Applied Mathematics, Modeling and Computer Simulation
C.-H. Chen et al. (Eds.)

doi:10.3233/ATDE221077

Approximation of Broad Learning System and Convolution Function Applied on Face Classification

Dong NAN, Zhengying LI, Luhong DIAO [1]
College of Applied Sciences, Beijing University of Technology, Beijing 100124, China

Abstract. Broad Learning System (BLS) is a very fast and effective discriminative learning which is developed by C. L. P. Chen, Z. Liu and others. It avoids the shortcomings of complex model design and large amount of calculation in deep learning. This paper studies the approximation capability of BLS for continuous functions defined on a compact set. It is proved that if the activation function of the enhancement node of BLS is not polynomial, for any continuous function $f(x)\in C(K)$ defined on the compact set K ,there is $\lim_{\substack{mq\to\infty\\ nk\to\infty}}\|f(x)-f_w(x)\|_2^2=0$,that is $\forall\varepsilon>0$, $\exists nk\in N, mq\in N'$,and parameter set w , so that $\|f(x)-f_w(x)\|_2^2<\varepsilon$. A reconstructed model of BLS which is combined the CNN network with the BLS is applied to numerical experiments. The semi-supervised broad learning system(SS-BLS) and its algorithm are proposed. Then, SS-BLS and convolution function are combined to establish SS-CBLS, the numerical experiments of SS-CBLS on face classification are carried out by ORL and Yale face database respectively

Keywords. Broad learning system, compact sets, approximation, face classification

1. Introduction

In 2018, C. L. P. Chen and Z. Liu [1] developed a very fast and effective learning system, that is, broad learning system (BLS). It is a fast and accurate learning without deep structure. BLS structure is very suitable for modelling and learning in big data environment [2-6].

The BLS contains a feature layer, an enhancement layer and an output layer. The following is the BLS build process [7-9].

(1) Giving the training data $\{X,Y\}\in R^{N\times(M+C)}$ .The nonlinear transform function mapping $\varphi_i(x), i=1,2,\cdots,n$ is used to generate the i th set of feature mappings $Z_i=\varphi_i(XW_{e_i}+\beta_{e_i})$,where W_{e_i},β_{e_i} is a randomly matrix and bias matrix. All feature nodes are combined and written as $Z^n=[Z_1,Z_2,\cdots,Z_n]\in R^{N\times nk}$, here N is the number of samples, nk is the number of all feature nodes.

(2) Nonlinear functional transformation $\xi_j(\cdot)$ is applied on Z^n to generate enhanced nodes $H_j=\xi_j(Z^nW_{h_j}+\beta_{h_j})$, where $W_{h_j},\beta_{h_j}\,(j=1,2,\cdots,m)$ are randomly matrix and

1 Corresponding Author, Luhong DIAO, College of Applied Sciences, Beijing University of Technology, Beijing 100124, China; E-mail: diaoluhong@bjut.edu.cn.

bias matrix. m is the number of groups of enhanced nodes. All enhanced nodes are combined and written as $H^m = [H_1, H_2, \cdots, H_m] \in R^{N\times mq}$,where mq is the number of nodes.

(3) Let $A = [Z^n \mid H^m] \in R^{N\times(nk+mq)}$. Here $nk + mq$ is the number of all nodes.

(4) Thus, the output of BLS $Y = AW^m$, where W^m is the weight connecting the feature nodes and enhanced nodes to the output layer. A^+ denotes the pseudo-inverse operation, which is calculated by the following equation $A^+ = \lim_{\lambda\to 0}[(\lambda I + AA^T)^{-1} A^T]$.

The BLS network becomes a linear transformation from the input (feature + enhanced) layer to the output layer, as shown in figure 1.

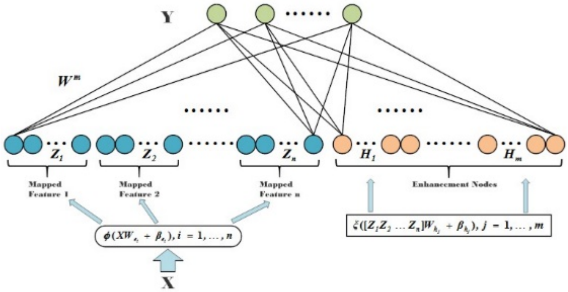


Figure 1. Construction of BLS

2. Approximation Capability of BLS

C. L. P. Chen et al. used the method of probability expectation to obtain that BLS is a nonlinear function approximation on measurable sets [10]. They also gave the framework of several BLS variants and models [11-13]. In this paper, it is proved that for any continuous function which is defined on any compact set, it can be approximated by BLS with any given accuracy if the activity function of BLS is not a polynomial. A conclusion is drawn that on a measurable set, any measurable function which can be approximated by BLS if the activity function of BLS is not a polynomial.

Suppose that φ is bounded feature mapping function and ξ is activity function. Then, the output function of BLS has the following form

$$f_w(X) = \sum_{i=1}^{nk} w_i \varphi(XW_{e_i} + \beta_{e_i}) + \sum_{i=1}^{mq} w_{nk+j} \xi(ZW_{h_i} + \beta_{h_i}) \tag{1}$$

Theorem: For any continuous function $f(x) \in C(K)$ which is defined on a compact set K , there is a the output function of BLS $f_w(X)$ such that

$$\lim_{\substack{nk\to\infty \\ mq\to\infty}} \left\| f(x) - f_w(X) \right\|_2^2 = 0 \tag{2}$$

That is, for any $\forall \varepsilon > 0$, there exist $\exists nk \in N$ and $mq \in N'$ such that $\left\| f(x) - f_w(X) \right\|_2^2 < \varepsilon$

Proof. Supposing that $w_a = [w_{a_1}, \cdots, w_{a_{nk}}]$ is the weight matrix which connects feature nodes Z^n to the output layer, and that $w_b = [w_{b_1}, \cdots, w_{b_{mq}}]$ is the weight matrix which connects enhanced nodes H^m to the output layer. Let $f_{w_a}(X)=\sum_{i=1}^{nk} w_{a_i}\varphi(XW_{e_i}+\beta_{e_i})$, $f_{w_b}(X)=\sum_{j=1}^{mq} w_{b_j}\xi(ZW_{h_j}+\beta_{h_j})=\sum_{j=1}^{mq} w_{b_j}H_j(X,W_{h_j},\beta_{h_j})$.

Taking that $H_j(X,W_{h_j},\beta_{h_j})=\xi(\sum_{i=1}^{nk}\varphi(XW_{e_i}+\beta_{e_i})w_{h_j}^i+\beta_{h_j})$ into account. Therefore

$$f_w(X)=f_{w_a}(X)+f_{w_b}(X) \tag{3}$$

Then

$$\|f(x)-f_w(X)\|_2^2=\|f(x)-f_{w_a}(X)-f_{w_b}(X)\|_2^2 \tag{4}$$

Let that $f_a(x)=f(x)-f_{w_a}(X)$,then $\|f(x)-f_w(X)\|_2^2=\|f_a(x)-f_{w_b}(X)\|_2^2$

Considering that the feature mapping function φ is bounded, and that $f(x)\in C(K)$. Then $f_a(x)=f(x)-f_{w_a}(X)$ which is bounded and integrable function on a compact set K . According to [14,15] ,for any $\forall\varepsilon>0$,there is $f_b(x)\in C(K)\subset L^2(K)$,such that

$$\|f_a(x)-f_b(x)\|_2^2<\frac{\varepsilon}{2} \tag{5}$$

$f_b(x)$ has the following form [16], $f_b(x)=\sum_{s=1}^{N}\alpha_s\prod_{t=1}^{d}g(xw_{st}+\eta)$. According to [16], for any $\forall\varepsilon>0$,if that $f_b(x)\in C(K)$ and that $H_j(X,W_{h_j},\beta_{h_j})$ is not a polynomial, there are w_{b_j} and h_j such that

$$\|f_b(x)-f_{w_b}(X)\|_2^2<\frac{\varepsilon}{2} \tag{6}$$

That is

$$\begin{aligned}\|f(x)-f_w(X)\|_2^2&=\|f_a(x)-f_{w_b}(X)\|_2^2\\&\le\|f_a(x)-f_b(x)\|_2^2+\|f_b(x)-f_{w_b}(X)\|_2^2<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon\end{aligned} \tag{7}$$

Thus, for any continuous function $f(x)$ defined on a compact set K ,it can be approximated by $f_w(X)$,which means that $f_w(X)$ is dense in $C(K)$.

Corollary: For any $f(x)\in C(K)$,supposing that P is a measurable function defined on Ω and $0\le P(x)\le 1$, there is that $\lim_{n,m\to\infty}E(\|f(x)-f_w(X)\|_2)=0$.

Proof: According to *Hölder* inequality [17],one can get that

$$\begin{aligned}E(\|f(x)-f_w(X)\|_2)&=\int_\Omega\|f(x)-f_w(X)\|_2P(x)dx\\&\le\left(\int_\Omega\|f(x)-f_w(X)\|_2^2dx\right)^{1/2}\left(\int_\Omega P^2(x)dx\right)^{1/2}\\&\le\left(\int_\Omega\|f(x)-f_w(X)\|_2^2dx\right)^{1/2}\\&\le\left(\int_\Omega\varepsilon^2dx\right)^{1/2}\le\varepsilon(\mu(\Omega))^{1/2}<\varepsilon_0\end{aligned} \tag{8}$$

Where, $\mu(\Omega)$ is a measure of the measurable set Ω .

Remark: this corollary is also the approximation conclusion of BLS on measurable sets which is proposed by C. L. P. Chen, Z. Liu and S. feng [12].

3. CNNBLS and Numerical Experiment

Since We combine broad learning system with convolution network. The principle vector of the image is used in the feature node of broad learning system. In the enhanced nodes of broad learning system, the main eigenvector corresponding to the convoluted image is adopted. The construction of CNNBLS is showed in figure 2.

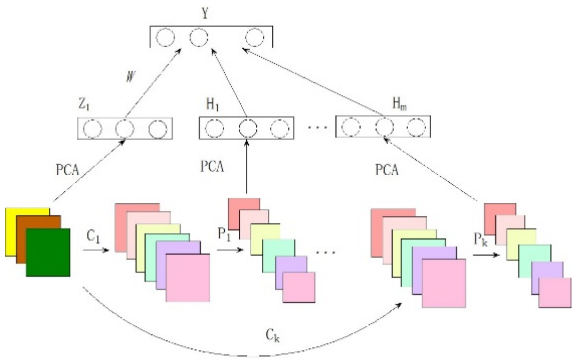


Figure 2. Construction of CNNBLS

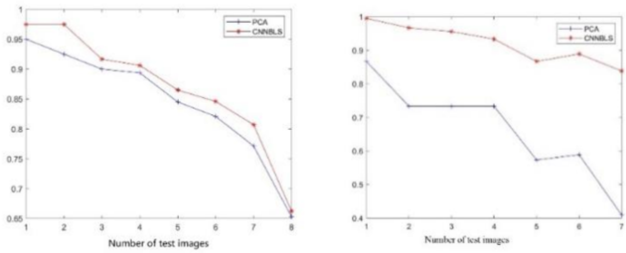
Table1. Algorithm of CNN Broad Learning System

1、PCA method is used to get features as feature nodes from input samples X ;
2、for i=1 : k
   (1) Convolute the input sample with the function;
   (2) Feature extraction in (1) by PCA as enhanced nodes
3、Use formula $W = (\lambda D + AA^T)^{-1} A^T Y$, to calculate W .

The ORL and Yale databases are adapted respectively: 1 photo per group for test data, the rest for training data; then 2 photos per group for test data, the rest for training data; and so on, until to 8 photos per group for test, the rest for training. Each test is for 1000 times, and the average is taken. Using these ways to compare CNNBLS and the traditional PCA method. It can be observed that the accuracy of the CNNBLS is much higher than the traditional PCA method. The algorithm of CNN-BLS is in table 1.

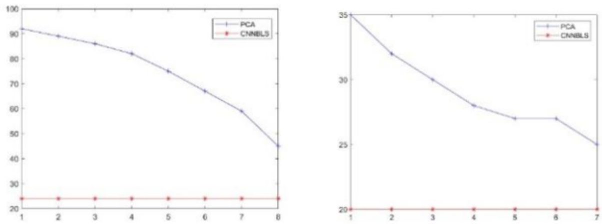
In the following, we compare the number of features required by the CNN-BLS and the traditional PCA method. As it can be seen from the figure 3 and 4 below, the number of features in CNN-BLS is much smaller than that of the traditional PCA method.

From the numerical experiments, it is clear that CNN-BLS is a fast and effective method for pattern recognition. It has much less number of features and the much higher accuracy are than the traditional PCA method.



(a) ORL database (b) Yale database

Figure 3. Accuracy in ORL and Yale



(a) ORL database (b) Yale database

Figure 4. Number of features in ORL and Yale

4. Semi-supervised CBLS on Manifold

The main idea of semi-supervised is to introduce a large number of unlabelled samples into the model training under the condition of few labelled samples, which can effectively avoid the performance problem of supervised learning [18]. In [18], Zhao H, Zheng J, Deng W and Song Y. proposed an improved the loss function of BLS.

$$L_{SS-BLS} = \theta\beta^* + A^T C(A\beta^* - \tilde{Y}) + \lambda A^T LA\beta^* \tag{9}$$

Here, $A = [Z^n \mid H^m]$, θ is the restriction of β, L is a Laplacian matrix, λ is a trade-off, C_i is punishment parameter. β^* is a superior solution which is satisfied the system of equations $e_i^T = a(x_i)\beta^* - y_i^T$,where y_i^T is the output of the labelled sample, $i = 1,\ldots,l$. β^* can be obtained for the derivation of the objective function

$$\beta^* = (\theta I_{n_h} + A^T CA + \lambda A^T LA)^{-1} A^T C\tilde{Y} \tag{10}$$

where, n_h For the sum of the number of enhancement nodes and feature nodes, I_{n_h} is unity matrix with n_h order. If A has more columns than rows, then

$$\beta^* = A^T(\theta I_{l+u} + CAA^T + \lambda LAA^T)^{-1} C\tilde{Y} \tag{11}$$

We combine the SS-BLS and convolution function to construct SS-CBLS. The algorithm of SS-BLS is in Table 2. The activation function of CNN hidden layer selects Softmax function, and the feature node of SS-CBLS feature selects the feature of labelled sample points. The enhanced node uses all samples and CNN convolution function to convolution and pooling, and finally uses Softmax activation function. We calculate Laplacian matrix L with original data and convolutional data of x_i and x_j , matrix A

with feature nodes, convolution and con-volution features as enhanced nodes. The structure of SS-CBLS is shown in figure 5.

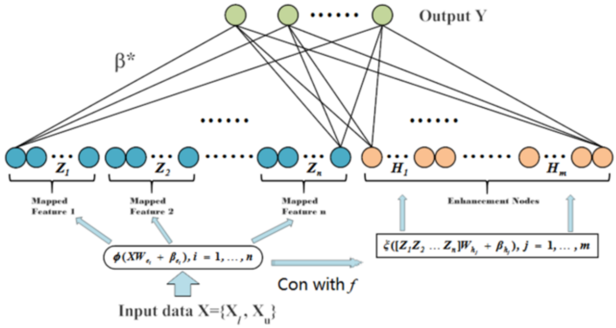


Figure 5. Structure of SS-CBLS

Table 2. Algorithms of SS-CBLS

Enter: Labelled Samples$\{X_l, Y_l\} = \{x_i, y_i\}_{i=1}^{l}$; Unlabelled Samples$X_u = \{x_i\}_{i=1}^{u}$
output: mapping functions for SS-CBLS$f: R^{n_i} \to R^{n_o}$

1. Sample with Labelled X_l and unlabeled X_u. Laplacian matrix is Calculated.
2. The feature nodes Z^n, the convolution and convolution features as enhanced nodes H^m are constructed, and the output matrix is calculated by $A = [Z^n|H^m] \in R^{(l+u)\times(nk+mq)}$.
3. Weights are obtained using the manifold regularization framework β^*.
4. Calculate the mapping function$f_i = a(x)\beta^*$ to estimate the unlabelled.

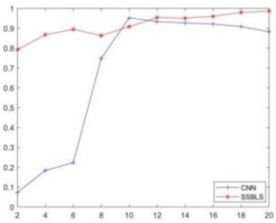
The ORL and Yale databases are adapted respectively to compare the performance of SS-CBLS and traditional CNN. Each method is run independently 100 times. The results of the two experiments on the Yale database are shown in table 2.

Table 3. The comparison of experimental results on Yale

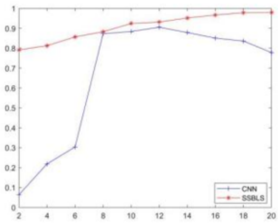
model	Test Precision (%)	Test Precision Variance (%)	Training Time (s)
CNN	86.67	7.78	4.27
SS-CBLS	93.33	1.10	0.42

From table 3, we can see that the average test accuracy, test variance and average training time of SS-CBLS are better than the traditional CNN. SS-CBLS has a good accuracy and efficiency for Yale database, and can achieve better semi-supervised classification of Yale data.

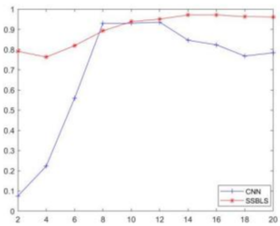
We select one photo in each group as the test data in the Yale database, and the rest is the training data; 2 photos for test data, the rest for training data, and so on, until each group selected 10 for the test, the rest for training. 1000 times per test, average the accuracy. In figure 6, the horizontal axis represents the number of convolution kernels, and the vertical axis represents the corresponding accuracy. It can be observed that the accuracy of SS-CBLS is generally higher than that of the corresponding traditional CNN in Yale database. And as the change of training samples and convolution kernels in each group, the accuracy of traditional CNN is more affected than it of SS-CBLS.



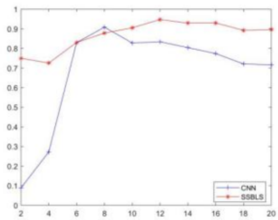
(1)



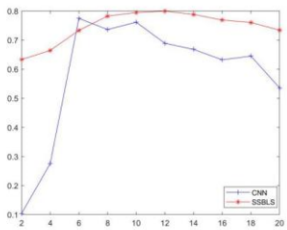
(2)



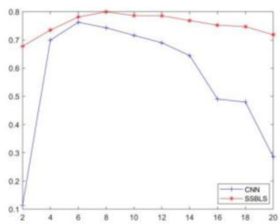
(3)



(4)



(5)



(6)

Figure 6. Accuracy between SS-CBLS and CNN on Yale

The experimental results show that SS-CBLS can more effectively extract data features from different data, so as to ensure the accuracy of the test and reduce the training time of SS-CBLS. It also shows that SS-CBLS has high stability and strong adaptability. Compared with the traditional CNN, SS-CBLS has a simple structure and only performs convolution, pooling and Softmax operations in the enhanced nodes. In addition, SS-CBLS not only retains the features of the convolutional image, but also retains the features of the original image at the feature nodes, while CNN only extracts features from the convolutional image.

5. Conclusion

BLS is based on RVFLNN. It is a fast and accurate learning without deep structure. It is obviously superior to the existing deep structure in learning accuracy and generalization ability. In this paper, the method of nonlinear functional approximation is adopted to prove that BLS can approximate a given function with arbitrary precision on a given compact set. The corollary is get that BLS can also be approximated by probability expectation on the measurable set. This inference is also a main conclusion of [12]. Numerical experiments are also carried out on CNN-BLS on ORL and Yale databases respectively. It is compared with traditional PCA methods. Then, a semi-supervised BLS (SS-BLS.) based on manifold regularization framework is proposed.

We combine SS-BLS and convolution function to establish SS-CBLS, and use ORL and Yale face database to study the problem of face classification on regular manifold. Experimental results show that SS-CBLS can perform better classification tasks in a simple semi-supervised environment than traditional CNNs.

Funding

This article is supported by the General Research Projects of Beijing Educations Committee in China under Grant KM 201910005013.

References

[1] Chen C L P and Liu Z. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, Jan. 29(1): 10-24.

[2] Yang F. A CNN-based broad learning system [C]. 2018 IEEE 4th International Conference on Computer and Communications (ICCC), 2018: 2105-2109.

[3] Chen C L P and Liu Z. Broad learning system: a new learning paradigm and system without going deep [C]. 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC), 2017: 1271-1276.

[4] Chen C L P and Liu Z. Broad learning system: structural extensions on single-layer and multi-layer neural networks [C]. 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2017: 136-141.

[5] Huang H, Zhang T, Yang C and Chen C L P. Motor learning and generalization using broad learning adaptive neural control [J]. IEEE Transactions on Industrial Electronics, 2020, Oct. 67(10): 8608-8617.

[6] LeCun Y, Huang F J and Bottou L. Learning methods for generic object recognition with invariance to pose and lighting [C]. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, 2: II-104.

[7] Peng X, Ota K and Dong M.A broad learning-driven network traffic analysis system based on fog computing paradigm [J]. China Communications, 2020,17(2): 1-13.

[8] Yuan L, Li T, Chen C L P, Shan Q and Han M. Broad learning system-based learning controller for course control of marine vessels[C]. 2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP), 2019: 133-136.

[9] Feng L, Zhao C, Chen C L P, Li Y, Zhou M, Qiao H and Fu C. BNGBS: an efficient network boosting system with triple incremental learning capabilities for more nodes, samples, and classes[J]. Neurocomputing, 2020, 412: 486-501.

[10] Chen C L P, Liu Z and Feng S. Universal approximation capability of broad learning system and its structural variations [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, Apr. 30(4): 1191-1204.

[11] Zhang L et al. Analysis and variants of broad learning system [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, Jan. 52(1): 334-344.

[12] Sui S, Chen C L P, Tong S and Feng S. Finite-time adaptive quantized control of stochastic nonlinear systems with input quantization: a broad learning system based identification method [J]. IEEE Transactions on Industrial Electronics, 2020, Oct. 67(10): 8555-8565.

[13] Zhao H, Zheng J, Deng W and Song Y. Semi-supervised broad learning system based on manifold regularization and broad network [J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2020, Mar. 67(3): 983-994.

[14] Jiang C H. Approximation capability of neural networks[J]. Chinese Annals of Mathematics, 1998(03): 12-17. (in Chinese)

[15] Jiang C H and Guo H B. Approximation by neural networks in and system identification [J]. Chinese Annals of Mathematics, 2000(04): 417-422. (in Chinese)

[16] Luo Y-H and Shen S Y. approximation of sigma-pi neural networks[J]. IEEE Transactions on Neural Networks, 2000, Nov. 11(6): 1485-1489.

[17] Rudin W. Real and Complex Analysis (Higher Mathematics Series) [M]. New York, NY, USA: McGraw-Hill, 1987: 61-112.

[18] Zhao H, Zheng J, Deng W and Song Y. Semi-supervised broad learning system based on manifold regularization and broad network [J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2020, 67(3): 983-994.