

常见伤口感染细菌高场不对称波形离子迁移谱识别算法

陆彦邑,尹军,毕玉田,严军,黎敏,何庆华

第三军医大学第三附属医院野战外科研究所/国家创伤烧伤复合伤重点实验室,重庆 400042

【摘要】目的:使用模式识别算法对常见伤口感染细菌(大肠杆菌、金黄色葡萄球菌、铜绿假单胞菌)TH肉汤培养液及纯TH培养液的高场不对称波形离子迁移谱(FAIMS图谱)进行分类识别。**方法:**使用FAIMS分析仪收集了4种样品的训练及测试样本。预处理后,用主成分分析和线性判别分析对样本进行降维和特征提取,得到了训练集和测试集的空间三维分布模型,再用最近邻规则算法进行样本识别。**结果:**特征提取后,正负模式样本均具有良好的分离效果,并且正模式可分性明显优于负模式。当 K 取合适的值时,正负模式识别率分别可达90%和70%以上。对于本文的数据模型, K 取值等于或最接近每类样本数的奇数为最佳。**结论:**该种算法可用于常见伤口感染细菌肉汤培养液FAIMS图谱的分类及识别。

【关键词】高场不对称波形离子迁移谱;常见伤口感染细菌;模式识别算法

【中图分类号】TN911.73

【文献标志码】A

【文章编号】1005-202X(2018)02-0205-05

Algorithmic identification for high-field asymmetric ion mobility spectra of common wound-infecting bacteria

LU Yanyi, YIN Jun, BI Yutian, YAN Jun, LI Min, HE Qinghua

State Key Laboratory of Trauma, Burns and Combined Injury/Surgery Institute of the Third Military Medical University, Chongqing 400042, China

Abstract: Objective To classify and recognize the high-field asymmetric ion mobility spectra (FAIMS) of common wound-infecting bacteria (*escherichia coli*, *staphylococcus aureus* and *pseudomonas aeruginosa*) cultured in thioglycolate (TH) broth and bare TH broth by pattern recognition algorithm. **Methods** The spectra of training set and testing set of four types of samples were obtained by FAIMS analyzer. After pretreatment, principal component analysis and linear discriminant analysis were used to reduce the dimension and extract features for obtaining the three-dimensional spatial distribution patterns of training set and testing set. Finally, k-nearest neighbors algorithm was used to identify the four types of samples. **Results** After feature extraction, the samples showed good separation effects under both positive and negative modes, and the separability of positive mode was superior to that of negative mode. When K was appropriately assigned, the recognition rate under positive and negative modes could be up to 90% and 70%, respectively, or even higher. For the data model in this study, the optimal K was equal to the odd number closest to sample number of each type. **Conclusion** The proposed algorithm can be used to classify and recognize the FAIMS of common wound-infecting bacteria cultured in TH broth.

Keywords: high-field asymmetric ion mobility spectra; common wound-infecting bacteria; pattern recognition algorithm

前言

高场不对称波形离子迁移谱 (high-Field Asymmetric Ion Mobility Spectrometry, FAIMS) 又称

差分离子迁移谱,是一种利用离子在高低电场作用下运动轨迹的不同来检测离子的新型气相离子分离技术^[1]。

气态分子被离子化源电离后,离子进入电极沟道,射频电压在高电场和低电场下交替振荡。离子迁移率在高电场下的差别,决定了离子在沟道内的运动轨迹,不同种类的离子沿着不同轨迹运动。此时,在射频电场上施加一个横向的直流补偿电压,特定的离子团在特定的补偿电压下通过电极沟道到达前方的检测极板,则实现对该种离子的检测分离。因此,不同样品其FAIMS图谱形状不同,可通过模式识

【收稿日期】2017-10-17

【基金项目】国家国际科技合作专项(2014DFA31560);重庆市科技研发基地建设计划(cstc2013ghz10003)

【作者简介】陆彦邑,硕士,研究实习员,研究方向:生物医学信号处理, E-mail: 90911xbh@163.com

【通信作者】何庆华,博士,副研究员,研究方向:生物医学信号处理, E-mail: qinghuahe@126.com

别算法来鉴别样品种类。常用的FAIMS图谱处理算法有线性判别分析(Linear Discriminant Analysis, LDA)^[2-4]、主成分分析(Principal Component Analysis, PCA)^[5]、小波分析^[6]、交叉验证及支持向量机^[7]等。

虽然FAIMS技术已在危险品与毒品分析^[8-9]、化学检测^[10-12]、环境检测^[13-14]及生物医药^[15]等诸多领域得到了广泛应用,但在医学上还具有更为广阔的发展前景。因伤口感染细菌检测在临床工作中的重要性,本文将FAIMS应用于伤口感染常见细菌的检测。对3种细菌培养液及纯培养液的FAIMS图谱使用PCA和LDA进行降维和特征提取,得到了这4种样本在新模式空间中的分布特征。后使用最近邻规则(K-Nearest Neighbor, KNN)算法进行分类识别,具有较高的识别率。本研究为FAIMS在细菌检测及识别方面的应用提供了一定的理论与实践基础。

1 实验部分

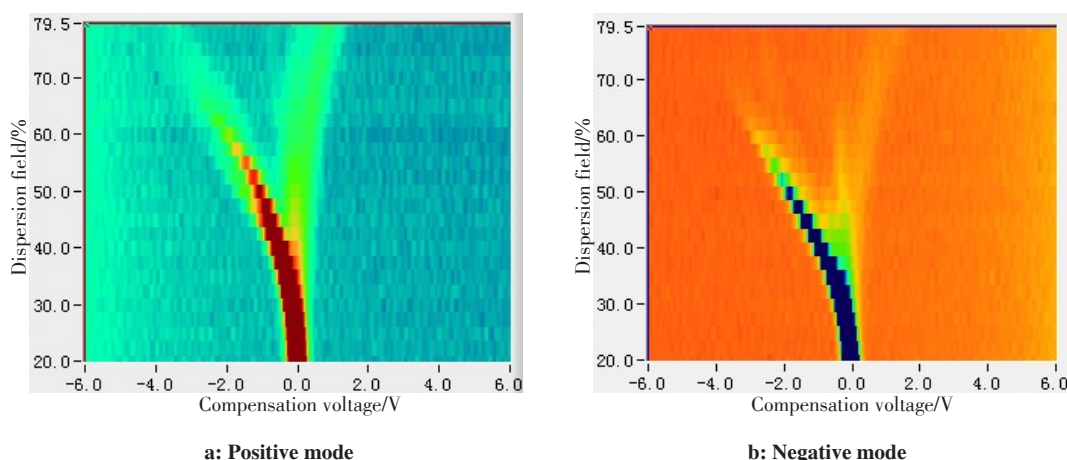


图1 TH培养液正负模式FAIMS图谱

Fig.1 High-field asymmetric ion mobility spectra of thioglycolate (TH) broth under positive and negative modes

2 细菌识别算法

2.1 特征提取

本文经过实验得到的满足4种样品检测的最佳场强范围为20%~80%,扫描线数为25条。检测所得的FAIMS图谱大小为25×512,对于识别算法来说,维度较高会大大增加计算量。而一幅图中不仅包含着特征信息,也包含了噪音和低能量特征,因此算法的第一步是提取图像特征,同时也达到降维与降噪的目的。

PCA是目前常用的一种基于变量协方差矩阵的数据特征提取方法,可以将多个数据特征映射为少数几个综合特征。协方差矩阵中主对角线元素是各个维度上的方差,方差越大包含的原矩阵信息越大,

本文采用英国 Owlstone 公司的 Lonestar 作为 FAIMS 分析仪^[16],实验样品为 TH 肉汤管转种的 3 种细菌培养液:大肠杆菌(ATCC25922)、金黄色葡萄球菌(ATCC25923)、铜绿假单胞菌(ATCC27853),以及未转种的纯 TH 肉汤培养液作为对照,总计 4 种样品。每一次实验,3 种细菌均为同一时间转种、相同时长培养以保持一致性。

检测样品时,设定一个场强范围和扫描线数量,则场强增加的步长即可确定。Lonestar 扫描过程为:纵坐标从最小场强开始扫描,横坐标补偿电压从-6 V 扫到+6 V,扫描点数为 512,得到第一条扫描线;然后场强以设定的步长增加,补偿电压再次从-6 V 扫到+6 V,得到第二条扫描线;场强继续增加,直至扫描完整个设定的场强范围。即可得样品正负模式下的 FAIMS 图谱,图 1 所示为检测 TH 培养液所得 FAIMS 图谱。场强单位为百分比,即占 Lonestar 所能达到的最高场强的比例。

降维即是保留方差最大的少数维度;非对角线元素为两两维度间的协方差,代表原矩阵维度间的相关性,降噪即是使保留下来的维度间相关性最小,也就是使非对角线元素尽可能都为零。通过对协方差矩阵对角化,可使非对角线元素全部为零。对角化后的矩阵,对角线上为原协方差矩阵的特征值,选取较大特征值对应的特征向量组成投影矩阵,对原始样本矩阵进行投影,即可得到降维后的新样本矩阵,降维后的维数为样本总数-类别数。

2.2 样本分类

降维后的样本矩阵包含了各类样本的特征信息,此时需要将各类信息进行分类用于后续识别。

作为经典的模式识别算法,LDA应用非常广泛。其基本原理是将高维的模式样本投影到维度更低的空间中,使投影后的样本按类别分开。本文使用LDA来对样本矩阵进行分类,同时也是进一步的降维。将原样本投影到新的样本空间,使得新的样本空间具有最大的类间距离 S_b 和最小的类内距离 S_w ,则在新的样本空间中各类具有最大的可分性。投影函数如下:

$$y_i = w^T x_i \quad (1)$$

其中, x_i 为原始样本, y_i 为新样本空间, w^T 为投影方向。投影向量的获取方法为:对 $S_w^{-1}S_b$ 进行特征值分解,将最大特征值对应的特征向量作为最佳投影方向 w 。对于多类问题,需要类别数减一个投影向量组成投影矩阵,因此投影后的新样本维度=类别数(需要进行分类的类别数)-1。

2.3 分类识别

完成了特征提取与样本分类,本文选取KNN算法作为分类识别算法。作为实践中比较流行的一种识别方法,KNN的原理并不复杂。如图2所示,中心问号圈代表待识别样本,不同颜色方框代表已知类别。搜索待识别样本周围 K (K 为奇数)个点,统计这 K 个点中已知类别分别有多少个,哪一类个数最多,就是与待识别样本最邻近的类,则将待识别样本归于这一类。

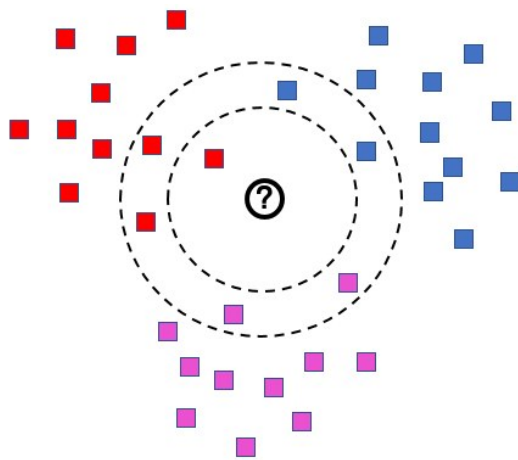


图2 KNN原理图

Fig.2 Schematic diagram of K -nearest neighbor (KNN)

2.4 算法流程

模式识别中,经常用到训练集与测试集这两个概念。训练集是一个已知的样本集,用于开发模式识别分类器;测试集则是用来检验模式识别算法的正确性。在本文的研究中,也将检测到的所有样本FAIMS图谱随机分为训练集与测试集,两个集合各 N

个样本,每一个类别各 N/C 个样本, C 为类别数。算法流程图如图3所示。其中,将训练集和测试集所有FAIMS图谱转换为行向量,具体做法为:将每一幅图谱的每一行按场强从低到高依次连接,得到一个样本,大小为 $1 \times 12\ 800$ 。

3 结果与讨论

经过实验,本文得到了TH培养液培养的大肠杆菌、金黄色葡萄球菌和铜绿假单胞菌3种菌液以及纯TH培养液的FAIMS特征图谱^[16]。共采集了训练集与测试集各204个样本数据,两个集合中每类样本各51个,一个样本即一幅FAIMS图谱。

3.1 特征提取和样本分类

本文使用Matlab进行算法实现。模式训练中,对正负模式,用PCA对训练集样本进行变换后可得按贡献值由大到小排列的 $204 \times 12\ 800$ 维数据,本文选取了前200维组成投影矩阵,其贡献率达到了99.99%,得到了 204×200 维的新样本数据。再用LDA将新样本数据进行分类,本文为4分类问题,因此需要 200×3 维的投影矩阵,最终得到了 204×3 维的样本特征矩阵。图4为LDA后的正负模式下样本特征矩阵的三维空间分布图。正模式中,用训练集得到的4类样本的训练模型,在三维空间中有着很好的分离特点;使用同样的投影矩阵对测试集进行投影得到的测试集样本分布依然具有良好的分离效果,这对于之后的分类十分有利。而负模式中,训练模型的聚合程度不如正模式强,各类测试样本的分离程度也弱于正模式。因此,正模式的样本类别可分性明显高于负模式。

3.2 分类识别

使用KNN算法作分类识别时,计算测试模型的每个样本到训练模型各样本的距离,从小到大进行排序,统计前 K 个样本中各类别的数目。将该测试样本归为类别数最多所属的那一类,若该类别与原测试样本类别标签吻合,则识别数目加1,识别率即为识别数目除以总样本数。

识别过程中, K 值的选择会影响样本识别率,不同的数据模型需要的 K 值不同。本文的训练集模型样本总数为204个,为确定其合适的 K 值大小,将 K 取从1到204中的所有奇数,可得正负模式下识别率随 K 值变化的曲线图(图5)。图5中数字为识别率明显降低时 K 的前后取值,可以看出, K 的两次陡降值为103和153,正好在样本类别数(51)的2倍和3倍处。出现此种现象的原因主要是, K 值若大于样本类别数的2倍,距测试样本的前 K 个最小距离的训练模型中,可能会包括2个及以上类别的全部测试样本,此时统计到的这2个及以

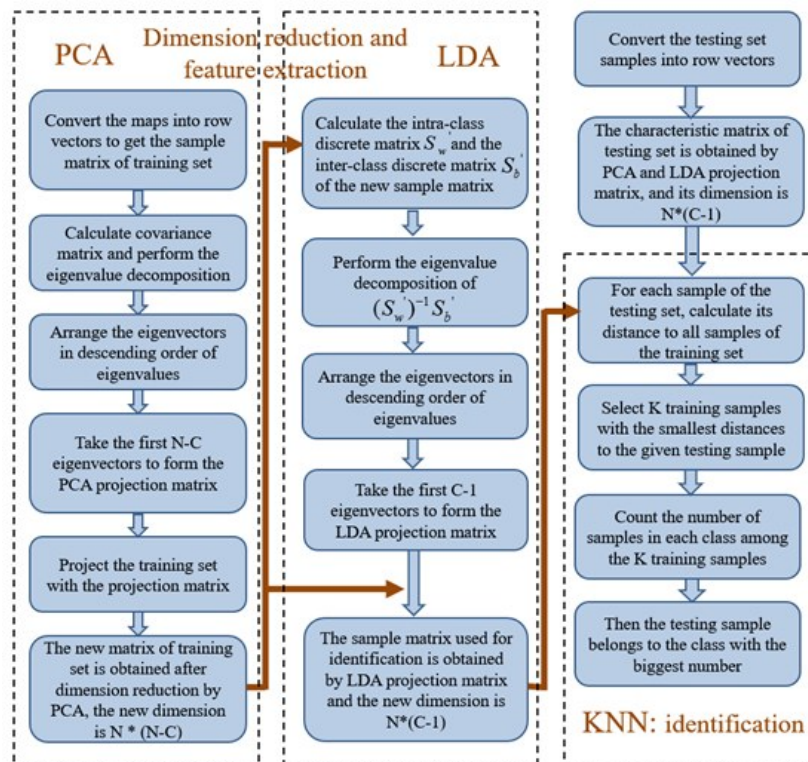


图3 算法流程图

Fig.3 Algorithm flowchart

PCA: Principal component analysis; LDA: Linear discriminant analysis

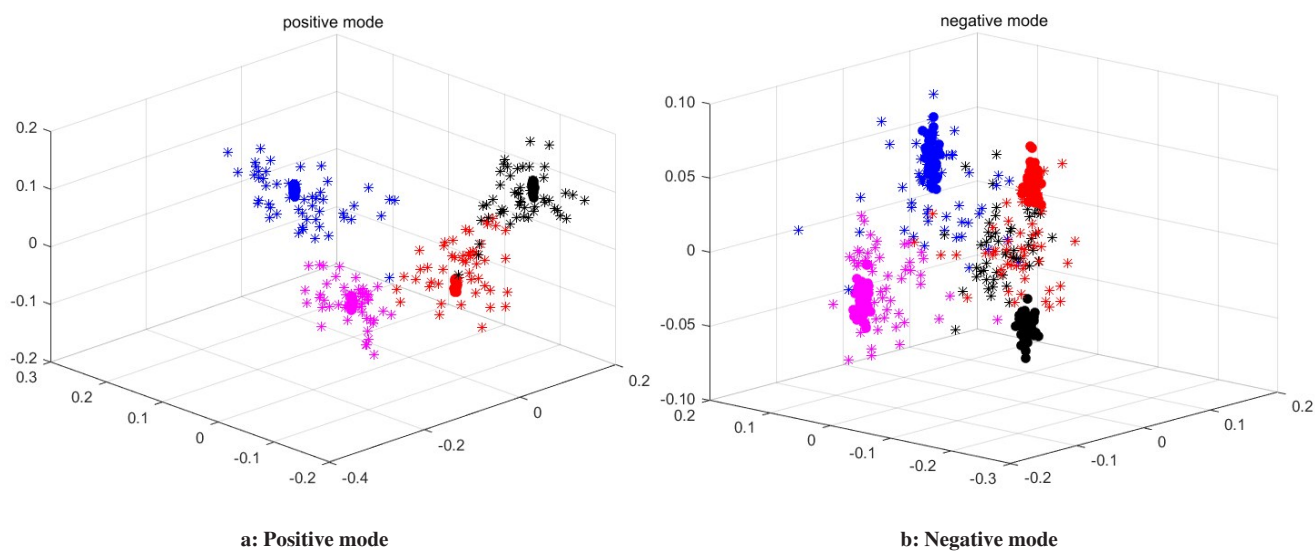


图4 LDA后的样本三维特征空间分布图

Fig.4 Three-dimensional feature spatial distribution after LDA

Blue: *Escherichia coli*; purple: *Staphylococcus aureus*; red: *Pseudomonas aeruginosa*; black: TH broth; ●: Training set; *: Testing Set

上类别的类别数相同,算法则把该测试样本判别为2个及以上类别中的第一个,导致识别失败。同时,也可看出当 K 值小于样本类别数的2倍时,正负模式都具有好的识别率,尤其正模式,能达到90%以上。

基于以上分析可知,4种细菌样本的识别率与训

练集样本数的选择和 K 值大小有关。实践中,为了保证识别率尽量准确, K 值应尽量不小于训练样本数,但也不能大于训练样本中两类样本数的和。对于本文的数据模型,若每类样本数为偶数, K 取值等于每类样本数减1为最佳;若每类样本数为奇数,则 K 取

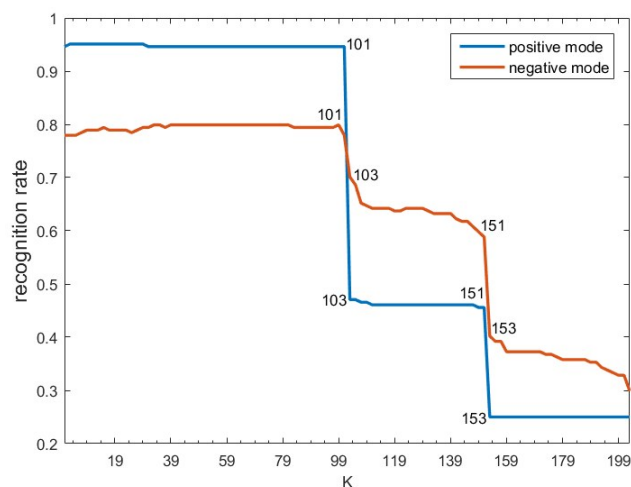


图5 识别率随K值变化曲线

Fig.5 Curve of recognition rate varying with K value

值等于每类样本数为最佳。

4 结论

本文利用Lonestar FAIMS分析仪检测得到了大肠杆菌、金黄色葡萄球菌、铜绿假单胞菌3种细菌培养液及纯TH肉汤培养液的FAIMS图谱。收集4种图谱样本,将正模式和负模式样本均分为了训练集与测试集。用PCA和LDA进行降维及特征提取,从训练集中提取出了投影矩阵,得到了训练完成后的样本空间分布模型,再用该投影矩阵同样对测试集样本进行投影,得到了测试样本的空间分布。训练集与测试集在新样本空间中的分布表明,正模式较负模式具有更好的分离效果。后用KNN对正负模式进行分类识别,结果显示,当K取值小于训练样本中两类别样本数的和时,样本具有良好的识别率,正模式识别率高于负模式。可知,从模型训练到样本识别,本文中4种样本的正模式样本可分性优于负模式。本文的实验结果为该种算法用于细菌FAIMS图谱的分类识别提供了一定的理论与实践基础,也为FAIMS技术用于伤口感染细菌的快速检测也提供了潜在的可能性。

【参考文献】

- [1] BURYAKOV I A, KRYLOV E V, NAZAROV E G, et al. A new method of separation of multi-atomic ions by mobility at atmospheric pressure using a high-frequency amplitude-asymmetric strong electric field [J]. Int J Mass Spectrom Ion Processes, 1993, 128(3): 143-148.
- [2] ARASARADNAM R P, OUARET N, THOMAS M G, et al. A novel tool for non invasive diagnosis and tracking of patients with inflammatory bowel disease (IBD)[J]. The University Warwick, 2013, 19(5): 999-1003.
- [3] COVINGTON J A, WESTENBRINK E W, OUARET N, et al. Application of a novel tool for diagnosing bile acid diarrhoea[J]. Sensors, 2013, 41(7): 11899-11912.
- [4] ARASARADNAM R P, MC FARLANE M J, RYAN-FISHER C, et al. Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis[J]. PLoS One, 2014, 9(9): e108750.
- [5] 李娟, 郭亚洁, 赵伟军, 等. 气相色谱-高场非对称离子迁移谱(GC-FAIMS)法应用于白酒的风味研究[J]. 食品安全质量检测学报, 2014, 5(6): 1687-1694.
- [6] LI J, GUO Y J, ZHAO W J, et al. Study on liquor flavor by gas chromatography-field asymmetric ion mobility spectrometry method [J]. Journal of Food Safety and Quality, 2014, 5(6): 1687-1694.
- [7] COVINGTON J A, WEDLAKE L, ANDREYEV J, et al. The detection of patients at risk of gastrointestinal toxicity during pelvic radiotherapy by electronic nose and faims: a pilot study[J]. Sensors, 2012, 12(10): 13002-13018.
- [8] ARASARADNAM R P, WESTENBRINK E, MCFARLANE M J, et al. Differentiating coeliac disease from irritable bowel syndrome by urinary volatile organic compound analysis-a pilot study[J]. PLoS One, 2014, 9(10): e107312.
- [9] ZHANG J, LI L F, GUO D P, et al. Determination of hazardous chemicals by microchip-based field asymmetric waveform ion mobility spectrometric technology[J]. Chinese J of Anal Chem, 2013, 41(7): 986-992.
- [10] CUMERAS R, GRÀCIA I, FIGUERAS E, et al. Finite- element analysis of a miniaturized ion mobility spectrometer for security applications[J]. Sens Actuators B Chem, 2012, 170(8): 13-20.
- [11] WOJCIECH G, KENNETH L F. Rapid and sensitive differentiation of anomers, linkage, and position isomers of disaccharides using high-field asymmetric waveform ion mobility spectrometry (FAIMS)[J]. J Am Soc Mass Spectrom, 2003, 14(3): 265-277.
- [12] TAYLOR G N, MATHERLY R. Use of portable analytical methods to determine the stoichiometry of reaction for hexahydrotriazine-based hydrogen sulfide scavenger operations[J]. Anal Chem, 2014, 86(10): 4879-4882.
- [13] 李庄, 林丙涛, 孔德义, 等. 痕量挥发性有机物的高场不对称波形离子迁移谱研究[J]. 光谱学与光谱分析, 2011, 31(1): 12-15.
- [14] LI Z, LIN B T, KONG D Y, et al. FAIMS of trace volatile organic compounds[J]. Spectroscopy and Spectral Analysis, 2011, 31(1): 12-15.
- [15] ZHUANG L I, LIN B T, CHEN C L, et al. Volatile phenolic compounds detection by 2D scanning high- field asymmetric waveform ion mobility spectrometry [J]. Chinese J Anal Chem, 2011, 39(1): 107-110.
- [16] 陶羽宇, 李灵锋, 李鑫, 等. 用高场非对称波形离子迁移谱技术快速检测二乙醇胺[J]. 高等学校化学学报, 2013, 34(11): 2499-2504.
- [17] TAO Y Y, LI L F, LI X, et al. Rapid determination of diethanolamine by field asymmetric ion mobility spectrometry(FAIMS) technique[J]. Chemical Journal of Chinese Universities, 2013, 34(11): 2499-2504.
- [18] COVINGTON J A, VAN DER SCHEE M P, EDGE A S, et al. The application of FAIMS gas analysis in medical diagnostics[J]. Analyst, 2015, 140(20): 6775-6781.
- [19] 陆彦邑, 毕玉田, 尹军, 等. 基于高场不对称波形离子迁移谱技术的细菌培养液检测[J]. 中国医学物理学杂志, 2017, 34(3): 297-301.
- [20] LU Y Y, BI Y T, YIN J, et al. Detection of bacterial culture broth based on high- field asymmetric ion mobility spectrometry [J]. Chinese Journal of Medical Physics, 2017, 34(3): 297-301.

(编辑:薛泽玲)