Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 19–32
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Image classification with the use of radial basis function neural networks and the minimization of the localized generalization error

Wing W.Y. Ng[a,b,*], Andres Dorado[c], Daniel S. Yeung[a,b], Witold Pedrycz[d], Ebroul Izquierdo[c]

[a]*Department of Computing, The Hong Kong Polytechnic University, China*
[b]*Media and Life Science Computing Lab, Shenzhen Graduate School, Harbin Institute of Technology, China*
[c]*Department of Electronic Engineering, Queen Mary, University of London, UK*
[d]*Department of Electrical and Computer Engineering, University of Alberta, Canada*

Received 31 October 2005; accepted 6 July 2006

Abstract

Image classification arises as an important phase in the overall process of automatic image annotation and image retrieval. In this study, we are concerned with the design of image classifiers developed in the feature space formed by low level primitives defined in the setting of the MPEG-7 standard. Our objective is to investigate the discriminatory properties of such standard image descriptors and look at efficient architectures of the classifiers along with their design pursuits. The generalization capabilities of an image classifier are essential to its successful usage in image retrieval and annotation. Intuitively, it is expected that the classifier should achieve high classification accuracy on unseen images that are quite "similar" to those occurring in the training set. On the other hand, we may assume that the performance of the classifier could not be guaranteed in the case of images that are very much dissimilar from the elements of the training set. To follow this observation, we develop and use a concept of the localized generalization error and show how it guides the design of the classifier. As image classifier, we consider the usage of the radial basis function neural networks (RBFNNs). Through intensive experimentation we show that the resulting classifier outperforms other classifiers such as a multi-class support vector machines (SVMs) as well as "standard" RBFNNs (viz. those developed without the guidance offered by the optimization of the localized generalization error). The experimental studies reveal some interesting interpretation abilities of the RBFNN classifiers being related with their receptive fields.

Keywords: Image classification; Radial basis functions neural networks; MPEG-7; Support vector machines; Generalization error

1. Introduction

A vast amount of digital images are become omnipresent these days call for an intensified effort towards building efficient needs for their automatic annotation and retrieval mechanisms. Classification of digital images becomes one of the fundamental activities one could view as the fundamental prerequisite for all other image processing pursuits. In image classification we could follow the general paradigm of pattern recognition. In pattern recognition, each object is described by a collection of features that forms a multidimensional space in which all discrimination activities take place. Various classifiers, both linear and nonlinear, become available at this stage including support vector machines (SVM), linear classifiers, polynomial classifiers, radial basis function neural networks (RBFNNs), fuzzy rule-based systems, etc. No matter what classifier has been chosen, a formation of a suitable feature becomes of paramount relevance. The problem of forming of the feature space in the case of images is even more complicated. On one hand, we have a lot of different alternatives. On the other

* Corresponding author. Department of Computing, The Hong Kong Polytechnic University, China. Tel.: +852 2766 4901; fax: +852 2774 0842.
E-mail addresses: cswyng@comp.polyu.edu.hk (W.W.Y. Ng), andres.dorado@elec.qmul.ac.uk (A. Dorado), csdaniel@comp.polyu.edu.hk (D.S. Yeung), pedrycz@ee.ualberta.ca (W. Pedrycz), ebroul.izquierdo@elec.qmul.ac.uk (E. Izquierdo).

0031-3203/$30.00
doi:10.1016/j.patcog.2006.07.002

hand, the diversity of images contributes to the elevated level of complexity and difficulty. In images, we encounter a variety of images showing different shapes, colors, texture, etc yet belonging to the same class. An image could be described by an intensity of color of each pixel or even better by some descriptors. In this study, our objective is to explore and quantify the discriminatory properties of the MPEG-7 image descriptors in classification problems. Those are explored in conjunction to two main categories of classifiers such as SVMs and RBFNNs.

Unfortunately, it becomes obvious that any classifier requiring high training accuracy may not achieve good generalization capability. Since both target outputs and distributions of the unseen samples are unknown, it is impossible to compute the generalization error in a direct way. There are two major approaches to estimate the generalization error, namely, analytical model and cross-validation (CV). In general, analytical models bound above the generalization error for any unseen samples and do not distinguish trained classifiers with the same number of effective parameters but different values of parameters. Thus, the error bounds given by those models are usually loose [1]. The major problem of analytical models is the estimation of the number of effective parameters of the classifier, which could be solved by using the VC-dimensions [2]. The VC-dimension of a classifier is defined as the largest number of samples that can be shattered by this classifier [2]. However, only loose bound of VC-dimensions could be found for nonlinear classifiers, e.g. neural networks, and this puts a severe limitation on the applicability of analytical models to nonlinear classifiers, except the SVM [3]. Although CV uses true target outputs for unseen samples, it is time consuming for large datasets and *CL* classifiers must be trained for *C*-fold CV and *L* choices of classifier parameters. CV methods estimate the expected generalization error instead of its bound, thus they do not guarantee the finally built classifier to have good generalization capability [1].

In image classification, one may not expect a classifier trained using one category of images (say, animals) to correctly classify images coming from some other categories (e.g. vegetables). In this case, one may revise the training dataset by adding training samples of vegetables and re-train the classifier to include the new class of images. For example, in our dataset we have images of cow but not airplane, thus we could not expect the classifier trained using our dataset to correctly recognize an airplane. It is expected that an image classifier work well for those classes that have been used to train it assuming images belonging to the same class are conceptually similar and such that their descriptor values should also be similar. That is, unseen samples similar to the training samples, in terms of sup-type of distance in the feature space is smaller than a given threshold, are considered to be more important. Thus, in the evaluation of the generalization capabilities of the image classifiers, one may ignore those images that are totally dissimilar to those existing in the training set.

In general, image classification problems are multi-class classification problems and difficult to find a classifier with good generalization properties. In this work, we aim to find an image classifier featuring better generalization capability and interpretability with respect to domain knowledge in image classification. We concentrate on finding an optimal number of receptive fields for RBFNNs to classify the images with lower generalization error to unseen images.

We organize the study in the following manner. The starting point is a discussion on the formation of the feature space based upon the framework of descriptors being available in the MPEG-7 standard. These issues are covered in Section 2. We provide a brief introduction to image classifiers in Section 3. The localized generalization error model (R^*_{SM}) and the corresponding approach to the selection of the architecture of the network are described in Sections 4 and 5, respectively. We present a comprehensive suite of experimental studies in Section 5. Concluding comments are covered in Section 6.

2. MPEG-7 feature space

In this section, we elaborate on the feature space arising within the framework of MPEG-7. The MPEG-7 descriptors are useful for low-level matching and provide a great flexibility for a wide range of applications.

2.1. MPEG-7 descriptors

MPEG-7, formally known as multimedia content description interface is an ISO/IEC standard developed by the moving picture experts group (MPEG) for description and search of audio and video content; refer to www.chiariglione.org/mpeg/. In contrast with the earlier standards known as MPEG-1, MPEG-2, and MPEG-4 that are focused on coding and representation of audio–visual content, On the other hand, MPEG-7 moves forward and becomes more general by embracing a description of multimedia content [4].

MPEG-7 has emerged as a cornerstone of the development of a wide spectrum of applications dealing with audio, speech, video, still pictures, graphics, 3D models, and alike. In a nutshell, the MPEG-7 environment delivers a comprehensible metadata description standard that is interoperable with other leading standards such as SMPTE Metadata Dictionary, Dublin Core, EBU P/Meta, and TV Anytime; refer to www.ebu.ch/trev_284-mulder.pdf. Initially, MPEG-7 was focused more on web-based applications and annotation tools (e.g. Refs. [5,6]). Nowadays, it is being drifted to other domains such as education, video surveillance, entertainment, medicine and biomedicine.

The ultimate objective of MPEG-7 is to provide interoperability among systems and applications used in generation, management, distribution, and consumption of audio–visual content descriptions. Such descriptions of streamed (live) or