Information Sciences 316 (2015) 567–581

Contents lists available at ScienceDirect

Information Sciences

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

Evolutionary compact embedding for large-scale image classification

Li Liu [a,b], Ling Shao [a,b,*], Xuelong Li [c]

[a] College of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China
[b] Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, UK
[c] State Key Laboratory of Transient Optics and Photonics, XIOPM, Chinese Academy of Sciences, Xi'an 710119, Shannxi, China

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form 1 May 2014
Accepted 25 June 2014
Available online 10 July 2014

Keywords:
Dimensionality reduction
Large-scale image classification
Evolutionary compact embedding
Genetic programming
AdaBoost

ABSTRACT

Effective dimensionality reduction is a classical research area for many large-scale analysis tasks in computer vision. Several recent methods attempt to learn either graph embedding or binary hashing for fast and accurate applications. In this paper, we propose a novel framework to automatically learn the task-specific compact coding, called evolutionary compact embedding (ECE), which can be regarded as an optimization algorithm combining genetic programming (GP) and a boosting trick. As an evolutionary computation methodology, GP can solve problems inspired by natural evolution without any prior knowledge of the solutions. In our evolutionary architecture, each bit of ECE is iteratively computed using a binary classification function, which is generated through GP evolving by jointly minimizing its empirical risk with the AdaBoost strategy on a training set. We address this as greedy optimization leading to small Hamming distances for similar samples and large distances for dissimilar samples. We then evaluate ECE on four image datasets: USPS digital hand-writing, CMU PIE face, CIFAR-10 tiny image and SUN397 scene, showing the accurate and robust performance of our method for large-scale image classification.

1. Introduction

Dimensionality reduction has been a critical preprocessing step in many fields of information processing and analysis, such as data mining [5,12,18–20], information retrieval [14,16], and pattern recognition [11,21,29,37]. Recently, with the advances of computer technologies and the development of the World Wide Web, a huge amount of digital data, including text, images and videos, is generated, stored, analyzed, and accessed every day. To overcome the shortcomings of text-based image retrieval, content-based image classification and retrieval has attracted substantial attention. The most basic but essential scheme for image classification is the nearest neighbor search: given a query image, to find an image that is most similar to it within a large database and assign the same label of the nearest neighbor to this query image. However, greedily searching a dataset with N samples is infeasible because linear complexity $O(N)$ is not scalable in practical applications. To overcome this kind of computational complexity problem, many other methods have been proposed to index the data for fast query responses, such as K-D tree and R tree [9]. However, these methods can only operate with small dimensionality, typically less than 100 [2,3]. Besides, most of the vision-based applications also suffer from the curse of dimensionality problems,[1] because visual descriptors usually

* Corresponding author. Addresses: College of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China, Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, UK.

E-mail address: ling.shao@ieee.org (L. Shao).

[1] The effectiveness and efficiency of these methods drop exponentially as the dimensionality increases, which is commonly referred to as the "curse of dimensionality".

http://dx.doi.org/10.1016/j.ins.2014.06.030
0020-0255/

have hundreds or even thousands of dimensions. Therefore, to make large-scale search or classification practical, some methods have been proposed to effectively reduce the dimension of data and increase the classification speed and accuracy.

One of the most baseline dimensionality reduction algorithms might be principal component analysis (PCA), which is used to explain the variance–covariance structure of a set of variables through linear combinations of those variables. PCA is most commonly applied to condense the information contained in a large number of original variables into a smaller set of new composite variables or dimensions, at the same time ensuring a minimum loss of information. Another effective scheme for dimensionality reduction is linear discriminant analysis (LDA). LDA is a supervised method that has been proved successful on classification problems [4,8]. Following the Fisher discriminant criterion, the projection vectors are commonly obtained by maximizing the between-class covariance and simultaneously minimizing the within-class covariance. However, the classical LDA is a linear method and cannot tackle nonlinear problems. In order to overcome this limitation, kernel discriminant analysis (KDA) [22] is then developed. KDA is the nonlinear extension of LDA using the kernel trick that can be implicitly performed in a new feature space, which allows non-linear mappings to be learned. Beyond that, some other dimension reduction methods can also achieve promising results for different applications. Locality preserving projections (LPP) [12] are linear projective maps that are obtained by solving a variational problem that preserves the neighborhood structure of the data set. LPP aims to find the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. In addition, another popular method, termed discriminative locality alignment (DLA) [38], is also been used as dimensionality reduction algorithms for classification. All the above methods can be thought as the direct graph embedding or its linear/kernel/tensor extensions of a specific intrinsic graph that describes certain desired statistical or geometric properties of a data set, with constraints from scale normalization or a penalty graph [36]. With the need for fast search and classification in large-scale vision applications, some recent effort has been turned to applying binary hashing techniques, which explore the approximate similarity search based on Hamming distance to effectively reduce the indexing time. Among this kind of methods, kernelized locality-sensitive hashing (KLSH) [18] has been successfully utilized for large-scale image retrieval and classification. KLSH is essentially a kernelized method of performing probabilistic dimension reduction of high-dimensional data. The basic idea is to hash the input items so that similar items are mapped to the same buckets with high probability. Beyond that, some deep learning methods are also used to learn binary coding unsupervised, e.g., restricted boltzmann machine (RBM) [6].

Although the existing dimensionality reduction methods achieve promising results in a variety of applications, they basically rely on complex and advanced mathematical knowledge to optimize the pre-defined object functions. However, for some optimization problems, direct solutions cannot always be found. Besides, in large-scale settings, matrix factorization techniques used in the above methods can also cause a heavy computational burden. So how to automatically generate better solutions to optimization problems becomes an interesting topic for real-world vision applications. In this work, we propose evolutionary compact embedding (ECE), which applies genetic programming (GP) in combination with AdaBoost to automatically evolve dimensionality reduction. ECE is demonstrated to enable accurate and robust large-scale image classification. Fig. 1 shows the working flow of ECE.

Genetic programming (GP) [26] simulates the Darwinian principle of natural selection to solve optimization problems. Different from hand-crafted techniques based on deep domain knowledge, GP is inspired by natural evolution and can be employed to automatically solve problems without prior knowledge of the solutions. Users can utilize GP to solve a wide range of practical problems, producing human-competitive results and even patentable inventions. Relying on natural and random processes, GP can escape traps by which deterministic methods may be captured. Because of this, usage of GP is not limited to any research domain and creates relatively generalized solutions for any target tasks. In GP, a group of primitive operators is first adopted to randomly assemble computational programs which are regarded as the GP initial population. This population is then allowed to 'evolve' (using crossover and mutation) through sexual reproduction with single or pair parents chosen stochastically while biased in their fitness on the task at hand. In this way, the general fitness of population tends to improve over time. Finally, the obtained individual that achieves best performance is taken as the final solution. A typical GP procedure can be included in Algorithm 1.

Algorithm 1. Genetic Programming

Start
  Initialization Randomly create an initial population of computer programs from the available primitives (terminal set & function set).
  Repeat
    (1) Execute each program and evaluate its fitness.
    (2) Choose one program from the population with a particular probability based on the fitness to involve genetic operations
    (3) Create a new generation programs by applying genetic operations.
  If An acceptable solution is found or reach the maximum number of generations defined by user.
  Stop
  Return The best-so-far solution selected by Genetic Programming.
End
