An Online Multiple Instance Learning System for Semantic Image Retrieval

Chengcui Zhang, Xin Chen, and Wei-Bang Chen
Department of Computer and Information Sciences
University of Alabama at Birmingham
Birmingham, AL 35294
{ zhang, chenxin, wbc0522}@cis.uab.edu

1. System overview

The proposed system is a region based semantic image retrieval system. The retrieval is realized through mapping and solving a Multiple Instance Learning (MIL) problem [1]. MIL is a supervised learning problem in which the interest is to know the "instance" label according to the known labels of the "bag" containing the instances. This problem can be mapped exactly to a region-based Content Based Image Retrieval (CBIR) scenario, in which it is assumed that the user is only interested in one particular region of the query image instead of the image as a whole. By incorporating Relevance Feedback (RF) [2] technique in CBIR, in each retrieval iteration the user is asked for the relativity of each retrieved image to the query image region. Therefore, the label of the whole image (bag) is known. The learning algorithm then needs to find out the labels of the unseen image regions (instances) in the database. In the proposed system, the One-class Support Vector Machine (SVM) [3] is applied as the learning algorithm. The overview of the system is illustrated in Figure 1.

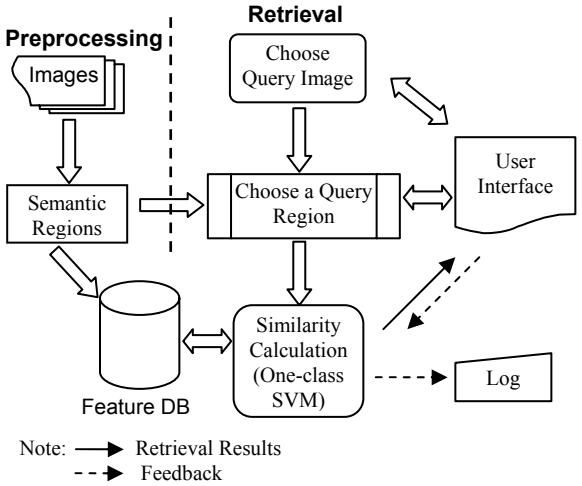

Figure 1. System Architecture

In the preprocessing phase, all images in the image database are segmented into image regions. This phrase is performed by Blob-World [4]. After segmentation, 8 features are extracted from each image region (three color features, three texture features, and two shape features), which are stored in the database.

In the image retrieval phase, the user specifies a query image of interest. Then the user is asked for the specific region in that query image that he/she is interested, which is then set as the query target. In the initial retrieval, there is no user feedback information. 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. The training set is fed into One-class SVM to compute the similarity score of each image region in the database. The similarity score of each individual image is then set to the highest score of all its regions. The top *n* highest-score images are returned to the user for feedback. Retrieval iteration then starts with these new feedbacks. The whole process goes through several iterations until a satisfactory result is obtained.

The technical details of the proposed system's kernel mechanism can be found in our previous publication [5]. The system that we implement and will demonstrate in this proposal is an online version which is based on a Model-View-Controller architecture. In the View module, user interaction with the retrieval system is realized through browser interface. The backend Model is the One-class SVM learning algorithm with an SQL enabled database that stores the essential data. The Controller module is in charge of listening to the user's feedbacks and returning to the user the retrieval results obtained from the Model module.

2. Merits of the proposed system

The proposed system successfully maps a Content Based Image Retrieval (CBIR) problem to a MIL problem. In CBIR, we have two types of labels

0-7695-3084-2/07 $25.00 © 2007 IEEE
DOI 10.1109/ISM.Workshops.2007.23

– Positive and Negative. Each image is considered a bag of semantic regions (instances). A user labels an image positive if it contains the region of interest; otherwise, it is labeled negative. Therefore relation between the bag (image) label and the instance (region) label can be defined as – if the bag label is positive, there exists at least one instance that is labeled positive; otherwise, if the bag label is negative, labels of all the instances in that bag are 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.

The MIL problem is solved by the One-class SVM. One-class SVM concentrates on hose 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. The idea of One-class SVM is to model the dense region as a hyper-sphere so that all the positive samples are inside the sphere and the negative ones are outside.

The proposed system is operated online with user feedbacks collected in the database log. The user can perform queries online by choosing query targets and providing feedbacks to the retrieval results. All these information are then stored in the log file. The system is implemented based on JSP-Servlet technology and the back-end database is MySQL.

3. Comparison

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 [6]; 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 2 shows the result from the First Query and Figure 3 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.

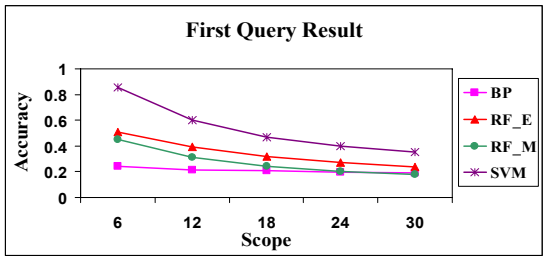

Figure 2. Retrieval Accuracy after the 1st Query

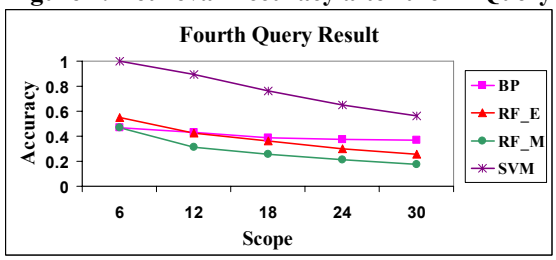

Figure 3. Retrieval Accuracy after the 4th Query

It can be seen that the proposed algorithm is superior to the other systems. Its retrieval accuracy among top 30 returned images increases around 15% from the first query iteration to the fourth one.

4. Acknowledgement

The research of Dr. Chengcui Zhang is supported in part by the UAB ADVANCE program through the sponsorship of the National Science Foundation.

5. References

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

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

[3] 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.

[4] 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.

[5] C. Zhang, X. Chen, M. Chen, S.-C. Chen, and M.-L. Shyu, "A Multiple Instance Learning Approach for Content Based Image Retrieval Using One-class Support Vector Machine," in *Proceedings of the IEEE International Conference on Multimedia and Expo. (ICME),* Amsterdam, Netherlands, 2005.

[6] 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," in *Proc. of the 3rd International Workshop on Multimedia Data Mining (MDM/KDD'2002)*, pp. 100-108, 2002.