

基于多光谱应用 BP 神经网络预测血糖

李东明^{1,2}, 贾书海²

¹西安石油大学理学院, 陕西 西安 710065;

²西安交通大学机械工程学院, 陕西 西安 710049

摘要 利用反向传播(BP)神经网络预测方法,通过光纤将红外光谱仪、拉曼光谱仪和旋光测量系统结合在一起,建立了基于多光谱测量血糖含量的分析模型,提出了数据融合的处理方法。选择了30个人体血液样品,分别测量旋光光谱、红外光谱、拉曼光谱。将光谱数据进行了预处理与归一化处理,建立BP神经网络模型,预测血液样品的糖含量值。使用克拉克误差网格分析法分别分析了三种测量方法和数据融合后的血糖值,结果应用BP神经网络模型预测血糖值的拟合精度为0.9992,预测误差低于0.2 mmol/L,满足临床医学的精度要求,并且具有较高的稳健性和较强的容错能力。

关键词 医用光学; 血糖预测; 多光谱; BP神经网络; 克拉克误差网格分析

中图分类号 O433; TP212 **文献标识码** A

doi: 10.3788/LOP54.031703

Application of BP Artificial Neural Network in Blood Glucose Prediction Based on Multi-Spectrum

Li Dongming^{1,2}, Jia Shuhai²

¹ School of Science, Xi'an Shiyou University, Xi'an, Shaanxi 710065, China;

² School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

Abstract Based on the back-propagation (BP) neural network prediction method and combined with the infrared spectrometer, Raman spectrometer and polarimetry analysis system through the optical fiber, a multi-spectral blood glucose measurement system is developed and a processing method of data fusion is proposed. 30 human blood samples were measured to obtain the optical rotatory dispersion spectrum, infrared spectrum and Raman spectrum, respectively. Spectral data was preprocessed and normalized. The BP neural network model was established to predict the blood glucose content. We use the Clarke error grid analysis to compare the blood glucose content obtained by the three measurement methods and by data fusion. Results show that the fitting precision of the fusion data is 0.9992, and the prediction error is lower than 0.2 mmol/L, which can meet the accuracy of clinic medicine. This method also has high robustness and strong tolerance.

Key words medical optics; blood glucose prediction; multi-spectrum; BP neural network; Clarke error grid analysis

OCIS codes 170.1470; 170.6510; 170.4580

1 引言

糖尿病是由于内分泌代谢调节紊乱所造成的疾病,它对循环系统和血管的影响将会伤害到身体的组织和器官,已经严重地威胁人类的健康。Guariguata等^[1]调查发现全世界大约3.8亿人患有糖尿病,到2035年将达到5.9亿人。为了避免血糖过高或过低,糖尿病患者需要坚持控制饮食,包括含有限量蔗糖的碳水化合物和单饱和脂肪的摄入,并和胰岛素药物治疗相结合。为了准确了解药物的用量,需要不断检测血糖值,正常监测周期大约为24 h。目前临床使用的方法是获取少量的血样测量血糖。这种方法不够方便并且容

收稿日期: 2016-11-11; 收到修改稿日期: 2016-11-18

基金项目: 国家自然科学基金委员会与中国民航局联合资助(U1233116)、西安石油大学博士启动基金

作者简介: 李东明(1970—),男,博士,副教授,主要从事光学测量及计算模拟方面的研究。E-mail: dongming-li@126.com

易感染。研发一种无痛且方便无创的测量设备可作为解决这个问题的一种方法。

目前无创伤血糖检测方法主要有微波检测法、旋光法、光声法、拉曼光谱法、光散射系数法、红外光谱法等^[2]。近红外光谱法是研究最广泛的血糖无创检测技术^[3]。Goodarzi 等^[4]利用葡萄糖分子在红外区域(1500~1800 nm 和 2050~2300 nm)的特征吸收谱,并利用红外特征光谱之间的多变分析模型,实现对血糖浓度的无创检测。Vezouviou 等^[5]利用近红外光全息技术测量血糖,在 0~70 mmol/L 范围内得到较好的结果。在过去几十年中有许多组织和机构研究无创测量的方法,但是到目前为止还没有达到满意的测量精度^[6]。

人工神经网络(ANN)是模拟人类中枢神经网络的概念,模拟人思维的一种方式,是一个非线性动力学系统。特别适合需要同时考虑许多因素和条件的、不精确和模糊的信息处理问题^[7]。段程鹏等^[8]利用近红外透射光谱结合 ANN 检测了乳制品中蛋白质与脂肪的含量,获得了良好的结果。程宇琼等^[9]利用激光诱导击穿光谱结合 ANN 能有效解决土壤间存在的基体效应,实现不同类型土壤中 Cu 元素的定量检测。崔日鲜等^[10]将 ANN、支持向量回归等融合方法用于可见光光谱估算冬小麦叶片氮积累量。Caduff 等^[11]利用数据融合技术将 5 个不同频率传感器的信号与温度信号融合预测血糖值,精度达到 0.29 mmol/L。

针对目前光学血糖无创检测精度低、重现性和可靠性差,难以满足实际使用要求的问题,本文提出将红外光谱仪、拉曼光谱仪和旋光测量系统结合在一起,利用 ANN 数据融合技术,对这三种光谱数据综合处理、分析来提高血糖光学无创监测的精度和可靠性。

2 原理与测量

2.1 原理

当光和生物组织相遇时,可能发生反射、散射或透射,这和样品的结构以及化学成分有关。旋光法是利用光学活性化合物溶液,如血液中葡萄糖的浓度和通路的长度与旋光角度的变化成正比来检测血糖。由此可见,所需测量的血糖浓度与偏转角度呈线性关系,但由于血液和组织中含有其他的旋光物质,会影响测量结果。吸收光谱测量是基于葡萄糖分子在红外区域具有的特征吸收来实现血糖检测,光谱的强度和葡萄糖浓度有关,但由于血液和组织复杂的成分及结构会使光谱重叠,影响测量的精度。拉曼光谱与吸收光谱有着本质的差别,拉曼光谱被认为是分子结构的指纹谱,用于物质的鉴定和分子结构谱线特征的研究,因此可用于测量血糖浓度,但拉曼光谱的缺点是信号微弱,测量时间长,会受到温度的影响而产生漂移。数据融合在处理分析多源不确定信息时有着特定优势,将三种光学测量方法有机地结合起来,提高信息的获取量,有利于增加数据探测渠道和数据量,并充分利用各种光谱数据间的互补性及合作性,让各种光谱信息有效地关联起来、互相印证,提高光学无创血糖检测的精度和可靠性。

2.2 测量

血液样品使用某医院的查体样品,在测量前 30 min 在室温下晃动。样品的测量设备为日立 7180 全自动生化仪。拉曼光谱使用 HR800 光谱仪测量,激光的波长为 514 nm;吸收光谱采用 Perkin-Elmer 公司的 lambda 750S 分光光度计测量;旋光光谱使用锁相放大器测量^[12],将 30 个样品的测量数据分为训练数据(表 1)和测试数据(表 2)。

表 1 三种光谱法检测训练血液样品的浓度结果及误差

Table 1 Concentration and error of the training blood samples detected by the three spectrum methods

No.	Standard Value	Raman spectrum		Infrared spectrum		Polarimetry	
		Value	Error	Value	Error	Value	Error
1	4.38	4.21	0.17	3.81	0.57	3.12	1.26
2	5.22	5.28	0.06	5.04	0.18	3.56	1.66
3	5.49	5.41	0.08	5.81	0.32	3.24	2.25
4	5.56	5.53	0.03	5.91	0.35	6.98	1.42
5	5.81	5.79	0.02	6.59	0.78	5.96	0.15
6	5.96	5.84	0.12	5.18	0.78	7.32	1.36

续表 1

No.	Standard	Raman spectrum		Infrared spectrum		Polarimetry	
	Value	Value	Error	Value	Error	Value	Error
7	6.19	6.15	0.04	6.57	0.38	7.85	1.66
8	6.48	6.36	0.12	4.73	1.75	7.27	0.79
9	6.89	7.11	0.22	2.52	4.37	9.32	2.43
10	8.42	8.39	0.03	15.42	7.00	8.34	0.08
11	8.83	8.81	0.02	9.82	0.99	9.76	0.93
12	9.21	9.42	0.21	10.29	1.08	15.12	5.91
13	9.67	9.53	0.14	9.78	0.11	10.54	0.87
14	10.26	10.52	0.26	10.79	0.53	11.37	1.11
15	12.33	12.25	0.08	13.40	1.07	12.58	0.25

表 2 三种光谱法检测血液样品的浓度结果及误差

Table 2 Concentration and error of the test blood samples detected by the three spectrum methods

mmol/L

No.	Standard	Raman spectrum		Infrared spectrum		Polarimetry	
	Value	Value	Error	Value	Error	Value	Error
1	4.98	5.02	0.04	6.09	1.11	1.43	3.55
2	5.33	5.39	0.06	5.04	0.29	4.87	0.46
3	5.51	5.61	0.10	4.55	0.96	4.32	1.19
4	5.77	5.62	0.15	6.77	1.00	6.03	0.26
5	5.90	5.70	0.20	5.54	0.36	6.21	0.31
6	6.16	6.27	0.11	6.26	0.10	6.20	0.04
7	6.47	6.58	0.11	6.86	0.39	6.78	0.31
8	6.78	6.91	0.13	5.83	0.95	6.83	0.05
9	7.85	7.65	0.20	9.69	1.84	7.43	0.42
10	8.75	8.90	0.15	8.75	0.00	9.23	0.48
11	8.93	9.14	0.21	5.65	3.28	8.67	0.26
12	9.24	9.38	0.14	8.42	0.82	8.89	0.35
13	9.74	10.10	0.36	9.78	0.04	8.52	1.22
14	11.58	11.66	0.08	11.14	0.44	10.99	0.59
15	13.48	13.57	0.09	15.18	1.70	15.36	1.88

通过表 1 和表 2 可以看出,拉曼光谱的测量结果较好,但也有 5 个数据的相对误差超过了 0.2 mmol/L;吸收法和旋光法得到结果更差。

3 血糖预测的 BP 神经网络模型

采用 BP 神经网络模型进行血糖光谱数据融合时一般用于特征级融合,获取原始数据,按血糖值随机分成等分的 n 个训练样本,对数据进行预处理,融合模型见图 1。具体过程为:将旋光法、吸收法以及拉曼光谱法的测量数值进行融合,得到新的数值。将三种测量结果的数据(测量值)和临床测量的数据(标准值)选取 1/2 作为训练数据,剩下的作为测试数据。这些测量值经过归一化处理后都处在 $[0, 1]$ 区间,所以,输入样本是 $[0, 1]$ 区间的三维矩阵。

BP 神经网络模型组成及工作过程如下。1) 输入层神经元个数。输入层神经元的数量即输入参数的数量,模型是对三种方法的融合,故输入层含 3 个神经元,分别代表不同方法测得的血糖的浓度值。2) 输出层神经元个数。输出层神经元个数由研究对象及研究目的决定。本模型输出层有 1 个神经元,即融合后的血糖浓度值。3) 隐层层数和隐层节点数。理论上已证明,在不限制隐层节点数的情况下,3 层 BP 神经网络可以实现任意非线性映射,最适合用于模拟输入输出关系,所以,本文 BP 网络采用 3 层结构。本模型根据收敛步骤和误差选择节点数。4) 网络传递函数的确定。在本模型中,输入层与隐层的传递函数选为 tansig ,

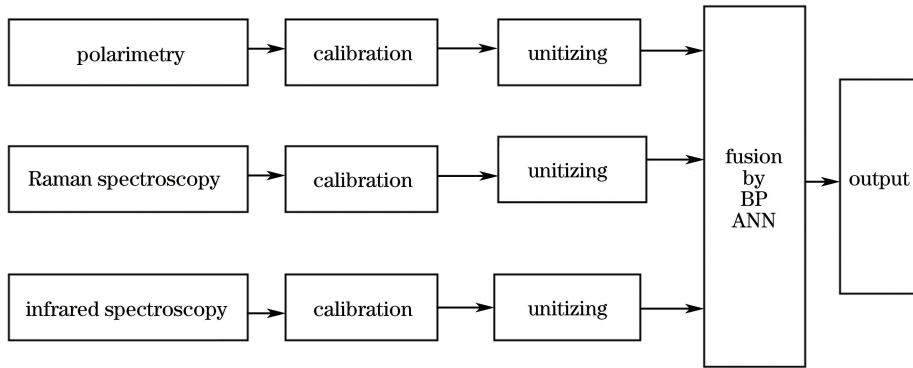


图 1 BP 神经网络进行血糖光谱数据融合

Fig. 1 Data fusion of the blood glucose spectrum with BP neural network

隐层与输出层之间选用 purelin 函数。5) 训练方法的选择。BP 神经网络常用的训练算法包括：自调整学习率与添加动量项相结合的方法，贝叶斯法规范化算法，LevenLerg-Marquardt(L-M)算法等。其中，L-M 算法在避免陷入局部极小值、收敛快速性以及网络泛化能力上明显优于其他算法而得到广泛应用。本模型的训练方法选为 L-M 算法。6) 训练过程。设定隐层的节点数为 6，则函数为 $\text{net} = \text{newff}(\text{threshold}, [6, 1], \{\text{tansig}, \text{purelin}\}, \text{trainlm}, \text{learn}g\text{d})$ ；设定训练误差为 10^{-4} ，最大迭代次数为 800 次。经过 800 次迭代后没有达到误差要求[图 2(a)]。设定隐层的节点数为 10 时，训练过程误差分析图[图 2(b)]，由图可以看出，经过 283 次迭代后达到误差的要求。设定隐层的节点数为 11 时，训练过程误差分析图见图 2(c)，由图可以看出，经过 94 次迭代后达到误差的要求。若节点数为 12，训练次数会增加，故本模型选节点数为 11。7) BP 神经网络融合预测糖值。将表 1 的数据输入建立的模型，使用表 2 的数据作为检验数据，计算融合的结果见表 3。

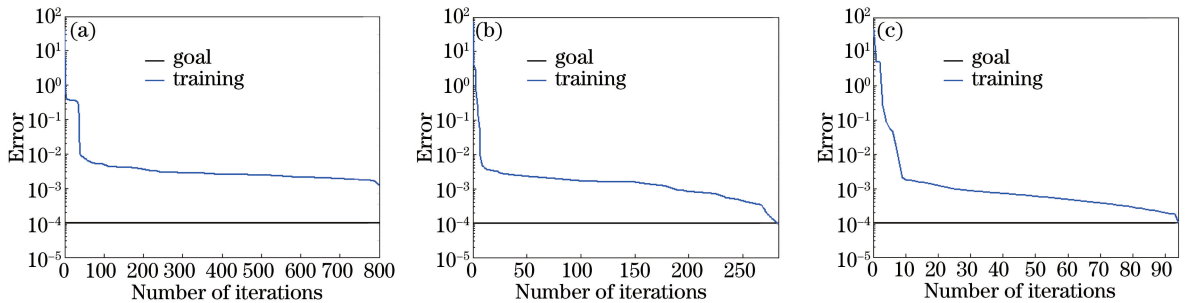


图 2 不同迭代次数下的训练过程误差分析。(a) 800;(b) 283;(c) 94

Fig. 2 Error analysis of the training process for different number of iterations. (a) 800; (b) 283; (c) 94

表 3 BP 神经网络对血液样品的预测浓度结果

Table 3 Concentration results of the blood samples predicted by BP neural network

No.	Standard value/ (mmol /L)	Fusion value/ (mmol /L)	Error / (mmol /L)			
			BP ANN	Raman spectrum	Infrared spectrum	Polarimetry
1	4.98	4.981	0.001	0.04	1.11	3.55
2	5.33	5.337	0.007	0.06	0.29	0.46
3	5.51	5.506	0.004	0.10	0.96	1.19
4	5.77	5.772	0.002	0.15	1.00	0.26
5	5.90	5.883	0.017	0.20	0.36	0.31
6	6.16	6.173	0.013	0.11	0.10	0.04
7	6.47	6.460	0.010	0.11	0.39	0.31
8	6.78	6.784	0.004	0.13	0.95	0.05
9	7.85	7.847	0.003	0.20	1.84	0.42
10	8.75	8.749	0.001	0.15	0.00	0.48

No.	Standard value/ (mmol /L)	Fusion value/ (mmol /L)	Error / (mmol /L)			
			BP ANN	Raman spectrum	Infrared spectrum	Polarimetry
11	8.93	8,920	0,010	0,21	3,28	0,26
12	9.24	9,238	0,002	0,14	0,82	0,35
13	9.74	9,738	0,002	0,36	0,04	1,22
14	11.58	11,581	0,001	0,08	0,44	0,59
15	13.48	13,481	0,001	0,09	1,70	1,88

通过对表 1~3 中数据进行对比可以得出,由 BP 神经网络将三种方法测量的血糖值经过融合后所有的结果误差值都小于 0.2 mmol/L。由预测结果可以得出 BP 神经网络融合方法是一种简单有效的实用数据融合方法;该方法克服单一测量方法带来的不确定性,有效解决了单一方法信息量不充分的问题^[13],提高了测量血糖的预测精度,对离线血液样品的测量精度达到了临床精度的要求。若长期测量,外界环境温度发生变化,将温度也作为输入参数,BP 神经网络融合可以消除温度变化的影响^[14]。

4 克拉克误差网格分析

克拉克误差网格分析是为了评估病人的测量血糖值和参考血糖的临床精确度而于 1987 年开发的分析方法。今天克拉克方法已经是判断血糖测量值和参考值的金判据^[15]。克拉克误差网格将估计的血糖值和参考血糖的分布图分为五个区域,即 A、B、C、D、E。A 区的值代表临床精确度,B 区域代表良好,C、D、E 区域的值可能是不适当的处理结果。将三种测量方法得到的结果分别使用克拉克分析方法进行分析,其结果如图 3 所示。图 3(a)为吸收光谱法克拉克误差网格分析图,图 3(b)为旋光法克拉克误差网格分析图,图 3(c)为拉曼光谱法克拉克误差网格分析图,其曲线拟合系数分别为 0.7718,0.7129,0.9933。从图中可以看出大部分数据都在 A 区和 B 区,但在 C 区和 D 区也有部分数据。在低浓度时,测量可能会受样品池光学性能的影响,从图可以看出,低浓度数据分布要差些。图 3(d)为 BP 神经网络融合后测量的克拉克误差网格分析,其曲线拟合系数为 0.9992,从图中可以看出测量结果都在 A 区内,并且所有数据都在临床要求的精度内,其误差小于 0.2 mmol/L。

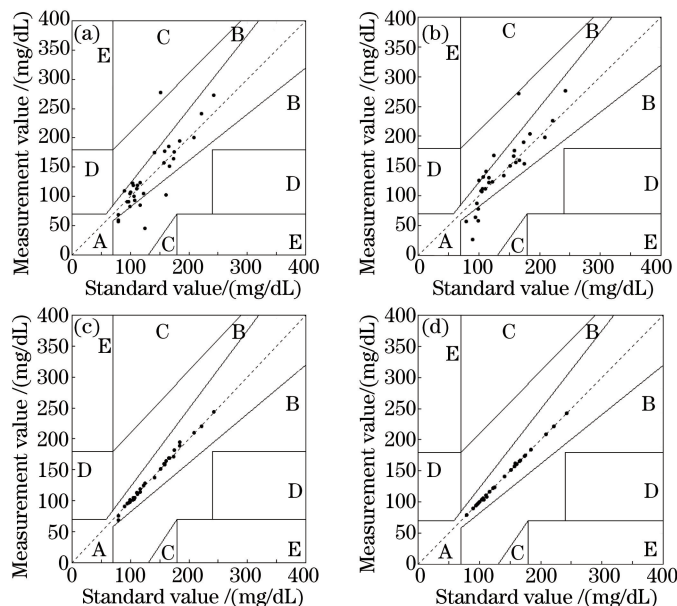


图 3 (a)吸收光谱法;(b)旋光法;(c)拉曼光谱法;(d) BP 神经网络融合后测量的的克拉克误差网格分析图

Fig. 3 Clarke error grid analysis of (a) absorption spectrum method; (b) polarimetry; (c) Raman spectrum method; (d) BP neural network fusion

5 结 论

利用多种光谱数据和 BP 神经网络建立了血糖含量的预测模型,利用 BP 神经网络可以规避仪器、测量方法、人为等因素引起的异常数据对模型的不良影响,使得分析结果达到实用的预测精度。使用 BP 神经网络模型可将血糖的多光谱测量数据有机地融合,测量结果可达到临床精度的要求,预测误差可低于 0.2 mmol/L。使用 BP 神经网络建立模型,当样本足够多时,即使保留极少数几个异常值,也不会影响网络模型的精度,稳健性较好。

参 考 文 献

- [1] Guariguata L, Whiting D R, Hambleton I, *et al.* Global estimates of diabetes prevalence for 2013 and projections for 2035[J]. *Diabetes Research and Clinical Practice*, 2014, 103(2): 137-149.
- [2] So C F, Choi K S, Wong T K, *et al.* Recent advances in noninvasive glucose monitoring[J]. *Medical Devices: Evidence and Research*, 2012, 5: 45-52.
- [3] Goodarzi M, Saeys W. Selection of the most informative near infrared spectroscopy wavebands for continuous glucose monitoring in human serum[J]. *Talanta*, 2016, 146(1): 155-165.
- [4] Goodarzi M, Sharma S, Ramon H, *et al.* Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring[J]. *TrAC Trends in Analytical Chemistry*, 2015, 67: 147-158.
- [5] Vezouviou E, Lowe C R. A near infrared holographic glucose sensor[J]. *Biosensors and Bioelectronics*, 2015, 68(15): 371-381.
- [6] Yadava J, Rania A, Singh V, *et al.* Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy[J]. *Biomedical Signal Processing and Control*, 2015, 18: 214-227.
- [7] Cheng C, Cheng X S, Dai N, *et al.* Prediction of facial deformation after complete denture prosthesis using BP neural network[J]. *Computers in Biology and Medicine*, 2015, 66(1): 103-112.
- [8] Duan Chengpeng, Liu Wei, Chen Yaohong, *et al.* Multiple background sampling adaptive non-uniform correction algorithm[J]. *Acta Optica Sinica*, 2016, 36(10): 1020001.
段程鹏, 刘 伟, 陈耀弘, 等. 多本底采样自适应非均匀校正算法[J]. *光学学报*, 2016, 36(10): 1020001.
- [9] Cheng Yuqiong, Lu Wei, Luo Hui, *et al.* Study on prediction of rice seed germination rate by using continuous polarization spectroscopy and inlaid grey neural network[J]. *Acta Optica Sinica*, 2015, 35(12): 1230001.
程宇琼, 卢 伟, 罗 慧, 等. 基于连续偏振光谱技术与嵌入型灰色神经网络的稻种发芽率检测方法研究[J]. *光学学报*, 2015, 35(12): 1230001.
- [10] Cui Rixian, Liu Yadong, Fu Jindong. Estimation of winter wheat leaf nitrogen accumulation using machine learning algorithm and visible spectral[J]. *Spectroscopy and Spectral Analysis*, 2016, 36(6): 1837-1842.
崔日鲜, 刘亚东, 付金东. 基于机器学习和可见光光谱的冬小麦叶片氮积累量估算[J]. *光谱学与光谱分析*, 2016, 36(6): 1837-1842.
- [11] Caduff A, Mueller M, Megej A, *et al.* Characteristics of a multisensor system for non invasive glucose monitoring with external validation and prospective evaluation[J]. *Biosensors and Bioelectronics*, 2011, 26(9): 3794-3800.
- [12] Li D M, Jia S H, Wang J. Study on continuous monitoring glucose concentration with terbium gallium garnet crystal [C]. *International Conference on Advanced Design and Manufacturing Engineering*, 2011, 317-319: 53-57.
- [13] Dong J, Zhuang D F, Huang Y H, *et al.* Advances in multi-sensor data fusion: algorithms and applications [J]. *Sensors*, 2009, 9(10): 7771-7784.
- [14] Wu B, Han S J, Xiao J, *et al.* Error compensation based on BP neural network for airborne laser ranging[J]. *Optik*, 2016, 127(8): 4083-4088.
- [15] Sivananthan S, Naumova V, Man C D, *et al.* Assessment of blood glucose predictors: the prediction-error grid analysis [J]. *Diabetes Technology Therapeutics*, 2011, 13(8): 787-796.