Neural Networks 81 (2016) 91–102

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Extreme learning machine and adaptive sparse representation for image classification

Jiuwen Cao [a,*], Kai Zhang [b], Minxia Luo [b], Chun Yin [c], Xiaoping Lai [a]

[a] Key Lab for IOT and Information Fusion Technology of Zhejiang, Hangzhou Dianzi University, Zhejiang, 310018, China

[b] Department of Mathematics, China Jiliang University, Hangzhou, Zhejiang Province, 310018, China

[c] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form 1 June 2016
Accepted 6 June 2016
Available online 23 June 2016

Keywords:
Extreme learning machine
Sparse representation
Image classification
Leave-one-out cross validation

ABSTRACT

Recent research has shown the speed advantage of extreme learning machine (ELM) and the accuracy advantage of sparse representation classification (SRC) in the area of image classification. Those two methods, however, have their respective drawbacks, e.g., in general, ELM is known to be less robust to noise while SRC is known to be time-consuming. Consequently, ELM and SRC complement each other in computational complexity and classification accuracy. In order to unify such mutual complementarity and thus further enhance the classification performance, we propose an efficient hybrid classifier to exploit the advantages of ELM and SRC in this paper. More precisely, the proposed classifier consists of two stages: first, an ELM network is trained by supervised learning. Second, a discriminative criterion about the reliability of the obtained ELM output is adopted to decide whether the query image can be correctly classified or not. If the output is reliable, the classification will be performed by ELM; otherwise the query image will be fed to SRC. Meanwhile, in the stage of SRC, a sub-dictionary that is adaptive to the query image instead of the entire dictionary is extracted via the ELM output. The computational burden of SRC thus can be reduced. Extensive experiments on handwritten digit classification, landmark recognition and face recognition demonstrate that the proposed hybrid classifier outperforms ELM and SRC in classification accuracy with outstanding computational efficiency.

1. Introduction

Image classification, with the goal of automatically assigning a certain category to the input image, has attracted intensive attention due to its high value in security systems, medical diagnosis, bioinformatics, human–computer interaction and a wide variety of other applications (Bai, Li, & Zhou, 2015; Chen & Yap, 2014; Han, Chen, & Xu, 2015; Wang et al., 2014; Zhu, Li, & Zhang, 2016). Within the past few years, various techniques developed from machine learning research have already had tremendous influence in the area of image classification. In fact, almost every method proposed in the past has its own merits and limitations. One of the inevitable problems is the compromise between computational complexity and classification accuracy. In other words, it is unlikely to design an overall winner that could achieve the best performance in terms of both speed and accuracy for all applications. Attempts to resolve this dilemma have resulted in the development of hybrid systems (Cao, Zhao, Lai, Ong et al., 2015; Chen & Girod, 2015; España-Boquera, Castro-Bleda, Gorbe-Moya, & Zamora-Martinez, 2011; Luo & Zhang, 2014; Yin, Cheng, Chen, Stark, & Zhong, 2015; Yin, Chen, & Zhong, 2014), which usually exploit the advantages from various techniques and unify these different methods in a more efficient manner.

One of the crucial factors for a successful image classification system is the classifier. A well designed classifier would not be sensitive to some of the other factors, such as feature extraction. In the past several decades, artificial neural networks benefit a lot from random generated parameters, not only in the learning speed but also in the generalization performance (Huang, Zhu, & Siew, 2006; Pao, Park, & Sobajic, 1994; Pao & Takefuji, 1992). Among them, ELM is popular for its fast learning speed, real time processing capability and is well recognized by the research community (Bai, Huang, Wang, Wang, & Westover, 2014; Cao, Chen, & Fan, 2016; Cao & Lin, 2015a; Cao, Lin, Huang, & Liu, 2012; Czarnecki, 2015; Deng, Zheng, & Wang, 2014; Heeswijk & Miche, 2015; Huang, 2015; Huang, Huang, Song, & You, 2015; Huang, Liu et al., 2015; Huang et al., 2006; Iosifidis, Tefas, & Pitas, 2016; Li, You, Guo, Luo, & Zhao, 2016; Luo, Vong, & Wong, 2014;

* Corresponding author.
E-mail address: jwcao@hdu.edu.cn (J. Cao).

http://dx.doi.org/10.1016/j.neunet.2016.06.001
0893-6080/

Zhang & Luo, 2015). Besides ELM, the sparse representation based classifier (SRC) is of particular interest to the pattern recognition community (Olshausen & Field, 1996; Vinje & Gallant, 2000; Wright, Yang, Ganesh, Sastry, & Ma, 2009). The SRC algorithm was first motivated by the sparsity of response in human visual cortex neuron (Olshausen & Field, 1996; Vinje & Gallant, 2000) and then found capabilities in face recognition, image processing, computer vision, direction estimation, etc. (Bai et al., 2015; Du & Cheng, 2014; Ghofrani, 2015; Han et al., 2015; Luo & Zhang, 2014; Wang et al., 2014; Wright, Yang et al., 2009; Yang, Sastry, Ganesh, & Ma, 2010; Yang, Zhou, Balasubramanian, Sastry, & Ma, 2013; Zhu et al., 2016). For image classification, SRC tries to exploit the relationship between image samples from the same category and build the sparse representation for the query image through linear regressions. Despite of the respective excellent characteristics of ELM and SRC, there also remain drawbacks that limit their practical applications. It is shown that ELM performs extremely quickly and cannot handle noise well, whereas SRC shows robustness to noise but suffers high computational cost in image recognition (Cao, Zhao, Lai, Chen et al., 2015; Cao, Zhao, Lai, Ong et al., 2015; Luo & Zhang, 2014; Wang et al., 2014). In addition, it should be noted that a well-designed image classifier should not only achieve high prediction accuracy but also be computationally efficient. As a consequence, ELM and SRC have complementary strengths in speed and accuracy, and it is reasonable to design a hybrid model through the fusion of those two methods. In our previous studies, we have presented a hybrid algorithm by combining ELM and SRC (ELM-SRC) for face and landmark image classification in Luo and Zhang (2014) and Cao, Zhao, Lai, Ong et al. (2015). It is shown that the hybrid classifier indeed outperforms ELM in recognition rate and SRC in computational complexity. However, ELM-SRC still suffers a high computational burden due to using the over-complete or highly redundant dictionary for linear representation (Cao, Zhao, Lai, Chen et al., 2015; Cao, Zhao, Lai, Ong et al., 2015).

In view of above considerations, we aim to develop an improved and efficient classifier which is a cascade of ELM and SRC for image recognition. Since ELM is not effective in handling noisy images, it is natural to pick up the noisy images and perform the classification with the more effective SRC classifier. Thus, in the first stage, all the query images are tested by the pre-trained ELM network. Then, a discriminative criterion proposed in Luo and Zhang (2014) that attempts to select the noisy images via estimating the ELM misclassified samples is employed. To enhance the performance of ELM and maximize the separation boundary, the regularized ELM adopting the leave-one-out cross validation (LOO) scheme for optimal regularization parameter selection is used. As pointed out previously, utilizing the redundant and over-complete dictionary for sparse representation, the existing SRC algorithm generally suffers the drawbacks of high computational complexity and lack of adaptability due to the negative effects of uncorrelated classes. To address this issue, in the second stage a sub-dictionary (sparse domain) selection strategy for each query image is proposed for sparse representation rather than using the whole dictionary. This is achieved under the principle that the best sub-dictionary should contain the most relevant classes to the query image. Accordingly, an adaptive sub-dictionary selection strategy based on the ELM output is presented to construct the sub-dictionary. To sum up, we propose a hybrid classifier combining improved ELM and adaptive SRC (referred to as EA-SRC) for image classification in this paper. As a reminder, the classifier presented in this paper is an evolution of our previous work (Cao, Zhao, Lai, Ong et al., 2015; Luo & Zhang, 2014), and the main contributions lie in optimal regularized ELM via LOO error in the first stage and the sub-dictionary adaption in the second stage, which result in, as we will see in the experiments, further improved accuracy and reduced computational complexity.

The rest of the paper is organized as follows. Section 2 briefly introduces the related work. Details on the proposed EA-SRC classifier are presented in Section 3. To verify the efficiency and effectiveness of the proposed method, extensive experiments concerning on typical image classification problems, including handwritten digit classification, landmark recognition and face recognition, are conducted in Section 4. Finally, Section 5 concludes the paper.

2. Related work

2.1. ELM

ELM was originally proposed for single hidden layer feedforward neural networks (SLFNs) and then extended to generalized feedforward networks. The notable merit of ELM is attributed to the random selection of the hidden node parameters (input weights and bias), whereby only the output weights need to be determined.

For a set of training samples $\{(\mathbf{x}_j, \mathbf{t}_j)\}_{j=1}^N$ with N samples and m classes, the SLFN with L hidden nodes and activation function $g(x)$ is expressed as

$$\sum_{i=1}^{L} \beta_i g_i(\mathbf{x}_j) = \sum_{i=1}^{L} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{o}_j, \quad j = 1, 2, \ldots, N \tag{1}$$

where $\mathbf{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]^T$, $\mathbf{t}_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]^T$, $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$, and b_i are the input, its corresponding desired output, the connecting weights of the ith hidden neuron to input neurons, and the bias of the ith hidden node, respectively; $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the connecting weights of the ith hidden neuron to the output neurons and $\mathbf{o}_j$ is the actual network output with respect to input $\mathbf{x}_j$. As the hidden parameters $\{\mathbf{w}_i, b_i\}$ can be randomly generated without tuning during training, ELM aims to solve the following compact model which minimizes the error between $\mathbf{t}_j$ and $\mathbf{o}_j$:

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\|_F \tag{2}$$

with

$$\mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_L, b_1, \ldots, b_L) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix},$$
$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix} \tag{3}$$

here $\mathbf{H}$ is called the hidden layer output matrix (Huang et al., 2006) while β is the output weight matrix. (2) is actually a least squares problem whose solution can be given by $\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}$, where $\mathbf{H}^{\dagger}$ is the pseudo-inverse of $\mathbf{H}$. For classification problems, ELM generally employs the one-against-all (OAA) label coding to transfer the multi-classification problem to a multi-output function regression. After the calculation of actual output for a query sample, the predicted class label is then given by the index of the actual output. The pseudo-code of ELM classifier is shown in Algorithm 1.

Due to the numerical instability of the pseudo-inverse, the widely-known ridge regression or regularized least squares in the following form are often used to optimize the solution (Heeswijk & Miche, 2015)

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\|_F + \frac{1}{\lambda}\|\beta\|_F \tag{4}$$