Neurocomputing 72 (2009) 3922–3929

Contents lists available at ScienceDirect

Neurocomputing

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

Image retrieval using nonlinear manifold embedding

Can Wang *, Jun Zhao, Xiaofei He, Chun Chen, Jiajun Bu

Zhejiang Key Laboratory of Service Robots, College of Computer Science, Zhejiang University, No. 38 Zheda Road, Hangzhou, Zhejiang 310027, China

ARTICLE INFO

Article history:
Received 15 January 2009
Received in revised form
28 March 2009
Accepted 9 April 2009
Communicated by X. Li
Available online 10 May 2009

Keywords:
Dimensionality reduction
Image retrieval
Relevance feedback
Kernel method

ABSTRACT

The huge number of images on the Web gives rise to the content-based image retrieval (CBIR) as the text-based search techniques cannot cater to the needs of precisely retrieving Web images. However, CBIR comes with a fundamental flaw: the semantic gap between high-level semantic concepts and low-level visual features. Consequently, relevance feedback is introduced into CBIR to learn the subjective needs of users. However, in practical applications the limited number of user feedbacks is usually overwhelmed by the large number of dimensionalities of the visual feature space. To address this issue, a novel semi-supervised learning method for dimensionality reduction, namely kernel maximum margin projection (KMMP) is proposed in this paper based on our previous work of maximum margin projection (MMP). Unlike traditional dimensionality reduction algorithms such as principal component analysis (PCA) and linear discriminant analysis (LDA), which only see the global Euclidean structure, KMMP is designed for discovering the local manifold structure. After projecting the images into a lower dimensional subspace, KMMP significantly improves the performance of image retrieval. The experimental results on Corel image database demonstrate the effectiveness of our proposed nonlinear algorithm.

1. Introduction

The Internet age sees a dramatic increase in the images stored, exchanged and shared on the Web. Like their experience with text search engines, Web users naturally require techniques and tools that help them locate precisely the images they search. However, image search is inherently a more difficult task than text search, as noted by Datta et al. that text is man's creation, while typical images are a mere replica of what man has seen, for which concrete descriptions are relatively elusive [4]. Furthermore, like other resources on the Web, image collections are semi-structured, nonhomogeneous, and massive in volume, making effective retrieval even more challenging.

To address this issue, content-based image retrieval (CBIR) was introduced in the early 1980s and from then on has attracted substantial research interest [20,22,16,17,26,25,2,9,27]. CBIR aims to organize image collections by their visual contents such as color, texture, shapes, etc. Image retrieval is thus performed in the feature space to find similar images to the example submitted by user. The most challenging problem in CBIR is the semantic gap between low-level visual features and high-level semantic concepts [4]. That is, the richness of the user's subjective needs is not matched by the simplicity of the available low-level visual features. One feasible way to address this problem is through learning from the user's relevance feedback [20].

While the relevance feedbacks provided by the user is often quite limited, typically less than 20, there might exist hundreds or even thousands of features to represent an image. This leads to a crucial problem in CBIR called as the "*curse of dimensionality*" [18], i.e. algorithms and procedures that are analytically and computationally effective in low-dimensional space become totally impractical in this case. Thus, various dimensionality reduction techniques have been introduced to ease this problem. Perhaps the most canonical dimensionality reduction algorithms are principal component analysis (PCA) [5,23] and linear discriminant analysis (LDA) [5,24]. PCA is unsupervised and it aims to find a projection on which the data variance is maximized. LDA is supervised. It aims to find a projection on which data points with the same label are clustered whereas data points with different labels are distant from one another. When relevance feedback is introduced, image retrieval is essentially a supervised learning problem. Therefore, LDA usually gives better performance than PCA.

The major disadvantage of PCA and LDA is that both of them see only Euclidean structure. However, many researcher have shown that the image space is probably a nonlinear manifold [10,8,15,7]. In order to discover the manifold structure, many manifold learning algorithms were developed in recent years. In [11] He et al. proposed locality preserving projections (LPP) to find a linear approximation of the intrinsic data manifold. Image retrieval using LPP [7,12] is then performed in the reduced

* Corresponding author. Tel.: +86 57187951431; fax: +86 57187953955.
E-mail address: wcan@zju.edu.cn (C. Wang).

0925-2312/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.04.011

subspace by using a Euclidean metric. Another work in image retrieval, augmented relation embedding (ARE) [15] learns a semantic manifold by taking into account the user's preferences. Most recently, He et al. proposed a novel image retrieval method called maximum margin projection (MMP) [8] that discovers the local manifold structure by maximizing the margin between positive and negative examples at each local neighborhood. These algorithms have demonstrated superior performance to traditional Euclidean based algorithms. However, they are still linear algorithms and may fail if the image space is highly nonlinear.

Some other nonlinear manifold learning techniques, including locally linear embedding (LLE) [19], Isomap [28], and Laplacian Eigenmaps [1], consider the case when the data lives on or close to a nonlinear submanifold of the ambient space. These methods attempt to discover the intrinsic manifold structure by estimating both the geometrical and the discriminant properties of the submanifold from random points lying on this unknown submanifold. However, these methods are defined only on the training data points, and it is unclear how the map can be evaluated for new test points. Therefore, they are not suitable for image retrieval. Besides, recently there has been a lot of interest in tensor-based dimensionality reduction, see [14].

In this paper, we propose a novel nonlinear manifold learning algorithm called kernel maximum margin projection (KMMP). KMMP is fundamentally based on maximum margin projection [8]. By using kernel techniques, the embedding can be performed in the reproducing kernel Hilbert space (RKHS) [21] which gives rise to nonlinear embedding. When a linear kernel is used, KMMP reduces to MMP. Similar to MMP, KMMP is also defined everywhere. KMMP finds a mapping function which can map any new test points to the lower dimensional Euclidean space. By using nonlinear kernel, the nonlinear structure of the image manifold can be well preserved, and therefore the image retrieval performance can be improved.

The rest of the paper is organized as follows. In Section 2, we give a brief review of the canonical dimensionality reduction algorithms, that is, PCA and LDA. We describe our proposed kernel maximum margin projection algorithm in Section 3. The experimental results are presented in Section 4. Finally, we give concluding remarks and suggestions for future work in Section 5.

2. Linear dimensionality reduction

In this section, we give a brief description of PCA and LDA.

2.1. Principal component analysis

The generic problem of linear dimensionality reduction is the following. Given a set $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ in $\mathbb{R}^n$, find a transformation matrix $A = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_l\}$ that maps these m points to a set of points $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ in $\mathbb{R}^l$ ($l \ll n$), such that $\mathbf{y}_i$ "represents" $\mathbf{x}_i$, where $\mathbf{y}_i = A^T\mathbf{x}_i$.

PCA can be viewed as finding the projecting axes which minimize the reconstruction error:

$$\begin{aligned} \min \quad & \sum_i \|(\mathbf{a}^T\mathbf{x}_i)\mathbf{a} - \mathbf{x}_i\|^2 \\ \text{s.t.} \quad & \|\mathbf{a}\|^2 = 1 \end{aligned} \tag{1}$$

Let $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$. This will yield the following eigen-problem:

$$XX^T\mathbf{a} = \lambda\mathbf{a} \tag{2}$$

The solutions of $\mathbf{a}$ correspond to the eigenvectors with the largest eigenvalues.

On the other hand, PCA can also be viewed as finding the directions where data scatters most, i.e. maximize the amount of scattering as follows:

$$\begin{aligned} \max \quad & \sum_i (\mathbf{a}^T\mathbf{x}_i)^2 \\ \text{s.t.} \quad & \|\mathbf{a}\|^2 = 1 \end{aligned} \tag{3}$$

which is equivalent to finding $\mathbf{a}$ subjects to

$$\begin{aligned} \max \quad & \mathbf{a}^T \sum_i \mathbf{x}_i\mathbf{x}_i^T\mathbf{a} \\ \text{s.t.} \quad & \|\mathbf{a}\|^2 = 1 \end{aligned} \tag{4}$$

which has exactly the same solution as the eigen-problem above. This is why the matrix $XX^T = \sum_i \mathbf{x}_i\mathbf{x}_i^T$ is called scattering matrix since its largest eigenvalues measure the amount of scattering of data points.

2.2. Linear discriminant analysis

In contrast to PCA, LDA aims at finding the projecting axes on which data points with the same label are mapped together whereas data points with different labels are mapped far apart. LDA is a supervised method. To be more specific, LDA maximizes the between-class scattering and minimizes the within-class scattering.

The within-class scattering is defined in the same way as the scattering in PCA but within each class:

$$s_W = \sum_{i=1}^{c} s_i = \sum_{i=1}^{c} \left(\mathbf{a}^T \sum_{j \in D_i} \mathbf{x}_j\mathbf{x}_j^T\mathbf{a} \right) \tag{5}$$

The between-class scattering is defined as the scattering of the mean vectors of each class:

$$s_B = \mathbf{a}^T \sum_{i=1}^{c} \bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T\mathbf{a} \tag{6}$$

LDA solves the following optimization problem:

$$\max \frac{s_B}{s_W} = \frac{\mathbf{a}^T S_B\mathbf{a}}{\mathbf{a}^T S_W\mathbf{a}} \tag{7}$$

where $S_B = \sum_{i=1}^{c} \bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T$ and $S_W = \sum_{i=1}^{c}\sum_{j \in D_i} \mathbf{x}_j\mathbf{x}_j^T$ are called between-class scattering matrix and within-class scattering matrix, respectively.

This yields a generalized eigen-problem:

$$S_B\mathbf{a} = \lambda S_W\mathbf{a} \tag{8}$$

which can be solved as a conventional eigen-problem since S_W is usually nonsingular. The solution of $\mathbf{a}$ corresponds to the eigenvectors with largest eigenvalues.

3. Kernel maximum margin projection

In this section, we introduce the kernel maximum margin projection algorithm. Since our algorithm is fundamentally based on maximum margin projection [8]. We begin with a brief description of the graph construction in MMP.

3.1. Graph based semi-supervised manifold learning

We consider the case of m data points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\} \in \mathbb{R}_n$ sampled from a underlying image submanifold $\mathscr{M}$, with the first point ($\mathbf{x}_1$, query example) labeled and the rest $m-l$ points