

基于 BP 神经网络的血红蛋白定量光学检测方法

王姗姗, 黄凯, 李铭, 陈平, 刘伟伟, 林列

南开大学现代光学研究所, 天津 300350

摘要 构建了基于 BP 神经网络的血红蛋白定量计算模型, 其既可用于血红蛋白检测, 又能成功地区分不同肿瘤疾病。建立了血红蛋白定量计算的双隐含层 BP 神经网络模型, 预测集的相关系数为 0.9838, 预测集的相对偏差为 2.532%; 利用该模型得到了乳腺肿瘤患者和白血病患者血清中血红蛋白的含量, 结果显示两者浓度间存在极其显著性差异 ($P < 0.001$), 预示该模型在区分肿瘤疾病方面具有潜在的应用价值。

关键词 医用光学; BP 神经网络; 血红蛋白; 光学检测; 定量检测

中图分类号 O433.4

文献标识码 A

doi: 10.3788/AOS201838.0717002

Quantitative Optical Detection Method of Hemoglobin Based on BP Neural Network

Wang Shanshan, Huang Kai, Li Ming, Chen Ping, Liu Weiwei, Lin Lie

Institute of Modern Optics, Nankai University, Tianjin 300350, China

Abstract In this paper, we establish a quantitative model based on BP neural network to detect hemoglobin in serum, and put it into practice of hemoglobin measurement, in the result of which various cancer diseases can be completely distinguished. A double-hidden-layer BP neural network model to quantify hemoglobin is established. The correlation coefficient of the prediction set in the model is 0.9838, and its relative deviation is 2.532%. The hemoglobin concentrations in serum samples of breast cancer patients and leukemia patients are obtained by the BP neural network model. A remarkable difference appears in the concentrations between these two sorts of patients ($P < 0.001$), indicating that the model has potential application in distinguishing tumor diseases.

Key words medical optics; BP neural network; hemoglobin; optical detection; quantitative detection

OCIS codes 170.1470; 170.1610; 170.4580

1 引 言

血红蛋白是血液中的一种机能性蛋白质, 在生物体新陈代谢过程中起着非常重要的作用。根据世界卫生组织统计, 人体的血红蛋白浓度会因个体的不同而有所不同, 除年龄、性别等因素以外, 与人的健康状况也密切相关^[1]。创伤性溶血或人体生理性失血等都会使人体血红蛋白含量偏低; 相反, 肿瘤疾病^[2-4]、肾脏病变、心肺病变以及慢性高原病等会使血红蛋白含量偏高^[5-6]。因此, 血红蛋白浓度的定量检测对生物体健康状况的判断具有重要意义。医学

上测定血红蛋白浓度的方法有: 比色法^[7-9]、测铁法、测氧法、比重法。测铁法和测氧法操作复杂, 不适于常规使用; 比重法准确性差; 比色法相对于生化检测方法精度高, 且具有操作简单、显色快等优点, 是临床上使用最广泛的方法。比色法中的氰化高铁分光光度法具有测量结果准确、重复性好等优点, 被认为是测量血红蛋白浓度最准确的方法^[10]。但是, 该方法需要对样品进行预标记, 而且仅能对特定波峰处的数据进行分析, 数据的利用率低。

人工神经网络算法可将大量的输入数据进行非线性运算后体现在一个输出结果中。近年来, 人工

收稿日期: 2017-11-24; 收到修改稿日期: 2018-03-05

基金项目: 天津市科技支撑重点项目(15ZCZDZX00250, 08ZCKFGX09400)、发光学及应用国家重点实验室开放基金资助项目

作者简介: 王姗姗(1993—), 女, 硕士研究生, 主要从事生物医学光学方面的研究。

E-mail: wangshanshan@mail.nankai.edu.cn

导师简介: 陈平(1971—), 女, 博士, 研究员, 博士生导师, 主要从事生物医学光子学、LSPR 微流控与柔性光电器件、生物光学信息处理方面的研究。E-mail: chping@nankai.edu.cn(通信联系人)

神经网络在医学领域的应用取得了一定进展,在疾病诊断^[11]、医学影像处理^[12]、临床决策分析^[13]等方面得到了广泛应用。文献^[14-15]利用误差反向传播(BP)神经网络对光谱数据与相应的血红蛋白浓度进行建模分析,利用训练好的网络模型能够较准确地检测血红蛋白的浓度;但其建立的BP神经网络模型采用单隐含层^[15],得到的预测集相关系数(0.9365)有待提高。鉴于此,本文基于BP神经网络算法构建了计算人血清中血红蛋白含量的双隐含层BP神经网络模型,预测集的相关系数和相对偏差分别为0.9838和2.532%,并将此计算模型运用于血红蛋白吸收光谱的数据处理中,实现了血红蛋白含量的快速、无标记、准确检测。本研究用上述方法得到了乳腺肿瘤和白血病患者血清中的血红蛋白含量,结果显示两者存在显著性差异($P < 0.001$),预示这种方法在区分肿瘤疾病方面具有潜在的应用价值。

2 实验原理

本实验主要利用血红蛋白的紫外可见吸收光谱获得实验数据,并采用BP神经网络处理实验数据,从含有多成分的血清吸收光谱中直接获得血红蛋白的含量。紫外可见吸收光谱法的基本原理是朗伯-比尔定律,光被透明介质吸收的比例与入射光的强度无关,在光程上等厚层介质吸收相同比例的光,数学表达式为

$$A = \lg(I_0/I_t) = Kbc, \quad (1)$$

式中: A 为吸光度; I_0 为入射光强; I_t 为透射光强; K 为摩尔消光系数; b 为光程; c 为浓度。

本研究采用BP神经网络算法处理数据,BP神经网络是人工神经网络的一种。人工神经网络的基本组成单元是人工神经元,原理如图1所示,其有多个输入 $x_i (i = 1, 2, \dots, n)$,经过非线性运算后只有单个输出结果 y 。

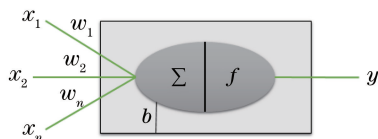


图1 人工神经元示意图^[16]

Fig. 1 Artificial neuron diagram

对于人工神经单元,其数学模型可以用数学表达式表示为

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right), \quad (2)$$

式中: y 为神经元的输出结果; f 为非线性或线性转

换函数; b 为偏置; $w_i (i = 1, 2, \dots, n)$ 为第 i 个信号的作用权重,即神经元与输入信号之间的连接强度; $x_i (i = 1, 2, \dots, n)$ 为神经元接收的第 i 个输入信号。

人工神经网络需要预先用已知对应关系的样本进行网络训练,通过预测样本的检测后再应用到检测未知样品上。BP神经网络模型将训练过程分为信号的正向传播与误差的反向传播两个阶段。在正向传播阶段,输入信息从输入层经隐含层传向输出层,在输出端产生输出信号。如果在输出层不能得到期望的输出,则转入误差信号反向传播。在反向传播阶段,未能满足精度要求的误差信号由输出端开始,以某种方式逐层向输入端传播,并将误差分摊给各层的所有单元,依据误差信号动态地调整各单元层的连接权重。经过反复的正向传播与反向调节后,当输出信号的误差满足精度要求时,神经网络学习停止,训练过程结束。

3 实 验

训练BP神经网络模型的标准样本所用血红蛋白和血清均购于Sigma公司(美国),纯度为99.7%。待检测样本为18名健康人、18名乳腺肿瘤患者和18名白血病患者血液样本,由吉林大学第一附属医院和天津医科大学附属医院总医院提供。

利用美国伯腾仪器有限公司的BioTek微量吸收光谱仪Epoch获得上述样品的光谱数据,波长扫描范围为200~999 nm,步长设置为1 nm,实验温度为25 °C,相对湿度为16%。

4 结果与讨论

采用BP神经网络模型处理数据,模型的输入数据是实验测得的吸收光谱的吸光度,目标数据是配制的血红蛋白溶液的参考浓度。

为了保证构建的网络模型达到高的相关系数且不冗余,建立模型前先对数据进行预处理。根据本实验测得的实验数据(见图2)可知,血清的吸收光谱峰值在280 nm处,血红蛋白的吸收光谱峰值在405 nm处。为了降低网络的复杂度和避免模型过拟合,仅选取血红蛋白光谱特征峰所在的波段(395~415 nm)对应的吸光度数据,以达到对数据库降维的目的,缩短训练时间,减小存储空间。另外,为了构建的模型能够准确预测出实际血清中血红蛋白的含量,将实验数据随机缩放到0.1~2.5 g/L范围内(待测样品的实际质量浓度在0.1~2.5 g/L范围内),使构建模型的数据范围覆盖后期

待测血红蛋白的浓度。由于实验获得的数据量有限,通过对实验数据附加均值为 0、方差为 0.01 的高斯噪声来增加训练数据量,使数据由原来的 84 组增加到 335 组,以提高模型的泛化能力。

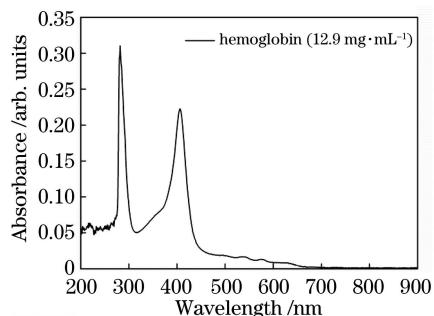


图 2 血清中血红蛋白的吸收光谱

Fig. 2 Absorption spectrum of hemoglobin in serum

BP 网络的拓扑结构和节点数对网络模型的预测精度非常重要,根据实验数据和要解决的问题确定 BP 神经网络拓扑结构^[17-18]。本网络模型采用双隐含层,即 BP 神经网络的拓扑结构包括 4 层:输入层、2 个隐含层和输出层。由于选择的是 395~415 nm(步长为 1 nm)波段的 21 个光谱数据,因此输入节点数设置为 21。输出结果为输入的光谱数据对应的血红蛋白浓度,则输出层设置为 1。采用网络结构增长型方法确定节点数,即先设置较少的节点数对网络进行训练,并测试训练误差,然后逐渐增加节点数,直到学习误差不再有明显降低为止。测试显示隐含层采用 10-5 结构的平均相对误差最小(0.09%),最终确定 BP 神经网络的拓扑结构为 21-10-5-1。

本模型学习率的初始值取经验值 0.01。对神经网络输入的样本数是 335 个,归一化处理后经由输入层随机给每个数据权重值 w_i 和偏置 b 赋值,然后被传输到隐含层^[19]。

为了选择传递函数,对获得的实验数据进行评估。首先,用 MATLAB 对已获得的光谱数据和对应的血红蛋白浓度分布进行评估,获得实验数据的散点分布如图 3 所示。

由图 3 可知,实验数据中的吸光度和浓度呈现出较好的线性关系,即传递函数选择线性函数。为了增加模型的非线性映射能力,本模型在 2 个隐含层之间设置了 1 个 Sigmoid 函数,因此,网络隐含层的传递函数选择 tansig 函数,即每个隐含层内神经元之间的传递函数采用正切函数,2 个隐含层之间的函数采用 Sigmoid 函数,经过 2 个隐含层的权重

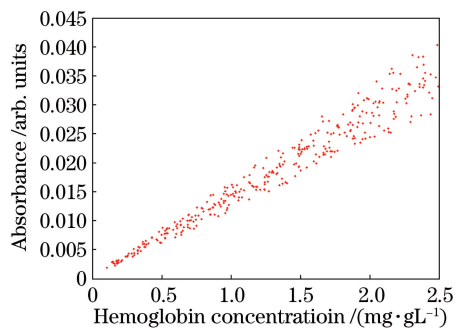


图 3 实验数据散点分布图

Fig. 3 Scatter diagram of experimental data

加权及转换,传输到输出层输出神经网络的预测值。

使用 MATLAB 自带的初始化函数逐次对网络的初始权值和偏置赋值,在此权值和偏置下对网络进行训练。调整权值和偏置的规则有随机梯度下降(SGD)法、最速下降法、Levenberg-Marquardt 法等。训练过程中部分算法会出现陷入局部最优或存储空间占用大等问题,因此选择随机梯度下降法对网络进行训练和结果预测,训练模型的结果如图 4 所示,图中蓝色曲线为由训练数据输入网络得到的训练误差,绿色曲线为验证样本输入网络后得到的验证误差,红色曲线为由预测数据得到的预测误差。

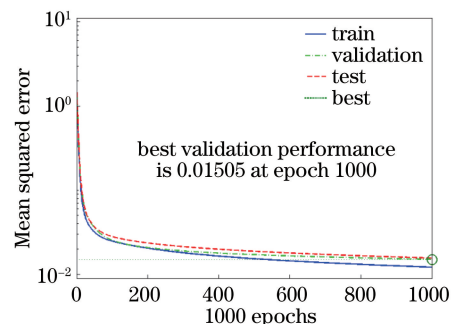


图 4 随机梯度下降法网络训练结果

Fig. 4 Network training results of stochastic gradient descent

由图 4 可知,网络模型经过了 1000 次训练后没有出现局部最优,且训练网络模型的速度较快,训练误差、验证误差和预测误差都具有较好的收敛性。采用此方法训练网络模型得到了模型的相关系数,如图 5 所示(Y 表示模型输出数据, T 表示目标数据)。

由图 5 可知,训练结果与目标数据的相关系数 $R=0.9900$,验证结果与目标数据的相关系数 $R=0.9857$,预测结果与目标数据的相关系数 $R=0.9838$,整体训练出来的网络模型的相关系数 $R=0.9882$ 。结果显示,模型的检测值与实际结果具有良好的相关性,可用于离体血红蛋白含量的定量分析。

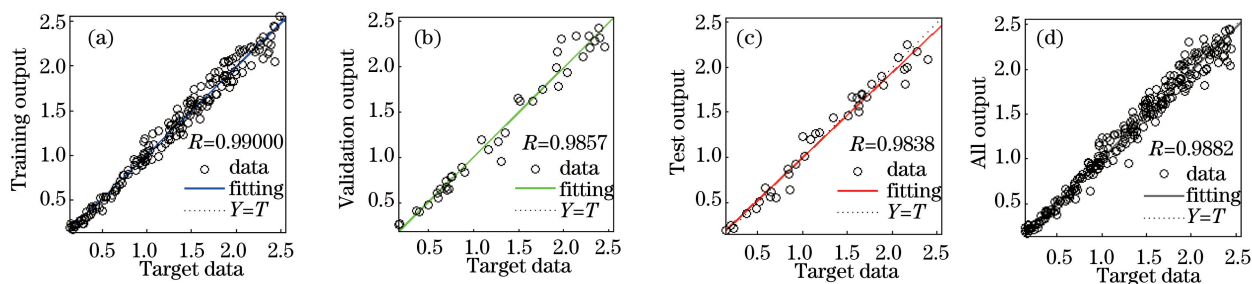


图 5 随机梯度下降法网络训练的相关系数。(a)训练结果;(b)验证结果;(c)预测结果;(d)整体训练结果

Fig. 5 Correlation coefficients of network training using stochastic gradient descent method.

(a) Training output; (b) validation output; (c) test output; (d) all output

利用测试集样本数据和随机梯度下降法对网络模型进行预测,预测结果如表 1 所示。

表 1 随机梯度下降法网络训练的部分结果

Table 1 Partial network training results of stochastic gradient descent method

Target data / $(\text{g}\cdot\text{L}^{-1})$	Predicted data / $(\text{g}\cdot\text{L}^{-1})$	Residual error / $(\text{g}\cdot\text{L}^{-1})$	Relative deviation /%
0.3256	0.3277	0.0021	0.6450
0.9457	0.9683	0.0226	2.3900
1.4900	1.4760	-0.0140	0.9200
2.3510	2.3450	-0.0060	0.2760
0.6827	0.7063	0.0236	3.4570
1.2980	1.2890	0.0090	0.6934
0.7420	0.7492	0.0072	0.9704
0.6337	0.6592	0.0255	4.0240
0.5208	0.4947	-0.0261	5.0120
2.1930	2.2380	0.0450	2.0560
1.1800	1.1410	-0.0390	2.5690
0.9967	1.0200	0.0233	2.3380
2.2710	2.2170	-0.0540	2.3600
1.0410	1.1020	0.0610	5.8880
1.2730	1.3290	0.0560	4.3740
Relative deviation /%			2.532

由训练算法的预测结果和残差分布可知,随机梯度下降法表现出较好的结果,相对偏差(2.532%)比较小。因此,选择随机梯度下降法作为网络训练算法建立的血红蛋白含量计算模型性能更加优异。

用本模型检测 18 名乳腺肿瘤患者、18 名白血病患者和 18 名健康人血清中血红蛋白的含量。实验测得此三者的平均吸收光谱如图 6 所示,由图 6 可知,乳腺肿瘤患者和白血病患者血清中存在游离的血红蛋白。利用此模型得到的实验结果如表 2 所示。

对表 2 进行统计分析可知,健康人血清中血红蛋白的质量浓度为 0.07~0.62 g/L,白血病患者血清中血红蛋白的质量浓度为 0.20~0.95 g/L,乳腺肿瘤患者血清中血红蛋白的质量浓度为 1.08~2.00 g/L。1、4、6、10、12、13 和 16 号乳腺肿瘤患者由于接受过药物治疗,血清中血红蛋白浓度表现出

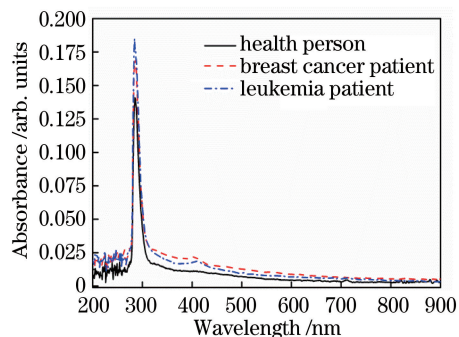


图 6 健康人、乳腺肿瘤患者和白血病患者血清的吸收光谱
Fig. 6 Absorption spectra of serum in healthy person, patients with breast cancer and leukemia

较低水平。经正态性检验后发现,健康人、白血病患者、乳腺肿瘤患者血清中血红蛋白含量服从正态分布,采用 t 检验 (student's t -test) 对这三者的差异

表 2 健康人、白血病患者和乳腺肿瘤患者血清中的血红蛋白含量

Table 2 Hemoglobin content in serum of healthy person, patients with breast cancer and leukemia

Serial number	Mass concentration of hemoglobin / (g·L ⁻¹)		
	Healthy person	Leukemia patient	Breast cancer patient
1	0.16	0.42	0.78
2	0.37	0.66	1.08
3	0.28	0.76	2.00
4	0.62	0.41	0.94
5	0.26	0.55	1.27
6	0.51	0.33	0.43
7	0.53	0.23	1.39
8	0.17	0.62	1.97
9	0.07	0.40	1.50
10	0.36	0.80	0.58
11	0.29	0.51	1.46
12	0.36	0.43	0.56
13	0.48	0.50	0.46
14	0.46	0.73	1.82
15	0.29	0.48	1.70
16	0.20	0.32	0.62
17	0.54	0.95	1.58
18	0.27	0.64	1.64

进行统计学分析,结果显示健康人与白血病患者血红蛋白浓度存在显著性差异($P=0.002<0.01$),健康人与乳腺肿瘤患者的血红蛋白浓度存在极其显著性差异($P<0.001$),白血病患者和乳腺肿瘤患者血清中血红蛋白的浓度存在极其显著性差异($P<0.001$)。由此可得,肿瘤患者血清中血红蛋白浓度比健康人血清中的普遍偏高,白血病患者血清中血红蛋白浓度比乳腺肿瘤患者的偏低,期望这一发现能对肿瘤病人疾病种类的判定起到辅助作用。当人体内发生病变时,血清中大分子物质(如蛋白质、糖类、核酸)的结构与含量会因病原产物与代谢产物进入血液而发生变化^[20-22],同时肿瘤患者组织癌变会引起人体内环境变化,进而导致血清中的游离血红蛋白增多^[23],并且肿瘤患者体内血红蛋白的浓度会随肿瘤组织体积的增大而逐渐增加^[3],这表明血红蛋白含量的光学检测对生物体健康状况的判断具有重要意义。

5 结 论

为了实现血红蛋白含量快速、无标记、准确检测,基于BP神经网络算法构建了计算人血清中血红蛋白含量的双隐含层BP神经网络模型,预测集的相关系数为0.9838,预测集的相对偏差为2.532%;将此模型运用于血红蛋白吸收光谱的数据处理,实现了血

蛋白含量快速、无标记、准确检测。采用上述方法测得了乳腺肿瘤患者和白血病患者血清中的血红蛋白含量,结果显示两者存在极其显著性差异($P<0.001$),可用于区分这2种肿瘤疾病,预示该方法在区分肿瘤疾病方面具有潜在的应用价值。

参 考 文 献

- [1] Li L Q, Wei H J, He B H, *et al.* Gastric cancer detection using diffuse reflectance spectral ratio R_{540}/R_{575} of oxygenated hemoglobin bands [J]. *Acta Optica Sinica*, 2009, 29(10): 2860-2865.
李兰权, 魏华江, 何博华, 等. 采用氧合血红蛋白吸收带的漫反射光谱比率 R_{540}/R_{575} 检测人胃癌 [J]. *光学学报*, 2009, 29(10): 2860-2865.
- [2] Croci S, Pedrazzi G, Passeri G, *et al.* Acetylphenylhydrazine induced haemoglobin oxidation in erythrocytes studied by Mössbauer spectroscopy [J]. *Biochimica et Biophysica Acta (General Subjects)*, 2001, 1568(1): 99-104.
- [3] Orlova A G, Kirillin M Y, Volovetsky A B, *et al.* Diffuse optical spectroscopy monitoring of oxygen state and hemoglobin concentration during SKBR-3 tumor model growth [J]. *Laser Physics Letters*, 2016, 14(1): 015601.
- [4] Boas D A, Franceschini M A. Haemoglobin oxygen saturation as a biomarker: the problem and a solution [J]. *Philosophical Transactions A*, 2011, 369

- (1955): 4407-4424.
- [5] Wang X M, Wang J H, Wang X D, *et al.* Application in diagnostic cancer by Fourier transform infrared spectrometry [J]. *Modern Instruments & Medical Treatment*, 2010, 16(4): 8-11.
王绪明, 王静怀, 王旭东, 等. 傅里叶变换红外光谱在癌症诊断中的应用进展 [J]. *现代仪器*, 2010, 16(4): 8-11.
- [6] Saetchnikov V A, Tcherniavskaia E A, Schweiger G. Combined fiber optical thermal sensor for noninvasive monitoring of blood and human tissue through diffuse scattering and metabolic parameters [J]. *Proceedings of SPIE*, 2007, 6628: 66280N.
- [7] McCall K B. Spectrophotometric determination of total hemoglobin in plasma [J]. *Analytical Chemistry*, 1956, 28(2): 189-191.
- [8] Rodkey F L, Hill T A, Pitts L L, *et al.* Spectrophotometric measurement of carboxyhemoglobin and methemoglobin in blood [J]. *Clinical Chemistry*, 1979, 25(8): 1388-1393.
- [9] Blakney G B, Dinwoodie A J. A spectrophotometric scanning technique for the rapid determination of plasma hemoglobin [J]. *Clinical Biochemistry*, 1975, 8(2): 96-102.
- [10] Schalm O W, Jain N C, Carroll E J. *Veterinary hematology* [M]. 3rd ed. Philadelphia: Lea and Febiger, 1975: 54.
- [11] Yang Y C, Chen H, Wang D, *et al.* Diagnosis of pancreatic carcinoma based on combined measurement of multiple serum tumor markers using artificial neural network analysis [J]. *Chinese Medical Journal*, 2014, 127(10): 1891-1896.
- [12] Muralidaran C, Dey P, Nijhawan R, *et al.* Artificial neural network in diagnosis of urothelial cell carcinoma in urine cytology [J]. *Diagnostic Cytopathology*, 2015, 43(6): 443-449.
- [13] Yun J, Mackenzie M, Rathee S, *et al.* An artificial neural network (ANN)-based lung-tumor motion predictor for intrafractional MR tumor tracking [J]. *Medical Physics*, 2012, 39(7): 4423-4433.
- [14] Ding H Q, Lu Q P, Gao H Z, *et al.* Non-invasive prediction of hemoglobin levels by principal component and back propagation artificial neural network [J]. *Biomedical Optics Express*, 2014, 5(4): 1145-1152.
- [15] Zhang Z Y, Men J L, Li G, *et al.* Non-invasive measurement of human hemoglobin concentration by dynamic spectrum method [J]. *Spectroscopy and Spectral Analysis*, 2010, 30(1): 150-153.
张志勇, 门剑龙, 李刚, 等. 动态光谱法用于人体血红蛋白浓度的无创测量 [J]. *光谱学与光谱分析*, 2010, 30(1): 150-153.
- [16] Han L Q. *Artificial neural network tutorial* [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2006: 29-44.
韩力群. *神经网络教程* [M]. 北京: 北京邮电大学出版社, 2006: 29-44.
- [17] Qi D H, Kang J C. On design of the BP neural network [J]. *Computer Engineering & Design*, 1998, 19(2): 48-50.
戚德虎, 康继昌. BP神经网络的设计 [J]. *计算机工程与设计*, 1998, 19(2): 48-50.
- [18] Zhang X C, Zhou J J. Improved artificial neural network algorithm(I): optimization and convergence criterion of network structure [J]. *Computers and Applied Chemistry*, 1995, 12(3): 186-191.
张晓晨, 周家驹. 改进的人工神经网络算法(I): 网络结构的优化和收敛判据 [J]. *计算机与应用化学*, 1995, 12(3): 186-191.
- [19] Sun S Q, Chu F H. Temperature compensation of fiber Bragg grating current sensor based on optimized neural network algorithm [J]. *Acta Optica Sinica*, 2017, 37(10): 1006001.
孙诗晴, 初凤红. 基于优化神经网络算法的光纤布拉格光栅电流传感器的温度补偿 [J]. *光学学报*, 2017, 37(10): 1006001.
- [20] Chen D R, Chen S T, Wang T W, *et al.* Characterization of estrogen quinone-derived protein adducts and their identification in human serum albumin derived from breast cancer patients and healthy controls [J]. *Toxicology Letters*, 2011, 202(3): 244-252.
- [21] Lokhov P G, Dashtiev M I, Moshkovskii S A, *et al.* Metabolite profiling of blood plasma of patients with prostate cancer [J]. *Metabolomics*, 2010, 6(1): 156-163.
- [22] Shaw R A, Kotowich S, Eysel H H, *et al.* Arthritis diagnosis based upon the near-infrared spectrum of synovial fluid [J]. *Rheumatology International*, 1995, 15(4): 159-165.
- [23] Gunasekaran S, Sankari D. FTIR and UV-visible spectral study on normal and jaundice blood samples [J]. *Asian Journal of Chemistry*, 2004, 16(3/4): 1779-1786.