Contents lists available at ScienceDirect

Neurocomputing

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

Margin-based two-stage supervised hashing for image retrieval

Ye Liu [a,b], Yan Pan [a,c,*], Hanjiang Lai [a,b,f], Cong Liu [a,e], Jian Yin [a,b,d,*]

[a] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
[b] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China
[c] School of Software, Sun Yat-sen University, Guangzhou, China
[d] SYSU-CMU Shunde International Joint Research Institute, Foshan, China
[e] School of Advanced Computing, Sun Yat-sen University, Guangzhou, China
[f] Department of Electronic and Computer Engineering, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
10 March 2016
Accepted 8 July 2016
Communicated by Liang Lin

Keywords:
Deep learning
Image retrieval
Image hashing
Neural network
Optimization algorithm

ABSTRACT

Similarity-preserving hashing is a widely used method for nearest neighbor search in large-scale image retrieval. Recently, supervised hashing methods are appealing in that they learn compact hash codes with fewer bits by incorporating supervised information. In this paper, we propose a new two-stage supervised hashing methods which decomposes the hash learning process into a stage of learning approximate hash codes followed by a stage of learning hash functions. In the first stage, we propose a margin-based objective to find approximate hash codes such that a pair of hash codes associating to a pair of similar (dissimilar) images has sufficiently small (large) Hamming distance. This objective results in a challenging optimization problem. We develop a coordinate descent algorithm to efficiently solve this optimization problem. In the second stage, we use convolutional neural networks to learn hash functions. We conduct extensive evaluations on several benchmark datasets with different kinds of images. The results show that the proposed margin-based hashing method has substantial improvement upon the state-of-the-art supervised or unsupervised hashing methods.

1. Introduction

Large-scale image retrieval, a task that finds images of containing a similar object or a scene as in a query image, has attracted increasing interest due to the explosive growth of available image data on the web. Approximate nearest neighbor (ANN) search has become a popular technique in image retrieval on datasets with millions or billions of images.

A notable stream of efficient ANN search methods is learning-to-hash, i.e., an approach of learning to compress data points (e.g., images) into binary representations such that semantically similar data points have nearby binary codes. The existing learning-to-hash methods can be divided into two main categories: unsupervised methods and supervised methods. Unsupervised methods (e.g., [1–3]) learn a set of hash functions from unlabeled data without any side information. Supervised methods (e.g., [4–7]) try to learn compact hash codes by leveraging supervised information on data points (e.g., similarities on pairs of images).

Among various supervised learning-to-hash methods for images, an emerging stream is the two-stage [4,7] methods, in which the hash learning process is divided into two stages: (1) learning approximate hash codes that preserves the similarities on pairs of images, (2) learning a set of hash functions from input images so that these hash functions can generate the learned approximate hash codes. Such methods are appealing in that the learning problem in their second stage is a standard multi-task binary classification problem that can be solved by off-the-shelf classifiers (e.g., kernel methods [7] or deep neural networks [4]). Particularly, with the advance of deep learning in the last few years, convolutional neural networks have made dramatically progress in image recognition and detection. The two-stage methods provide a way to boost the performance of hashing via leveraging successful deep models.

One of the main challenges in the two-stage methods is how to find approximate hash codes (each of which associates with an input image) that accurately preserves the similarities on pairs of images. The existing two-stage hashing methods (e.g., [7,4]) usually try to learn the approximate hash codes by pursuing the minimum (maximum) Hamming distance between a pair of hash codes that associates with a pair of semantically similar (dissimilar) images. More specifically, for a pair of images x and y whose

* Corresponding authors at: School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China.
E-mail addresses: panyan5@mail.sysu.edu.cn (Y. Pan), issjyin@mail.sysu.edu.cn (J. Yin).

http://dx.doi.org/10.1016/j.neucom.2016.07.024
0925-2312/

Please cite this article as: Y. Liu, et al., Margin-based two-stage supervised hashing for image retrieval, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.07.024

q-bit approximate hash codes are H_x and H_y, if x and y are semantically similar, the existing two-step methods try to make $\| H_x - H_y \|_{\mathcal{H}} \to 0$ (the minimum distance, where $\|\cdot\|_{\mathcal{H}}$ represents the Hamming distance); if x and y are dissimilar, the existing methods try to make $\| H_x - H_y \|_{\mathcal{H}} \to q$ (the maximum distance). However, pursuing minimum (maximum) Hamming distance for similar (dissimilar) pairs may be too harsh and possibly degrade the quality of hashing. In this paper, we propose to learn the approximate hash codes that satisfies a more flexible setting: for two similar (dissimilar) data points, the Hamming distance between their corresponding hash codes only needs to be sufficiently small (large), i.e., if x and y are similar, one has $\| H_x - H_y \|_{\mathcal{H}} \le r_1$; if x and y are dissimilar, one has $\| H_x - H_y \|_{\mathcal{H}} \ge r_2$, where $0 \le r_1 < r_2 \le q$. This new setting is a generalization of the setting used in [5,7,4] (with $r1 = 0$ and $r2 = q$).

Under this new setting, the corresponding objective for approximate hash code learning is difficult to optimize, due to its non-convexity and the existence of an element-wise multiplication operator. To address this issue, we propose a novel random coordinate descent algorithm to efficiently solve the corresponding optimization problem. After obtaining the approximate hash codes, we learn a set of hash functions via deep convolutional networks.

We conduct extensive evaluations of the proposed method on several benchmarks datasets with different kinds of images. The experimental results indicate that the proposed method has superior performance against several state-of-the-art supervised or unsupervised methods.

2. Related work

Due to the encouraging efficiency in search speed, hashing has become a popular ANN search method in large-scale image retrieval. Hashing methods can be divided into data-independent hashing and data dependent hashing. The early efforts mainly focus on data-independent hashing. For example, the notable Locality-Sensitive Hashing (LSH) [8] methods construct hash functions by random projections or random permutations that are independent on the data points. The main limitation of data-independent methods lies in that they usually requires long hash codes to obtain good performance. However, long hash codes lead to inefficient search due to large storage space and low recall rates.

Data-dependent hashing, a.k.a. learning-to-hash, pursues a compact binary representation with fewer bits from the training data. According to whether using side information, learning-to-hash methods can be divided into two sub-categories: unsupervised methods and supervised methods.

Unsupervised methods try to learn a set of similarity-preserving hash functions only from the unlabeled data. Here are some representative methods in this sub-category. Kernelized LSH (KLSH) [2] generalizes LSH to accommodate arbitrary kernel functions, making it possible to learn hash functions which preserve data points' similarity in a kernel space. Semantic hashing [9] generates hash functions by a deep auto-encoder via stacking multiple restricted Boltzmann machines (RBMs). Graph-based hashing methods, such as Spectral hashing [10] and Anchor Graph Hashing [3], learn non-linear mappings as hash functions which try to preserve the similarities within the data neighborhood graph. In order to reduce the quantization errors, Iterative Quantization (ITQ) [1] seeks to learn an orthogonal rotation matrix which is applied to the data matrix after principal component analysis projection.

Supervised methods aim to learn better bitwise representations by incorporating supervised information. The commonly used supervised information is in the form of pairwise labels which indicates the semantical similarity/dissimilarity between data points. Binary Reconstruction Embedding (BRE) [6] learns hash functions by explicitly minimizing the reconstruction error between the original distances of data points and the Hamming distances of the corresponding binary codes. Minimal Loss Hashing (MLH) [11] learns similarity-preserving hash codes by minimizing a hinge-like loss function which is formulated as structured prediction with latent variables. Supervised Hashing with Kernels (KSH) [5] is a kernel-based supervised method which learns to hash the data points to compact binary codes whose Hamming distances are minimized on similar pairs and maximized on dissimilar pairs. Column Generation Hash (CGHash) [12] is a column generation based method to learn hash functions with proximity comparison information. Semi-Supervised Hashing (SSH) [13] learns hash functions via minimizing similarity errors on the labeled data while simultaneously maximizing the entropy of the learnt hash codes over the unlabeled data.

In supervised hashing methods, an emerging stream is the two-stage methods [7,4] in which the learning process is decomposed into two stages: (1) learning approximate hash bits (each piece of hash bits associates with a data point in the training set) from the supervised information, and (2) learning hash functions that can generate the learned approximate hash bits from the training data. The main advantage of two-stage methods is that the learning problem in their second stage can be regarded as multi-task binary classification which can be solved by off-the-shelf classifiers. For example, Two-Step Hashing (TSH) [7] uses kernel methods to learn hash functions in its second stage; CNNH [4] uses deep convolutional networks to learn hash functions for images in its second stage. However, the existing two-stage methods [7,4] learn the approximate hash codes by minimizing (maximizing) the Hamming distance on similar (dissimilar) data points, which may be overly restrictive and degrade the hashing performance. The proposed method in this paper belongs to two-stage methods. Compared to the existing two-stage methods, the proposed method learns approximate hash codes by minimizing a margin-based objective that pursues sufficient small (large) Hamming distance on similar (dissimilar) data points. We empirically show that such a new setting can boost the quality of hashing.

3. The approach

In this paper, we follow the notations used in [4]. Given a training set of n images $\mathcal{I} = \{I_1, I_2, \ldots, I_n\}$ and its corresponding pairwise similarity matrix S defined by:

$$S_{ij} = \begin{cases} +1, & I_i,\ I_j \text{ are semantically similar} \\ -1, & I_i,\ I_j \text{ are semantically dissimilar,} \end{cases}$$

where $S_{ij} \in [-1, 1]$ represents the similarity between the ith and the jth images. The goal of supervised hashing is to learn a set of q hash functions based on S and $\mathcal{I}$. Each hash function is a mapping of an input image to $\{1, -1\}$, which accounts for generating one binary hash bit.

Some supervised hashing methods formulate the whole learning problem as a single objective, resulting in a complex and difficult optimization problem. On the other hand, the two-stage methods (e.g., [4,7]) decompose the learning process into two stages: an approximate hash code learning stage followed by a hash function learning stage, where the corresponding optimization problem in each stage is relatively simple. The proposed method belongs to the two-stage methods.

Please cite this article as: Y. Liu, et al., Margin-based two-stage supervised hashing for image retrieval, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.07.024