

近红外光谱法定性描述酵母菌的生长过程

王 玮 江 辉* 刘国海 梅从立 吉 奕

(江苏大学电气信息工程学院, 镇江 212013)

摘 要 提出了一种基于近红外光谱分析技术的酵母菌生长过程描述方法。利用 Antaris II 型傅里叶变换近红外光谱仪获取酵母菌培养过程中, 发酵物样本在 $10000 \sim 4000 \text{ cm}^{-1}$ 范围内的光谱数据, 同时采用光电比浊法测定各样本的光密度 (Optical density, OD) 值; 运用竞争性自适应重加权采样 (Competitive adaptive reweighted sampling, CARS) 算法优选特征光谱, 再利用极限学习机 (Extreme learning machine, ELM) 建立酵母菌生长过程 4 个阶段的分类模型。研究结果显示, 参与 CARS-ELM 模型建立的波长个数为 30, 其 10 次运行在训练集和测试集中的平均识别率分别为 98.68% 和 97.37%。研究结果表明, 利用近红外光谱分析技术结合适当的化学计量学方法描述酵母菌生长过程是可行的。

关键词 酵母菌; 近红外光谱; 竞争性自适应重加权采样法; 极限学习机

1 引言

在全球能源逐渐匮乏的大环境下, 利用酵母菌发酵生物质产酒精作为能源替代品越来越引起重视。在工业生产过程中, 酵母菌生长过程的测定对发酵具有重要的指导作用^[1~3]。目前, 酵母菌检测的方法主要有血球计数板计数法和平板菌落计数法等, 这些方法虽然检测过程直观、快速, 但检测结果受操作人员因素影响较大, 其稳定性和均一性难以保证^[4]。

近红外光是介于紫外可见光和中红外光之间的电磁波, 其光谱信息主要来自有机物中含氢基团倍频与合频的吸收, 不同基团或同一基团在不同化学环境中的近红外吸收波长与强度都有明显差别, 适用于有机化合物理化参数的间接测量^[5~7]。在酵母菌培养过程中, 基质中的有机物大分子包含了大量的含氢基团^[8]。近年来, 已有一些学者利用近红外光谱分析技术对微生物发酵过程中的底物、产物和生物量浓度进行检测, 取得了较理想的结果^[9~11]。而在利用近红外光谱分析技术对微生物生长过程动态跟踪方面的研究却鲜有报道。大量研究表明, 近红外光谱信息是由很多弱的、宽的、非特征重叠谱带所构成^[12], 这些光谱信息包含了很大数目的波长变量, 这些波长变量有些是无信息变量和冗余变量, 波长变量之间也存在着很严重的线性关系^[13]。这些波长变量不仅会增大计算量, 而且对光谱的有用信息进行干扰, 从而降低模型的预测能力。因此, 在利用近红外光谱分析技术对酵母菌生长过程进行定性分析时, 光谱特征和化学计量学模型的选择与优化至关重要, 直接影响最终检测结果的精度^[12]。因此, 本研究提出基于近红外光谱分析技术的酵母菌生长过程快速描述方法。采用竞争性自适应重加权采样 (Competitive adaptive reweighted sampling, CARS) 算法筛选预处理后的近红外光谱特征波长, 然后利用极限学习机 (Extreme learning machine, ELM) 建立酵母菌生长过程 4 个时期的识别模型, 实现了利用近红外光谱分析技术高精度检测酵母菌生长状态。

2 实验部分

2.1 酵母菌培养及数据采集

2.1.1 酵母菌的扩大培养 从上海瑞楚生物科技有限公司购买工业发酵酵母菌种 1 mL, 再配制麦芽汁培养基, 然后将原代酵母菌菌种接种到培养基中做平行扩大培养, 每次取生长状况良好的菌种作为下一次接种母菌, 直到培养得到酵母菌种 40 mL 结束。

2017-02-18 收稿; 2017-06-07 接受

本文系江苏省自然科学基金项目 (No. BK20140538)、中国博士后科学基金 (No. 2016M600381)、江苏省高校自然科学研究面上项目 (No. 16KJB210003) 和江苏省高校研究生实践创新计划项目 (No. SJZZ16_0193) 资助

* E-mail: h. v. jiang@ujs. edu. cn

2.1.2 酵母菌的分装培养 酵母菌扩大培养结束后,分别在 3 个 250 mL 容量瓶内装入 125 mL 无菌麦芽汁培养基和 0.5 mL 酵母菌悬液,分别将 3 个容量瓶标记为 I、II 和 III,放入恒温振荡培养箱中连续培养 72 h,温度设置为 28℃,转速为 110 r/min。按以上方案共进行 6 批酵母菌培养实验。

2.1.3 数据采集 在酵母菌培养过程中,从接种开始每隔 4 h 采样一次,共有 19 个采样时间点(即 0, 4, 8, ……72 h)。为了避免采样次数过多而引起容量瓶内发酵污染,采样时,将 19 个采样时间点分为三部分,即 0~24 h、28~48 h 和 52~72 h 时间内的采样分别在容量瓶 I、II 和 III 中进行。这样每批实验可获得 19 个样本数据。共进行 6 批实验,可获得 114 个样本。

2.2 光谱采集

采用 Antaris II 傅里叶变换近红外光谱仪(美国 Thermo Scientific 公司)的透射模式采集各样本的近红外光谱数据。光谱采集时,室内温度保持在 25℃ 左右,湿度基本恒定。样品池采用光程 5 mm 标准管,扫描次数为 32 次,分辨率为 8 cm⁻¹,扫描波数范围为 10000~4000 cm⁻¹。每个样本采集 3 次,取其平均光谱作为该样本的原始光谱。

2.3 光电比浊法测定 OD 值

在样本光密度(Optical density, OD)值测定时,首先将 UV-2204PC 型紫外可见分光光度计的波长设置为 600 nm,透光率调为 100%,取光程为 1 cm 比色皿装入 3.5 mL 无菌麦芽汁培养基为对照组;样本溶液经 0.45 μm 滤膜过滤后,移入比色皿测量其 OD 值。每个样本测量 3 次,再取其平均值。若样本溶液过稠,需稀释后再进行测量,使得 OD 值保持在 0.1~0.65 之间^[14,15]。

2.4 数据分析方法

2.4.1 竞争性自适应重加权采样算法 为了消除光谱变量之间的冗余和共线性信息,需要对光谱数据进行变量筛选。竞争性自适应重加权采样(Competitive adaptive reweighted sampling, CARS)是模拟达尔文进化论中“适者生存”原则^[16],通过蒙特卡罗采样法随机选择 80% 的样本,建立偏最小二乘法(Partial least square, PLS)模型,保留回归系数绝对值大的波长点,同时去除权重小的波长点,多次重复筛选后,选出交叉验证均方根误差 RMSECV 最小的变量子集,即为特征波长变量^[17,18]。

2.4.2 极限学习机判别分析法 极限学习机算法(Extreme learning machine, ELM)是由新加坡南洋理工大学的 Huang 等^[19]提出的一种针对单隐含层前馈神经网络的学习算法。该算法能随机产生输入层与隐含层之间的连接权值和隐含层神经元的阈值,并且在训练过程中无需调整,只需设置隐含层神经元的个数,就可以获得唯一的最优解,克服了传统神经网络训练速度慢、易陷入局部最优的问题^[20],并以其学习速度快、泛化性能好等优势被广泛应用于模式分类领域。

3 结果与讨论

3.1 光谱分析

酵母菌发酵液主要是由蛋白质、碳水化合物等大分子化合物和乙醇组成^[21]。其中,碳水化合物的主要吸收波段在 6298.4~5650.4 cm⁻¹之间,蛋白质的主要吸收波段在 6506.4~6776.6 cm⁻¹之间,乙醇的主要吸收波段在 7154.6~6954.1 cm⁻¹和 9997.2~9981.7 cm⁻¹之间^[22]。图 1A 为所有酵母菌发酵液样本的原始光谱图。从图 1 可见,不同时间段获取的酵母菌发酵液样本的光谱吸收峰基本与文献^[22]描述的大分子化合物的吸收波段范围吻合,很好地反映了酵母菌培养过程中大分子有机化合物的微量变化,这为近红外光谱分析技术用于酵母菌发酵过程定性分析提供了理论依据。

为了消除发酵液中固态颗粒及光散射等因素对采集光谱的影响,研究采用标准正态变量变换(Standard normal variate transformation, SNV)对原始光谱进行预处理,该方法可有效消除液态样品中悬浮颗粒及光程变化等外部因素对光谱采集的影响^[23,24]。SNV 预处理光谱图如图 1B 所示。

3.2 酵母菌生长曲线及样本集划分

为了直观地反映酵母菌的动态生长过程,本研究根据各采样时间点所测的样本 OD 值拟合出酵母菌的生长曲线。如图 2 所示,0~8 h 为酵母菌生长的迟滞期,8~28 h 为酵母菌生长的对数期,28~60 h 为酵母菌生长的稳定期,60~72 h 为酵母菌生长的衰亡期,很好地反映了酵母菌的 4 个生长阶段。

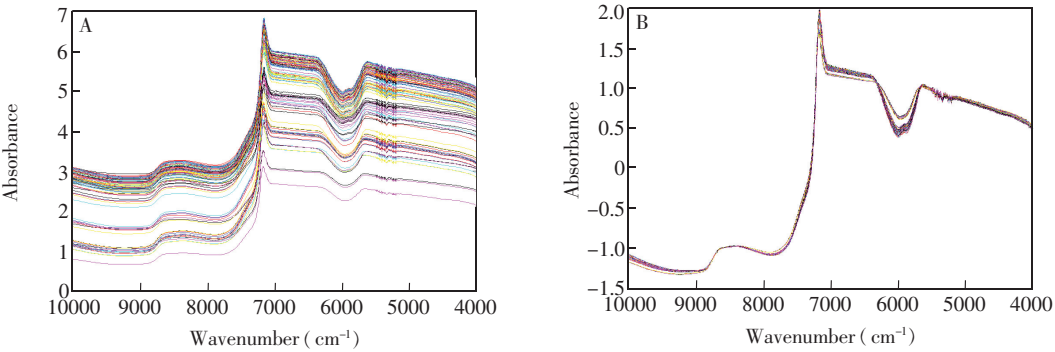


图1 酵母菌发酵液的原始光谱图 (A) 和标准正态变量变换预处理光谱图 (B)

Fig.1 Raw near-infrared (NIR) spectra (A) and standard normal variate transformation (SNV) preprocessing spectra (B) of yeast cultural media

在模型校正过程中,将前 4 批实验获取的样本作为训练集,后两批实验获取的样本作为独立测试集,用于校正模型的验证。表 1 列出了酵母菌生长过程中采集的所有样本的 OD 值在训练集和测试集中的分布情况。

3.3 光谱变量筛选

图 3 呈现了应用 CARS 算法对预处理后的光谱进行特征波长筛选的过程。从图 3A 可见,随着采样次数增加,被保留的波长变量呈指数规律衰减,较好地反映了 CARS 算法在执行时对变量粗选和精选的过程。图 3B 为 CARS 采样过程中,交互验证均方根误差 (RMSECV) 随采样次数的变化情况。从图 3B

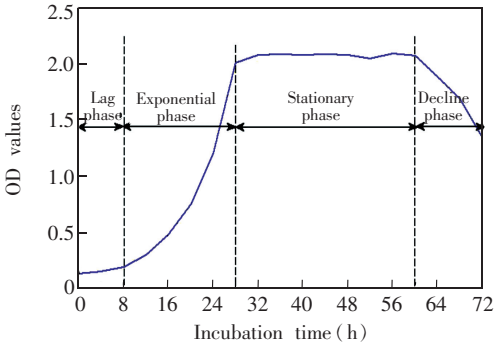


图2 酵母菌生长曲线

Fig.2 Curve of yeast growth. OD, optical density

表 1 样本 OD 值在训练集和测试集中的统计描述

Table 1 Descriptive statistics of OD in the training and testing sets

样本集 Sample set	样本数 Sample size	光密度值 OD value			
		平均值 Mean value	最大值 Maximum value	最小值 Minimum value	标准偏差 Standard deviation
训练集 Training set	76	1.397	2.150	0.123	0.780
测试集 Testing set	38	1.408	2.150	0.125	0.792

可见,当采样次数为 28 时,RMSECV 值达到最小(0.1736);此时,入选的波长变量数为 30,它们在全光谱区域的分布如图 4 所示。对上述筛选变量的 NIR 吸收谱带解析为:5650.4, 5932.0, 5935.8, 5939.7, 5943.5, 6059.2, 6159.5 和 6298.4 cm^{-1} 是位于 C—H 基团一级倍频振动吸收的波段范围;6506.4, 6564.2, 6583.8, 6587.6, 6595.4, 6599.2, 6618.2, 6695.6, 6699.5, 6768.9, 6772.8 和 6776.6 cm^{-1} 为胺基 N—H 键伸缩振动的一级倍频附近;6954.1, 6957.9, 6996.5, 7089.0, 7131.5, 7154.6 cm^{-1} 和 9981.7, 9985.6, 9989.5, 9997.2 cm^{-1} 分别位于醇类 O—H 基团伸缩振动的一级和二级倍频附近。分析上述筛选的 30 个波长变量的波数可知,这些波数光谱基本都在酵母菌发酵液中的蛋白质、碳水化合物等大分子和乙醇的光谱主要吸收波段内。因此,利用 CARS 算法筛选的特征波长变量能较好地反映酵母菌培养过程中基质中有机的微量变化。

3.4 ELM 判别模型建立及预测

选用经 CARS 算法优选后的 30 个特征波长变量建立 ELM 判别模型,完成酵母菌生长阶段的定性描述。在 ELM 模型建立过程中,其隐含层神经元个数 K 是影响其性能的重要参数。因此,在 ELM 模型建立过程中需对其进行优化。研究初始化 $K=10$,并以 11 为间隔增加,依据模型在训练集和测试集中

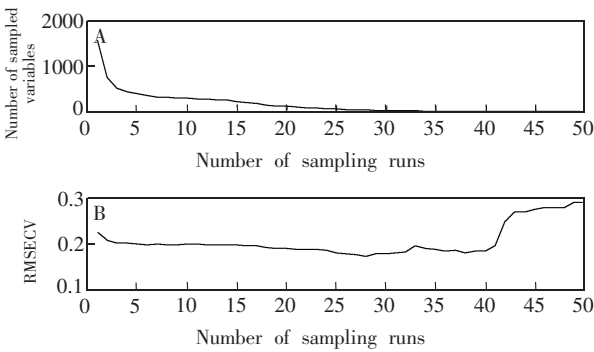


图3 竞争性自适应重加权采样的变量筛选过程图
Fig. 3 Diagram of competitive adaptive reweighted sampling (CARS) variable selecting process (A) change of the number of selected wavelength variables and (B) change of root mean square error of cross validation (RMSECV)

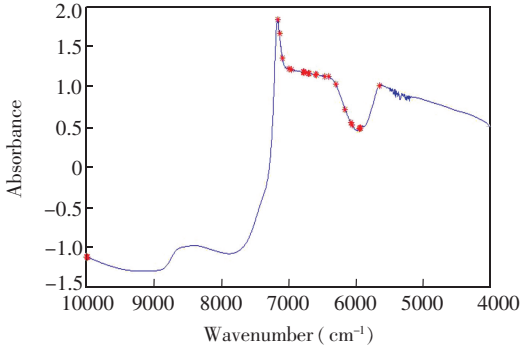


图4 CARS 变量筛选分布图
Fig.4 Distribution of variables (show by ‘*’) selected by competitive adaptive reweighted sampling (CARS) method

的预测正确率来确定最佳的隐含层神经元个数。由于 ELM 算法权重初始化时具有随机性,因此,针对每个 K ,研究均运行 10 次,取其中 5 次预测效果较好模型的记录于表 2 中。从表 2 可知,当 $K=43$ 时,训练集和测试集的平均预测准确率分别达到了 100% 和 99.47%,性能最佳。因此,最终确定为 $K=43$ 。确定 ELM 算法关键参数后,10 次运行 ELM,其在训练集中的平均预测准确率为 98.68%,在训练集中的平均预测正确率为 97.37%,很好地对酵母菌生长的 4 个阶段进行有效区分。

表2 隐含层神经元个数对 ELM 性能的影响
Table 2 Effect of the number of hidden layer neurons on the performance of extreme learning machine (ELM)

隐含层神经元数目 NNHL	预测准确率 Prediction accuracy (%)											
	第 1 次 The first time		第 2 次 The second time		第 3 次 The third time		第 4 次 The fourth time		第 5 次 The fifth time		平均值 Mean value	
	训练集 TRS	测试集 TES	训练集 TRS	测试集 TES	训练集 TRS	测试集 TES	训练集 TRS	测试集 TES	训练集 TRS	测试集 TES	训练集 TRS	测试集 TES
10	93.42	97.37	96.05	97.37	94.74	92.11	90.79	86.84	88.16	89.47	92.63	92.63
21	100	97.37	100	97.37	100	100	98.68	94.74	100	94.74	99.74	96.84
32	100	97.37	100	97.37	100	97.37	100	97.37	100	100	100	97.89
43	100	100	100	100	100	97.37	100	100	100	100	100	99.47
54	100	100	100	94.74	100	94.74	100	97.37	100	97.37	100	96.84
65	100	97.37	100	100	100	97.37	100	97.37	100	94.74	100	97.37
76	100	86.84	100	86.84	100	100	100	100	100	92.11	100	93.16

NNHL; Number of neurons in hidden layer; TRS; training set; TES; Testing set.

4 结论

本研究利用近红外光谱分析技术实现酵母菌培养过程的动态监测。利用 CARS 算法对预处理后的光谱进行特征波长筛选,优化 ELM 隐含层神经元数,最后建立酵母菌生长过程定性识别模型。结果表明,建立在由 CARS 筛选法 30 个特征波长变量基础上的最佳 ELM 识别模型,10 次运行在测试集中的平均识别率达到 97.37%。因此,利用近红外光谱分析技术结合合适的化学计量学方法快速监测酵母菌生长过程是可行的。本研究结果为酵母菌生长过程的快速在线监测提供了技术支持。

References

1 Yu J, Xu Z, Tan T. *Fuel Process. Technol.*, **2008**, 89(11): 1056–1059
2 Sablayrolles J M. Pandey A, Rao L V, Soccol C R. *Food Res. Int.*, **2009**, 42(4): 418–424

- 3 Doran J B, Cripe J, Sutton M, Foster B. *Appl. Biochem. Biotechnol.*, **2000**, 84-86(1): 141-152
- 4 Wu Y J, Jin Y, Li Y R, Sun D, Liu X S, Chen Y. *Vib. Spectrosc.*, **2012**, 58(1): 109-118
- 5 TAO Lin-Li, YANG Xiu-Juan, DENG Jun-Ming, ZHANG Xi. *Spectroscopy and Spectral Analysis*, **2013**, 33(11): 3002-3009
陶琳丽, 杨秀娟, 邓君明, 张 曦. 光谱学与光谱分析, **2013**, 33(11): 3002-3009
- 6 ZHANG Li-Juan, WU Wei, QIU Lin, LIU Ying. *Modern Instrument & Medical Treatment*, **2012**, 18(4): 76-79
张丽娟, 吴 炜, 邱 琳, 刘 莹. 现代仪器与医疗, **2012**, 18(4): 76-79
- 7 CHEN Huan-Wen, HU Bin, ZHANG Xie. *Chinese J. Anal. Chem.*, **2010**, 38(8): 1069-1088
陈焕文, 胡 斌, 张 燮. 分析化学, **2010**, 38(8): 1069-1088
- 8 XU Bao-Cheng, LIU Jian-Xue, YI Jun-Peng, ZHONG Xian-Feng, CUI Guo-Ting. *China Brewing*, **2007**, 26(3): 8-10
徐宝成, 刘建学, 易军鹏, 钟先锋, 崔国庭. 中国酿造, **2007**, 26(3): 8-10
- 9 HUANG Chang-Yi, FAN Hai-Bin, LIU Fei, XU Gan-Rong. *Journal of Instrumental Analysis*, **2014**, 33(5): 520-526
黄常毅, 范海滨, 刘 飞, 许赣荣. 分析测试学报, **2014**, 33(5): 520-526
- 10 ZHANG Shu-Ming, YANG Yang, NI Yuan-Ying. *Spectroscopy and Spectral Analysis*, **2012**, 32(11): 2997-3001
张树明, 杨 阳, 倪元颖. 光谱学与光谱分析, **2012**, 32(11): 2997-3001
- 11 PENG Bang-Zhu, YUE Dian-Li, YUAN Ya-Hong, GAO Zhen-Peng. *Spectroscopy and Spectral Analysis*, **2009**, 29(3): 652-655
彭帮柱, 岳田利, 袁亚宏, 高振鹏. 光谱学与光谱分析, **2009**, 29(3): 652-655
- 12 LIU Guo-Hai, XIA Rong-Sheng, JAING Hui, MEI Cong-Li, HUANG Yong-Hong. *Spectroscopy and Spectral Analysis*, **2014**, 34(8): 2094-2097
刘国海, 夏荣盛, 江 辉, 梅丛立, 黄永红. 光谱学与光谱分析, **2014**, 34(8): 2094-2097
- 13 Blanco M, Coello J, Iturriaga H, MasPOCH S, González Bañó R. *Analyst*, **2000**, 50(1): 75-82
- 14 LI Qin. *Light Ind. Sci. Technolo.*, **2014**, (8): 7-8
李 勤. 轻工科技, **2014**, (8): 7-8
- 15 LI Li, YANG Ze-Xian, WANG Su-Xia, DU Jin-Min. *Feed Res.*, **2015**, (1): 71-73
李 丽, 杨泽贤, 王素霞, 杜进民. 饲料研究, **2015**, (1): 71-73
- 16 Fan W, Shan Y, Li G Y, Lyu H Y, Li H D, Liang Y Z. *Food Anal. Method*, **2012**, 5(3): 585-590
- 17 Xie C, Ning X, Shao Y, He Y. *Spectrochim. Acta A*, **2015**, 149: 971-977
- 18 Tang G, Huang Y, Tian K D, Song X Z, Yan H, Hu J, Xiong Y M, Min S G. *Analyst*, **2014**, 139(19): 4894-4902
- 19 Huang G B, Zhu Q Y, Siew C K. *Neurocomputing*, **2006**, 70(1-3): 489-501
- 20 SHI Feng, WANG Hui, YU Lei, HU Fei. *MATLAB Intelligence Algorithm-30 Case Analysis*. Beijing: Beihang University Press, **2011**: 290-302
史 峰, 王 辉, 郁 磊, 胡 斐. MATLAB 智能算法-30 个案例分析. 北京: 北京航空航天大学出版社, **2011**: 290-302
- 21 HAN Run-Ping, BAO Gai-Ling, ZHU Lu. *Spectroscopy and Spectral Analysis*, **2004**, 24(7): 820-822
韩润平, 鲍改玲, 朱 路. 光谱学与光谱分析, **2004**, 24(7): 820-822
- 22 LU Wan-Zhen. *Modern Near Infrared Spectroscopy Analytical Technology (Second Edition)*. China Petrochemical Press, **2007**: 30-32
陆婉珍. 现代近红外光谱分析技术 (第 2 版). 中国石化出版社, **2007**: 30-32
- 23 Bi Y M, Yuan K L, Xiao W Q, Wu J Z, Shi C Y, Xia J, Chu G H, Zhang G X, Zhou G J. *Anal. Chim. Acta*, **2016**, 909: 30-40
- 24 Barnes R J, Dhanoa M S, Lister S J. *Appl. Spectrosc.*, **1989**, 43(5): 772-777

Qualitative Prediction of Yeast Growth Process Based on Near Infrared Spectroscopy

WANG Wei, JIANG Hui*, LIU Guo-Hai, MEI Cong-Li, JI Yi

(School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract To improve the yield of industrial fermentation, a method based on near infrared spectroscopy was presented to predict the growth of yeast. The spectral data of fermentation sample were measured by Fourier-transform near-infrared (FT-NIR) spectrometer in the process of yeast culture. Each spectrum was acquired over the range of $10000 - 4000 \text{ cm}^{-1}$. Meanwhile, the optical density (OD) of fermentation sample was determined with photoelectric turbidity method. After that, a method based on competitive adaptive reweighted sampling (CARS) was used to select characteristic wavelength variables of NIR data, and then extreme learning machine (ELM) algorithm was employed to develop the categorization model about the four growth processes of yeast. Experimental result showed that, only 30 characteristic wavelength variables of NIR data were selected by CRAS algorithms, and the prediction accuracies of training set and test set of the CARS-ELM model were 98.68% and 97.37%, respectively. The research showed that the near infrared spectrum analysis technology was feasible to predict the growth process of yeast.

Keywords Near infrared spectroscopy; Growth of yeast; Competitive adaptive reweighted sampling; Extreme learning machine

(Received 18 February 2017; accepted 7 June 2017)

珀金埃尔默发布全新 Avio® 500 ICP-OES

作为原子光谱领域的领导者和创新者,珀金埃尔默公司于2017年7月6日发布全新 Avio® 500 电感耦合等离子体发射光谱仪(ICP-OES)。这款 ICP-OES 是专为高通量检测实验室打造的多元素无机分析设备,可以应对各个领域的多种类型样品。

作为一款真正的同步检测 ICP-OES,Avio 500 具备同步背景校正功能,可提供更高的样品通量和数据可靠性;超群的基体耐受性;以及业内最低的氦气消耗量。无论待测元素的种类和浓度范围再宽,Avio 500 都可以有效应对,使用户的检测工作符合行业规范。Avio 500 的可兼容多个应用领域,包括环境、石化(尤其是润滑油服役情况分析)、地矿、食品、制药、制造业(包括电池制造)等。为何选择 Avio 500 ICP-OES?

“随着标准、法规的要求越来越复杂,待测元素种类越来越多,检测实验室技术人员不得不面对日益增加的工作挑战。这时,提升检测效率、拓宽检测范围就显得尤为重要。”珀金埃尔默公司执行副总裁、DAS(探索与分析事业部)总裁 Jim Corbett 先生表示。“为了满足这一需求,我们推出了全新的 Avio 500 ICP-OES,集多项创新技术于一身,适用于多个应用领域,可有效提升检测效率,同时降低运行成本。”

关键功能

以下功能使得 Avio 500 具备卓越的性能:为提升样品通量而设计的同步数据采集功能、为节约使用成本设计的业内最低的氦气消耗量、为缩短样品前处理流程设计的超宽线性范围。无论样品多么复杂,这些功能都可确保检测数据的准确性。

带有快速拆装炬管座的垂直炬焰设计:带来强悍基体耐受性,有效缩短样品预处理时间; Flat Plate™ 平板等离子体技术:仅消耗同类产品一半的氦气就可以生成稳定的等离子体炬,无论样品基体多么复杂;双向观测技术:可同时等离子体进行全波长轴向和径向观测,高低含量元素一次进样同时分析;全谱全读技术:对全波长进行同步数据采集,无需重复进样就可以获得全部光谱数据;空气刀(PlasmaShear™)技术:无需氦气,即可消除等离子体冷尾焰产生的基体干扰,而且完全无需维护;全彩等离子体观测(PlasmaCam™)技术:提供全彩色的等离子体实时影像,简化方法开发工作,并使远程诊断成为可能。