ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

IMAGE RETRIEVAL THROUGH LOW RANK ONLINE MULTI MODAL DISTANCE METRIC LEARNING

M.HEMA, A. NISHAR FATHIMA, S.MALATHI, S.LIVIYA
GUIDED BY
D.GOKUL PRASATH (AP/CSE)

Abstract—Distance metric learning (DML) is an important technique to improve similarity search in content-based image retrieval. Despite being studied extensively, most existing DML approaches typically adopt a single-modal learning framework that learns the distance metric on either a single feature type or a combined feature space where multiple types of features are simply concatenated. Such single-modal DML methods suffer from some critical limitations: (i) some type of features may significantly dominate the others in the DML task due to diverse feature representations; and (ii) learning a distance metric on the combined high-dimensional feature space can be extremely time-consuming using the naive feature concatenation approach. To address these limitations, in this paper, we investigate a novel scheme of online multi-modal distance metric learning (OMDML), which explores a unified two-level online learning scheme: (i) it learns to optimize a distance metric on each individual feature space; and (ii) then it learns to And the optimal combination of diverse types of features. To further reduce the expensive cost of DML on high-dimensional feature space, we propose a low-rank OMDML algorithm which not only significantly reduces the computational cost but also retains highly competing or even better learning accuracy. We conduct extensive experiments to evaluate the performance of the proposed algorithms for multi-modal image retrieval, in which encouraging results validate the effectiveness of the proposed technique.

1 INTRODUCTION

ONE of the core research problems in multimedia ret- rieval is to seek an effective distance metric/function for computing similarity of two objects in content-based multimedia retrieval tasks [1], [2], [3]. Over the past decades, multimedia researchers have spent much effort in designing a variety of low-level feature representations and different distance measures [4], [5], [6]. Finding a good distance metric/function remains an open challenge for content-based multimedia retrieval tasks till now. In recent years, one promising direction to address this challenge is to explore distance metric learning (DML) [7], [8], [9] by applying machine learning techniques to optimize distance metrics from training data or side information, such as historical logs of user relevance feedback in content-based image retrieval (CBIR) systems [6], [7]. Although various DML algorithms have been proposed in literature [7], [10], [11], [12], [13], most existing DML methods in general belong to single-modal DML in that they learn a distance metric either on a single type of feature or on a combined feature space by simply concatenating multiple types of diverse features together. In a real-world application, such approaches may suffer from some practical limitations: (i) some types of features may significantly dominate the others in the DML task, weakening the ability to exploit the potential of all features; and (ii) the naive concatenation approach may result in a combined high- dimensional feature space, making the subsequent DML task computationally intensive. To overcome the above limitations, this paper investigates a novel framework of Online Multi-modal Distance Metric Learning (OMDML), which learns distance metrics from multi-modal data or multiple types of features

via an efficient and scalable online learning scheme. Unlike the above concatenation approach, the key ideas of OMDML are two- fold: (i) it learns to optimize a separate distance metric for each individual modality (i.e., each type of feature space), and (ii) it learns to and an optimal combination of diverse distance metrics on multiple modalities. Moreover, OMDML takes advantages of online learning techniques for high efficiency and scalability towards large-scale learning tasks. To further reduce the computational cost, we also propose a Low-rank Online Multi-modal DML (LOMDML) algorithm, which avoids the need of doing intensive positive semi-definite (PSD) projections and thus saves a significant amount of computational cost for DML on high-dimensional data. As a summary the major contributions of this paper include: We present a novel framework of Online Multi- modal Distance Metric Learning, which simultaneously learns optimal metrics on each individual modality and the optimal combination of the metrics from multiple modalities via efficient and scalable online learning.

_ We further propose a low-rank OMDML algorithm which by significantly reducing computational costs for high-dimensional data without PSD projection We offer theoretical analysis of the OMDML method.We conduct an extensive set of experiments to evaluate the performance of the proposed techniques for CBIR tasks using multiple types of features. The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 first gives the problem formulation, and then presents our method of online multi-modal metric learning, followed by proposing an improved low-rank algorithm. Section 4 provides theoretical analysis for the proposed algorithms, Section 5 discusses our experimental results, and finally Section 6 concludes this work.

2 RELATED WORK

Our work is related to three major groups of research: content-based image retrieval, distance metric learning, and online learning. In the following, we briefly review the closely related representative works in each group.

2.1 Content-Based Image Retrieval

With the rapid growth of digital cameras and photo sharing websites, image retrieval has become one of the most important research topics in the past decades, among which content-based image retrieval is one of key challenging problems [1], [2], [3]. The objective of CBIR is to search images by analyzing the actual contents of the image as opposed to analyzing metadata like keywords, title and author, such that extensive efforts have been done for investigating various low-level feature descriptors for image representation [14]. For example, researchers have spent many years in studying various global features for image representation, such as color features [14], edge features [14], and texture features [15]. Recent years also witness the surge of research on local feature based representation, such as the bag-of-words models [16], [17] using local feature descriptors (e.g., SIFT [18]). Conventional CBIR approaches usually choose rigid distance functions on some extracted low-level features for multimedia similarity search, such as the classical Euclidean distance or cosine similarity. Hence, recent years have witnesses a surge of active research efforts in design of various distance/similarity measures on some low-level features by exploiting machine learning techniques [19], [20], [21], among which some works focus on learning to hash for compact codes[19], [22], [23], [24], [25], and some others can be categorized into distance metric learning

that will be introduced in the next section. Our work is also related to multimodal/multitier studies, which have been widely studied on image classification and object recognition fields [26], [27], [28], [29]. However, it is usually hard to exploit these techniques directly on CBIR because (i) in general, image classes will not be given explicitly on CBIR tasks, (ii) even if classes are given, the number will be very large, (iii) image datasets tend to be much larger on CBIR than on classification tasks. We thus exclude the direct comparisons to such existing works in this paper. There are still some other open issues in CBIR studies, such as the efficiency and scalability of the retrieval process that often requires an effective indexing scheme, which are out of this paper's scope.

2.2 **Distance Metric Learning**

Distance metric learning has been extensively studied in both machine learning and multimedia retrieval communities [7], [30], [31], [32], [33], [34], [35], [36]. The essential idea is to learn an optimal metric which minimizes the distance between similar/related images and simultaneously maximizes the distance between dissimilar/unrelated images. Existing DML studies can be grouped into different categories according to different learning settings and principles. For example, in terms of different types of constraint settings, DML techniques are typically categorized into two groups. Global supervised approaches [7], [30]: to learn a metric on a global setting, e.g., all constraints will be satisfied simultaneously. Local supervised approaches[32],[33]:to learn a metric in the local sense, e.g., the given local constraints from neighboring information will be satisfied. Moreover, according to different training data forms, DML studies in machine learning typically learn metrics directly from explicit class labels [32], while DML studies in multimedia mainly learn metrics from side information, which usually can be obtained in the following two forms-T denotes the cardinality of entire triplet set. When only explicit class labels are provided, one can also construct side information by simply considering relation- ships of instances in same class as related, and relationships of instances belonging to different classes as unrelated. In our works, we focus on triple constraints. Finally, in terms of learning methodology, most existing DML studies generally employ batch learning methods which often assume the whole collection of training data must be given before the learning task and train a model from scratch, except for a few recent DML studies which begin to explore online learning techniques [37], [38]. All these works generally address single-modal DML, which is different from our focus on multi-modal DML. We also note that our work is very different from the existing multi- view DML study [26] which is concerned with regular
Classification tasks by learning a metric on training data with explicit class labels, making it difficult to be compared with our method directly. We note that our work is different from another multimodal learning study in [39] which addresses a very different problem of search-based face annotation where their multimodal learning is formulated with a batch learning task for optimizing a specific loss function tailored for search-based face annotation tasks from weakly labeled data. Finally, we note that our work is also different from some existing distance learning studies that learn nonlinear distance functions using kernel or deep learning methods [21], [35], [40]. In comparison to the linear distance metric learning methods, kernel methods usually may achieve better learning accuracy in some scenarios, but falls short in being difficult to scale up for large-scale applications due to the curse of kernelization, i.e., the learning cost increases dramatically when the number of training instances increases. Thus, our empirical study is focused on direct comparisons to the family of linear methods.

2.3 Online Learning

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Our work generally falls in the category of online learning methodology, which has been extensively studied in machine learning [41], [42]. Unlike batch learning methods that usually suffer from expensive re-training cost when new training data arrive, online learning sequentially makes a highly efficient (typically constant) update for each new training data, making it highly scalable for large-scale applications. In general, online learning operates on a sequence of data instances with time stamps. At each time step, an online learning algorithm processes an incoming example by first predicting its class label after the prediction, it receives the true class label which is then used to measure the suffered loss between the predicted label and the true label; at the end of each time step, the model is updated with the loss whenever it is nonzero. The overall objective of an online learning task is to minimize the cumulative loss over the entire sequence of received instances. In literature, a variety of algorithms have been proposed for online learning [43], [44], [45], [46], [47]. Some well- known examples include the Hedge algorithm for online prediction with expert advice [48], the Perceptron algorithm [43], the family of passive-Aggressive (PA) learning algorithms [44], and the online gradient descent (OGD) algorithms [49]. There is also some study that attempts to improve the scalability of online kernel methods, such as [50] which proposed a bounded online gradient descent for addressing online kernel-based classification tasks. In this work, we apply online learning techniques, i.e., the Hedge, PA, and online gradient descent algorithms, to tackle the multi-modal distance metric learning task for content- based image retrieval. Besides, we note that this work was partially inspired by the recent study of online multiple kernel learning which aims to address online classification tasks using multiple kernels [51]. In the following, we give a brief overview of several popular online learning algorithms.

2.3.1 Hedge Algorithms

The Hedge algorithm [48], [52] is a learning algorithm which aims to dynamically combine multiple strategies in an optimal way, i.e., making the final cumulative loss asymptomatically approach that of the best strategy. Its key idea is to maintain a dynamic weigh-distribution over the set of strategies. During the online learning process, the distribution is updated according to the performance of those strategies. Specifically, the weight of every strategy is decreased exponentially with respect to its suffered loss, making the overall strategy approaching the best strategy.

2.3.2 Passive-Aggressive Learning

As a classical well-known online learning technique, the Perceptron algorithm [43] simply updates the model by adding an incoming instance with a constant weight whenever it is misclassified. Recent years have witnessed a variety of algorithms proposed to improve Perceptron [44], [53], which usually follow the principle of maximum margin learning in order to maximize the margin of the classifier. Among them, one of the most notable approaches is the family of Passive-Aggressive learning algorithms [44], which updates the model whenever the classifier fails to produce a large margin on the incoming instance. In particular, the family of online PA learning is formulated to trade off the minimization of the distance between the target classifier and the previous classifier, and the minimization of the loss suffered by the target classier on the current instance. The PA algorithms enjoy good efficiency and scalability due to

their simple closed-form solutions. Finally, both theoretical analysis and most empirical studies demonstrate the advantages of the PA algorithms over the classical Perceptron algorithm.

2.3.3 Online Gradient Descent

Besides Perceptron and PA methods, another well-known online learning method is the family of Online Gradient Descent algorithms, which applies the family of online convex optimization techniques to optimize some particular objective function of an online learning task [49]. It enjoys solid theoretical foundation of online convex optimization, and thus works effectively in empirical applications. When the training data is abundant and computing resources are comparatively scarce, some existing studies showed that a properly designed OGD algorithm can asymptotically approach or even outperform a respective batch learning algorithm [54].

3 ONLINE MULTI-MODAL DISTANCE METRIC LEARNING

3.1 Overview

In literature, many techniques have been proposed to improve the performance of CBIR. Some existing studies have made efforts on investigating novel low-level feature descriptors in order to better represent visual content of images, while others have focused on the investigation of designing or learning effective distance/similarity measures based on some extracted low-level features. In practice, it is hard to and a single best low-level feature representation that consistently beats the others at all scenarios. Thus, it is highly desirable to explore machine learning techniques to automatically combine multiple types of diverse features and their respective distance measures. We refer to this open research problem as a multi-modal distance metric learning task, and present two new algorithms to solve it in this section. Fig. 1 illustrates the system low of the proposed multi-modal distance metric learning scheme for content based image retrieval, which consists of two phases, i.e., learning phase and retrieval phase. The goal is to learn the distance metrics in the learning phase in order to facilitate the image ranking task in the retrieval phase. We note that these two phases may operate concurrently in practice, where the learning phase may never stop by learning from endless stream training data. During the learning phase, we assume triplet training data instances arrive sequentially, which is natural for a real-world CBIR system. For example, in online relevance feedback, a user is often asked to provide feedback to indicate if a retrieved image is related or unrelated to a query; as a result, users' relevance feedback log data can be collected to generate the training data in a sequential manner for the learning task [55]. Once a triplet of images is received, we extract different low-level feature descriptors on multiple modalities from these images. After that, every distance function on a single modality can be updated by exploiting the corresponding features and label information. Simultaneously, we also learn the optimal . During the retrieval phase, when the CBIR system receives a query from users, it first applies the similar approach to extract low-level feature descriptors on multiple modalities, then employs the learned optimal distance function to rank the images.

MODULE DESCRIPTION

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

INPUT DATASET

We adopt four publicly-available image data sets in our experiments, which have been widely adopted for the benchmarks of content-based image retrieval, image classification and recognition tasks. The first test bed is the "caltech101" 2 , which has been widely adopted for object recognition and image retrieval. This dataset contains 101 object categories and 8,677 images. The second test bed is the "indoor" dataset, 3 which was used for recognizing indoor scenes. The third test bed is the "Image CLEF" dataset. The fourth tested is the "Corel" dataset, which consists of photos from COREL image CDs.

IMPLEMENTING OMDML ALGORITHM

We present an online learning algorithm to tackle the multi-modal distance metric learning task. The key challenge to online multi-modal distance metric learning tasks is to develop an efficient and scalable learning scheme that can optimize both the distance metric on each individual modality and meanwhile optimize the combinational weights of different modalities. To this end, we propose to explore an online distance metric learning algorithm, i.e., a variant of OASIS and PA, to learn the individual distance metric, and apply the well-known Hedge algorithm to learn the optimal combinational weights.

PERFORMANCE EVALUATION

We conduct an extensive set of experiments to evaluate the efficacy of the proposed algorithms for similarity search with multiple types of visual features in CBIR.

QUALITATIVE COMPARISON

Finally, to examine the qualitative retrieval performance, we randomly sample some query images from the query set, and compare the qualitative image similarity search by different algorithms. From the visual results, we can see that LOMDML generally returns more related results than the other base lines.

FUTURE WORK

Future work can be extend our framework in resolving other types of multi modal data analytics tasks beyond image retrieval.