J. Vis. Commun. Image R. 25 (2014) 1112–1117

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Multiple kernel-based multi-instance learning algorithm for image classification

Daxiang Li [a,c,*], Jing Wang [b], Xiaoqiang Zhao [a], Ying Liu [a,c], Dianwei Wang [a,c]

[a] School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
[b] Computer Graphics, Imaging and Vision (CGIV) Research Group, School of Computing and Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK
[c] Lab of Image Processing, Crime Scene Investigation Unit of Shaanxi Province, Xi'an 710121, China

ARTICLE INFO

Article history:
Received 19 July 2013
Accepted 24 March 2014
Available online 3 April 2014

Keywords:
Multi-instance learning (MIL)
Image classification
Affinity propagation (AP)
Multiple kernel learning (MKL)
Image retrieval
Visual words
Cluster analysis
Support vector machines

ABSTRACT

In this paper, a novel multi-instance learning (MIL) algorithm based on multiple-kernels (MK) framework has been proposed for image classification. This newly developed algorithm defines each image as a bag, and the low-level visual features extracted from its segmented regions as instances. This algorithm is started from constructing a "word-space" from instances based on a collection of "visual-words" generated by affinity propagation (AP) clustering method. After calculating the distance between a "visual-word" and the bag (image), a nonlinear mapping mechanism is introduced for registering each bag as a coordinate point in the "word-space". In this case, the MIL problem is transformed into a standard supervised learning problem, which allows multiple-kernels support vector machine (MKSVM) classifiers to be trained for the image categorization. Compared with many popular MIL algorithms, the proposed method, named as MKSVM-MIL, shows its satisfactorily experimental results on the COREL dataset, which highlights the robustness and effectiveness for image classification applications.

1. Introduction

Image semantic categorization is the use of image analysis and computer technology to assign an image into a pre-defined semantic category based on its dominated objects or scene type. With the increasing demand from digital image used for online and offline purpose, automatic image categorization method becomes increasingly important [1]. In order to bridge the gap between images and semantics, it is required to extract the global visual features (i.e. color, texture & shape, etc.), intermediate semantics features [2] or key-points feature [3] from images, and then combining supervised learning methods (i.e. SVM) to carry on image classifications. However, there are two major tasks should be considered during these operations: (1) Because of the existence of "semantic gap", seeking effective semantics representing model to describe image is very important. (2) Before a learning machine can perform classification, the training samples need to be accurately labeled. Unfortunately, manually collecting those training examples (and possibly further annotating, aligning, cropping, etc.) is a tedious job, which is both time consuming and error-prone [4].

One possible solution is regards every image as a bag and the low-level visual feature vectors of its segmented regions as instances, images can be labeled as positive or negative bags based on whether it contains an interesting semantic for user. Therefore, the semantic-based image categorization problem can be then transformed into a multi-instance learning (MIL) problem [5]. This is because the training samples of MIL allows for coarse labeling at image level rather than fine labeling at their region level, so the efficiency of labeling processes can be improved significantly.

Many multi-instance learning algorithms have been intensively studied during this decade, such as the Diverse Density (DD) algorithm [6], multi-label multi-instance learning (MLMIL) [7], and neural network algorithm [8]. It is difficult to list all existing MIL algorithms. Here, we mainly focus on the methods based on support vector machines (SVM), which have been successfully used in many machine-learning problems. Andrews et al. [9] first modified the SVM formulation, and presented mi-SVM and MI-SVM algorithms. However, unlike the standard SVM, they lead to non-convex optimization problems, which suffer from local minima. Therefore, Gehler et al. [10] applied deterministic annealing for solving so-called non-convex optimization problem [9]. This method is able to find better local minimum of the objective function. Gartner [11] designed kernels directly on the bags by using a standard SVM to solve MIL problem. Since instance labels are unavailable, the MI kernel defines a crude assumption that all

* Corresponding author at: School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China. Fax: +86 29 88166381.
E-mail addresses: 35108809@qq.com (D. Li), jing.wang.ac@gmail.com (J. Wang).

http://dx.doi.org/10.1016/j.jvcir.2014.03.011
1047-3203/

instances in a bag are equally important. Based on Gartner's work, Kwok et al. [12] designed marginalized multi-instance kernels by highlighting that the contribution from different instances should be different. Chen et al. [13] proposed a DD-SVM method employed Diverse Density to learn a set of instance prototypes and then mapped the bags to a new feature space based on the instance prototypes. Recently, Chen et al. [14] also devised a new algorithm called Multi-Instance Learning via Embedded Instance Selection (MILES) to solve multiple instance problems. In addition, many multiple-instance semi-supervised learning algorithms have been presented during this decade, such as MissSVM [15], MISSL [16] and LSA-MIL [17] algorithms.

Recently, converting every bag in the MIL problem into a single representation vector, and then using a standard supervised learning method to solve the MIL problem, is a kind of very effective MIL algorithms. However, most of the existing feature representation methods are not effective to describe the bags. This makes it difficult to adapt some well-known supervised learning methods for MIL problems. For example, DD-SVM [13] method must learn a collection of instance prototypes to construct a new feature space using Diverse Density (DD) function. So its representation features are very sensitive to noise and incur very high computation cost. Therefore, MILES [14] method uses all the instances from the training bags instead of the prototypes used with DD-SVM to construct a new feature space. Although MILES is less sensitive to noise and more efficient than DD-SVM, but the feature space for representing bags is of very high dimensionality because it contains too many irrelevant features, so 1-norm SVM is applied to perform standard supervised learning as it can select important features and construct classifiers simultaneously.

Inspired by DD-SVM [13] and MILES [14] methods, in this paper, we present an affinity propagation (AP) clustering [18,19] based method to convert the MIL problem into a standard supervised learning problem, then use the multiple kernels learning (MKL) method [20] to solve the MIL problem. The main contributions are summarized as follows:

- AP clustering-based feature representation method is proposed to convert the MIL problem into a standard supervised learning problem. To the best of our knowledge, this is the first inductive AP method for MIL problem. While comparing with other feature representation methods, AP does not need user to specify the number of clusters, so the proposed method has stronger robustness and adaptability.
- After the MIL problem be converted to a standard supervised learning problem, based on multiple kernel support vector machine (MKSVM), we present a new MKSVM-based MIL algorithm, named as MKSVM-MIL, for image classification problems. Based on the experiment analysis, the proposed algorithm shows its promising robustness when facing clutter classification problems. In the reference [21], Sun et al. propose a visual object detection method based on multiple-kernel, multiple-instance similarity features, where the multi-kernel is used to calculate the similarity feature vector between two instances, and obtaining the multiple-kernel similarity features (MKSF) for all images, then combined with linear or Gaussian SVM to achieve object detection. However, in our proposed MKSVM-MIL algorithm, multi-kernels is used to measure the similarity between two bags' representation vectors while train the bag-level MIL classifier. Therefore, the purpose of using multi-kernels is different between MKSVM-MIL algorithm and MKSF [21].

The rest of the paper is organized as follows: Section 2 introduces the affinity propagation (AP) clustering and the multiple kernels SVM learning; Section 3 introduces detailed description of the proposed MKSVM-MIL algorithm based on concept from "word-space", nonlinear mapping and the multiple kernels support vector machine (MKSVM). The system evaluation and experimental results on COREL data set are presented in Section 4. Section 5 concludes the paper.

2. Preliminaries

2.1. Affinity propagation (AP) clustering

AP clustering is an innovative clustering algorithm introduced by Frey and Dueck [18]. Compared with traditional clustering algorithms, AP algorithm shows many advantages [19]: (a) According to the similarity between the data, this algorithm can determine the number of clusters automatically. (b) Do not need to specify the initial cluster centers. (c) Low error rate and less time-consuming calculations. In this paper, the AP clustering has been used for grouping InstSet into many clusters and locating each cluster centroid as a "visual-word".

The mathematical model of the AP approach can be briefly described as following [18,19]. Given an N data point's dataset, x_i and x_j are two objects in it. The similarity $s(i,j)$ indicates how well s_i is suited to be the exemplar for s_j, which can be initialized as $s(i,j) = 1/\|s_i - s_j\|_2, \quad i \neq j$. If there is no heuristic knowledge, self-similarities are constant values and can be named as preference in [18] as:

$$s(n,n) = \frac{\sum_{i=1;i\neq j}^{N} s(i,n)}{N-1}, \quad n = 1,2,\ldots,N \tag{1}$$

The AP approach computes two kinds of messages exchanged between data points. The first one is called "responsibility" $r(i,j)$: It is sent from data point i to candidate exemplar point j which reflects the accumulated evidence for how well-suited point j is to serve as the exemplar for point i. The second message is called "availability" $a(i,j)$: It is sent from candidate exemplar point j to point i reflecting the accumulated evidence for how appropriate it would be for point i to choose point j as its exemplar. At the beginning, the availabilities are initialized to zero: $a(i,j) = 0$. The update equations for $r(i,j)$ and $a(i,j)$ can be written as:

$$r(i,j) = s(i,j) - \max_{j'\neq j}\{a(i,j') + s(i,j')\} \tag{2}$$

$$a(i,j) = \begin{cases} \min\{0, r(j,j) + \sum \max\{0, r(i',j)\}\} & i \neq j \\ \sum_{i'\neq j} \max\{0, r(i',j)\} & i = j \end{cases} \tag{3}$$

In addition, during each message's exchange between data points, a damping factor $\lambda \in [0,1]$ is added to avoid numerical oscillations that may arise in some circumstances:

$$\begin{aligned} R_{t+1} &= (1-\lambda)R_t + \lambda R_{t-1} \\ A_{t+1} &= (1-\lambda)A_t + \lambda A_{t-1} \end{aligned} \tag{4}$$

where $R = [r(i,j)]$ and $A = [a(i,j)]$ represents the responsibility matrix and availability matrix, respectively. t indicates the iteration times.

The above two messages are updated iteratively, until they reach some specified values or the local decisions stay constant after iterations. At this stage, availabilities and responsibilities can be combined to identify exemplars:

$$c_i \Leftarrow \arg\max_{1\leq j\leq N}\left[r(r,j) + a(i,j)\right] \tag{5}$$