Neurocomputing ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Neurocomputing

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

An image classification method that considers privacy-preservation

Chongwen Liu [a,b], Zhaowei Shang [a,b,*], Yuan Yan Tang [c]

[a] Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education, China
[b] College of Computer Science, Chongqing University, Chongqing, China
[c] Faculty of Science and Technology, University of Macau, Macau, China

ARTICLE INFO

Article history:
Received 26 September 2015
Received in revised form
25 March 2016
Accepted 1 April 2016

Keywords:
Image classification
Privacy-preservation
Ensemble system

ABSTRACT

Using images that are dispersed in a network can improve image classification performance; however, it is easy to leak the image holders' privacy. In this paper, we present a novel method that considers both image classification performance and the image holders' privacy-preservation. Our method contains four stages; the feature extraction stage follows the Bag-of-Features (BoF) framework, but it introduces significantly novel stages in the training, classification and ensemble stages. In the training stage, we split the image classification into many one-vs.-one classification problems; a set of binary classifiers as weak learners are learned on a local machine, to classify each image in an autonomous system. In the classification stage, a local machine sends features of unlabeled images to each image autonomous system, and then, each image autonomous system and local machine finish a prediction task under their own trained model. Afterward, the image autonomous systems send their prediction labels back to the local system. In the ensemble stage, we gather all of the prediction labels from the image autonomous systems and design an ensemble system to ensemble them into final classification results. This ensemble system must consider both the efficiency of the different trained models in each image autonomous system and the efficiency of the trained model on each class in the one-vs.-one classification problems. During the whole process, only the features of unlabeled images and prediction labels are transmitted in the network, which largely reduces the risk of privacy leakage. We conducted experiments on the network environment, to prove that our method has preferable performance. Our first experiment demonstrates that the performance of our method achieves state-of-the-art performance on four public dataset benchmarks, which shows that our method has competitive performance. The second experiment shows that our method achieves state-of-the-art performance with even fewer training samples, which means that there is less communication overhead and better privacy-preservation.

1. Introduction

Scene categorization aims to classify an input image into a specific scene category (e.g., city, mountain, street). This capability has extensive applications in image retrieval, such as content-based web search engines, especially with the growing popularity of digital photography, which leads to a tremendous number of images collected in personal photo albums and online image repositories. Despite the great progress made over the past decade, the classification performance still cannot fulfill many real-world applications, and there are many unresolved problems.

The existing methods have good performance in image classification [1,2], but most of these methods consider that all the training and testing images are stored in local storage systems, and the image classification task is completed in the local machine. During the development of a network, image data are usually distributed in storage [3], some network image classification methods had been proposed. In [4], a distributed approach is developed for achieving large-scale classifier training and image classification considers two critical issues of huge computational cost and huge storage/memory cost for large-scale classifier training and image classification. We analyze the characters of network structure first. Fig. 1 shows the network structure of a distributed image storage strategy, which is commonly used in current image data storage strategies in a network environment. In Fig. 1, we define an autonomous system (AS) as a collection of connected hosts under the control of one or more network operators, and an image autonomous system (IAS) is a network

* Corresponding author at: College of Computer Science, Chongqing University, Chongqing, China.
E-mail addresses: liuchongwen16@gmail.com (C. Liu), szw@cqu.edu.cn (Z. Shang), yytang@umac.mo (Y.Y. Tang).

http://dx.doi.org/10.1016/j.neucom.2016.04.056
0925-2312/

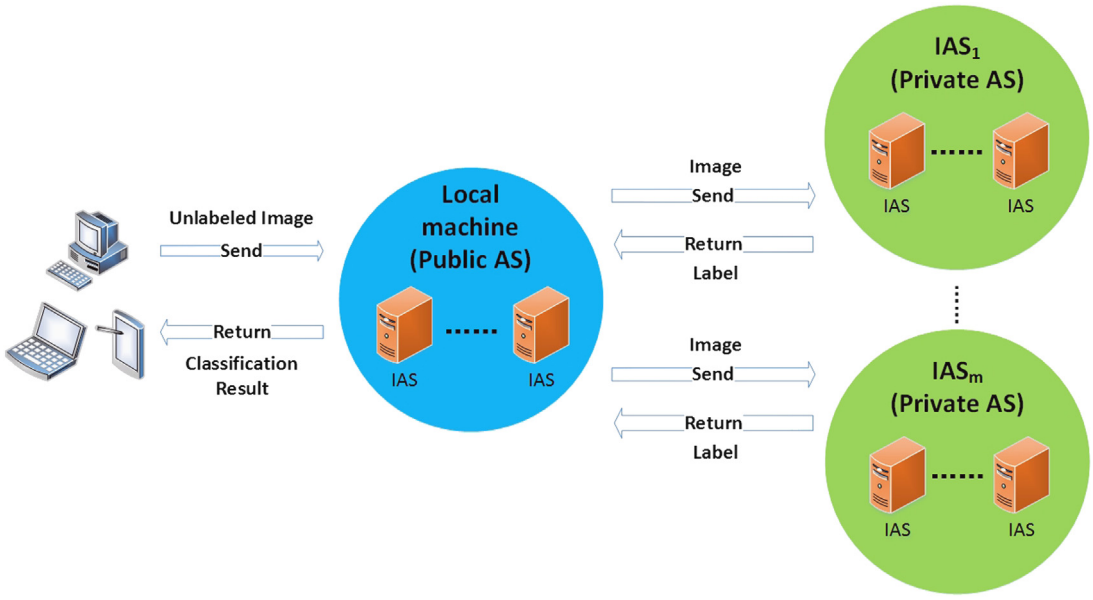

Fig. 1. Network structure.

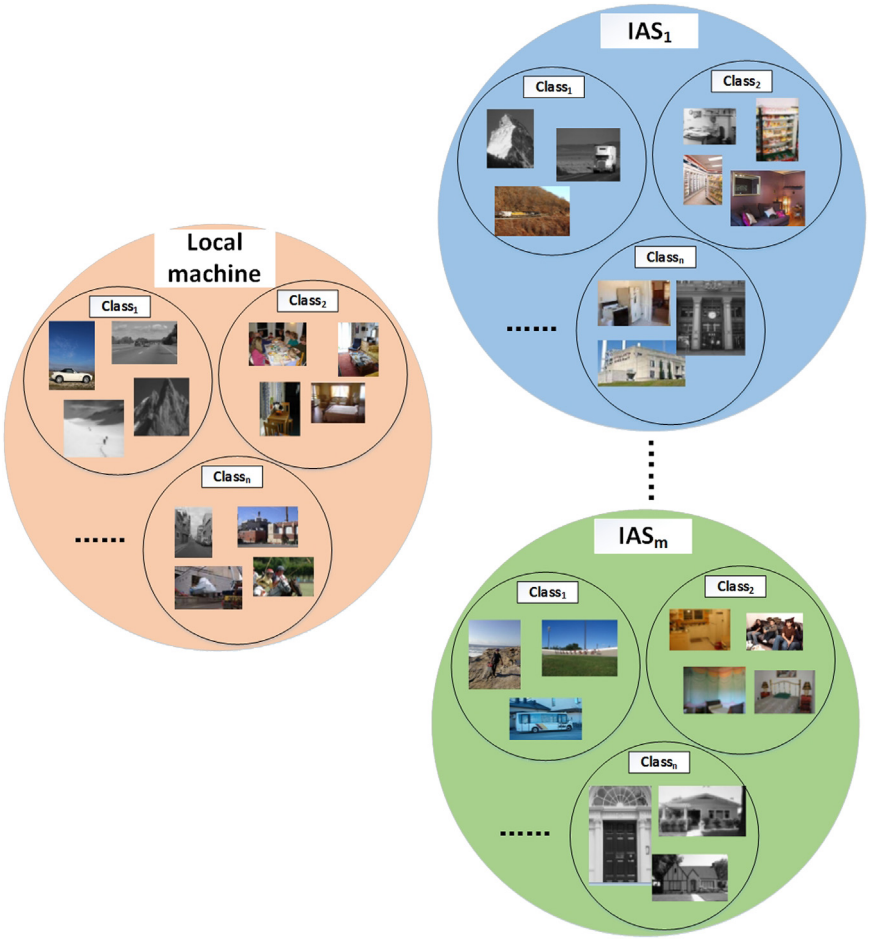

Fig. 2. Image storage logic structure.

unit in an AS that has free transmission and shares image data. However, the physical truth is that the images are stored in IAS, which public users have no right to use, and how to use a private image to enhance image classification performance with the premise of privacy-preservation is a major problem in the image classification field. Fig. 2 shows the image storage logic structure of an image distributed storage strategy. Because we have no right to access private data, only the public AS can be accessed. We designed our method to use private data with the premise of privacy-preservation, to improve the performance and reduce the

Please cite this article as: C. Liu, et al., An image classification method that considers privacy-preservation, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.04.056