Signal Processing 141 (2017) 137–143

Contents lists available at ScienceDirect

Signal Processing

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

LMAE: A large margin Auto-Encoders for classification

Weifeng Liu[a,*], Tengzhou Ma[a], Qiangsheng Xie[b], Dapeng Tao[c], Jun Cheng[d]

[a] China University of Petroleum (East China), China
[b] Shandong Institute for Food and Drug Control, China
[c] Yunnan University, China
[d] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Received 5 April 2017
Revised 23 May 2017
Accepted 30 May 2017

Keywords:
Auto-Encoder
Large Margin
kNN
Classification

ABSTRACT

Auto-Encoders, as one representative deep learning method, has demonstrated to achieve superior performance in many applications. Hence, it is drawing more and more attentions and variants of Auto-Encoders have been reported including Contractive Auto-Encoders, Denoising Auto-Encoders, Sparse Auto-Encoders and Nonnegativity Constraints Auto-Encoders. Recently, a Discriminative Auto-Encoders is reported to improve the performance by considering the within class and between class information. In this paper, we propose the Large Margin Auto-Encoders (LMAE) to further boost the discriminability by enforcing different class samples to be large marginally distributed in hidden feature space. Particularly, we stack the single-layer LMAE to construct a deep neural network to learn proper features. And finally we put these features into a softmax classifier for classification. Extensive experiments are conducted on the MNIST dataset and the CIFAR-10 dataset for classification respectively. The experimental results demonstrate that the proposed LMAE outperforms the traditional Auto-Encoders algorithm.

1. Introduction

Feature extraction plays a key role in computer vision applications such as image annotation, face recognition, action analysis, object detection, target tracking and video retrieval. Auto-Encoders, as one representative deep learning method, has demonstrated to achieve superior performance for feature representation learning [1–5]. Recently, various variants of Auto-Encoders have been brought up. They are Sparse Auto-Encoders (SAE) [6–8], Denoising Auto-Encoders (DAE) [9–13], Contractive Auto-Encoders (CAE) [14–16], Nonnegativity Constraints Auto-Encoders (NCAE) [17,18], Laplacian Regularized Auto-Encoders (LAE) [19,20], Hessian Regularized Sparse Auto-Encoders (HSAE) [21], Bayesian Auto-Encoder (BAE) [22], Coupled Deep Auto-Encoder (CDA) [23], Multimodal Deep Auto-Encoder (MDA) [24] and Discriminative Auto-Encoder [25]. SAE introduced the sparsity regularization into the code vector of the hidden layer [6,7] or the output layer [8]. DAE was trained to make the learned representations robust to partial corruption of the input pattern [9–12]. A step further, the Denoising Auto-Encoder was trained in a convolutional way that can abstract hierarchical feature representations from raw visual data [13]. CAE added a penalty term computed by the Forbenius norm of the Jacobian matrix on the hidden layer features [14]. And then Rifai et al. extended the method by adding a penalizing term on second order derivatives of the encoders' output with respect to the input [15]. NCAE trained the Sparse Auto-Encoder by applying nonnegativity constraints [18] on the weight matrix [17]. LAE added a Laplacian regularization penalty term to enhance the locality-preserving property of learned encoders for data points [19]. HSAE applied the Hessian regularization to SAE, which can well preserve local geometry for data points [21]. In the algorithm of BAE, the author combined Auto-Encoder with Bayesian Net, and constructed multi-layer Bayes Net as a recognition system [22]. CDA was based on an individual architecture that can simultaneously learns the intrinsic representations of low-resolution and high-resolution image patches for single image super-resolution [23]. MDA extracted features with multimodal fusion and back-propagation deep learning [24].

All the abovementioned variants of Auto-Encoders learnt the feature representation without considering the label information in the pre-training phase. It is undeniable that the optimization process will be more effective with an unsupervised pre-training to initialize the model [26]. And if we enforce the discriminability of the features by using the labels, that will promote the efficiency of the classifier [27,28]. The discriminative Auto-Encoder aimed to boost the discriminability of the hidden layer features by minimizing the within class scatter and maximizing the between class scatter of samples [25], the method was efficacious. In this paper, we propose a Large Margin Auto-Encoders (LMAE) to fur-

* Corresponding author.
E-mail addresses: liuwf@upc.edu.cn, liuwfxy@gmail.com (W. Liu).

http://dx.doi.org/10.1016/j.sigpro.2017.05.030
0165-1684/

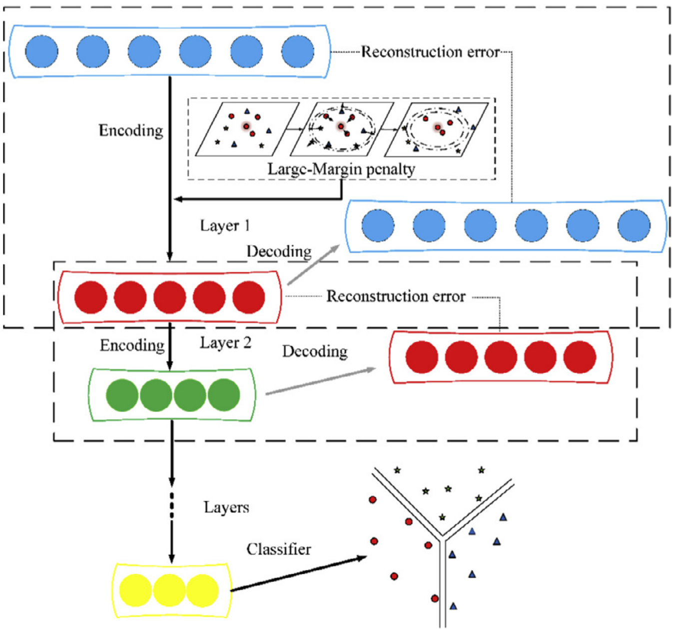

Fig. 1. The framework of LMAE for classification.

ther enforce the discriminability by enforcing different class samples to be large marginally distributed in hidden feature space. Particularly, we employ a large margin penalizing term that constrains the samples with different class labels to be distanced by an safety margin in the k-nearest neighborhood [29–32]. Fig. 1 shows the deep architecture by stacking multiple layers of LMAE for classification. In each layer, LMAE attempts to minimize the reconstruction error between the inputs and outputs while separating the different class samples with a large margin. Then, a deep architecture is constructed by stacking multiple layers of LMAE. Finally, the learnt feature representations are put into a classifier for recognition.

To assess the effectiveness of our method, we apply LMAE on two popular datasets including the MNIST dataset and the CIFAR-10 dataset for classification. And we also compare the proposed LMAE with the traditional Auto-Encoders. The experimental results demonstrate that LMAE always outperforms the traditional Auto-Encoders algorithm. In summary, our contribution in this paper is threefold: (1) we integrate the large margin penalty into the framework of Auto-Encoders that boost the discriminability significantly, (2) we provide the optimization of the proposed Large Margin Auto-Encoders (LMAE) algorithm, and (3) we conduct comparing experiments on two popular datasets respectively to demonstrate the advantages of LMAE.

We organize the rest of this paper as follows. In Section 2, we briefly review the related works including the traditional Auto-Encoders, the Discriminative Auto-Encoder and the Large-Margin kNN classification method. In Section 3, we present the proposed LMAE and the corresponding optimization. In Section 4, we describe our extensive experiments and discuss the experimental results. Finally, in Section 5, we conclude the paper with some discussions and propose possible extensions of our current method.

2. Related works

In this section, we briefly review the related works including the traditional Auto-Encoders, Discriminative Auto-Encoders and the Large-Margin Nearest Neighbor classification.

2.1. Auto-Encoders

The basic Auto-Encoder [33] aims to find a parameter vector θ for the encoder and decoder by minimizing the reconstruction error $J_{AE}(\theta)$. The objective function can be expressed as the following problem:

$$\min_{\theta} J_{AE}(\theta) = \min_{\theta} \sum_{k} L(x_k, g_\theta(f_\theta(x_k))), \tag{1}$$

where x_k is a training sample, $L(x, r) = \|x - r\|^2$ is the reconstruction error of the input and output data. $f_\theta(x) = s_f(b^e + Wx)$ and $g_\theta(x) = s_g(b^z + W'x)$ are the encoder and decoder mapping functions respectively. Usually, s_f and s_g can be the general activation functions such as the sigmoid function. The parameter vector $\theta = \{b^e, W, b^z, W'\}$, where b^e and b^z are bias vectors of the encoder and decoder, and W and W' are weight matrices of the encoder and decoder.

For the traditional Auto-Encoder, a weight decay term J_{wd} can be added into the overall objective function to control the decreasing of the weight magnitudes [34,36]. Significantly, it can improve the generalization and avoid the overfitting in the neural network by suppressing the effects of static noise on the targets and the irrelevant components of the weight vector [35].

A deep architecture can be constructed by stacking the above-mentioned basis Auto-Encoders, in which the output of the hidden layer in the first encoder is treated as the input of the second encoder. And then the last layer of the deep architecture obtains the final representation of the samples that can be used for classification tasks.

2.2. Discriminative Auto-Encoders

The Discriminative Auto-Encoder tries to boost the discriminative of the hidden layer features [25]. Denote $e_{i,j}$ as the hidden layer features of the j^{th} sample from the class i. The discriminative Auto-Encoders simultaneously minimizes the within-class scatter $S_w(e)$ and maximizes the between-class scatter $S_b(e)$. The $S_w(e)$ and $S_b(e)$ can be defined as:

$$S_w(e) = \sum_{i=1}^{c} \sum_{e_{i,j} \in i} \left(e_{i,j} - \bar{e}_i\right)\left(e_{i,j} - \bar{e}_i\right)^T, \tag{2}$$

$$S_b(e) = \sum_{i=1}^{c} m_i(\bar{e}_i - \bar{e})(\bar{e}_i - \bar{e})^T, \tag{3}$$

where $\bar{e}_i$ and $\bar{e}$ are denoted as the mean vector of e_i and e, respectively. And m_i is the samples number of class i. The discriminative regularization term can be defined as:

$$L(e) = tr(S_w(e)) - tr(S_b(e)). \tag{4}$$

Then, the objective function of the Discriminative Auto-Encoder can be expressed as the following problem:

$$\min_{\theta} J_{Dis-AE}(\theta) = \min_{\theta} \left[J_{AE}(\theta) + \frac{1}{2}\lambda J_{wd} + \frac{1}{2}\gamma L(e)\right], \tag{5}$$

where λ and γ are parameters to balance the different penalty terms respectively.

2.3. Large-Margin kNN classification

The Large-Margin kNN classification method (LMNN) attempts to shrink distances of neighboring same labeled points and to separate points in different classes [30,37]. The LMNN optimization problem can be formulated as the following minimization problem: