A. chamaeleon</i>	100	100
<i>Protaetia</i> 属 2 个种 2 species of genus <i>Protaetia</i>	100	0.000001	<i>P. cathaica</i>	100	100
			<i>P. brevitarsis</i>	100	100

科的鉴别模型中所有样本的识别率为 96.04%，只有 4 条光谱被错误识别；鳃金龟、丽金龟和花金龟的识别率分别达到了 98%、97%和 95%；犀金龟的识别率虽然为 86%，是由于这个模型中犀金龟样本量相对较少。鳃金龟和丽金龟的拒绝率均大于 95%，犀金龟和花金龟的拒绝率均为 100%，这说明金龟科 4 个亚科的鉴别模型自身评价性能很好。

鳃金龟亚科 4 个属、丽金龟亚科 3 个属、*Holotrichia* 属 2 个种、*Maladera* 属 2 个种、*Anomala* 属 2 个种、*Protaetia* 属 2 个种这 6 个模型的自身验证结果对各自样本的识别率和拒绝率均达到了 100%，可知这些模型的性能很好。

花金龟亚科 3 个属的鉴别模型自身检验时发现 *Cetonia* 属和 *Protaetia* 属的识别率均为 100%，但 *Gametis* 属的识别率只有 33%，*Gametis*

属的样本被错误识别到了 *Protaetia* 属中，原因可能是这个属的样本量太少，因此花金龟亚科 3 个属的鉴别模型可以有效鉴别 *Cetonia* 属和 *Protaetia* 属的花金龟，但对能否准确鉴别 *Gametis* 属花金龟尚存疑问。

2.3 模型预测集验证结果

由表 5 中各个鉴别模型的预测结果可以看出：绝大多数金龟子均能被准确鉴别，但鳃金龟亚科 4 个属的鉴别模型中对 *Polyphylla* 属（云斑鳃金龟 *P. laticollis*）未能够正确识别，这个属的样本被错误识别到 *Hoplosternus* 属中一个原因是样本量太少（1 头）而无法准确鉴别，但自身验证结果显示 *Polyphylla* 属的识别率和拒绝率都为 100%，因此可以认为鳃金龟亚科 4 个属的鉴别模型可以应用于准确鉴别。

在花金龟亚科 3 个属的鉴别模型虽然识别率和拒绝率都为 100%, 但自身验证时未能完全准确鉴别 *Gametis* 属花金龟, 因此花金龟亚科 3 个属的鉴别模型可以推广应用于鉴别 *Cetonia* 属和 *Protaetia* 属花金龟, 但对能否准确鉴别 *Gametis* 属花金龟有待商榷。除此之外所建的各个模型的预测均显示了理想的结果, 识别率和拒绝率均为 100%。

基于已定的金龟子所建立的模型库均可很好地鉴别除云斑鳃金龟 *P. laticolli* 和小青花金龟 *G. jucunda* 之外的 13 种金龟子。因此采取近红外光谱扫描技术结合支持向量机得出的植食性金龟子鉴别模型具有很强的推广能力。

模型的自身交叉验证与预测集验证显示 4 个属内不同种的模型准确率均达到 100%, 这说明此方法对属内不同种的鉴别更具有优势。因此

采取近红外光谱扫描技术结合支持向量机算法得出的植食性金龟子鉴别模型具有很强的推广能力。

3 讨论

近红外光谱 (NIR) 分析技术是研究昆虫代谢物质的重要方法, 与液相色谱和质谱等技术相比近红外光谱分析技术具有快速、无污染、无损坏、和成本低的特点(Paliwal *et al.* 2004 ;Munck , 2007)。支持向量机算法能较好地解决非线性、小样本、高维数和局部极小值的模式识别与评价预测等问题。本文采用近红外光谱技术结合支持向量机判定方法对 15 种为害较重的植食性金龟子建立了鉴别模型库, 该模型能正确识别大部分种类, 植食性金龟子鉴别模型的建立不仅满足了害虫鉴别的简便、快速、准确的要求, 而且大大节

表 5 基于支持向量机的金龟子近红外光谱鉴别模型的预测
Table 5 NIR discriminant models' projections based on SVM of chafers

模型 Model	模型的预测 Forecast of the models		
	分类单元 Taxon	识别率 (%) Recognition rate	拒绝率 (%) Rejection rate
金龟科 4 个亚科 4 subfamilies of family Scarabaeidae	Melolonthinae	100	100
	Dynastinae	100	100
	Rutelinae	100	100
	Cetoniinae	100	100
	<i>Polyphylla</i>	0	100
丽金龟亚科 3 个属 3 genera of subfamily Rutelinae	<i>Hoplosternus</i>	100	92
	<i>Holotrichia</i>	100	100
	<i>Maladera</i>	100	100
	<i>Blitopertha</i>	100	100
鳃金龟亚科 4 个属 4 genera of subfamily Melolonthinae	<i>Mimela</i>	100	100
	<i>Anomala</i>	100	100
	<i>Cetonia</i>	100	100
	<i>Protaetia</i>	100	100
花金龟亚科 3 个属 3 genera of subfamily Cetoniinae	<i>Gametis</i>	100	100
	<i>H. parallela</i>	100	100
	<i>H. diomphalia</i>	100	100
<i>Maladera</i> 属 2 个种 2 species of genus <i>Maladera</i>	<i>M. verticalis</i>	100	100
	<i>M. orientalis</i>	100	100
<i>Anomala</i> 属 2 个种 2 species of genus <i>Anomala</i>	<i>A. mongolica</i>	100	100
	<i>A. chamaeleon</i>	100	100
<i>Protaetia</i> 属 2 个种 2 species of genus <i>Protaetia</i>	<i>P. cathaica</i>	100	100
	<i>P. brevitarsis</i>	100	100

省了人工与时间,降低了鉴别成本。因此为解决害虫快速检测与检疫提供了一种新的研究方法。

本文所建立的植食性金龟子鉴别模型中,鳃金龟亚科 4 个属的鉴别模型对 *Polyphylla* 属(云斑鳃金龟 *P. laticollis*) 未能够正确识别,样本被错误识别到 *Hoplosternus* 属中,其中一个原因是预测集样本量只有 1 头而无法准确鉴别,另外 *Hoplosternus* 属与 *Polyphylla* 属金龟子的鞘翅都覆有黄色或白色的鳞片,这 2 个类群的鞘翅形态结构很像也有可能造成被错误识别。

花金龟亚科 3 个属的鉴别模型自身检验时对 *Gametis* 属的识别率只有 33%,此处小青花金龟的样本量足够,交叉验证显示只准确识别了 33%的样本量,而形态上 *Protaetia* 属和 *Gametis* 属的差异比较大,亲缘关系也不近,因此此处可能的原因是模型所确定的核参数不适用与 *Gametis* 属与其他 2 个属的区分,需要进一步分析。

本文中利用近红外光谱技术扫描昆虫时需要注意光纤探头要与昆虫的体壁贴紧以免有空隙影响光谱吸光度。光谱预处理和所选取的分析方法是影响鉴别模型的最主要的因素,需要根据标本量的多少和种类的不同来确定。标本保存的质量也是影响扫描结果的因素之一。实验中所用的金龟子都是随机选取的标本,对于鉴别过程中雌雄标本、不同羽化时期、不同地理种群等因素的影响需要继续研究分析。

参考文献 (References)

- Aldrich BT, Maghirang EB, Dowell FE, Kambhampati S, 2007. Identification of termite species and subspecies of the genus *Zootermopsis* using near-infrared reflectance spectroscopy. *Journal of Insect Science*, 7(1): 1–18.
- Antoniali WF Jr, Suárez YR, Izida T, Andrade LHC, Lima SM, 2008. Intra-and interspecific variation of cuticular hydrocarbon composition in two *Ectatomma* species (Hymenoptera: Formicidae) based on fourier transform infrared photoacoustic spectroscopy. *Genetics and Molecular Research*, 7(2): 559–566.
- Baker JE, Dowell FE, Throne JE, 1999. Detection of parasitized rice weevils in wheat kernels with near-infrared spectroscopy. *Biological Control*, 16(1): 88–90.
- Barnes RJ, Dhanoa MS, Lister SJ, 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5): 772–777.
- Carlson DA, 1979. Differentiation between species of the *Anopheles gambiae* Giles complex (Diptera: Culicidae) by analysis of cuticular hydrocarbons. *Annals of Tropical Medicine and Parasitology*, 73(6): 589–592.
- Chen T, Liu QJ, Wu YW, Peng SL, Jiang A, 2015. A study of edible oil infrared spectroscopy method based on statistical learning theory. *Journal of Food Safety and Quality*, 6(3): 836–842. [陈婷, 刘清珺, 武彦文, 彭思龙, 姜安, 2015. 基于统计学习理论的食用油红外光谱分析方法研究. 食品安全质量检测学报, 6(3): 836–842.]
- Chu XL, Yuan HF, Lu WZ, 2004. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique. *Progress in Chemistry*, 4(16): 528–542. [褚小立, 袁洪福, 陆婉珍, 2004. 近红外分析中光谱预处理及波长选择方法进展与应用. 化学进展, 4(16): 528–542.]
- Cole TJ, Ram MS, Dowell FE, Omwega CO, Overholt WA, Ramaswamy SB, 2003. Near-infrared spectroscopic method to identify *Cotesia flavipes* and *Cotesia sesamiae* (Hymenoptera: Braconidae). *Annals of the Entomological Society of America*, 96(6): 865–869.
- Feng GH, 2011. Parameter optimizing for support vector machines classification. *Computer Engineering and Applications*, 47(3): 123–124. [奉国和, 2011. SVM 分类核函数及参数选择比较. 计算机工程与应用, 47(3): 123–124.]
- Gao MY, 2011. Application of cuticular hydrocarbon analysis to insect taxonomy. *Acta Entomologica Sinica*, 44(1): 119–122. [高明媛, 2001. 昆虫表皮碳氢化合物在昆虫分类中的应用. 昆虫学报, 44(1): 119–122.]
- Howard RW, Blomquist GJ, 2005. Ecological, behavior, and biochemical aspects of insect hydrocarbons. *Annual Review of Entomology*, 50: 371–393.
- Huang HB, Yu HY, Xu HR, Ying YB, 2008. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *Journal of Food Engineering*, 87(3): 303–313.
- Jaime IR, Claudio JB De C, Celio P, Kássio MG De L, Mauricio OM, Gregorio CA, 2011. Barcoding without DNA? Species identification using near infrared spectroscopy. *Zootaxa*, 2011(4): 1–9.
- Ju LF, Wang QJ, Li GL, Hu CG, Qian Z, 2014. Parameter optimization for support vector machine model of permanent magnet spherical motors. *Transactions of China Electrotechnical Society*, 29(1): 85–90. [鞠鲁峰, 王群京, 李国丽, 胡存刚, 钱喆, 2014. 永磁球形电机的支持向量机模型的参数寻优. 电工技术学报, 29(1): 85–90.]

- Klarica J, Bittner L, Pallua J, Pezzeri C, Huck-Pezzeri V, Dowell F, Schied J, Bonn GK, Huck C, Schlick-Steiner BC, Steiner FM, 2011. Near-infrared imaging spectroscopy as a tool to discriminate two cryptic *Tetramorium* ant species. *Journal of Chemical Ecology*, 37(6): 549–552.
- Lazzari SMN, Ceruti FC, Rodriguez-Fernandez JJ, Opat G, Lazzari FA, 2010. Intra and interspecific variation assessment in Psocoptera using near spectroscopy. *Julius-Kühn-Archiv*, 425: 139–144.
- Liu GR, Zhang YW, Wang R, 1997. The Colour Illustrated of Common Lamellicornia Beetles of Northern China. Beijing: China Forestry Publishing House. 1–98. [刘广瑞, 章有为, 王瑞, 1997. 中国北方常见金龟子彩色图鉴. 北京: 中国林业出版社. 1–98.]
- Lu WZ, 2007. Modern Near Infrared Spectroscopy Analytical Technology. Beijing: China Petrochemical Press. 226–249. [陆婉珍, 2007. 现代近红外光谱分析技术(第二版). 北京: 中国石化出版社. 246–249.]
- Maghirang EB, Dowell FE, Baker JE, Throne JE, 2003. Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy. *Transactions of the ASAE*, 46(4): 1277–1282.
- Mayagaya VS, Michel K, Benedict MQ, Killeen GF, Wirtz RA, Ferguson HM, Dowell FE, 2009. Non-destructive determination of age and species of *Anopheles gambiae* s.l. using near-infrared spectroscopy. *American Journal of Tropical Medicine and Hygiene*, 81(4): 622–630.
- Munck L, 2007. A new holistic exploratory approach to systems biology by near infrared spectroscopy evaluated by chemometrics and data inspection. *Journal of Chemometrics*, 21(10/11): 406–426.
- Paliwal J, Wang W, Symons SJ, Karunakaran C, 2004. Insect species and infestation level determination in stored wheat using near-infrared spectroscopy. *Canadian Biosystems Engineering*, 46(7): 17–24.
- Perez-Mendoza J, Dowell FE, Broce AB, Throne JE, Wirtz RA, Xie F, Fabrick JA, Baker JE, 2002. Chronological age-grading of house flies by using near-infrared spectroscopy. *Journal of Medical Entomology*, 39(3): 499–508.
- Perez-Mendoza J, Throne JE, Dowell FE, Baker JE, 2004. Chronological age-grading of three species of stored-product beetles by using near-infrared spectroscopy. *Journal of Economic Entomology*, 97(3): 1159–1167.
- Sieglwart M, Bouvier F, Maugin S, Lecomte A, Lavigne C, 2015. Differentiating oriental fruit moth and codling moth lepidoptera: tortricidae larvae using near-infrared spectroscopy. *Journal of Economic Entomology*, 108(1): 219–227.
- Teng LS, Wang D, Song J, Zhang YB, Guo WL, Teng LR, 2008. Application of near infrared spectroscopy in rapid and simultaneous determination of essential components in five varieties of anti-tuberculosis tablets. *Spectroscopy and Spectral Analysis*, 28(8): 1814–1818. [滕乐生, 王迪, 宋佳, 张益波, 郭伟良, 滕利荣, 2008. 近红外光谱法同时快速测定五种抗结核药物的主成分含量. 光谱学与光谱分析, 28(8): 1814–1818.]
- Tian YW, Niu Y, 2009. Applied research of support vector machine on recognition of cucumber disease. *Journal of Agricultural Mechanization Research*, 31(3): 36–39. [田有文, 牛妍, 2009. 支持向量机在黄瓜病害识别中的应用研究. 农机化研究, 31(3): 36–39.]
- Tian YW, Zhang CS, Li CH, 2004. Application of support vector machine to shape recognition of plant disease spot. *Transactions of The Chinese Society of Agricultural Engineering*, 20(3): 134–136. [田有文, 张长水, 李成华, 2004. 支持向量机在植物病斑形状识别中的应用研究. 农业工程学报, 20(3): 134–136.]
- Wang L, He Y, Liu F, Ying XF, 2008. Rapid detection of sugar content and pH in beer by using spectroscopy technique combined with support vector machines. *Journal of Infrared and Millimeter Waves*, 27(1): 51–55. [王莉, 何勇, 刘飞, 应霞芳, 2008. 应用光谱技术和支持向量机分析方法快速检测啤酒糖度和 pH 值. 红外与毫米波学报, 27(1): 51–55.]
- Wolf M, Ferrari M, Quaresima V, 2007. Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *Journal of Biomedical Optics*, 12(6): 062104. Doi: 10.1117/1.2804899.
- Yan YL, Chen B, Zhu DZ, 2013. Near Infrared Spectroscopy-Principles, Technologies and Applications. Beijing: China Light Industry Press. 1–48. [严衍禄, 陈斌, 朱大洲, 2013. 近红外光谱分析的原理、技术与应用. 北京: 中国轻工业出版社. 1–48.]
- Zhao QJ, Cao Y, Zhou YL, Li CY, Zhang YJ, Guo YY, 2014. Using spectral reflectance in the field to determine the resistance of cotton varieties to *Apolygus lucorum*. *Chinese Journal of Applied Entomology*, 50 (6): 1660–1664. [赵秋剑, 曹杨, 周延乐, 李长友, 张友军, 郭予元, 2013. 棉花种抗绿盲蝽特性的田间光谱诊断. 应用昆虫学报, 50(6): 1660–1664.]
- Zheng F, Sheinberg R, Yee MS, Ono M, Zheng YY, Hogue CW, 2013. Cerebral near-infrared spectroscopy NIRS monitoring and neurologic outcomes in adult cardiac surgery patients and neurologic outcomes: a systematic review. *Anesthesia and Analgesia*, 116(3): 663–676.