SCIENCE CHINA
Information Sciences

• LETTER •

April 2021, Vol. 64 149205:1–149205:3
https://doi.org/10.1007/s11432-018-9630-0

# Performance analysis of fuzzy BLS using different cluster methods for classification

Shuang FENG[1,2] & C. L. Philip CHEN[2*]

[1] *School of Applied Mathematics, Beijing Normal University, Zhuhai* 519087*, China;*
[2] *Faculty of Science and Technology, University of Macau, Macau* 999078*, China*

Received 23 August 2018/Accepted 23 October 2018/Published online 14 May 2020

**Citation** Feng S, Chen C L P. Performance analysis of fuzzy BLS using different cluster methods for classification. Sci China Inf Sci, 2021, 64(4): 149205, https://doi.org/10.1007/s11432-018-9630-0

Dear editor,

Neural networks (NNs) and fuzzy systems are commonly used computational intelligence techniques, each with their own merits in terms of applications. The integration of NNs and fuzzy systems, which leads to a hybrid framework known as neuro-fuzzy systems, inherits the useful properties of its constituents: the learning power of an NN and the knowledge representation of a fuzzy inference system, which makes their combination a powerful tool for machine learning. Numerous structures and models of neuro-fuzzy systems have been proposed and applied to real-world problems [1–5].

The fuzzy broad learning system (Fuzzy BLS) [6] is a recently proposed neuro-fuzzy model which is constructed by replacing the feature layer of a BLS [7] with Takagi-Sugeno-Kang (TSK) fuzzy sub-systems (see Figure 1(a)). It retains the main architecture of a BLS and employs the k-means algorithm to cluster the input data to reduce computation complexity. Fuzzy BLS achieves higher accuracy in classification and regression problems compared with current state-of-the-art neuro-fuzzy models.

To further investigate the performance of Fuzzy BLS, we employ fuzzy c-means to cluster the input data and determine the number of fuzzy rules as well as the centers of Gaussian membership functions. We also randomly generate centers in the input data domain to perform a comprehensive comparison between the case of using cluster methods (i.e., k-means and fuzzy c-means) and the case without clustering. The three variants of Fuzzy BLS (denoted by FBLS-KM, FBLS-FCM, and FBLS-RAND) are evaluated and compared using some popular data sets for classification.

*Fuzzy broad learning system.* We consider a simplified version of Fuzzy BLS, which has $m$ enhancement nodes instead of $m$ groups of enhancement nodes (see Figure 1(a)). Given the following training data: $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^{\mathrm{T}} \in \mathbb{R}^{N\times D}$, and the corresponding targets $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)^{\mathrm{T}} \in \mathbb{R}^{N\times C}$ ($N$ is the number of samples, $D$ represents the input dimension, and $C$ is the output dimension), we employ $n$ TSK fuzzy sub-systems with $K_i$ fuzzy rules in the $i$th fuzzy sub-system for Fuzzy BLS, where the fuzzy rules have the following form ($k = 1, 2, \ldots, K_i$): If $x_1$ is $A^i_{k1} \cdots$ and $x_D$ is $A^i_{kD}$, then $r^i_k = \sum_{t=1}^D a^i_{kt} x_t$, where $A^i_{kt}$ is a fuzzy set associated with the Gaussian membership function $\mu^i_{kt} = \exp(-(\frac{x-c^i_{kl}}{\sigma^i_{kl}})^2)$, the coefficients $a^i_{kt}$ in the consequent part are random numbers generated from $[0, 1]$, and the width $\sigma^i_{kl}$ is usually 1, while the center $c^i_{kl}$ is determined by k-means.

Suppose that $\boldsymbol{x}_s = (x_{s1}, x_{s2}, \ldots, x_{sD})$ is one training sample and $\boldsymbol{y}_s = (y_{s1}, y_{s2}, \ldots, y_{sC})$ is the corresponding target ($s = 1, 2, \ldots, N$). Then the weighted activation degree of the $k$th fuzzy rule in the $i$th fuzzy sub-system is $\lambda^i_{sk} = \tau^i_{sk} / \sum_{k=1}^{K_i} \tau^i_{sk}$, where $\tau^i_{sk} = \prod_{t=1}^D \mu^i_{kt}(x_{st})$.

For the $i$th fuzzy sub-system and the $s$th training sample $\boldsymbol{x}_s$, we define its weighted rule outputs as a vector $\boldsymbol{R}_{si} = (\lambda^i_{s1} r^i_{s1}, \lambda^i_{s2} r^i_{s2}, \ldots, \lambda^i_{sK_i} r^i_{sK_i})^{\mathrm{T}}$.

For all training samples $\boldsymbol{X}$, the weighted rule output matrix of the $i$th fuzzy sub-system is $\boldsymbol{R}_i = (\boldsymbol{R}_{1i}, \boldsymbol{R}_{2i}, \ldots, \boldsymbol{R}_{Ni})^{\mathrm{T}} \in \mathbb{R}^{N\times K_i}$.

Then, we combine every weighted rule output matrix of each fuzzy sub-system into one matrix, which is $\boldsymbol{R} = (\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_n) \in \mathbb{R}^{N\times \sum_{i=1}^n K_i}$. These intermediate values denoted by $\boldsymbol{R}$ will be fed into the enhancement layer for further nonlinear transformation.

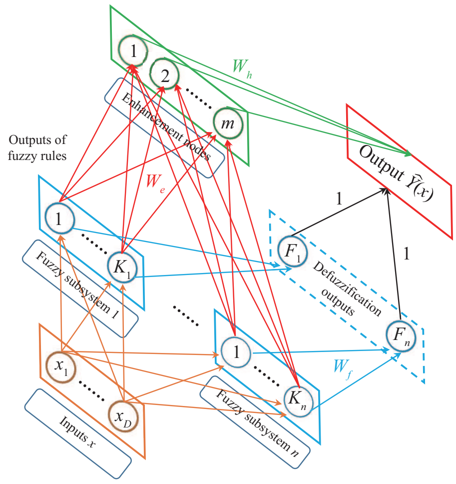
We now calculate the defuzzification output of every fuzzy sub-system. The output of the $i$th fuzzy sub-system for training sample $\boldsymbol{x}_s$ is $\boldsymbol{F}_{si} = (\sum_{t=1}^D a^i_{kt} x_{st} (\lambda^i_{s1}, \ldots, \lambda^i_{sK_i}) \boldsymbol{w^i})^{\mathrm{T}}$, where we include a new parameter $\boldsymbol{w}^i = (w^i_{kc})_{K_i\times C}$ in the consequent part of each fuzzy rule in the $i$th fuzzy sub-system ($i = 1, 2, \ldots, n$).

For all training data $\boldsymbol{X}$, the defuzzification outputs of the $i$th fuzzy sub-system are as follows: $\boldsymbol{F}_i = (\boldsymbol{F}_{1i}, \boldsymbol{F}_{2i}, \ldots, \boldsymbol{F}_{Ni})^{\mathrm{T}} \triangleq \boldsymbol{B}\Lambda^i \boldsymbol{w}^i$, where $\boldsymbol{B} = \mathrm{diag}\{\sum_{t=1}^D a^i_{kt} x_{1t}, \ldots, \sum_{t=1}^D a^i_{kt} x_{Nt}\}$, and $\Lambda^i = (\lambda^i_{sk})_{N\times K_i}$.

Hence, the defuzzification output matrix of $n$ fuzzy sub-systems is $\boldsymbol{F} = \sum_{i=1}^n \boldsymbol{F}_i \triangleq \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{W_f}$, where $\boldsymbol{\Lambda} \in \mathbb{R}^{N\times \sum_{i=1}^n K_i}$ is the activation degree matrix consisting of $\lambda^i_{sk}$, and $\boldsymbol{W_f} = (\boldsymbol{w}^1, \ldots, \boldsymbol{w}^n)$.

* Corresponding author (email: philip.chen@ieee.org)

(a)




(b)

| Models | OS-F-ELM | | F-ELM | | IF-ELM | | FBLS-KM | | FBLS-FCM | | FBLS-RAND | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Training | Testing | Training | Testing | Training | Testing | Training | Testing | Training | Testing | Training | Testing |
| Diabetes | – | – | 75.35% | 74.09% | – | – | 82.03% | 78.65% | 83.59% | **81.25%** | 78.68% | 79.69% |
| Breastcancer | – | – | 93.94% | 94.17% | 96.93% | 97.41% | 96.57% | **99.29%** | 95.14% | **99.29%** | 96.57% | 99.29% |
| Segmentation | 95.84% | 94.41% | 98.71% | 95.60% | – | – | 98.60% | 96.79% | 98.93% | **97.78%** | 97.67% | 96.30% |
| Satellite | 92.84% | 89.41% | 93.20% | 90.19% | 93.95% | 89.22% | 92.76% | 90.50% | 92.81% | **91.20%** | 92.06% | 89.35% |
| DNA | 96.51% | 94.21% | 96.85% | 94.65% | – | – | 96.15% | **95.21%** | 97.60% | 94.50% | 86.90% | 83.11% |
| Pageblock | 96.72% | 95.99% | – | – | – | – | 96.96% | 96.36% | 97.44% | **96.65%** | 97.93% | 96.47% |
| Spamemails | 93.75% | 91.88% | – | – | – | – | 94.43% | **92.82%** | 93.33% | 92.63% | 94.03% | 92.50% |
| Letter | 96.09% | 93.62% | – | – | – | – | 99.30% | **95.66%** | 94.04% | 90.99% | 99.15% | 95.28% |

**Figure 1** (Color online) Simplified structure of Fuzzy BLS.

Using the intermediate matrix $\boldsymbol{R}$, we can compute the outputs of the enhancement layer $\boldsymbol{H} = \xi(\boldsymbol{R}\boldsymbol{W}_e + \boldsymbol{\beta}_e) \in \mathbb{R}^{N\times m}$, where $\boldsymbol{W}_e$ represents the weights between the fuzzy sub-systems and their corresponding enhancement nodes with bias terms $\boldsymbol{\beta}_e$, which are initialized as random numbers from $[0, 1]$, and $\xi$ is the activation function.

The weights between the enhancement layer and top layer are denoted by $\boldsymbol{W}_h \in \mathbb{R}^{m\times C}$. Hence, we have the following model output

$$\hat{\boldsymbol{Y}} = \boldsymbol{F} + \boldsymbol{H}\boldsymbol{W}_h = (\boldsymbol{B}\boldsymbol{\Lambda}, \boldsymbol{H})\begin{pmatrix}\boldsymbol{W}_f\\ \boldsymbol{W}_h\end{pmatrix} \triangleq (\boldsymbol{B}\boldsymbol{\Lambda}, \boldsymbol{H})\boldsymbol{W},$$

where $\boldsymbol{W}$ includes all parameters.

Given the training targets $\boldsymbol{Y}$, we can calculate $\boldsymbol{W}$ directly by the ridge regression approximation of the pseudoinverse of $\boldsymbol{E} \triangleq (\boldsymbol{B}\boldsymbol{\Lambda}, \boldsymbol{H})$, i.e.,

$$\boldsymbol{E}^+ = \lim_{\alpha\to 0}(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{E} + \alpha\boldsymbol{I})^{-1}\boldsymbol{E}^{\mathrm{T}}, \tag{1}$$

where $\alpha$ is the regularization coefficient.

Then we can obtain

$$\boldsymbol{W} = \boldsymbol{E}^+\boldsymbol{Y}. \tag{2}$$

*Fuzzy BLS based on fuzzy c-means and random centers.* Now, we consider replacing k-means with fuzzy c-means in the Fuzzy BLS. For the $i$th fuzzy sub-system, the training data is divided into $K_i$ clusters with center $C = (\boldsymbol{c}_k)_{K_i\times D}$, and we apply the fuzzy c-means method by minimizing the objective function $J_f(U, C) = \sum_{s=1}^{N}\sum_{k=1}^{K_i} u_{sk}^f \|\boldsymbol{x}_s - \boldsymbol{c}_k\|_2^2$, where $f = 2$ is a fuzzification coefficient, $U = (u_{sk})_{N\times K_i}$ is a membership matrix, $u_{sk}$ is the degree of $\boldsymbol{x}_s$ belonging to cluster $k$, and $\boldsymbol{c}_k$ is the $D$-dimension center of cluster $k$.

The above objective function $J_f(U, C)$ can be solved by iteratively updating the membership $u_{sk}$ and center $\boldsymbol{c}_k$ by

$$u_{sk} = \left(\sum_{p=1}^{K_i}\left(\frac{\|\boldsymbol{x}_s - \boldsymbol{c}_k\|}{\|\boldsymbol{x}_s - \boldsymbol{c}_p\|}\right)^{\frac{2}{f-1}}\right)^{-1}, \tag{3}$$

and $\boldsymbol{c}_k = \sum_{s=1}^{D} u_{sk}^f \cdot \boldsymbol{x}_s / \sum_{s=1}^{D} u_{sk}^f$.

Unlike the k-means method, we can directly set the weighted fire strength in the $i$th fuzzy sub-system as $\lambda_{sk}^i = u_{sk}$ without defining extra membership functions for fuzzy rules.

Because it is time-consuming to run cluster algorithms on large-scale data with relatively high dimension, alternatively we can randomly generate the centers for Gaussian membership functions within the domain of input variables to speed up the learning process, and follow the same remaining steps discussed above for Fuzzy BLS.

The algorithms for training Fuzzy BLS with different cluster methods are summarized in Algorithm A1.

We use FBLS-KM, FBLS-FCM, and FBLS-RAND to represent the Fuzzy BLS using k-means, fuzzy c-means, and random centers, respectively. We select eight classification benchmarks from the UCI repository to analyze the performance of the three variants of Fuzzy BLS and compare them with OS-F-ELM [1], Fuzzy ELM (F-ELM) [2], and improved fuzzy ELM (IF-ELM) [3].

The details of the selected data sets are displayed in Table A1. The optimal parameters for the models are listed in Table A2. The classification performances of the different approaches on the eight data sets are summarized in Figure 1(b). The running time of the three variants of FBLS are compared in Table A3.

We can conclude from the above tables that: (1) In terms of classification performance, FBLS-KM, FBLS-FCM, and FBLS-RAND obtain the same accuracy on one data set, and FBLS-KM achieves the highest classification accuracy on three data sets, while FBLS-FCM performs the best on the remaining four data sets. Therefore, FBLS-RAND does not provide any advantages over FBLS-KM and FBLS-FCM, which indicates that Fuzzy BLS can benefit by clustering input data to generate appropriate centers for fuzzy membership functions rather than by randomly selecting them. (2) FBLS-FCM slightly outperforms FBLS-KM in some cases; however, the advantage is not significant enough to draw the conclusion that the fuzzy c-means algorithm is generally better than k-means, especially when we note that the performance of FBLS-FCM on the Letter data set is much worse than either FBLS-KM or FBLS-RAND. (3) Compared with OS-F-ELM, F-ELM, and IF-ELM, the FBLS-KM, FBLS-FCM, and FBLS-RAND algorithms achieve higher accuracy on the eight data sets.

*Conclusion*. We carry out a comprehensive comparison among the variants of Fuzzy BLS, i.e., FBLS-KM, FBLS-FCM, and FBLS-RAND. They are evaluated using certain UCI benchmarks for classification. The results reveal that FBLS-KM and FBLS-FCM can generally obtain higher accuracy than FBLS-RAND, which demonstrates that Fuzzy BLS does benefit from clustering training data. However, there is no significant difference between FBLS-KM and FBLS-FCM in our experiments.

**Acknowledgements** This work was supported in part by National Natural Science Foundation of China (Grant Nos. 61751202, 61751205, 61572540), Macau Science and Technology Development Fund (Grant Nos. 019/2015/A1, 079/2017/A2, 024/2015/AMJ), Multiyear Research Grants of University of Macau, and Teacher Research Capacity Promotion Program of Beijing Normal University, Zhuhai.

**Supporting information** Appendix A. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

**References**

1 Rong H J, Huang G B, Sundararajan N, et al. Online sequential fuzzy extreme learning machine for function approximation and classification problems. IEEE Trans Syst Man Cybern B, 2009, 39: 1067–1072

2 Wong S Y, Yap K S, Yap H J, et al. On equivalence of FIS and ELM for interpretable rule-based knowledge representation. IEEE Trans Neural Netw Learn Syst, 2015, 26: 1417–1430

3 Ouyang C S, Kao T C, Cheng Y Y, et al. An improved fuzzy extreme learning machine for classification and regression. In: Proceedings of International Conference on Cybernetics, Robotics and Control (CRC), Hong Kong, 2016. 91–94

4 Feng S, Chen C L P. A fuzzy restricted Boltzmann machine: novel learning algorithms based on the crisp possibilistic mean value of fuzzy numbers. IEEE Trans Fuzzy Syst, 2018, 26: 117–130

5 Segatori A, Bechini A, Ducange P, et al. A distributed fuzzy associative classifier for big data. IEEE Trans Cybern, 2018, 48: 2656–2669

6 Feng S, Chen C L P. Fuzzy broad learning system: a novel neuro-fuzzy model for regression and classification. IEEE Trans Cybern, 2020, 50: 414–424

7 Chen C L P, Liu Z L. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. IEEE Trans Neural Netw Learn Syst, 2018, 29: 10–24