

Bayesian Multi-instance Learning for Image Retrieval with Unlabeled Data

Tao CHEN

School of Computer Science & Engineering,
South China University of Technology
Guangzhou, China
ccthentao@tom.com

Huifang DENG

School of Computer Science & Engineering,
South China University of Technology
Guangzhou, China
hdeng2008@gmail.com

Abstract—To deal with the two problems in image retrieval, i.e., the small number of query images, the ambiguity of an image - the image consists of many regions with different semantic meaning, in this paper, we proposed a novel method for image retrieval based on Bayesian multi-instance learning using unlabeled data, termed as Bayesian-MIL method, which treats the image retrieval as a binary classification problem. In this method, to obtain an approximate estimation of the class-conditional probability of positive images, a multi-instance learning algorithm is adopted to filter out background regions in positive images, and then a Bayesian classifier is constructed to rank the images from a large digital repository according to their score of posterior probability. Finally, the ranking top k images will be returned to users. Experimental results on COREL image data set have demonstrated the effectiveness and efficiency of the proposed approach.

Keywords- content-based image retrieval; Bayesian classifier; multi-instance learning; machine learning

I. INTRODUCTION

Over the past few decades, with the explosive growth of digital images, the content-based image retrieval (CBIR) technique has attracted great research attention [1]. Rather than the text retrieval, in the CBIR systems, a query is composed of one or a set of example images where the images are represented with visual features [2], such as color, texture, and shape, and users receive many similar images from an image repository.

One of the most popular supervised learning models used in the machine learning domain is the Bayesian classifier, which has been applied to dealing with various real problems. However, the Bayesian classifier has two main drawbacks when used for image retrieval. First, it often suffers from learning with a small number of labeled examples, as in the case of query images. Since few users will be so patient to label a large number of images, the number of training images is limited. While the rich image repository is indeed available during learning, which is helpful in the retrieval process. Therefore, it is natural to consider using the large amount of unlabeled data to improve retrieval accuracy. Second, under Bayesian learning framework, the ambiguity of image has impact on the approximate estimation of the class-conditional probability of positive images. Specially, an image is composed of several regions, and each region has its own semantic content. In the process of image retrieval, user is usually given only query images, without indicating which portion of the images being of interest. If

the class-conditional probability of positive images is estimated in the collection consisting of all regions in positive images, due to the existence of background regions, the estimation may be inaccurate. Naturally, the performance of the Bayesian classifier will be influenced.

In order to overcome these problems, we design a new scheme for image retrieval based Bayesian classifier and multi-instance learning using the unlabeled data, termed as Bayesian-MIL. The approach has two features as follows: First, it solves the problem of small training size by utilizing the information of unlabeled images for approximately estimating the class-conditional probability of positive images; Second, it addresses the issue of the image ambiguity by adopting multi-instance learning algorithm to filter out the background regions in positive images, which makes the approximation of the class-conditional probability more accurately. Our experimental study on COREL image dataset shows that this new method is effective and efficient.

The rest of this paper is organized as follows: The related works are discussed in Section 2; the problem formulation and our solution in more detail are given in Section 3; the experimental evaluations are conducted in Section 4; and the conclusions and future works are given in Section 5.

II. RELATED LEARNING METHODS

It is well known that the Bayesian classifier is optimal method in minimizing the classification error probability. Cox and Miller et al.[3] proposed a PicHunter framework which used a weighted pairwise distance measure to model the similarity between images under the Bayesian rule, with weights chosen by maximum likelihood. Yavlinsky et al.[4] described a system for both retrieval and annotation of images where the class density was modeled using a non-parameter kernel density estimator, with the Earth Mover's Distance being used as the kernel. Vasconcelos [5] adopted the minimum probability of error as the optimality criterion and retrieval was formulated as a problem of statistical classification, where the feature distribution in each class was modeled using a Gaussian mixture projected down to a low dimensional space. The parameters of model were estimated by using the expectation maximization algorithm. Gosselin and Cord [6] introduced an active learning strategy to select the most difficult images to classify with only few training data. Hoi and Lyu [7] proposed a semi-supervised active learning framework comprising a fusion of support vector machines and semi-supervised learning with Gaussian fields and harmonic functions.

Multi-instance learning (MIL) was first presented in drug activity prediction by Dietterich et al. [8], which was a type of learning algorithms to tackle the problems with coarsely labeled information on bags. In the framework of MIL, each sample is called a bag which contains several instances, label information is assigned in bag level, and instances do not have label. A bag is labeled positive if at least one instance in it is positive, or labeled negative if and only if all of its instances are negative [9]. For being able to identify ambiguous object effectively, MIL has been successfully introduced into the field of image retrieval [10]. Maron and Lozano-perez [11] proposed the diverse density (DD) algorithm, which searched a point with maximum of diverse density function (DD) in instance feature space. Zhang and Goldman [12] developed an algorithm which combined the idea of expectation-maximization with DD algorithm (EM-DD). Andrews et al. [13] modified the SVM formulation, proposed two algorithms: mi-SVM and MI-SVM. Chen and Wang proposed two algorithms: DD-SVM framework [14] using instance prototypes and MILES algorithm [15] in which a bag is embedded in a feature space defined by the instances in all the training bags.

To exploit useful information from unlabeled images, Rouhollah R. et al. [16] combined DD with graph-based semi-supervised algorithm, proposed MISSL algorithm and applied it to image retrieval. Wang et al. [17] presented a graph-based multi-instance semi-supervised learning (GMIL) for object-based image retrieval, which generated the instance-level graph by selecting instance prototypes.

III. THE PROPOSED APPROACH (BAYESIAN-MIL)

In this section, we formulate the proposed Bayesian-MIL approach, which integrates MIL into the framework of Bayesian classification. The ranking criterion is the posterior probability score corresponding to the positive class for each unlabeled image.

Given a set of query images consisting of $L = \{(I_1, y_1), \dots, (I_l, y_l)\}$, and u unlabeled images from an digital repository $U = \{I_{l+1}, \dots, I_n\}$, where $l = |L|$ is the number of query images, $u = |U| = n - l$ is the number of unlabeled images; $y_i \in \{-1, 1\}$ is the class label of image I_i , specifically, according to the user's preference, the query images are partitioned into two classes: either positive images or negative images. If the image I_i has a region whose semantic meaning is similar to the common semantic concepts in the positive images, then y_i is set to 1, otherwise it is set to 0. Let $I_i = \{r_{ij} \mid j = 1, \dots, n_i\}$, where $r_{ij} \in \mathbb{R}^d$ represents a region in the image I_i as a d -dimensional vector, n_i is the number of regions in the image I_i . Our goal is to get a similarity

function $f(I_i \in U)$, which can generate a probability score f_i corresponding to the positive class for every unlabeled image I_i . Finally, the top k images are returned to user according to the probability score.

A. Bayesian learning for image retrieval

Under the framework of Bayesian learning, the similarity score f_i is converted to the posterior probability $p(c_p | I_i)$ of an image I_i with respect to the positive class. From the Bayesian rule, the posterior probability $p(c_p | I_i)$ can be calculated based on the class-conditional probability $p(I_i | c_p)$ and prior probability $p(c_p)$ by:

$$p(c_p | I_i) = \frac{p(I_i | c_p)p(c_p)}{p(I_i)}, \quad (1)$$

Where $p(I_i)$ is the probability density function of image I_i , which is the same for each image and does not affect the decision.

In our method, an image is divided into many regions, and each region is represented with a 36-dimensional feature vector [18]. There are two methods to compute the class-conditional probability according to the contribution of regions relative to the image. One is the mean region contribution:

$$p(I_i | c_p) \propto \frac{1}{m} \sum_{j=1}^m p(r_{ij} | c_p) \quad (2)$$

Another is that the region with maximum of posteriori probability represents the image:

$$p(I_i | c_p) \propto \max_j p(r_{ij} | c_p) \quad (3)$$

From equations (1), (2) and (3), we have

$$p(c_p | I_i) \propto \frac{1}{m} \sum_{j=1}^m p(r_{ij} | c_p)p(c_p) \text{ or } p(c_p | I_i) \propto \max_j p(r_{ij} | c_p)p(c_p),$$

Where the prior probability $p(c_p)$ is computed by $p(c_p) = |c_p| / |c|$, $|c_p|$ is the number of the positive images, $|c|$ is the number of all query images.

We use a Gaussian function to approximate the positive class distribution:

$$p(x | c_p) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{(x - \mu)^T \Sigma (x - \mu)}{2}\right) \quad (4)$$

where feature vector x is a region in positive images, d is its dimensionality, $\mu = E[X]$ is the mean value vector,

$\Sigma = E[(X - \mu)(X - \mu)^T]$ is the $d \times d$ covariance matrix, $|\Sigma|$ denotes the determinant of Σ , X is all regions in positive images.

However, the difficulty arises here from the fact that positive images may have dozens of regions, but maybe only one of which really interests to us when estimating the positive class-conditional probability. For example, the images with the semantic meaning of the “tiger”, which may contain some background regions corresponding to the concepts “grass”, “river”, or “bush”. The low-level features of these background regions are completely different from “tiger” regions. Thus the Gaussian function in Equation (4) usually can not approximate the positive class-conditional distribution. Moreover, the small size of query examples can also increase the difficulty in the estimation of class conditional probability. Therefore, it is necessary to take into consideration the useful information from unlabeled images and reducing the background regions in positive images.

B. Use multi-instance learning to filter out the background regions in positive images with the unlabeled data

In the framework of MIL, an image I_i is seen as a bag B_i , a region r_{ij} is seen as an instance B_{ij} in the bag. The label information is assigned in bag level, and instances do not have label. A bag is labeled positive if at least one instance in it is positive, or labeled negative if and only if all of its instances are negative. The diverse density (DD) approach [11] is one of MIL algorithms, which searches for a point with maximum of diverse density function (DD) in the instance feature space. The DD function value of a point is defined as a measure of how many different positive bags containing instances near this point, and meanwhile how far the negative instances being from this point. The point t with maximum DD value can be determined by the following formula:

$$DD(t|L) = \arg \max_t \prod_i \Pr(t|B_i^+) \prod_i \Pr(t|B_i^-) \quad (5)$$

Where L is the training set, B_i^+ denotes the i -th positive bag. Using the “noisy-or” assumption, the product in (5) can be transformed to:

$$\Pr(t|B_i^+) = \Pr(t|B_{i1}^+, B_{i2}^+, \dots, B_{im}^+) = 1 - \prod_j (1 - \Pr(t|B_{ij}^+)) \quad (6)$$

$$\Pr(t|B_i^-) = \Pr(t|B_{i1}^-, B_{i2}^-, \dots, B_{nm}^-) = \prod_j (1 - \Pr(t|B_{ij}^-)) \quad (7)$$

The causal probability of the instance appearing at the potential targets is computed by

$$\Pr(t|B_{ij}) = \exp(-\|B_{ij} - t\|^2).$$

The DD algorithm uses a gradient ascent method to find the point with maximum of DD value. In order to avoid falling into the local optimal and guarantee finding the global optimum, the starting points are from every instance

from all positive bags. The procedure of maximization is often very time-consuming.

The difference between our proposed method and DD algorithm lies in two aspects: first, our approach does not need to search for the maximum of DD value in the instance feature space by using the gradient ascent method; second, the unlabeled data take part in the learning process.

With the observation that the DD value of positive instance is large, and the DD value of negative instance is small, intuitively, we treat the instances with large DD value as the positive instances. Since there have some positive images in unlabeled images, intuitively, we add them into the process of learning.

For each instance $t \in L^+ + U$, its DD value can be computed by:

$$DD(t) = \prod_i \Pr(t|B_i^+) \prod_i \Pr(t|B_i^-)$$

However, how large a value is regarded as a positive instance and how small a value as a negative sample requires a threshold to determine. According to the assumption in MIL that each positive bag at least has a positive instance, we set the decision threshold DD_{thre} to be the maximum of satisfying that one instance at least is larger than or equal to it.

$DD_{thre}^* = \max\{DD^* | \forall B_i \in L^+, s.t. \exists DD_{B_{ij}} \geq DD^*\}$, where $DD_{B_{ij}}$ denotes the DD value of instance B_{ij} in bag B_i .

Once the DD values of all instances $t \in L^+ + U$ are obtained, we select the instances satisfying $DD_{B_{ij}} \geq DD_{thre}^*$ as the positive instance to group into a collection, and estimate the positive class-conditional probability in the instance collection.

C. The implementation of Bayesian-MIL method

The flowchart of the Bayesian-MIL method is described in Figure 1.

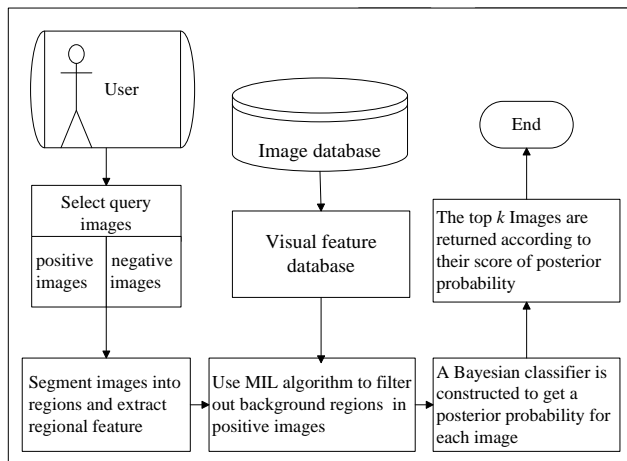


Figure 1: The flowchart of the Bayesian-MIL method

The procedure for implementing Bayesian-MIL is as below:

Input: A set of query images and u unlabeled images $L = \{(I_1, y_1), \dots, (I_{|L|}, y_{|L|})\}$, $U = \{I_{|L|+1}, \dots, I_n\}$

Output: The ranking top k images relevant to the positive images

1. Users select a set of relevant images and irrelevant images as query images;
2. Segment each image into several regions and extract out a 36-dimensional feature vector for each region, treat image as bag and regions as instances in the bag, respectively;
3. Use a multi-instance learning algorithm to filter out the background regions of positive images, group these positive regions into a collection, and estimate the positive class-conditional probability in the region collection.
4. A Bayesian classifier is constructed to get a posterior probability corresponding to positive class for each image from digital repository. Their score of posterior probability are as the ranking criterion, and the top k images are returned to user as the retrieval results.

IV. EXPERIMENTAL RESULTS

A. Dataset and feature extraction

The experiments are conducted on COREL 2000 image data set consisting of 2000 images, which has 20 categories representing a distinct concept, each of which contains 100 images. The format of Images is of size 384×256 or 256×384 in JPEG. Images are segmented using normalized cuts [18], where each image is divided into 5-10 regions, and each region is characterized by a 36-dimensional vector including the color, texture, and shape properties of region.

B. Performance Comparison

Experiment 1: To investigate the performance of the Bayesian-MIL method, we make a comparison with two related algorithms: naïve Bayesian classifier, DD method

[11]. The Bayesian-MIL method has two versions: Bayesian-MIL-mean, Bayesian-MIL-max. Bayesian-MIL-mean denotes that an image is represented with the mean region contribution. Bayesian-MIL-max denotes that an image is represented with the region with maximum of posteriori probability. The precision-recall curve is used to appraise the effectiveness of the algorithm performance. Precision is represented as the percent of the relevant images returned to all images returned. The average precision is the mean value of precision on 20 kinds of images. Recall is the ratio of the number of relevant images returned to the number of all relevant images in the image datasets. The comparison of precision-recall curve between the Bayesian-MIL and the other methods is presented in Figure 3.

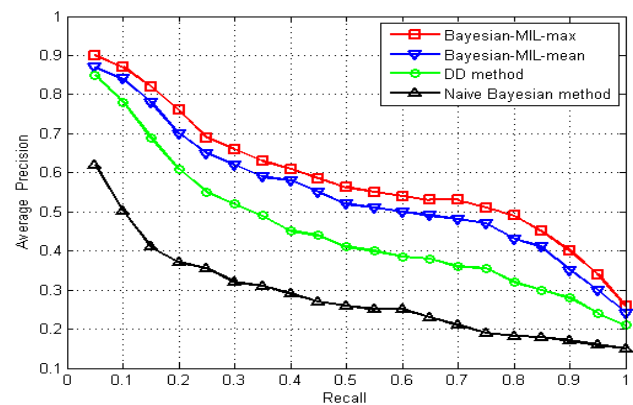


Figure 2: Precision-recall curves of the Bayesian-MIL, the DD method, and the Naïve Bayesian method

As can be seen from Figure 2, the precision-recall curves of two versions of Bayesian-MIL method are significantly better than the other two approaches. In the naïve Bayesian classifier, the positive class-conditional probability is estimated based on the collection consisting of all regions of the positive images. Due to the existence of background regions and small size of query images, the positive class-conditional probability is difficult to be approximated closely. In the DD approach, these images are sorted by using a distance similarity between images and the concept with maximum DD value, which is sensitive to query images. While the Bayesian-MIL takes into consideration the available information of unlabeled images, and uses MIL to filter out background regions, which makes the Gaussian function more approximately closer to the estimation of positive class-conditional probability. Its ranking criterion is the posterior probability score corresponding to the positive class for each unlabeled image. This makes the ranking score be insensitive to query images. Another observation is its two versions; the performance of Bayesian-MIL-max is slightly superior to the Bayesian-MIL-mean.

Experiment 2: The computing time is used to evaluate the efficiency of the algorithm performance. It is the time

that was taken from the user's requesting for retrieval images to the first return of images. The experiments are conducted in the computer of CPU 2.2GHz and 2GB memory. The computing times of three approaches are listed in Table 1.

Table 1: The computing time (in seconds)

Naïve Bayesian	2.6
DD method	10.4
Bayesian-MIL-max	3.7

From Table 1 we can see that Bayesian-MIL-max method is slower than the naïve Bayesian classifier, this is because Bayesian-MIL-max has an extra overhead cost on using MIL to filter out the background regions in positive class. Albeit at the expense of computing time, better retrieval results are obtained in Bayesian-MIL-max. Compared with DD method, it is nearly three times faster than DD method. This is because a maximization procedure is involved in DD algorithm using a gradient ascent method for every instance in positive bags as starting search points, which consumes a lot of time.

V. CONCLUSIONS AND FUTURE WORKS

In this paper we proposed a novel image retrieval method based on Bayesian multi-instance learning using unlabeled data, termed as Bayesian-MIL, which transforms image retrieval to a binary classification problem. To accurately approximate the positive class-conditional probability, the Bayesian-MIL exploits the available information of unlabeled data, and takes the advantages of MIL's capability of dealing with ambiguous object. The experiment results on the COREL image data set have demonstrated the effectiveness and efficiency of Bayesian-MIL. An extensional direction is introducing the relevance feedback technique to further improve the retrieval accuracy.

REFERENCES

- [1] Lew Michael S., sebe Nicu, Djeraba Chabane, et al., Content-based multimedia information retrieval: State of the art and challenges, ACM Transactions on Multimedia Computing, Communications and Applications, 2006, 2(1): 1-19
- [2] Zhang Dan, Wang Fei, et al., Localized content based image retrieval by multiple instance active learning, Proceedings - International Conference on Image Processing, ICIP, 2008: 921-924.
- [3] Cox I.J., Miller M.L., Minka T.P., et al. The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments, IEEE Transactions on Image Processing, 2000, 9(3): 20-37
- [4] Yavlinsky A., Schofield E. Rüger S., automated image annotation using global features and robust nonparametric density estimation, the Proceeding of Image and Video Retrieval: 4th International Conference, 2005, 3568: 507-517
- [5] Vasconcelos N., Minimum probability of error image retrieval, IEEE Transactions on Signal Processing, 2004, 52(8): 2322-2336
- [6] Gosselin P.H., Cord M., A comparison of active classification methods for content-based image retrieval, Proceedings of the 1st International Workshop on Computer Vision Meets Databases, CVDB 2004, 66:51-58.
- [7] Hoi Steven C.H., Lyu Michael R., A semi-supervised active learning framework for image retrieval, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005,2:302-309
- [8] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple-instance problem with axis-parallel rectangles. Artificial Intelligence, 1997, 89(1-2): 31-71.
- [9] Foulds James, Frank Eibe, A review of multi-instance learning assumptions, Knowledge Engineering Review, 2010,25(1):1-25.
- [10] Gondra Iker, Xu Tao, A multiple instance learning based framework for semantic image segmentation. Multimedia Tools and Applications, 2010, 48(2), 339-365.
- [11] Maron O, Lozano-Pérez T., A framework for multiple-instance learning. Advances in Neural Information Processing Systems, Cambridge, MA: MIT Press, 1998, 10: 570-576.
- [12] Qi Zhang, Goldman S.A., EM-DD: An improved multiple-instance learning technique. Advances in Neural Information Processing Systems, 2002, 14(1):1073-1080.
- [13] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple instance learning. Advances in Neural Information Processing Systems, Cambridge, MA: MIT Press, 2003, 15:561-568.
- [14] Chen YX, Wang JZ, Image categorization by learning and reasoning with regions. Journal of Machine Learning Research, 2004, 5: 913-939.
- [15] Chen YX, Bi J, Wang JZ. MILES: Multiple-instance learning via embedded instance selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12):1931-1947.
- [16] Changhu Wang, Lei Zhang, Hongjiang Zhang, Graph-based multiple-instance learning for object-based image retrieval, Proceedings of the 1st International ACM Conference on Multimedia Information Retrieval, MIR 2008, 156-163.
- [17] Rouhollah R., Sally A., MISSL: Multiple-instance semi-supervised learning. Proceedings of the 23rd International Conference on Machine Learning, ICML 2006:705-712.
- [18] Jianbo Shi, Malik J., Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.