Neurocomputing 275 (2018) 448–461

Contents lists available at ScienceDirect

Neurocomputing

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

A hybrid deep learning CNN–ELM for age and gender classification

Mingxing Duan[a], Kenli Li[b,c,*], Canqun Yang[a], Keqin Li[b,c]

[a] College of Computer Science, National University of Defense Technology, Changsha 410073, China
[b] College of Information Science and Engineering, Hunan University, Changsha, Hunan 410082, China
[c] National Supercomputing Center in Changsha, Changsha, Hunan 410082, China

ARTICLE INFO

Article history:
Received 13 March 2017
Revised 24 August 2017
Accepted 27 August 2017
Available online 8 September 2017

Communicated by Dr. G.-B. Huang

Keywords:
Classification
Convolutional Neural Network
Extreme Learning Machine
Image
Overfitting

ABSTRACT

Automatic age and gender classification has been widely used in a large amount of applications, particularly in human-computer interaction, biometrics, visual surveillance, electronic customer, and commercial applications. In this paper, we introduce a hybrid structure which includes Convolutional Neural Network (CNN) and Extreme Learning Machine (ELM), and integrates the synergy of two classifiers to deal with age and gender classification. The hybrid architecture makes the most of their advantages: CNN is used to extract the features from the input images while ELM classifies the intermediate results. We not only give the detailed deployment of our structure including design of parameters and layers, analysis of the hybrid architecture, and the derivation of back-propagation in this system during the iterations, but also adopt several measures to limit the risk of overfitting. After that, two popular datasets, such as, MORPH-II and Adience Benchmark, are used to verify our hybrid structure. Experimental results show that our hybrid architecture outperforms other studies on the same datasets by exhibiting significant performance improvement in terms of accuracy and efficiency.

1. Introduction

1.1. Motivation

Age and gender classification play a very important role in our social lives, by which we can find whether the persons we contact are “sir” or “madam” and young or old. These behaviors are heavily dependent on our ability to estimate these individual traits: age and gender, which are from facial appearances [1]. These attributes are important in our lives while the ability to estimate them accurately and reliably from facial appearance is still far from satisfying the needs of commercial applications [2].

In order to enhance the ability to estimate or classify these attributes from face images, many methods have been put forward in the past years. Based on cranio-facial changes in feature-position rotation and on skin wrinkle analysis, these attributes have been classified from facial images [3] while a methodology is proposed to classify age and gender automatically from facial images through feature extraction including primary and secondary features [4]. However, these approaches mentioned above have been designed particularly for processing constrained age or gender tasks which are not suitable for practical applications including unconstrained image classification tasks.

The accuracy of age and gender classification depends on two aspects: feature extraction and classification, while feature extraction is a crucial factor for the success of classification. It not only demands the features having the most differentiable characteristics among different classes, but also retains unaltered characteristics within the same class. In recent years, due to its good feature extraction ability, CNN has been highlighted in machine learning and pattern recognition fields. It has achieved state-of-the-art performance in image recognition and can automatically extract the features.

With full consideration of what mentioned above, CNN has been introduced to classify unconstrained age and gender tasks automatically and significant performance has been obtained [2]. More importantly, the unconstrained images are without prior manual filtering, which are as true as real-world applications. CNN has shown great advantages in image recognition while it is the first time to use CNN to process these unconstrained tasks so that we can further improve the accuracy of classification through the fine tuning of its structure or its parameters.

With more discriminative features and more powerful classifier, higher recognition rate will be obtained. In a plain CNN, the full-connection layers are as same as a general single

* Corresponding author at: College of Information Science and Engineering, Hunan University, Changsha, Hunan 410082, China.
E-mail addresses: duanmingxing16@nudt.edu.cn (M. Duan), lkl@hnu.edu.cn (K. Li), canqun@nudt.edu.cn (C. Yang), lik@newpaltz.edu (K. Li).

http://dx.doi.org/10.1016/j.neucom.2017.08.062
0925-2312/

hidden layer feedforward neural network (SLFN) and trained through back-propagation (BP) algorithm. On the one hand, BP algorithm is sensitive to local minima of training errors. On the other hand, SLFN is likely to be over-trained leading to degradation of its generalization performance when it performs BP algorithm [5]. Therefore, the generalization performance of the fully connection layers in the network is probably sub-optimal and they cannot make full use of discriminative features extracted by convolutional layers.

In order to deal with the problems, it is urgent to find a new classifier which owns the similar ability as the full-connection layers or softmax classifier, while it can make full use of the discriminative features. Niu and Suen [6] proposed a hybrid model which integrated the synergy of two superior classifiers including CNN and Support Vector Machine (SVM), and got a better results compared with a plain CNN. In general, the design of SVM is so complicated that is important to find other classifiers with least needing tuning parameters, good classification performance, and high generalization ability to process the same tasks mentioned above. To the best of our knowledge, SVM, Naive Bayes [7], and Extreme Learning Machine (ELM) [8] are three important classification algorithms at present while ELM has been proved to be an efficient and fast classification algorithm because of its good generalization performance, fast training speed, and little human intervene [9]. What's more, ELM and improved ELM, including mixing with other methods, have been widely used to process pattern recognition tasks and obtain a good performance [10].

1.2. Our contributions

In order to make full use of the advantages of CNN and ELM, we propose a hybrid recognition architecture, called CNN–ELM, which is used to process age and gender classification tasks. It not only sufficiently exploits the excellent feature extraction ability of CNN and the outstanding classification property of ELM, but also is used to classify the popular human facial image datasets. At the same time, different effective approaches are adopted to reduce overfitting. With lower time complexity, the hybrid architecture gets a better performance compared with a plain CNN structure which contains the identical convolutional layers. The major contributions of this paper are summarized as follows:

- We propose a new hybrid CNN–ELM method to process age and gender classification aiming at image tasks. It combines Convolutional Neural Networks and Extreme Learning Machine in a hierarchical fashion which is sufficient in applying the advantages of CNN and ELM.
- We present the process of integrating the synergy of hybrid structure in detail, including the design of the layers in CNN, the selection of parameters in hybrid structure, the realization of back-propagation process in this hybrid model, and so on.
- Finally, two popular datasets, such as MORPH-II and Adience Benchmark, are used to verify our hybrid structure. Experiments show that our hybrid structure gets better performance compared with other studies on the same image datasets and also can fulfill the requirements of many real-world application.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 gives preliminary information. Section 4 discusses architecture of our hybrid CNN–ELM model. Section 5 describes merits of hybrid CNN-ELM model. We also analyze the time complexity of hybrid classification in Section 6. The experiments and results are illustrated in Section 7. Finally, we make a conclusion in Section 8.

2. Related work

2.1. Hybrid neural network system

CNN has been successfully applied to various fields, and specially, image recognition is a hot research field. However, few researchers have paid attention on hybrid neural network. Lawrence et al. [11] presented a hybrid neural-network solution for face recognition which made full use of advantages of self-organizing map (SOM) neural network and CNN. That approach showed a higher accuracy compared with other methods used for face recognition at that time. In 2012, Niu and Suen [6] introduced a hybrid classification system for objection recognition by integrating the synergy of CNN and SVM, and experimental results showed that the method improved the classification accuracy. Liu et al. [12] used CNN to extract features while Conditional Random Rield (CRF) was used to classify the deep features. With extensive experiments on different datasets, such as Weizmann horse, Graz-02, MSRC-21, Stanford Background, and PASCAL VOC 2011, the hybrid structure got better segmentation performance compared with other methods on the same datasets. In [13], Xie et al. used a hybrid representation method to process scene recognition and domain adaption. In that method, CNN was used to extract the features meanwhile mid-level local representation (MLR) and convolutional Fisher vector representation (CFV) made the most of local discriminative information in the input images. After that, SVM classifier was used to classify the hybrid representation and achieved better accuracy. Recently, Tang et al. [14] put forward a hybrid structure including Deep Neural Network (DNN) and ELM to detect ship on spaceborne images. In this time, DNN was used to process high-level feature representation and classification while ELM was worked as effective feature pooling and decision making. What is more, extensive experiments were presented to demonstrate that the hybrid structure required least detection time and achieved highter detection accuracy compared with existing relevant methods. Based on the analysis above, we can integrate CNN with other classifiers to improve the classification accuracy. In Sections 4–6, we will present our hybrid CNN-ELM in detail and show its better performance compared with other methods to process the same tasks.

2.2. Age classification

Recently, age and gender classification has received huge attention, which provides direct and quickest way for obtaining implicit and critical social information [15]. Fu et al. [16] made a detailed investigation of age classification and we can learn more information about recent situation from Ref. [2]. Classifying age from the human facial images was first introduced by Kwon et al. [3] and it was presented that calculating ratios and detecting the appearance of wrinkles could classify facial features into different age categorization. After that, the same method was used to model craniofacial growth with a view to both psychophysical evidences and anthropometric evidences [17] while this approach demanded accurate localization of facial features.

Geng et al. [18] proposed a subspace method called AGing pattErn Subspace which was used to estimate age automatically while age manifold learning scheme was presented in [19] to extract face aging features and a locally adjusted robust regressor was designed to predict human ages. Although these methods have shown many advantages, the requirement that input images need to be near-frontal and well-aligned is their weakness. It is not difficult to find that the datasets in their experiments are constrained, so that these approaches are not suited for many practical applications including unconstrained image tasks.