

支持向量机算法在激光诱导击穿光谱技术塑料识别中的应用研究*

于洋¹⁾ 郝中骥¹⁾ 李常茂¹⁾ 郭连波¹⁾ 李阔湖¹⁾ 曾庆栋¹⁾
李祥友^{1)†} 任昭²⁾ 曾晓雁¹⁾

1) (华中科技大学, 武汉光电国家实验室(筹) 激光与太赫兹功能实验室, 武汉 430074)

2) (武汉新瑞达激光工程有限责任公司, 武汉 430074)

(2013年6月24日收到; 2013年7月25日收到修改稿)

基于支持向量机 (support vector machines, SVM) 算法采用激光诱导击穿光谱技术对 11 种塑料进行了识别. 每种塑料各采集 100 个光谱, 其中 50 个光谱作为训练集, 用于建立支持向量机模型, 剩下的 50 个光谱作为测试集, 用于测试所建立支持向量机模型的识别精度. 结果表明测试集 550 个光谱中有 543 个光谱识别正确, 算术平均识别精度达到了 98.73%. 其中有 6 个聚氨酯 (PU) 光谱被误判为有机玻璃 (PMMA), 原因主要是受空气中氮气的影响, 使得有机玻璃和聚氨酯两种塑料在氮元素含量上的差异不能通过 NI 746.87 nm, C-N(0,0) 388.3 nm 两条谱线的强度准确表征. 本结果为 LIBS 技术塑料分类提供了方法和数据参考.

关键词: 支持向量机, 激光诱导击穿光谱, 塑料识别

PACS: 52.38.Mf, 61.80.Ba, 61.41.+e, 61.82.Pv

DOI: 10.7498/aps.62.215201

1 引言

随着塑料制品消费量不断增大, 废弃塑料也不断增多, 为回收资源和保护环境, 塑料分类成为亟需解决的问题. 目前, 主要塑料分类方法有人工分类^[1]、X 射线荧光光谱法 (XRF)^[2,4] 和近红外光谱法 (NIR)^[2-4] 等. 其中, 人工分类法成本较高、识别速度慢并且容易出现错误; X 射线荧光光谱法能有效探测到氯元素, 因此被特别用于识别聚氯乙烯 (PVC) 塑料^[4]; 近红外光谱法是通过探测红外光作用到塑料上所产生的反射光或者透射光来决定塑料的分子结构^[3,5], 从而实现塑料种类的识别, 具有精度高、速度快的优点. 但是这种方法对黑色塑料的识别效果较差并且光谱信号容易受到塑料表面杂质的影响^[1,5]. 作为一种新的塑料识别方

法, 激光诱导击穿光谱 (LIBS) 技术^[6-16] 具有识别速度快、样品不需要预处理、对样品损伤小、不受限于塑料的颜色等优点. 但是目前 LIBS 技术本身的稳定性不高, 利用该技术进行塑料识别, 其精度还不能满足工业应用的需求. 这使得采用辅助方法提高 LIBS 技术塑料识别精度成为当前研究的热点. 例如, 王茜萋等^[17] 将主成分分析法 (principle component analysis, PCA) 与反向传播人工神经网络 (back-propagation artificial neural network) 两种算法相结合, 对 7 种塑料的平均识别精度为 97.5%. Sattmann 等^[18] 将 LIBS 技术与前馈神经网络 (feed forward artificial neural networks) 相结合对 4 种塑料进行识别, 平均识别精度为 77.2%. Boueri 等^[19] 采用宽谱带的 LIBS 系统 (240—820 nm) 与人工神经网络 (ANNs) 算法相结合对 8 种塑料进行识别, 平均识别精度为 95.87%. 以上工作获得了较高的塑

* 国家重大科学仪器设备开发专项 (批准号: 2011YQ160017)、国家自然科学基金 (批准号: 51128501) 和中央高校基本科研业务费 (批准号: CXY13Q022, CXY13Q021) 资助的课题.

† 通讯作者. E-mail: xyli@mail.hust.edu.cn

料识别精度,但是所用塑料种类普遍较少,且未涉及支持向量机 (support vector machines, SVM) 算法结合 LIBS 用于塑料分类的研究.

相对于 PCA 法, SVM 算法和 ANNs 算法都属于有目标输出的学习型算法,分类精度较高. 其中 ANNs 算法根据分类问题的复杂程度,可以选择相应的网络层数和神经元的数量,具有自学习和自适应的能力. 但针对不同分类任务, ANNs 算法存在过学习、欠学习等问题,并且针对相同数据进行多次训练的结果有较大差异,因此对测试集的预测结果不稳定. 而 SVM 算法不存在网络结构选择、过学习和欠学习等问题;算法的实现只需要通过求解简单的优化问题,训练速度快;每次训练-预测的结果较为稳定;在解决小样本、非线性及高维模式识别问题中表现良好. 但当样本规模较大、不同类别样本的数量分布很不均衡——数据集偏斜,会导致 SVM 分类效果较差. 本文将利用 SVM 算法结合 LIBS 技术对 11 种塑料进行分类.

2 实验方法与支持向量机 (SVM) 算法

2.1 实验方法

实验装置如图 1 所示,采用的激光器为调 Q 开关 Nd:YAG 脉冲激光器 (激光波长 532 nm, 最大重复频率 10 Hz, 脉宽 5 ns). 激光束经反射镜和平凸透镜 (焦距 $f = 150$ mm) 后聚焦到塑料样品表

面. 所激发等离子体发射光由光收集器收集并耦合到直径 50 μm 、长 2 m 的光纤中,经光纤传输至光谱仪 (Andor ME 5000, 配备 Andor DH334T ICCD, 波长范围 230—850 nm, 分辨率 $\lambda/\Delta\lambda = 5000$) 中进行谱线分析. 样品放置在电脑控制的二维电动平台上,在光谱采集过程中保持样品做“弓”字形运动,确保激光脉冲均作用到样品表面的一个新位置.

实验第一步是采用标准汞光源对光谱仪进行校准,校准后波长误差小于 0.05 nm. 实验第二步是以获得最佳光谱信号背景比为标准对实验参数进行优化:激光单脉冲能量优化后设置为 49 mJ (激光能量为 49 mJ 时,统计 300 个激光脉冲获得的 RSD 为 1.977),考虑到反射镜和聚焦透镜对激光的反射,最终到达样品表面的激光能量为 44 mJ. 激光束焦点位于样品表面下方 1 mm 处,烧蚀坑直径约为 100 μm ,能量密度约为 140 J/cm². 优化后的 ICCD 门宽和延时分别为 1 μs 和 1.1 μs . 每种塑料采集 100 个光谱,每个光谱累积 30 个激光脉冲,以降低激光脉冲能量波动对谱线强度的影响.

实验样品包括 11 种塑料:聚四氟乙烯 (PTFE),聚苯乙烯 (PS),聚碳酸酯 (PC),聚氨酯 (PU),工程塑料 (ABS),聚氯乙烯 (PVC),有机玻璃 (PMMA),尼龙 -6 (PA-6),聚丙烯 (PP),聚乙烯 (PE) 和聚甲醛 (POM). 表 1 列出了上述 11 种塑料的分子式和颜色.

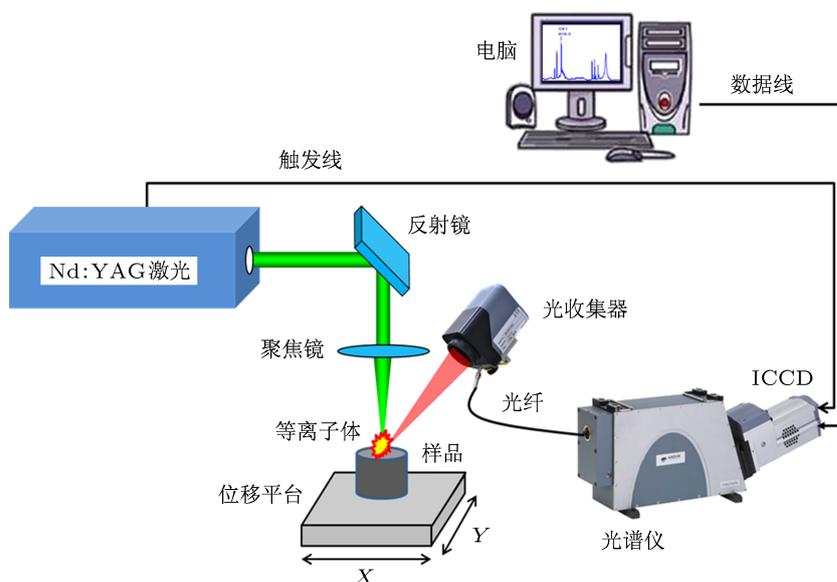


图 1 实验装置简图

表 1 11 种塑料材料的分子式和颜色

样品	分子式	颜色	样品	分子式	颜色
PTFE	[C ₂ F ₄] _n	白色不透明	PMMA	[C ₅ H ₈ O ₂] _n	无色透明
PU	[CHNO ₂] _n	黄褐色透明	PA-6	[C ₆ H ₁₁ ON] _n	淡黄色透明
PS	[C ₈ H ₈] _n	无色透明	PP	[CH ₂ CH(CH ₃)] _n	白色不透明
PC	[C ₁₆ H ₁₄ O ₃] _n	淡黄色透明	PE	[C ₂ H ₄] _n	白色不透明
ABS	[C ₃₈ H ₆₄ N ₂] _n	黄色不透明	POM	[CH ₂ O] _n	白色不透明
PVC	[C ₂ H ₃ Cl] _n	黑色不透明			

2.2 SVM 算法

SVM^[20,21,23] 算法是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的一种可训练的机器学习算法. 本文采用 Lin 等^[22] 开发的支持向量机软件工具箱 (A Library for Support Vector Machines, LIBSVM) 进行数据处理. 将分类数据分成训练集和测试集两部分, 为训练集和测试集中同种塑料的每组属性值设定相同的标签, 训练集中每个实例就包括一组属性值和一个标签. SVM 的识别过程是先根据训练集中的属性值和标签建立 SVM 识别模型, 然后, 依据 SVM 识别模型由测试集属性值预测出其对应标签. SVM 训练的数学过程如下: 设训练集中每个实例为 \mathbf{X}_i, y_i , 其中 $i = 1, \dots, 550$ $\mathbf{X}_i(x_{i1}, x_{i2}, \dots, x_{il}) \in \mathbf{R}^n$ 对应于属性值, $y_i \in \{1, 2, \dots, 10, 11\}$ ⁵⁵⁰ 对应于每组属性值的标签. 低维空间中线性不可分的属性值向量 $\mathbf{X}_i(x_{i1}, x_{i2}, \dots, x_{il})$ 通过函数 $\varphi(x) = (\varphi_1(x), \varphi_2(x), \dots)$ 映射到一个高维空间中. 这里 $\varphi(x)$ 由径向基函数 (radial basis function, RBF) 确定:

$$K(\mathbf{X}_i, \mathbf{X}_j) = \varphi(\mathbf{X}_i)^T \varphi(\mathbf{X}_j) = \exp(-g \|\mathbf{X}_i - \mathbf{X}_j\|^2), g > 0, \quad (1)$$

g 是 RBF 核函数的变量. 在该高维空间中, 映射后的不同塑料属性值向量 $\varphi(\mathbf{X}_i)$ 变得线性可分. SVM 在高维空间中构造最优分类超平面, 使得不同塑料 $\varphi(\mathbf{X}_i)$ 的分布区之间的间隔最大化. 因此, SVM 需要求解下面拉格朗日函数 (Lagrangian) 的最优化问题:

$$\min_{W, b, \xi} \left(\frac{1}{2} W^T W + C \sum_{i=1}^{550} \xi_i \right), \quad (2)$$

$$\text{约束条件: } \begin{cases} y_i \cdot (W^T \varphi(\mathbf{X}_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, l, \end{cases}$$

其中, W, b 分别为最优分类超平面的法向量和常数项. c, ξ_i, g 分别为惩罚参数、松弛变量和可变常数项. (2) 式的对偶形式:

$$\min_{\alpha} \left(\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \varphi(\mathbf{X}_i)^T \varphi(\mathbf{X}_j) - \sum_{i=1}^l \alpha_i \right), \quad (3)$$

$$\text{约束条件: } \begin{cases} 0 \leq \alpha_i \leq C, i = 1, \dots, l, \\ \sum_{i=1}^l \alpha_i y_i = 0, \end{cases}$$

其中, $W = \sum_{i=1}^l \alpha_i y_i \varphi(\mathbf{X}_i)$, α_i 为拉格朗日乘子. 最终, 非线性 SVM 分类器方程

$$F(\mathbf{X}) = \text{sign} \left(\sum_{i=1}^l y_i \alpha_i \varphi(\mathbf{X}_i)^T \varphi(\mathbf{X}) + b \right), \quad (4)$$

参数 C, g 对最优分类超平面的求解有重要影响, 故选择合适的 C, g 参数才能获得最佳的预测结果. 找到最优分类超平面, 即建立了 SVM 识别模型, 基于该识别模型可对测试集每组属性值的标签进行预测.

2.3 交互校验法 (cross-validation method)

根据训练集数据, 采用交互校验算法来获得合适的 C, g 参数. 设置 C, g 的选择范围: $C = 2^i, g = 2^j, i, j \in \{-10, -9, \dots, 9, 10\}$. 在某组 (C_i, g_j) 参数下, 将训练集平均分成 5 个子集, 其中 4 个子集用于训练 SVM 模型, 并用所建立 SVM 模型对剩下 1 个子集的标签进行预测, 共有 5 种“4+1”的组合方式, 分别得出每种组合方式下 4 个子集建立的 SVM 模型对剩下 1 个子集标签的识别精度并取平均, 从而获得对应于该组 (C_i, g_j) 数值的一个平均识别精度. 在 (C_i, g_j) 的取值范围内, 最大的平均识别精度对应的 (C_i, g_j) 数值便是获得的最佳 C, g 参数.

3 实验结果与分析

3.1 特征光谱的选取

以美国 NIST 基本原子光谱数据库为参照, 结合 Andor 软件自带的光谱数据库, 对 11 种塑料的等离子体发射光谱 (见图 2) 做定性分析. 在光谱仪系统探测波长 (230—850 nm) 范围内, 能够探测到主要的 12 种元素和两种分子谱带, 分别是 C, H, O, N, Mg, Ti, Ca, Al, K, Na, Cl, F, C-N 和 C-C. 由于不同塑料中以上元素的含量不同, 相应特征谱线的强度也有差异, 这样就可以通过特征谱线强度把不同塑料区分开. 特征谱线的选取需要满足两个条件:

1) 谱线强度大; 2) 自吸收小、谱线重叠少. 最初选择满足条件的 24 条谱线^[24] 采用自主开发的读光谱软件, 以美国 NIST 的标准谱线波长为中心波长, 读取 ± 0.1 nm 范围内 (考虑到温度波动引起谱线波长微小偏移) 的最大强度作为谱线强度. 采用交互验证法逐个测试这些谱线对分类结果的影响, 去除 8 条对分类结果影响最小的谱线, 剩余的 16 条谱线作为特征谱线 (见表 2). 每个光谱中 16 条特征谱线的强度均除以 C I 247.86 nm 谱线强度来做归一化处理. 最终得到对应于 11 种塑料的 1100 组分类数据 (每种塑料 100 组分类数据), 每组分类数据包括 15 个属性 (见表 2), 即对应于 SVM 的输入: $X_i(x_{i1}, x_{i2}, \dots, x_{i14}, x_{i15})$.

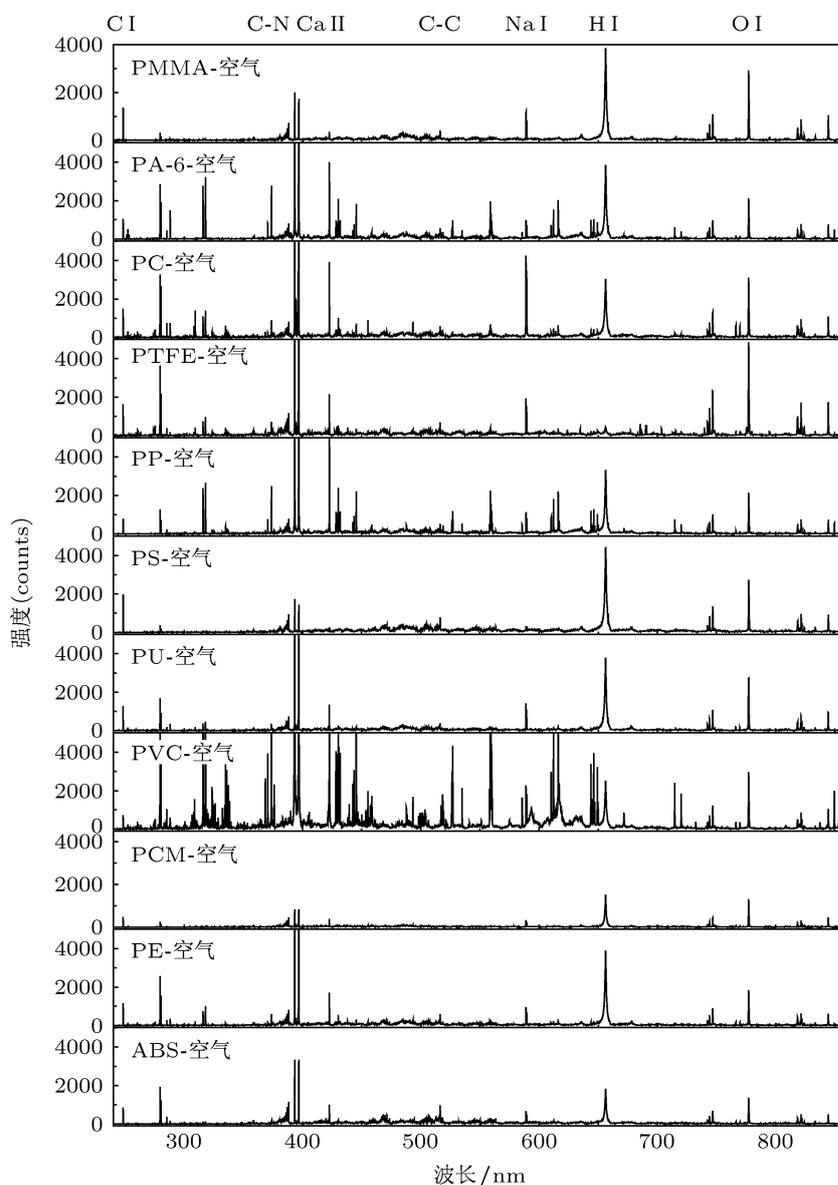


图2 空气环境下 11 种塑料的等离子体发射光谱

表 2 特征谱线及属性

特征谱线	波长/nm	属性	特征谱线	波长/nm	属性
C I	247.86		H I	656.3	H I/C
Mg II	279.55	Mg II/C	FI	685.7	FI: 685.7/C
Al I	309.27	Al/C	Cl I	725.7	Cl I: 725.7/C
Ti II	334.94	Ti II/C	FI	739.9	FI: 739.9/C
C-N(0, 0)	388.3	C-N/C	NI	746.9	NI/C
Ca II	393.34	Ca II/C	K I	766.5	K I/C
C ₂ (0, 0)	516.5	C ₂ (0,0)/C	O I	777.3	O I/C
Na I	589.06	Na/C	Cl I	837.59	Cl I: 837.59/C

3.2 SVM 算法塑料识别

使用交互校验法对 C, g 参数进行优化后分别

设置为 1024 和 0.25. 将每种塑料 100 组 LIBS 光谱数据平均分成测试集和训练集, 共 11 种塑料, 故训练集和测试集的 LIBS 光谱数据各为 550 组. 用训练集所建立的 SVM 模型对测试集 550 组光谱数据的标签进行预测. 结果如图 3 所示, “○” 代表每组数据的实际标签, “*” 代表预测标签, “*” 与 “○” 重合表示预测正确. 结果表明, 测试集 550 个光谱中有 6 个 PU 光谱被误判为 PMMA, 1 个 PA-6 光谱被误判为 PE, 其余 543 个光谱均预测正确, 因此 PU 和 PA 的预测准确率分别为 88% 和 98%, 其他 9 种塑料的预测准确率均为 100%, 算术平均识别精度达到了 98.73%.

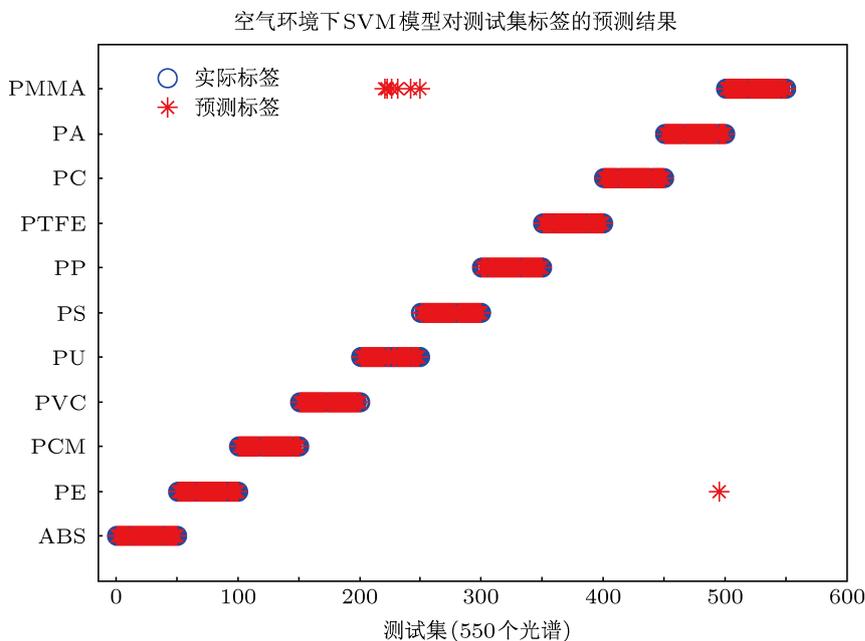


图 3 SVM 模型对测试集 11 种塑料共 550 个光谱的识别结果 (横轴代表测试集 550 个光谱, 纵轴代表每个光谱的标签)

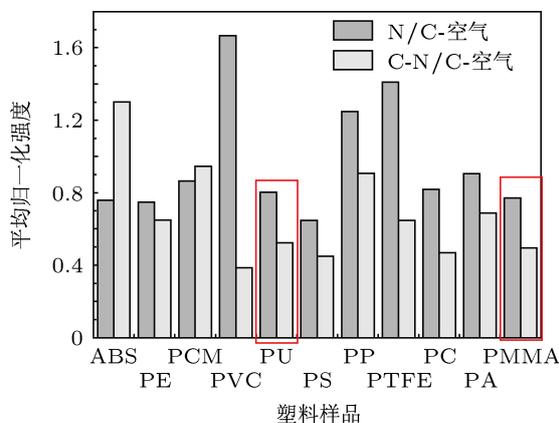


图 4 空气环境下 11 种塑料等离子体发射谱中 NI 746.87 nm 和 C-N(0, 0) 388.3 nm 两条谱线的归一化强度

尽管 PU 和 PMMA 在组成成分上有较大差别

(PU 中含有氮元素, 而 PMMA 中不含氮元素), 6 个 PU 光谱误判为 PMMA 的原因主要是受空气中氮气的影响, 使得它们在氮元素含量上的差别不能通过 C-N(0, 0) 388.3 nm 和 NI 746.87 nm 两条谱线的强度准确表征出来. 图 4 所示聚氨酯和有机玻璃的等离子体发射谱中, C-N(0, 0) 388.3 nm 和 NI 746.87 nm 两条谱线的归一化强度几乎相同, 故难以通过以上两条谱线对 PU 和 PMMA 进行区分.

4 结论

基于 SVM 算法采用激光诱导击穿光谱技术对 11 种塑料进行了分类识别, 测试集 550 个光谱中有

543 个识别正确, 算术平均识别精度为 98.73%. PU 和 PMMA 两种塑料在金属元素含量上的差异较小, 尽管在氮元素含量上的差异较大, 但受空气中氮气的影 响, 使得 C-N (0, 0) 388.3 nm 和 N I 746.87 nm 两条谱线的强度不能有效表征两种塑料在氮元素含量上的差异. 最终导致两种塑料的属性向量在高维特征空间中的分布区较为接近, 因此导致 SVM

将 6 个 PU 光谱误判为 PMMA. 如何进一步消除误判、提高分类精度, 将是下一步研究的重点. SVM 算法计算简单、训练速度快并且每次训练的结果较为稳定, 对 11 种塑料的分类精度较高. 由此可见, 结合 SVM 算法的 LIBS 技术应用于塑料识别具有良好的前景.

- [1] Bruno E A <http://infohouse.p2ric.org/ref/09/08620.pdf> [accessed April 26 2013]
- [2] Chanda M, Roy S K 1997 *Polymers Technology Handbook* 4th Edn. (New York: Marcel Dekker)
- [3] Van Den Broek W H A M, Wienke D, Melssen W J, Buydens L M C 1998 *Anal. Chim. Acta* **361** 161
- [4] Dinger, Peter 1992 *Journal of Composting & Organics Recycling* **33** 80
- [5] Eisenreich Norbert, Rohe Thomas 1996 *Kunststoffe* **2** 222
- [6] Cremers D A, Radziemski L J 2006 *Handbook of Laser-Induced Breakdown Spectroscopy* (Chichester: Wiley)
- [7] Singh J P, Thakur S N 2007 *Laser-induced breakdown spectroscopy* (The Netherlands: Elsevier)
- [8] Zhang D C, Ma X W, Wen W Q, Zhang P J, Zhu X L, Li B, Liu H P 2010 *Chin. Phys. Lett.* **27** 063202
- [9] Nakimana Agnes, Tao H Y, Hao Z Q, Sun C K, Gao X, Lin J Q 2013 *Chin. Phys. B* **22** 014209
- [10] Ma X W, Zhu X L, Zhang D C, Zu K L, Li B 2008 *Acta Phys. Sin.* **57** 6348 (in Chinese) [马新文, 朱小龙, 张大成, 祖凯玲, 李斌 2008 物理学报 **57** 6353]
- [11] Rusak D A, Weaver K D, Taroli B L 2008 *Appl. Spectrosc.* **62** 773
- [12] Wang C L, Liu J G, Zhao N J, Ma M J, Wang Y 2013 *Acta Phys. Sin.* **62** 125201 (in Chinese) [王春龙, 刘建国, 赵南京, 马明俊, 王寅 2013 物理学报 **62** 125201]
- [13] Stepputat M, Noll R 2003 *Applied Optics* **42** 6210
- [14] Moench I, Sattmann R, Noll R 1997 *Proc. SPIE 3100, Sensors, Sensor Systems, and Sensor Data Processing* Munich, Germany, June 16, 1997 p64
- [15] Guo L B, Hu W, Zhang B Y, He X N, Li C M, Zhou Y S, Cai Z X, Zeng X Y, Lu Y F 2011 *Opt. Express* **19** 14067
- [16] Du C, Gao X, Shao Y, Song X W, Zhao Z M, Hao Z Q, Lin J Q 2013 *Acta Phys. Sin.* **62** 045202 (in Chinese) [杜闯, 高勋, 邵妍, 宋晓伟, 赵振明, 郝作强, 林景全 2013 物理学报 **62** 045202]
- [17] Wang Q Q, Huang Z W, Liu K, Li W J, Yan J X 2012 *Spectroscopy and Spectral Analysis* **32** 3179 (in Chinese) [王茜茜, 黄志文, 刘凯, 李文江, 阎吉祥 2012 光谱学与光谱分析 **32** 3179]
- [18] Sattmann R, Monch I, Krause H, Noll R, Couris S, Hatziaepostolou A, Mavromanolakis A, Fotakis C, Larrauri E, Miguel R 1998 *Appl. Spectrosc* **52** 456
- [19] Boueri M, Motto-Ros V, Lei W Q, Ma Q L, Zheng L J, Zeng H P, Yu J 2011 *Appl. Spectrosc* **65** 307
- [20] Boser B E, Guyon I, Vapnik V 1992 *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory* USA, 1992, p144
- [21] Cortes C, Vapnik V 1995 *Machine Learning* **20** 273
- [22] Chang C C, Lin C J 2011 *ACM Transactions on Intelligent Systems and Technology* **2** 1
- [23] We C H, Tzeng G H, Goo Y J, Fang W C 2007 *Expert. Syst. Appl.* **32** 397
- [24] Boueri M 2010 *Ph.D. Dissertation* (Lyon: Universite de Lyon) (in France)

Identification of plastics by laser-induced breakdown spectroscopy combined with support vector machine algorithm*

Yu Yang¹⁾ Hao Zhong-Qi¹⁾ Li Chang-Mao¹⁾ Guo Lian-Bo¹⁾ Li Kuo-Hu¹⁾
Zeng Qing-Dong¹⁾ Li Xiang-You^{1)†} Ren Zhao²⁾ Zeng Xiao-Yan¹⁾

1) (Wuhan National Laboratory for Optoelectronics, Laser and Tera-Hertz Technology Division, Huazhong University of Science and Technology, Wuhan 430074, China)

2) (Wuhan New Research and Development Laser Engineering Co., Ltd, Wuhan 430074, China)

(Received 24 June 2013; revised manuscript received 25 July 2013)

Abstract

Laser-induced breakdown spectroscopy (LIBS) combined with support vector machine (SVM) algorithm was used to identify 11 kinds of plastics. For each plastic, 100 spectra recorded by the spectrometer system were divided equally into training set and test set, and the former was used to train SVM model while the latter was used to validate SVM model created by the training set. Result shows that 543 of 550 test set spectra are identified correctly with the average correct identification rate 98.73%. However, there are six spectra of PU misidentified as PMMA. This is because the difference of nitrogen content in 11 plastics cannot be reflected by the intensities of N I 746.87 nm and C-N (0,0) 388.3 nm due to the influence of ambient air. Methods and reference data are provided for further study of plastics identification by laser-induced breakdown spectroscopy technique.

Keywords: laser-induced breakdown spectroscopy, support vector machine, plastics identification

PACS: 52.38.Mf, 61.80.Ba, 61.41.+e, 61.82.Pv

DOI: 10.7498/aps.62.215201

* Project supported by the National Special Fund for the Development of Major Scientific Instruments and Equipments, China (Grant No. 2011YQ160017), the National Natural Science Foundation of China (Grant No. 51128501), and both the Fundamental Research Funds for the Central Universities, HUST, China (Grant Nos. CXY13Q022, CXY13Q021).

† Corresponding author. E-mail: xyli@mail.hust.edu.cn