A MULTIPLE INSTANCE LEARNING APPROACH FOR CONTENT BASED IMAGE RETRIEVAL USING ONE-CLASS SUPPORT VECTOR MACHINE

Chengcui Zhang[1], *Xin Chen*[1], *Min Chen*[2], *Shu-Ching Chen*[2], *Mei-Ling Shyu*[3]

[1]Department of Computer and Information Sciences, University of Alabama at Birmingham
[2]School of Computer Science, Florida International University
[3]Department of Electrical and Computer Engineering, University of Miami

ABSTRACT

Multiple Instance Learning (MIL) is a special kind of supervised learning problem that has been studied actively in recent years. In this paper, we propose an approach based on One-Class Support Vector Machine (SVM) to solve MIL problem in the region-based Content Based Image Retrieval (CBIR). Relevance Feedback technique is incorporated to provide progressive guidance to the learning process. Performance is evaluated and the effectiveness of our retrieval algorithm has been shown through comparative studies.

1. INTRODUCTION

Relevance feedback (RF) technique is used to incorporate user's concept with the learning process [2] [4] for Content-Based Image Retrieval (CBIR). The existing RF-based approaches consider each image as a whole. However, user's query interest is often just one part of the query image. Therefore it is more reasonable to view it as a set of semantic regions. In this context, the goal of image retrieval is to find the semantic region(s) of user's interest. Since each image is composed of several regions and each region can be regarded as an instance, CBIR is then transformed into a Multiple Instance Learning (MIL) problem [5]. Maron et al. applied MIL into natural scene image classification [5]. Each image is viewed as a bag of semantic regions (instances). In the scenario of MIL, the labels of individual instances in the training data are not available, instead the bags are labeled. When applied to RF-based CBIR, this corresponds to the scenario that the user gives feedback on the whole image (bag) although only a specific region (instance) is of user's interest. The goal of MIL is to generate labels for unseen bags (images) based on the user's interest on a specific region.

The key idea of our algorithm is to apply MIL to learn the region of interest from users' relevance feedback on the whole image. The proposed learning algorithm concentrates on those positive bags (images) and uses the learned region-of-interest to evaluate all the other images in the image database. The motivation comes from the fact that positive samples are all alike, while negative samples are each bad in their own way. It makes more sense to assume that all positive regions are in one class while the negative regions are outliers of the positive class. Therefore, we applied One-Class Support Vector Machine (SVM) [1] to solve the MIL problem in CBIR. Chen et al. [6] uses One-Class SVM in image retrieval but, again, it is applied to the image as a whole. In our approach, One-Class SVM is used to model the non-linear distribution of image regions and separate positive regions from negative ones. Each semantic region of the test images is given a score by the evaluation function built from the model. The images with the highest scores are returned to the user as query results. Experiments show that high retrieval accuracy can be achieved usually within 4 iterations.

In Section 2, the detailed learning and retrieval approach will be presented. In Section 3, the overall system is illustrated and the experimental results are presented. Section 4 concludes the paper.

2. THE PROPOSED LEARNING APPROACH

In this study, we assume that user is only interested in one semantic region of the query image. The goal is to retrieve those images that contain similar semantic regions. In the proposed CBIR system, we adopted the automatic image segmentation method as proposed in the Blobworld [3] to segment each image into a set of regions.

After segmentation, 8 global features (three texture features, three color features and two shape features [3]) for each semantic region are extracted. Hence, each image region is represented by an 8 dimensional feature vector and serves as an instance in Multiple Instance Learning.

2.1 Multiple Instance Learning with Relevance Feedback

In Multiple Instance Learning (MIL), the label of an instance, i.e. object, is unknown. Only the label of a set of instances is known, which is called the label of the bag. MIL problem needs to map an instance to its label according to the information learned from the bag labels.

0-7803-9332-5/05/$20.00 ©2005 IEEE

In CBIR, we have two types of labels – *Positive* and *Negative*. Each image is considered a bag of semantic regions (instances). User labels an image positive if it contains the region of interest; otherwise, it is labeled negative. The goal of MIL is to learn the label of each semantic region in the training set and use this information to estimate the similarity scores of the test image regions. In this way, the single object based CBIR problem can then be transformed to a MIL problem as defined below.

Definition 1. *Given a set of training examples T=<B,L> where* $B=B_i(i=1,\dots,n)$ *is a set of n bags and* $L=L_i(i=1,\dots,n)$ *is a set of labels of the corresponding bags.* $L_i \in \{1(Positive), 0(Negative)\}$ *The goal of MIL is to identify the label of a given instance in a given bag.*

The relation between a bag (image) label and the labels of all its instances (regions) is defined as below.

$$L_i = 1 \quad if \quad \exists_{j=1}^{m} l_{ij} = 1 \tag{1}$$

$$L_i = 0 \quad if \quad \forall_{j=1}^{m} l_{ij} = 0 \tag{2}$$

Suppose there are m instances in B_i. l_{ij} is the label of the j^{th} instance in the i^{th} bag. If the bag label is positive, there exists at least one positive instance in that bag. If the bag label is negative, all instances in that bag are negative.

Given a query image, user's feedbacks on the whole images in the training set are fed into our learning algorithm. In this study, the One-Class SVM is adopted as the underlying learning algorithm.

2.2 One-Class SVM

One-class classification is a kind of unsupervised learning mechanism, which trains on unlabelled data, trying to assess whether a test point is likely to belong to the distribution underlying the training data. It has so far been studied in the context of SVMs [1].

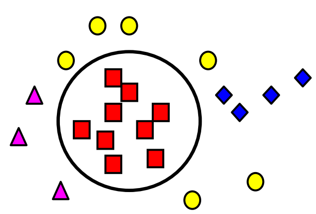

Figure 1. One-Class Classification

The idea is to model the dense region as a ball. In MIL problem, positive instances are inside the ball and negative instances are outside. If the center of the ball is $\vec{\alpha}$ and the radius is r, a point $\vec{x_i}$, in this case an instance represented by an 8-feature vector, is inside the ball *iff* $\left\| \vec{x_i} - \vec{\alpha} \right\| \le r$. This is shown in Figure 1 with red rectangles inside the circle being the positive instances. This "ball" is actually a hyper-sphere. The goal is to keep this hyper-sphere as "pure" as possible and include most of the positive objects. Since this is a non-linear distribution, the strategy of Schölkopf's one-class SVM is first to do a mapping θ to transform the data into the feature space F corresponding to the kernel k.

$$\theta(u) \cdot \theta(v) \equiv k(u,v) \tag{3}$$

where u and v are two data points. In this study, we choose to use Radial Basis Function (RBF) Machine as below.

$$k(u,v) = \exp\left(\frac{\|u - v\|^2}{2\sigma} \right) \tag{4}$$

Mathematically, one-class SVM solves the following quadratic problem:

$$\min_{w,\xi,\rho} \frac{1}{2}\|w\| - \alpha\rho + \frac{1}{n}\sum_{i=1}^{n} \xi_i \tag{5}$$

subject to

$$(w \cdot \theta(x_i)) \ge \rho - \xi_i, \quad \xi_i \ge 0 \text{ and } i = 1,\dots,n \tag{6}$$

where ξ_i is the slack variable, and $\alpha \in (0,1)$ is a parameter that controls the trade off between maximizing the distance from the origin and containing most of the data in the region created by the hyper-sphere and corresponds to the ratio of "outliers" in the training dataset. When it is applied to the MIL problem, Equation (5) is also subject to Equations (1) and (2).

If w and ρ are a solution to this problem, then the decision function is $f(x) = sign(w \cdot \theta(x) - \rho)$ and it will be 1 for most examples x_i contained in the training set.

2.3 Learning and Retrieval Process

In initial query, the user needs to identify a semantic region that he/she is interested in. Since no training sample is available at this point, we simply compute the Euclidean distances between the query semantic region and all other semantic regions. The similarity score for each image is set to the inverse of the minimum distance between its regions and the query region. We then construct the training sample set. If an image is labeled positive, its semantic region that is the least distant from the query region is labeled positive. For some images, Blob-world may "over-segment" such that one semantic region is segmented into two or more "blobs". Therefore, we cannot assume that only one region in each image is positive. Suppose the number of positive images is h and the number of all semantic regions in the training set is H. Then the ratio of "outliers" in the training set is set to:

$$\alpha = 1 - \left(\frac{h}{H} + z\right) \tag{7}$$

z is a small number used to adjust the α in order to alleviate the above mentioned problem. Our experiment results show that z= 0.01 is a reasonable value.

The training set as well as the parameter α are fed into one-class SVM to obtain w and ρ, which are used to compute the decision function for the test data. Each image region will be assigned a "score" by $w \cdot \theta(x) - \rho$ in the decision function. The similarity score of each image is then set to the highest score of all its regions.

3. EXPERIMENT

3.1 CBIR System Construction

Figure 2 shows the architecture of our system.

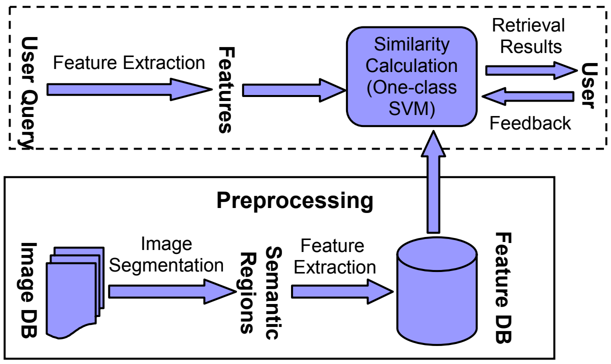

Figure 2. CBIR System Architecture

Figure 3 shows the initial query interface. The leftmost image is the query image. This image is segmented into 8 semantic regions (outlined by red lines). User identifies the "horse" region as the region of interest.

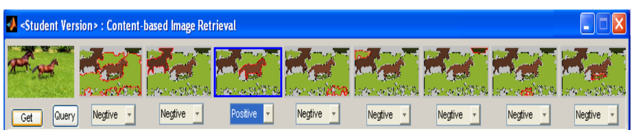

Figure 3. Initial Query Interface

After the initial query, user gives feedback to the retrieved images. One-Class SVM based algorithm learns from these feedbacks and starts another round of retrieval.

3.2 System Performance Evaluation

The experiment is conducted on a Corel image database of 9,800 images. We randomly choose 65 query images of 22 categories and compare our retrieval results with two other RF algorithms: 1) Neural Network based MIL algorithm [7]; 2) General feature re-weighting [2] algorithm.

Five rounds of relevance feedback are performed for each query image - Initial (no feedback), First, Second, Third, and Fourth. The accuracy rates, i.e. the percentage of positive images within the top 6, 12, 18, 24 and 30 images, are calculated. Figure 4 shows the result from the First Query and Figure 5 shows the result after the Fourth Query. "BP" is the Neural Network based MIL. "RF_E" is feature re-weighting with Euclidean Distance while "RF_M" uses Manhattan Distance, and "SVM" refers to the proposed algorithm.

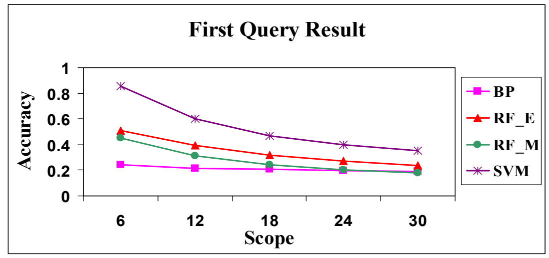

Figure 4. Retrieval Accuracy after the First Query

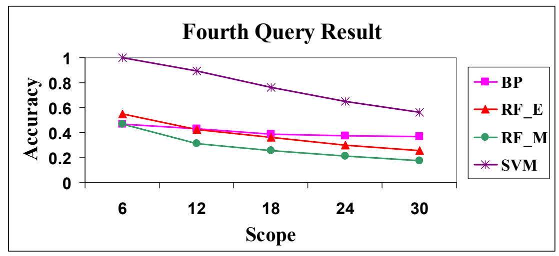

Figure 5. Retrieval Accuracy after the Fourth Query

It can be seen that the accuracy of our algorithm is superior to the others. We further compare these two by examining the exact image regions learned. Figures 6(a) and 7(a) show the Third Query results of "SVM" and "BP", respectively, given the query image and query region as in Figure 3. Figure 6(b) and 7(b) are the corresponding regions learned by the two algorithms, respectively. It can be seen that in this example, although Neural Network based method seems to successfully find several horse images, the regions it retrieved are actually the "grassland" rather than "horse". Consequently, the horse images in Figure 7(a) will be labeled positive by the user. This will definitely affect the next round of learning.

Figure 6(a). Third Query Results by One-Class SVM

Figure 6(b). Retrieved Regions of Figure 6(a)

Figure 7(a). Third Query Results by "BP"

Figure 7(b). Retrieved Regions of Figure 7(a)

We can see from Figure 8 below that accuracy rate of our algorithm across 5 iterations increases steadily.

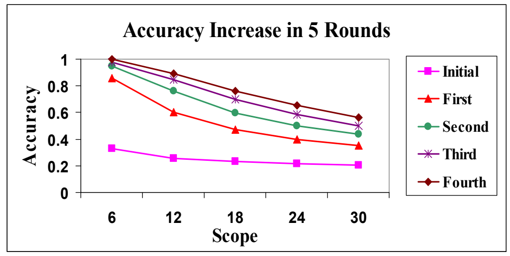

Figure 8. Retrieval Results of SVM across 5 Iterations

4. CONCLUSION

In this paper, we proposed a One-Class SVM based MIL framework for single object based CBIR systems. The advantage of our algorithm is that it is based on semantic region instead of the whole image, which is more reasonable since the user is often interested in only one region in the image. Our algorithm also transfers the One-Class SVM into MIL. Due to the robustness of One-Class SVM, the proposed system can better identify user's need.

5. ACKNOWLEDGEMENT

The work of Chengcui Zhang was supported in part by SBE-0245090 and the UAB ADVANCE program of the Office for the Advancement of Women in Science and Engineering.

6. REFERENCES

[1] B. Schölkopf, J.C. Platt, et al. "Estimating the support of a high-dimensional distribution," *Microsoft Research Corporation Technical Report* MSR-TR-99-87, 1999.

[2] Y. Rui, T.S. Huang, and S. Mehrotra, "Content-based Image Retrieval with Relevance Feedback in MARS," *Proc. of the International Conf. on Image Processing,* pp. 815-818, 1997.

[3] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No.8, 2002.

[4] Z. Su, H.J. Zhang, S. Li, and S.P. Ma, "Relevance Feedback in Content-based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressing Learning," *IEEE Trans. on Image Processing,* Vol. 12, No. 8, pp. 924-937, 2003.

[5] O. Maron, and T. Lozano-Perez, "A Framework for Multiple Instance Learning," *Advances in Natural Information Processing System 10.* Cambridge, MA, MIT Press, 1998.

[6] Y. Chen, X. Zhou, S. Tomas, and T.S. Huang, "One-Class SVM for Learning in Image Retrieval," *Proc. of IEEE International Conf. on Image Processing,* 2001.

[7] X. Huang, S.-C. Chen, M.-L. Shyu, and C. Zhang, "User Concept Pattern Discovery Using Relevance Feedback and Multiple Instance Learning for Content-Based Image Retrieval," *Proc. of the 3rd International Workshop on Multimedia Data Mining (MDM/KDD'2002)*, pp. 100-108, 2002.