

基于径向基函数神经网络的荧光光谱技术在菌种识别中的应用*

武浩¹⁾ 朱拓^{1)†} 孔艳¹⁾ 陈卫²⁾ 杨建磊³⁾

1)(江南大学理学院, 无锡 214122)

2)(江南大学食品科学与技术学院, 食品科学与技术国家重点实验室, 无锡 214122)

3)(江南大学通信与控制工程学院, 无锡 214122)

(2009 年 6 月 2 日收到; 2009 年 8 月 3 日收到修改稿)

对嗜酸乳杆菌、变异链球菌和保加利亚乳杆菌这三种菌的荧光光谱进行研究, 发现在紫外光的激励下, 益生菌溶液发出荧光. 在最佳激发波长 290 nm 的激励下, 荧光峰值在 300—650 nm 范围内. 采用小波变换对测得的 150 组光谱数据进行压缩, 压缩后每组数据由原来的 1341 个点减少为 168 个点, 既保留了原图谱的特征, 又提高了神经网络的处理速度. 径向基函数神经网络方法对压缩后的数据进行研究, 对每种菌的 40 组实验数据进行训练, 在此基础上对 30 组未知数据进行识别. 结果表明经过训练之后, 径向基函数神经网络能够准确预测未知菌种.

关键词: 荧光光谱, 径向基函数神经网络, 数据压缩

PACC: 3250D, 0650

1. 引言

益生菌是一类对人体有益的细菌, 服用含有益生菌的食品, 可以帮助人体维持肠道菌丛的平衡, 保持人体的健康. 益生菌的应用, 取得了很好的医疗和保健效果. 为了更好地应用益生菌造福人类, 需要从多方面研究益生菌^[1-4]. 如何实现对益生菌的快速精确的鉴别成为相关领域的学者关注的焦点. 传统的鉴别方法费时费力, 增加成本, 而快速的益生菌鉴别, 可以提高工业生产的效率, 节约时间, 减少成本.^[5-8] 荧光光谱技术具有灵敏度高、选择性强、样品用量少、方法简便、能提供较多物理参数等特点, 是一种十分有效的光谱化学分析手段, 在光谱分析领域发挥着重要的作用^[9-12]. 1996 年美国的 Chin 等^[13] 对空气、水中传播的致病菌的荧光做了研究, 并开发出可以检测这些致病菌的荧光寿命和进行初步鉴别的仪器; 1998 年德国的 Stephan 等^[14] 对海水中菌类的荧光激发、发射光谱与菌液浓度的关系做了研究报道; 2003 年 Héctor 等^[15] 对三类细菌

的自发荧光光谱进行研究, 并对三类细菌的荧光光谱进行鉴别, 以及使用主成分分析方法对菌种的荧光进行分类, 实验结果显示该方法相对传统的研究方法有快速、容易实现、成本低等优点. 2006 年英国的 Elliot 等^[16] 报道了水中浮游菌类的荧光光谱与温度的关系, 从温度对荧光强度、峰值位置的影响做了具体研究. 从以上报道中可以知道, 荧光分析法可用来研究微生物, 同样也可研究益生菌.

径向基函数(radial basis function, 简记为 RBF)神经网络, 是由 Moody 和 Darken 于 20 世纪 80 年代提出的一种神经网络, 具有其他神经网络所具有的自学习、自组织、自适应、高容错能力等优点, 此外 RBF 神经网络还被证明能够以任意的精度逼近任意的连续函数, 因此 RBF 神经网络在计算机科学、自动化控制等领域有着广泛的应用, 并取得了很好的效果^[17-21]. 本文对嗜酸乳杆菌、保加利亚链球菌和变异链球菌的荧光光谱进行研究, 采用多功能光谱仪测量三种不同菌溶液的荧光光谱; 采用小波变换的方法将数据进行压缩, 减少了光谱的数据量, 保留了光谱的特征, 节约了神经网络的处理时间;

* 江苏省博士后基金(批准号: 0802021C)、中国博士后科学基金(批准号: 20080440153)和国家高技术研究发展计划(批准号: 2007AA10Z353)资助的课题.

† 通讯联系人. E-mail: tzhu@jiangnan.edu.cn

神经网络的应用实现了识别过程中的计算机化和自动化,为快速、简便识别益生菌提供了一种全新的方法. 数据处理的结果表明将小波变换、RBF 神经网络与荧光光谱相结合,对三种菌溶液的荧光光谱进行鉴别,具有较高的精确度.

2. 实验部分

2.1. 仪器设备和测试条件

仪器设备:美国 Roper Scientific 公司的型号为 SP2558 多功能光谱测量系统.

测试条件:测量范围 300—650 nm. 积分时间为 2 s,三个狭缝的宽度分别为 3,2 和 1 mm. 激发波长采取 290 nm,为细胞的最佳激发波长之一.

2.2. 样品来源和处理

本工作选用三种益生菌:嗜酸乳杆菌、变异链球菌、保加利亚乳杆菌. 按照微生物的培养条件,将菌种接种到 MRS (de Man, Rogosa and Sharp) 培养基中,培养 12 h 左右,将益生菌从培养基中离心洗涤出来,放在生理盐水中,保存于 4 °C 冰箱中待用.

2.3. 实验方法

参考已有的文献可知,细菌溶液的最佳激发波长之一是 290 nm,荧光的范围为 300—650 nm,所以本文选择的是激发波长为 290 nm,荧光范围在 300—650 nm 的数据. 鉴于数据过多,为了减少网络训练的时间和避免可能发生的网络训练速度过慢,采用小波变换对原始数据进行压缩. 压缩前有 1341 个数据点,压缩后有 168 个数据点. 这样变量数大为减少,既能提高训练速度,又能够保持原有的特征峰. 三种样品共 150 组数据,每种菌有 50 组数据,其中 40 组用来训练,余下的 10 组用来作预测.

3. 结果和讨论

3.1. 三种益生菌的荧光光谱比较

益生菌以细胞为单位,细胞中的物质相似,不同菌液发出的荧光相互重叠,实验测得的三种菌液的荧光光谱如图 1 所示. 图中可以看出,三种菌在 340 nm 附近都有一个峰值,在 350—650 nm 范围

内,变异链球菌和嗜酸乳杆菌在 380 nm 附近有一明显峰值,但是保加利亚乳杆菌没有,它的峰值在 410 nm 附近. 此外变异链球菌在 460,520 nm 附近还有峰值,其他两种没有这种现象. 传统的荧光光谱分辨率低,选择性差,会产生很大的测量误差,加上三种菌的光谱较为相似,导致识别困难.

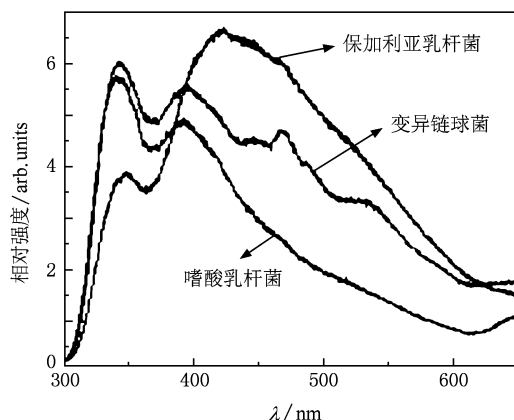


图 1 三种菌的荧光光谱

3.2. 小波变换对数据压缩

使用 Matlab 7.0 的内部函数 `appcoef` 对光谱仪采集到的数据进行小波变换,将原始的大量数据 (1341 个数据点) 进行压缩,压缩后的每组数据为 168 个数据点. 图 2 是根据压缩后的数据画出来的,与图 1 相比可以发现,压缩后的数据保留了原来光谱中的特征峰,但是压缩后的数据大量减少,可以方便后续的处理工作^[19,22,23].

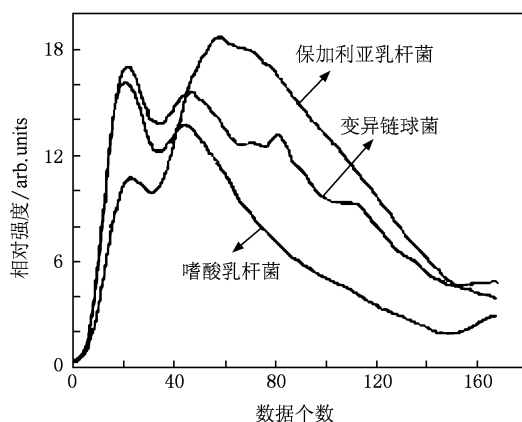


图 2 小波变换压缩后的数据图

3.3. 径向基函数神经网络训练

在运行程序前,为了统一量纲,有助于数据分

析和网络训练,分别对光谱数据进行归一化,为了加快径向基函数神经网络的收敛及保证网络的稳定性,界定输入和输出值均在 $[0,1]$ 的范围内,归一化公式为:

$$D'_i = \frac{D_i - D_{\min}}{D_{\max} - D_{\min}}, \quad (1)$$

其中 D_i 为归一化前数据, D_{\min} 为训练数据的最小值, D_{\max} 为训练数据的最大值, D'_i 为归一化后的数据.将归一化后的数据使用 Matlab 中的 newrb 函数进行网络训练,经过调试后得到目标误差为 $eg = 0.0001$,RBF 分布常数 $sc = 1$ 时训练的结果较好.设定三种菌输出值分别为:嗜酸乳杆菌为 0,保加利亚乳杆菌为 0.5,变异链球菌为 1.

从三种菌的光谱中各选十组数据作为预测集,采用 Matlab 编程,调用 Matlab 自带函数 appcoef 和 newrb 进行小波变换以及网络训练.设定目标误差为 0.0001,采用 Matlab 程序进行训练.图 3 为训练的误差收敛曲线,图中曲线 a 为使用原始数据训练所得的结果,b 是使用小波压缩后的数据,c 是目标误差.由图可得,a 曲线训练次数多,而且结果没有收敛.而曲线 b 训练次数少,收敛性好.在本文的训

练条件: $eg = 0.0001$, $sc = 1$ 下,使用原始数据,神经元迭代次数 $N_n = 55$,平方和误差 $SSE = 0.000149366$ 总时间为 13.60 s.使用压缩后的数据, $N_n = 25$, $SSE = 0.0000935188$,总时间为 1.71 s,很大程度上节约了总时间.所以使用小波压缩后的数据可以提高速度,节约时间.

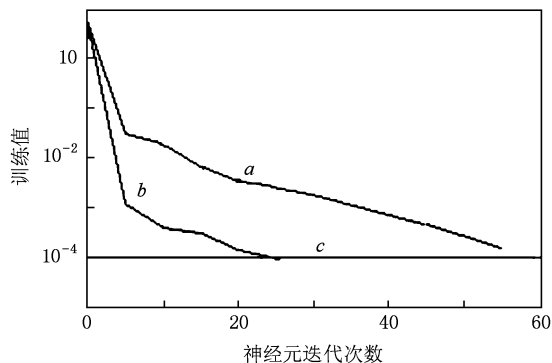


图 3 误差训练曲线

采用 RBF 神经网络方法,对 30 个未知菌种进行预测.表 1 是对 30 种菌预测的结果,由表中数据可以看出预测的结果和实际的值符合得比较好,说明 RBF 神经网络用于菌种的识别结果十分准确.

表 1 10 组未知菌种的预测值和实际值

序号	嗜酸乳杆菌		保加利亚乳杆菌		变异链球菌	
	预测值	实际值	预测值	实际值	预测值	实际值
#1	0.0013	0	0.4993	0.5	1.0025	1
#2	0.0016	0	0.5001	0.5	1.0025	1
#3	0.00145	0	0.5004	0.5	1.0023	1
#4	0.0028	0	0.5000	0.5	1.0024	1
#5	0.0057	0	0.4994	0.5	1.0032	1
#6	0.0038	0	0.4987	0.5	1.0009	1
#7	0.0032	0	0.4996	0.5	1.0040	1
#8	0.0023	0	0.4989	0.5	0.9997	1
#9	0.0046	0	0.4993	0.5	1.0017	1
#10	0.0028	0	0.5002	0.5	1.0018	1

4. 结 论

荧光分析法具有灵敏度高、简便快速、试剂需求量少的特点,但对于都是以细胞为基本结构的益生菌,它们的荧光光谱在较大波长范围内相互重叠,传统的荧光光谱法识别起来比较困难.神经网络

由于其自身的特点,使其适宜处理大量的模糊数据,但是由于荧光光谱数据非常多,采用小波函数对数据进行压缩,这样不但保留原光谱数据的特征,还能够加快了神经网络的运算速度.将小波分析、神经网络应用到益生菌的鉴别中,使得鉴别更加简便快速,而且精确度较高,该方法有望成为菌种鉴别的一种有效手段.

- [1] Yang J, Pan X H, Zheng Y Y, Dong M Z, Xiao P, Ye Y W, Zhong W J 2000 *Shanghai J. Prev. Med.* **12** 73 (in Chinese) [杨 隽、潘喜华、郑勇英、董妙珠、肖 萍、叶于薇、仲伟鉴 2000 上海预防医学杂志 **12** 73]
- [2] Li Q, Zhang S C, Zeng X Y, Zhang T T 2004 *Food Res. Dev.* **25** 106 (in Chinese) [李 琴、张世春、曾晓燕、张铁涛 2004 食品研究与开发 **25** 106]
- [3] Wei H, Liu C M, Cheng B C, Xu F 2007 *Food Sci.* **28** 618 (in Chinese) [魏 华、刘春梅、程波财、徐 峰 2007 食品科学 **28** 618]
- [4] Hu X Z 2001 *Shanghai Med. Pham. J.* **22** 356 (in Chinese) [胡学智 2001 上海医药 **22** 356]
- [5] Guo X H 2002 *Foundation and Application of Probiotics* (Beijing: Beijing Scientific and Technical Publishers) (in Chinese) [郭兴华 2002 益生菌基础与应用 (北京: 北京科学技术出版社)]
- [6] Zhang G 2007 *Lactic Acid Bacteria—Basis, Technology and Application* (Beijing: Chemical Industry Press) (in Chinese) [张 刚 2007 乳酸细菌——基础、技术和应用 (北京: 化学工业出版社)]
- [7] Zhang L, Tan B P, Mai K S, Ai Q H, Zhang W B, Xu W 2008 *Period. Ocean Univ. Chin.* **38** 225 (in Chinese) [张 玲、谭北平、麦康森、艾庆辉、张文兵、徐 玮 2008 中国海洋大学学报 **38** 225]
- [8] Li H B, Song X L, Wei S, Li Y, Lan P 2008 *Ocean Limin Sin.* **39** 374 (in Chinese) [李海兵、宋晓玲、韦 嵩、李 赟、兰 萍 2008 海洋与湖沼 **39** 374]
- [9] Liu Y, Song C Y, He W L, Luo X S, Lu J, Ni X W 2007 *Acta Phys. Sin.* **56** 2962 (in Chinese) [刘 莹、宋春元、何文亮、骆晓森、陆 建、倪晓武 2007 物理学报 **56** 2962]
- [10] Liu Y, Ni X W 2009 *Acta Phys. Sin.* **58** 3572 (in Chinese) [刘 莹、倪晓武 2009 物理学报 **58** 3572]
- [11] Zhang J F, Hu S S 2007 *Acta Phys. Sin.* **56** 713 (in Chinese) [张军峰、胡寿松 2007 物理学报 **56** 713].
- [12] Cheng L, Ding X L, Liu Z G, Pan Q L, Chu X L, Feng S L 2007 *Acta Phys. Sin.* **56** 6894 (in Chinese) [程 琳、丁训良、刘志国、潘秋丽、初学莲、冯松林 2007 物理学报 **56** 6894]
- [13] Chin R C, Mayo M W, Hallidy W, Salazar N, Cruce C 1996 *SPIE* **2836** 158
- [14] Stephan D, Jörg M L, Rainer R, Jürgen R 1998 *Marine Chem.* **62** 137
- [15] Héctor E G, Landolfo S J, Renato A Z, Marcos T T P 2003 *J. Fluorenc.* **13** 489
- [16] Elliott S, Lead J R, Baker A 2006 *Anal. Chim. Acta* **564** 219
- [17] Research and Development Center of Feisi Science and Technology Products 2005 *Neural Networks and Realized by Matlab7* (Beijing: Publishing House of Electronics Industry) (in Chinese) [飞思科技产品研发中心 2005 神经网络理论与 Matlab7 实现 (北京: 电子工业出版社)]
- [18] Dong C H 2007 *Matlab Neural Networks and Application* (2nd Ed.) (Beijing: National Defense Industry Press) (in Chinese) [董长虹 2007 Matlab 神经网络与应用 (第二版) (北京: 国防工业出版社)]
- [19] Ma S M, Liu S D, Zhang Z Y, Fan G Q 2005 *Spectrosc. Spectr. Anal.* **25** 874 (in Chinese) [马书民、刘思东、张卓勇、范国强 2005 光谱学与光谱分析 **25** 874]
- [20] Li Y B, Huang X Y, Sha M, Meng X S 2001 *Chin. J. Chromat.* **19** 112 (in Chinese) [李一波、黄小原、沙 明、孟宪生 2001 色谱 **19** 112]
- [21] Yu X H, Zhang Z Y, Ma Q, Fan G Q 2007 *Spectrosc. Spectr. Anal.* **27** 481 (in Chinese) [于晓辉、张卓勇、马 群、范国强 2007 光谱学与光谱分析 **27** 481]
- [22] Wang K, Zhang H, Chang S J, Shen J Y 2007 *Acta Phys. Sin.* **56** 3613 (in Chinese) [王 凯、张 会、常胜江、申金媛 2007 物理学报 **56** 3613]
- [23] Zhang H P, Zhao J B, Yang Z L 2003 *Comput. Modern.* **99** 600 (in Chinese) [张海鹏、赵景波、杨占录 2003 计算机与现代化 **99** 600]

Application of fluorescence spectra based on radial basis function neural network in identification of bacteria^{*}

Wu Hao¹⁾ Zhu Tuo^{1)†} Kong Yan¹⁾ Chen Wei²⁾ Yang Jian-Lei³⁾

1) (*School of Science, Jiangnan University, Wuxi 214122, China*)

2) (*State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, Wuxi 214122, China*)

3) (*School of Communication and Control Engineering, Jiangnan University, Wuxi 214122, China*)

(Received 2 June 2009; revised manuscript received 3 August 2009)

Abstract

The results of research on fluorescence spectra of three kinds of probiotic bacteria (lactobacillus acidophilus, streptococcus mutans and lactobacillus bulgaricus) show that the bacteria emit fluorescence when irradiated by ultraviolet light. The spectra are in the range of 300—650 nm with the excitation wavelength of 290 nm. Wavelet transform is used to compress the 150 groups of spectra data, and the number of data points in each group is reduced from 1341 to 168, which not only keeps the character of original spectra unchanged, but also improves the processing speed of neural network. Radial basis function neural network is applied for processing the compressed data. 40 groups of data of each strain are used to be trained, and the other 30 groups of data of which the kinds of bacteria are not given are used for prediction. The result shows that radial basis function neural network can identify the unknown bacterial strain accurately.

Keywords: fluorescence spectra, radial basis function neural network, data compression

PACC: 3250D, 0650

^{*} Project supported by the Jiangsu Province Postdoctoral Sustentation Fund, China (Grant No. 0802021C), the China Post-doctoral Science Foundation (Grant No. 20080440153) and the National High Technology Research and Development Program of China (Grant No. 2007AA10Z353).

[†] Corresponding author. E-mail: tzhu@jiangnan.edu.cn