Information Sciences 330 (2016) 199–210

Contents lists available at ScienceDirect

Information Sciences

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

Semantic Boosting Cross-Modal Hashing for efficient multimedia retrieval

Ke Wang[a,b,1], Jun Tang[a,b,1], Nian Wang[a,*], Ling Shao[c,d]

[a] Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Hefei 230039, China
[b] School of Electronics and Information Engineering, Anhui University, Hefei 230039, China
[c] Department of Computer Science and Digital Technologies, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
[d] College of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

ARTICLE INFO

Article history:
Received 31 May 2015
Revised 8 October 2015
Accepted 17 October 2015
Available online 22 October 2015

Keywords:
Cross-modal hashing
Multimedia retrieval
Boosting

ABSTRACT

Cross-modal hashing aims to embed data from different modalities into a common low-dimensional Hamming space, which serves as an important part in cross-modal retrieval. Although many linear projection methods were proposed to map cross-modal data into a common abstract space, the semantic similarity between cross-modal data was often ignored. To address this issue, we put forward a novel cross-modal hashing method named Semantic Boosting Cross-Modal Hashing (SBCMH). To preserve the semantic similarity, we first apply multi-class logistic regression to project heterogeneous data into a semantic space, respectively. To further narrow the semantic gap between different modalities, we then use a joint boosting framework to learn hash functions, and finally transform the mapped data representations into a measurable binary subspace. Comparative experiments on two public datasets demonstrate the effectiveness of the proposed SBCMH.

1. Introduction

With the prevalence of the Internet and social networks, a great number of multimedia data are uploaded online every day. Due to the heterogeneous nature of multimedia data, cross-modal retrieval is central to many practical applications, such as finding the relevant results of another modality given a query of one modality. However, the semantic gap between multi-modal data stands as a fundamental obstacle for successful cross-modal retrieval. To narrow this gap, a feasible way is to map multi-modal data into a common feature space. Over the past few decades, there have been considerable efforts dedicated to the development of methods for learning discriminative embedding for multi-modal data [28,29]. Efficient similarity search plays an important role in building scalable multimedia retrieval systems. Therefore, numerous hashing methods are proposed to transform high-dimensional original data into compact binary hash codes. On one hand, the storage is significantly reduced as the representations of instances are highly compressed as short binary codes. On the other hand, the searching process can be efficiently conducted by comparing the Hamming distances between two hash codes via bitwise XOR operation.

Most of the existing hashing methods focus on unimodal data. One of the most well-known unimodal hashing methods is Locality Sensitive Hashing (LSH) [1,10], which randomly selects linear projections to construct hash functions. However, LSH often produces ineffective long codes due to its data independence nature [35]. In order to generate compact and discriminative

* Corresponding author. Tel.: +86 13866154858.
E-mail addresses: wangke4747@126.com (K. Wang), tangjunahu@163.com (J. Tang), wangnianahu@163.com (N. Wang), ling.shao@ieee.org (L. Shao).
[1] The authors wish to be known that, in their opinion, the first two authors should be regarded as joint first authors.

http://dx.doi.org/10.1016/j.ins.2015.10.028
0020-0255/

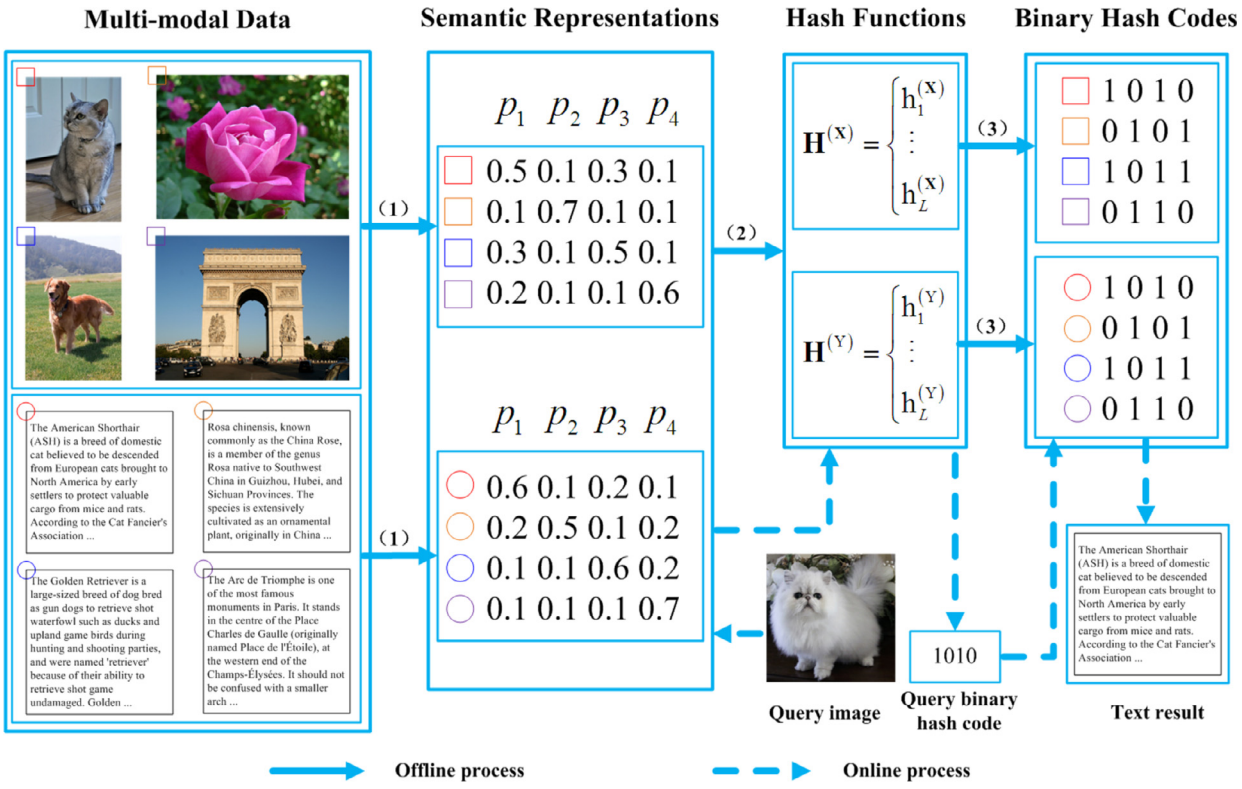

Fig. 1. Flowchart of the proposed approach.

hash codes, several learning-based hashing methods were proposed such as Nonnegative Sparse Locality Preserving Hashing [14], Spectral Hashing [27], PCA Hashing [25], Restricted Boltzmann Machines (i.e., semantic hashing) [19], Minimal Loss Hashing [17], and Kernel-Based Supervised Hashing [15].

According to the availability of prior knowledge, cross-modal hashing methods can be roughly categorized into two groups, i.e., unsupervised methods [5,22,37,39] and supervised methods [2,12,32,35,36,38]. As label information enables more discriminative hash function learning, supervised cross-modal hashing approaches can often yield better results. However, it is still a challenging problem for supervised cross-modal hashing approaches to manage the semantic similarity and the semantic gap across data from different modalities. When the semantic gap becomes large, the retrieval performance of most supervised methods often deteriorates sharply.

It is worth noting that the boosting technique is widely used to design effective hashing. In [2], Cross-Modal Similarity Sensitive Hashing (CMSSH) utilizes the boosting algorithm to learn two groups of weak binary classifiers (hash functions). In [23], Trzcinski et al. proposed a boosting framework named BinBoost to learn binary descriptors that are robust to illumination and viewpoint changes.

Inspired by the BinBoost method, we propose a supervised cross-modal hashing method, named Semantic Boosting Cross-Modal Hashing (SBCMH), to address the above-mentioned problem. The pipeline of our method is illustrated in Fig. 1. Firstly, in order to preserve the semantic similarity of heterogeneous data, we apply the multi-class logistic regression to map original data from different modalities into a semantic space, respectively. As a result, the features are represented as semantic concept probabilities. Secondly, we formulate learning binary codes as a binary classification problem. We integrate the boosting algorithm in [2] and BinBoost [23] into a joint boosting framework, which are used to learn both weak classifiers and strong classifiers. Finally, we convert the obtained semantic features into compact binary hash codes using the learned strong classifiers (hash functions). The main contributions of this paper are summarized as follows:

(1) We propose a novel supervised cross-modal hashing framework, which combines the logistic regression and the boosting algorithm together, to project high-dimensional feature vectors of multimedia data into a common low-dimensional Hamming space.
(2) We consider both the semantic similarity and the semantic gap between cross-modal data in the proposed approach, leading to highly discriminative binary codes.
(3) We conduct comprehensive experiments on two datasets to evaluate the proposed SBCMH. Experimental results verify that our approach can compete against the state-of-the-art methods.