Bio-Medical Materials and Engineering 26 (2015) S2025–S2032
DOI 10.3233/BME-151507
IOS Press

Hematocrit estimation using online sequential extreme learning machine

Hieu Trung Huynh[a], Yonggwan Won[b,*] and Jinsul Kim[b]
[a]*Faculty of Information Technology, Industrial University of Ho Chi Minh City, Viet Nam*
[b]*Department of Computer Engineering, Chonnam National University, Gwangju 500-757, Korea*

Abstract. Hematocrit is a blood test that is defined as the volume percentage of red blood cells in the whole blood. It is one of the important indicators for clinical decision making and the most effective factor in glucose measurement using handheld devices. In this paper, a method for hematocrit estimation that is based upon the transduced current curve and the neural network is presented. The salient points of this method are that (1) the neural network is trained by the online sequential extreme learning machine (OS-ELM) in which the devices can be still trained with new samples during the using process and (2) the extended features are used to reduce the number of current points which can save the battery power of devices and speed up the measurement process.

Keywords: hematocrit, extreme learning machine, online training, OS-ELM, neural network

1. Introduction

Hematocrit (HCT) is expressed as packed cell volume (PCV) which is a very useful clinical factor in hemodialysis, surgical procedure, and anemia. It could be used to estimate transfusions and determine the extent of anemia [1-3]. In addition, it is also a factor which highly affects the accuracy of glucose measurements [4-6]. Hence, estimating the hematocrit also plays an important role to improve the glucose measurement accuracy.

Hematocrit density is generally measured by centrifuging the blood in a capillary tube. With modern lab equipment, it can be measured indirectly by an automated analyzer in which hematocrit is calculated by multiplying the mean cell volume by the red blood cell count. Treo et al. proposed an approach for hematocrit estimation based on the dielectric spectroscopy, which is still complicated or requires special devices [7]. In our previous study, a neural network approach was developed for improving accuracy of the handheld glucose meters based on changing pattern of the transduced current curve [8].

The neural networks have been successfully applied to many fields including biology and medical data analysis [9-12]. Several network architectures have been proposed, however a single hidden layer feedforward neural network (SLFN) with the proper activation function of hidden layer can approximate any function and form the decision boundaries with arbitrary shapes. Traditionally, neural

* Address for correspondence: Yonggwan Won, Department of Computer Engineering, Chonnam National University, Korea. Tel.: 82-62-530-1804; Fax: 82-62-530-1759; E-mail: ykwon@chonnam.ac.kr.

networks could be trained by back-propagation (BP). This algorithm may have some obstacles such as slow convergence due to the improper learning rate, over-fitting or local minima. Although many researchers proposed several improvements to overcome the problems of BP algorithm [13, 14], up to now, training algorithms based upon the gradient descent approaches are still slow due to many iterative steps which are required in the training process. These problems may be overcome by an effective training algorithm which is called as extreme learning machine (ELM) proposed by Huang et al. [15, 16]. This algorithm can offer good performance with high learning speed in many applications.

Besides batch learning algorithms, the online sequential learning algorithms have shown their usefulness in applications where the full training set is not available or the memory costs are expensive. In addition, the network weights are updated corresponding to the new training data only. In this paper, an approach for estimating hematocrit which uses the neural network is proposed. Unlike previous studies [8], the network is trained by the online sequential extreme learning machine (OS-ELM) which is one of the effective sequential learning algorithms based on ELM [17]. Values of input features for the neural network are sampled from the transduced anodic current curves and the desired output values are obtained from the hospital analysis system. The extended features are also used to reduce the number of current points which can save the battery power of devices and speed up the measurement process. The rest of this paper is organized as follow. The online sequential extreme learning machine is reviewed in Section 2. The proposed approach for estimating hematocrit is presented in Section 3. Section 4 shows the experimental results and analysis. Finally, the conclusion is in Section 5.

2. Online sequential extreme learning machine

The neural networks are powerful in biology and medial data analysis. Several neural network architectures have been developed. However, it was shown that the SLFN can approximate any function if the activation function of hidden units and the number of hidden units are chosen properly. Therefore, in this study we focus on the SLFN. A typical architecture of the SLFN includes an input layer, an output layer, and a hidden layer which is between the input and output layer. The input layer collects information presented from surroundings and the output layer generates a response to a given input. For an input pattern $\mathbf{x}$, the SLFN with the activation of $\varphi(\cdot)$ in the hidden layer can be mathematically modeled as:

$$\mathbf{o} = \sum_{k=1}^{K} \mathbf{a}_k \varphi(\mathbf{x} \cdot \mathbf{w}_k + b_k), \tag{1}$$

where $\mathbf{o}$ is the output vector, $\mathbf{w}_k$ is the weight vector connecting from the input units to the k-th hidden unit, b_k is its bias, and $\mathbf{a}_k$ is the weight vector connecting from the k-th hidden unit to the output units.

An important operation in neural networks is training which tries to determine the network weights to optimize a predefined criterion. One of the most popular criteria is to minimize the error function defined by

$$E = \sum_{i=1}^{n} \left\| \mathbf{o}_i - \mathbf{d}_i \right\|^2, \tag{2}$$

where n is the number of training patterns, $\mathbf{d}_i=[d_{i1}\ d_{i2}\ ...\ d_{iC}]^T$ is the desired output vector corresponding to the i-th input pattern $\mathbf{x}_i=[x_{i1}\ x_{i2}\ ...\ x_{iN}]^T$. Traditionally, determining $\mathbf{w}$'s, $\mathbf{a}$'s, and b's can be performed by using the gradient-descent methods. One of the popular algorithms based on the gradient descent for training neural networks is back-propagation (BP). This algorithm may have problems such as learning rate, epochs, local minima, overtraining, etc. Several improvements have been proposed for BP algorithm. However, up to now it is still slow because it requires several iterative steps during the training process. These problems have been overcome by an algorithm called as extreme learning machine (ELM) which was proposed by Huang et al. [15, 16]. The ELM algorithm is applied for training SLFNs. It relies on a linear model given by

$$\mathbf{\Pi\Gamma}=\mathbf{D}, \tag{3}$$

where $\mathbf{\Pi}$ is the hidden layer output matrix of SLFN and defined by [16]:

$$\mathbf{\Pi}=\begin{bmatrix} \varphi(\mathbf{w}_1\cdot\mathbf{x}_1+b_1) & \cdots & \varphi(\mathbf{w}_K\cdot\mathbf{x}_1+b_K) \\ \vdots & \ddots & \vdots \\ \varphi(\mathbf{w}_1\cdot\mathbf{x}_n+b_1) & \cdots & \varphi(\mathbf{w}_K\cdot\mathbf{x}_n+b_K) \end{bmatrix}, \tag{4}$$

$$\mathbf{D}=[\mathbf{d}_1\ \mathbf{d}_2\ \ldots\ \mathbf{d}_n]^T, \tag{5}$$

and

$$\mathbf{\Gamma}=[\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_C]. \tag{6}$$

In ELM, the biases and input weights are randomly assigned, and the output weights are determined by

$$\hat{\mathbf{\Gamma}}=\mathbf{\Pi}^{\dagger}\mathbf{D}, \tag{7}$$

where $\mathbf{\Pi}^{\dagger}$ is the pseudo-inverse of $\mathbf{\Pi}$. Several improvements of ELM have also been developed by researchers [18, 19]. They can obtain good performance with high learning speed in many applications.

When the memory cost is expensive or when the training set is very large, the online training algorithms are preferred. An online training algorithm for SLFNs called as online sequential extreme learning machine (OS-ELM) was proposed by Liang et al. [17]. This algorithm is based on the ELM in which the input weights and hidden layer biases are assigned by the random values, the output weights are updated by non-iterative steps corresponding to an arrived training subset. In summary, the OS-ELM algorithm can be described as follows:

(1) Initialization:

For an initial training set $\mathbf{S}_0=\{(\mathbf{x}_i, \mathbf{d}_i) \mid i=1,\ldots, n_0\}$:

- Assign the random values for input weights ($\mathbf{w}_k$) and biases (b_k).
- Determine the hidden output matrix $\mathbf{\Pi}_0=[\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_{n0}]^T$, where $\mathbf{h}_i=[\varphi(\mathbf{w}_1\cdot\mathbf{x}_i+b_1)\ \varphi(\mathbf{w}_2\cdot\mathbf{x}_i+b_2)\ \ldots\ \varphi(\mathbf{w}_K\cdot\mathbf{x}_i+b_K)]^T$.
- Calculate $\mathbf{\Gamma}_0$ as

$$\mathbf{\Gamma}_0=\mathbf{L}_0^{-1}\mathbf{\Pi}_0^T\mathbf{D}_0, \tag{8}$$

where $\mathbf{L}_0 = \mathbf{\Pi}_0^T \mathbf{\Pi}_0$, and $\mathbf{D}_0 = [\mathbf{d}_1\ \mathbf{d}_2\ \ldots\ \mathbf{d}_{n0}]^T$

(2) Updating weights

For an arrived training set $\mathbf{S}_m = \{(\mathbf{x}_i, \mathbf{d}_i) \mid i = \sum_{i=0}^{m-1} n_i + 1, \ldots, \sum_{i=0}^{m} n_i\}$:

- Calculate the hidden output matrix $\mathbf{\Pi}_m = [\mathbf{h}_{\sum_{i=0}^{m-1} n_i + 1}\ \mathbf{h}_{\sum_{i=0}^{m-1} n_i + 2} \ldots \mathbf{h}_{\sum_{i=0}^{m} n_i}]^T$.
- Update the output weights $\mathbf{\Gamma}_m$ by

$$\mathbf{L}_m = \mathbf{L}_{m-1} + \mathbf{\Pi}_m^T \mathbf{\Pi}_m \tag{9}$$

$$\mathbf{\Gamma}_m = \mathbf{\Gamma}_{m-1} + \mathbf{L}_m^{-1} \mathbf{\Pi}_m^T (\mathbf{D}_m - \mathbf{\Pi}_m \mathbf{\Gamma}_{m-1}), \tag{10}$$

where $\mathbf{D}_m = [\mathbf{d}_{\sum_{i=0}^{m-1} n_i + 1}\ \mathbf{d}_{\sum_{i=0}^{m-1} n_i + 2} \ldots \mathbf{d}_{\sum_{i=0}^{m} n_i}]^T$.

3. Hematocrit estimation

3.1. Hematocrit estimation using the transduced current curve

In this study, the hematocrit estimation is performed by using SLFNs with the input features from the transduced anodic current curve. This current curve is generated by the chemical reaction between the glucose and enzyme during the glucose measurement process. In the electrochemical systems, the popular enzyme which is used in biosensors to detect the glucose levels is the glucose oxidase (GOD). It is utilized to catalyze the oxidation of glucose to produce gluconic acid and hydrogen peroxide. Its reduced form is oxidized to the original state by an electron mediator (ferrocence). The reduced mediator is then oxidized by the active electrode to release electrons. These electrons move between two electrodes to produce a transduced anodic current.

It has shown that the first eight seconds may be the incubation time which does not contain the hematocrit information. Hence, in our study, we address on the second part of the current curve during the next six seconds. In the next six seconds, the anodic current curve is sampled to produce current points. Let $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{iN}]$ be a vector of current points corresponding to the *i-th* current curve. In this study, the sampling frequency is 10Hz, so the number of current points is 59 (N=59). This vector of current points is used as the input features for estimating the hematocrit by using the neural network (SLFN). The neural network is trained by the online sequential learning algorithm.

3.2. Hematocrit estimation using the extended features

In order to reduce the number of sampled points which can reduce energy consumption and speed up the measurement time, we propose to use the additional features which can approximate the current trend. It is most likely that the current points are an exponential function of time. So they can be modeled by

$$x_t = \alpha p_t^{\beta},\ t = 1, 2, \ldots N, \tag{11}$$

where N is the number of sampled points on the current curve, p_t is the t-th time point, α and β are parameters which must be calculated. In order to simplify the calculation of these parameters, the model can be transferred to a linear one. We proceed by taking logarithm on both sides of Eq.11 to give

$$log(x_t)=log(\alpha)+\beta log(p_t),\ t=1,\ 2,\ \ldots,\ N. \tag{12}$$

Let $\omega_1=log(\alpha)$ and $\omega_2=\beta$, Eq. 12 can be rewritten as

$$log(x_t)=\omega_1+\omega_2 log(p_t),\ t=1,\ 2,\ \ldots,\ N. \tag{13}$$

In the matrix form, the linear model is given by

$$\mathbf{X}=\mathbf{P\Omega}, \tag{14}$$

where

$$\mathbf{X}=[log(x_1),\ log(x_2),\ \ldots,\ log(x_N)]^T,$$
$$\mathbf{\Omega}=[\omega_1\ \omega_2]^T$$

and $\mathbf{P}=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \log(p_1) & \log(p_2) & \cdots & \log(p_N) \end{bmatrix}^T$.

The least mean square solution for $\mathbf{\Omega}$ in the linear model (14) is given by

$$\hat{\Omega}=(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{X}. \tag{15}$$

Finally, the parameters α and β are determined by $\alpha=exp(\omega_1)$ and $\beta=\omega_2$. Thus, the parameters can be determined easily from the sampled points of the current curve by using the approximated model. These parameters can be considered as the extended features. They and the sampled points are used as the input features to estimate the hematocrit using the neural network which is trained by the online sequential learning algorithm.

4. Results and discussions

As done in our previous studies [8], our dataset consists of 199 blood samples which are from the randomly selected volunteers. Every sample is separated into two parts, the first part is used to determine the anodic current curve, and the second part is used to determine the accurate hematocrit by using the centrifugation method. In the second period which is after the incubation time, the current curve is sampled at a frequency of 10Hz to generate fifty-nine current points. In the first approach, these fifty-nine current points are used as input features for SLFN to estimate the hematocrit values. In the second approach, we have used a subset of current points (47 in our experiments) to determine two extended features. These two extended features and the selected subset of current points are used as input features for SLFN which can reduce the number of sampled points and save the battery consumption. The hematocrit values collected from centrifugation method have the mean of 36.02 and

Table 1
Comparison with reference hematocrit measurements using centrifugation (whole current curve)

Technique	Method	Training		Testing	
		RMSE	Std	RMSE	Std
Randomly selecting the training and testing sets	ELM (offline)	3.70	0.42	4.69	0.50
	OS-ELM (online)	3.69	0.25	4.37	0.37
10-fold cross validation	ELM (offline)	3.78	0.22	4.40	0.46
	OS-ELM (online)	3.72	0.36	4.28	0.30

the deviation of 6.39.

The dataset was divided into two subsets, training and testing. Two validation techniques are used; (1) the forty percent of the sample dataset which is selected randomly is used for training and the remaining sixty percent is used for testing, and (2) 10-fold cross validation. The input features are normalized into range [0, 1]. The neural networks are trained by the offline ELM algorithm and online sequential extreme learning machine (OS-ELM). The number of hidden units was 12 for both ELM and OS-ELM.

The average results of fifty trials with the input features of the whole current curve are shown in Table 1. The root mean square error (RMSE) is defined by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - r_i)^2} \tag{16}$$

where o_i is the estimated value and r_i is the reference value which is obtained from the centrifugation method.

From the Table 1 we can see that the accuracy of online training method corresponding to the testing sets is 4.37 and 4.28 for two validation techniques, these results are compatible to those of the offline training method for the same number of hidden nodes. Note that, for the online training method, the devices can be still trained with new samples during the using process which can expect to improve the performance further.

Table 2 shows the results for hematocrit estimation using the reduced number of current points. We have used 47 current points (instead of 59 points) and two additional features α and β. From the Table

Table 2
Comparison with reference hematocrit measurements using centrifugation (the reduced number of current points)

Technique	Method	Training		Testing	
		RMSE	Std	RMSE	Std
Randomly selecting the training and testing sets	ELM (offline)	3.45	0.26	4.11	0.49
	OS-ELM (online)	3.37	0.21	4.01	0.25
10-fold cross validation	ELM (offline)	3.51	0.27	4.38	0.60
	OS-ELM (online)	3.42	0.25	4.18	0.34

2 we can see that the accuracy of online training method is also compatible to those of the offline training method. The accuracy of the testing sets is 4.01 and 4.18 for the random selection of subsets and 10-fold cross validation, respectively. These results are compatible to those corresponding to using the whole current curve.

When the number of current points is reduced, the number of sampling times is reduced which can save the battery power on handheld devices. Moreover, reducing the last current points is expected to speed up the measurement process.

We have also experimented with another popular machine-learning approach that is support vector machine (SVM). In comparison with SVM, the proposed method can obtain better performance than SVM that has reported the RMSE of 3.74 and 4.50 for training and testing set, respectively.

5. Conclusion

The hematocrit is an important factor and significantly affects the accuracy of glucose measurement. Estimating hematocrit levels plays a crucial role to improve the accuracy of glucose measurement that uses portable devices. In this paper, we proposed an approach which applies the single-hidden layer feedforward neural network to estimate the hematocrit. The input features are obtained from the current curves; these curves are generated during chemical reaction of glucose oxidase in the electrochemical biosensors. The salient points of this method are that (1) the network is trained by the online sequential method based on extreme learning machine which can obtain good performance with high learning speed and (2) the extended features are used to reduce the number of current points which can save the battery power of devices and speed up the measurement process. The results of study in this paper can also be utilized to reduce the effects of hematocrit in glucose measurement which can enhance its accuracy.

Acknowledgment

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2013.47.

References

[1] J.Z. Ma, J. Ebben, H. Xia and A.J. Collins, Hematocrit level and associated mortality in hemodialysis patients, Journal of the American Society of the Nephrology **10** (1999), 610-619.
[2] A. Aris, J.M. Padro, J.O. Bonnin and J.M. Caralps, Prediction of hematocrit changes in open-heart surgery without blood transfusion, The Journal of Cardiovascular Surgery **25** (1984), 545-548.
[3] I.F. Bejarano, J.E. Dipierri, E.L. Alfaro, C. Tortora, T. García and M.C. Buys, Hematocrit values and prevalence of anemia in schoolchildren of Jujuy, Medicina **63** (2003), 288-292.
[4] R.F. Louie, Z. Tang, D.V. Sutton, J.H. Lee and G.J. Kost, Point of Care Glucose Testing: effects of critical variables, influence of reference instruments, and a modular glucose meter design, Archives of Pathology & Laboratory Medicine **124** (2000) 257-266.
[5] Z. Tang, J.H. Lee, R.F. Louie, G.J. Kost and D.V. Sutton, Effects of different hematocrit levels on glucose measurements with handheld meters for point of care testing, Archives of Pathology & Laboratory Medicine **124** (2000), 1135-1140.
[6] E.S. Kilpatrick, A.G. Rumley and H. Myin, The effect of variations in hematocrit, mean cell volume and red blood count on reagent strip tests for glucose, Annals of Clinical Biochemistry **30** (1993), 485-487.

[7] E.F. Treo, C.J. Felice, M.C. Tirado, M.E. Valentinuzzi and D.O. Cervantes, Hematocrit Measurement by Dielectric Spectroscopy, IEEE Transactions on Biomedical Engineering **25** (2005), 124-127.
[8] Hieu Trung Huynh, Yonggwan Won and Jung-ja Kim, Neural networks for the estimation of hematocrit from transduced current curves, The 2008 IEEE Int'l conference on Networking, Sensing and Control, 2008, pp. 1517-1520.
[9] P.J.G. Lisboa, E.C. Ifeachor and P.S. Szczepaniak, Artificial Neural Networks in Biomedicine, Springer-Verlag London Berlin Heidelberg, 2000.
[10] R.N.G. Naguib and G.V. Sherbet, Artificial Neural Networks in Cancer Diagnosis, Prognosis, and Patient Management, CRC Press, Washington D.C., 2001.
[11] Hieu Trung Huynh, Jungja Kim and Yonggwan Won, Performance comparison of SLFN training algorithms for DNA microarray classification, Advances in Experimental Medicine and Biology **696** (2011), 135-143.
[12] Hieu Trung Huynh, Jungja Kim and Yonggwan Won, Classification study on DNA microarray with feedforward neural network trained by singular value decomposition, International Journal of Bioscience and Biotechnology **1** (2009), 17-24.
[13] Syed Muhammad Aqil Burney, Tahseen Ahmed Jilani and Cemal Ardil, A comparison of first and second order training algorithms for artificial neural networks, International Journal of Computational Intelligence **1** (2004), 218-224.
[14] V.S. Asirvadam, S.F. McLoone and G.W. Irwin, Parallel and separable recursive Levenberg-Marquardt training algorithm, Proceedings of 12th IEEE Workshop Neural Network Signal Process **4-6** (2002), 129-138.
[15] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, Extreme learning machine: A new learning scheme for feedforward neural networks, Proceedings of International Joint Conference on Neural Networks **2** (2004), 985-990.
[16] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing **70** (2006), 489-501.
[17] N.-Y. Liang, G.-H. Huang, P. Saratchandran and N. Sundararajan, A fast and accurate online sequential learning algorithm for feedforward networks, IEEE Transaction on Neural Networks **17** (2006), 1411-1423.
[18] Hieu Trung Huynh, Yonggwan Won and Jung-ja Kim, An improvement of extreme learning machine for compact single-hidden-layer feedforward neural networks, International Journal of Neural Systems **18** (2008), 433-441.
[19] Hieu Trung Huynh and Yonggwan Won, Small number of hidden units for ELM with two-stage linear model, IEICE Transactions on Information and Systems **E91-D** (2008), 1042-1049.