ENTROPY-BASED ACTIVE LEARNING WITH SUPPORT VECTOR MACHINES FOR CONTENT-BASED IMAGE RETRIEVAL*

Feng Jing

State Key Lab of Intelligent Technology and Systems
Beijing 100084, China
Scenery_JF@hotmail.com

Mingjing Li, Hong-Jiang Zhang

Microsoft Research Asia
49 Zhichun Road
Beijing 100080, China
{mjli, hjzhang}@microsoft.com

Bo Zhang

State Key Lab of Intelligent Technology and Systems
Beijing 100084, China
dcszb@ mail.tsinghua.edu.cn

ABSTRACT

In this paper, an entropy-based active learning scheme with support vector machines (SVMs) is proposed for relevance feedback in content-based image retrieval. The main issue in active learning for image retrieval is how to choose images for the user to label in the next interaction. According to the information theory, we proposed an entropy-based criterion for good request selection. To apply the criterion with SVMs, probabilistic outputs are required. Since standard SVMs do not provide such outputs, two techniques are used to produce probabilities. One is to train the parameters of an additional sigmoid function. The other is to use the notion of version space. Experimental results on a database of 10,000 general-purpose images demonstrate the effectiveness of the proposed active learning scheme.

1. INTRODUCTION

After over a decade of intensified research, the retrieval result of content-based image retrieval (CBIR) is still not satisfactory. It is widely understood that the major bottleneck of CBIR approaches is the gap between visual feature representations and semantic concepts of images. To reduce the gap, relevance feedback (RF) initially developed in text retrieval [6] was introduced into CBIR during mid 1990's and has been shown to provide dramatic performance boost in retrieval systems [4][7]. The main idea of relevance feedback is to let user in the loop. During retrieval process, the user interacts with the system and rates the relevance of the retrieved images, according to his/her subjective judgment.

Relevance feedback in CBIR could be considered a classification problem. Under such consideration, the positive and negative examples provided by the user are used as training examples to train a classifier. Then, the classifier can separate other unlabelled images into relevant and irrelevant groups. Although many existing classifiers are suitable for this task, SVMs are preferred. As a core machine learning technology, SVMs have not only strong theoretical foundations but also excellent empirical successes [1][2]. SVMs have performed fairly well in the systems that use either global representations [7] or region-based ones [4].

In a practical system, the user might be impatient and provide relevance judgment for only few images. On the other hand, the learner, i.e. SVMs, has access to all images in the database. To make the most of the limited training data, choosing images for the user to label is a crucial issue for the learner to be efficient. Instead of passive learning in which the images are randomly selected, active learning could be employed here. The key issue in active learning is to find a way to choose good requests. Tong and Chang proposed a criterion for good request using the notion of version space [7]. According to their criterion, good requests should maximally reduce the size of the version space. Following the principle of maximal disagreement, the best strategy is to halve the version space each time. By taking advantage of the duality between the feature space and the parameter space, they showed that the points near the decision boundary can approximately achieve this goal. With this approximation, the criterion is simplified to be that good requests for the next interaction are the images nearest to the current decision boundary [7].

In this paper, an entropy-based criterion is proposed. According to the criterion, good requests should be informative requests. Entropy is used to characterize the information value of images based on information theory. Since standard SVMs do not provide calibrated probabilistic outputs that are needed for the calculation of entropy values, two techniques are considered to resolve the issue. One is to train the parameters of an additional

* This work was performed at Microsoft Research Asia.

0-7803-8603-5/04/$20.00 ©2004 IEEE.

sigmoid function to map the SVM outputs into probabilities. The other is to use Bayesian theory based on the notion of version space. With the latter technique, the proposed entropy-based criterion is shown to be consistent with the criterion in [7].

The remainder of the paper is organized as follows. In Section 2, an entropy-based active learning scheme is proposed by introducing a general criterion for good request selection. To apply the criterion with SVMs, two techniques that produce probabilistic outputs are adopted and discussed in Section 3. In Section 4, we provide experimental results that evaluate the active learning scheme by comparing it with a passive method. Finally, we conclude in Section 5.

2. ENTROPY-BASED ACTIVE LEARNING

The main issue with active learning for image retrieval is to choose images to request user in the next interaction. Since the goal of relevance feedback is to dynamically learn the user's query concept, a good request should provide additional information to the learner to accelerate the learning process. That is, a good request should also be an informative one.

Assume that current query concept learned by a relevance feedback learner, e.g. an SVM, is *C*. Also assume that the probability of an image *I* being relevant (irrelevant) with *C* is $P(C \mid I)$ ($P(\overline{C} \mid I) = 1 - P(C \mid I)$). From the information theory perspective, the entropy of this distribution is precisely the information value of image *I*. More specific, the entropy of *I* is:

$$En(I) = -P(C \mid I)\log P(C \mid I) - P(\overline{C} \mid I)\log P(\overline{C} \mid I) \qquad (1)$$

En(*I*) is maximized when $P(C \mid I) = 0.5$ and the smaller the difference between $P(C \mid I)$ and 0.5 the larger the value of *En*(*I*). Therefore, the criterion for choosing requests is: for an image *I*, the smaller the value of $|P(C \mid I) - 0.5|$, the better *I* will be a good request.

3. ACTIVE LEARNING WITH SUPPORT VECTOR MACHINES

As discussed in Section 1, relevance feedback could be considered as a classification problem. When SVM is used as the classifier, it captures the query concept *C* by separating relevant images from irrelevant ones with a maximal margin hyperplane in a projected feature space.

More specifically, suppose we are given *n* labeled images as the training set $\{(x_i, y_i)\}_{i=1}^{n} \subset (X \times \{-1,+1\})^n$ where x_i is the visual feature vector of image I_i and y_i is +1 (-1) if I_i is relevant (irrelevant). An SVM is a kernel classifier of the form:

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) \quad \alpha_i \in R \qquad (2)$$

where $x \in X$ is the feature vector of an unlabelled image *I* and *K* is referred to as kernel function. When $f(x) \geq 0$, *I* is classified as relevant, otherwise *I* is classified as irrelevant. Moreover, the larger the value of $f(x)$, the more relevant *I* is considered to be.

3.1. Producing Probabilities by Fitting Sigmoid

To perform the entropy-based active learning with SVMs, $f(x)$ is expected to be the probability of *I* being relevant with the query concept, i.e. $P(C \mid I)$. However, with standard training process, $f(x)$ is not a calibrated probability. A straightforward method to create probabilities is to directly train a kernel classifier with a logic link function and a regularized maximum likelihood score. However, such training will produce non-sparse kernel machines which will cost more training and testing time. Considering the real-time nature of relevance feedback, a more efficient algorithm proposed by Platt [5] is adopted. Instead of estimating the class-conditional densities, it utilizes a parametric model to fit the posterior directly. The parameters of the model are adapted to give the best probability outputs. The form of the parametric model is chosen to be sigmoid:

$$P(C \mid I) = \frac{1}{1 + \exp(Af(x) + B)} \qquad (3)$$

where *A*, *B* are the parameters to be fitted using maximum likelihood estimation from a training set. Three-fold cross-validation is used to form an unbiased training set.

We summarize the active selection strategy based on fitting sigmoid as follows:

- Learn an SVM on the current labeled images;
- Train the parameters, i.e. *A*, *B* of a sigmoid function with an unbiased training set;
- Use formula (3) to calculate the probability of other unlabeled images being relevant to the current query concept *C*, i.e. $P(C \mid I)$;
- The image with smallest $|P(C \mid I) - 0.5|$ is chosen as the next request for the user to label.

3.2. Producing Probabilities by Exploiting Version Space

Another way to produce probabilities is to use the notion of version space. If the kernel function in formula (2) satisfies the Mercer's condition, i.e. *K* is symmetric and positive definite, there exists a feature space *F* and a mapping $\phi : X \mapsto F$ such that *f* can be expressed as an inner product between the mapped point *x* and a vector $w \in F$, i.e.

$$f(x) = w \cdot \phi(x) \quad \text{where } w = \sum_{i=1}^{n} \alpha_i \phi(x_i) \qquad (4)$$

The version space *VS* is defined to be the set of all *w* consistent with the training set. More specific,

$$VS = \{w \mid y_i(w \cdot \phi(x_i)) > 0, i = 1,...,n\} \quad (5)$$

Given a new request image I_{n+1} with feature vector x_{n+1}, the version space is divided into two parts:

$$VS^+ = \{w \in VS \mid w \cdot \phi(x_{n+1}) \geq 0\} \quad (6)$$

$$VS^- = \{w \in VS \mid w \cdot \phi(x_{n+1}) < 0\} \quad (7)$$

Assume that the volume of *VS*, VS^+ and VS^- are *V*, V^+ and V^- respectively. According to the Bayesian theory, the probability of I_{n+1} being relevant is V^+ / V:

$$P(C \mid I_{n+1}) = V^+ / V. \quad (8)$$

Therefore, to maximize the information gain, the new request should halve the version space. However, it is not practical to explicitly compute the volumes. Considering that the goal of halving the version space is the same as that of [7], the simple margin rule could be adopted to approximate the probability estimation process. According to the rule, the closer an image is to the decision boundary, the more evenly the version space is expected to be split with that image being next request. Moreover, the more evenly the version space is split, the more informative the request is. Therefore, the image nearest to the boundary is considered to be the most informative one and should be used as the next request [7].

4. EXPERIMENTAL RESULTS

The active learning scheme was evaluated with a general-purpose image database consisting of 10,000 images from COREL. 1,000 images were randomly chosen from totally 79 categories as the query set. All the experimental results are averages of those 1,000 queries.

By properly incorporating spatial information into color histogram, correlogram has been proven to be one of the most effective features in CBIR [3] and therefore is used as the visual features in our implementation. In addition, the L_1 distance is used as the similarity measure.

As suggested by [1], the Laplacian kernel is chosen as the kernel of SVM, which is more appropriate for histogram-based features like the one we use. Moreover, Laplacian kernel satisfies the constraint: $\|w\| = \|\phi(x)\| = 1$, which is a requirement of the simple margin method [7].

Relevance feedback was simulated as follows. Given a query image, 5 most similar images were examined by the system. Images from the same (different) category as the initial query were used as positive (negative) examples. Based on the training set, an SVM classifier was learned to capture the query concept. During each of the following iterations, an unlabelled image was selected according to a certain criterion to request the user. After being labeled, the request image was added to the training set to refine the current learner. Currently, 11 iterations were carried out including the initialization one, which results in a final training set of 15 images. The SVM trained with this set is used to classify all other images into relevant and irrelevant classes. The most relevant images that are farthest from the SVM boundary on the relevant side are displayed as the final result.

To show the effectiveness of active learning, three selection strategies were compared: a random selection strategy (RD), the fitting sigmoid strategy (FS) and the simple margin strategy (SM). With the RD strategy, the next request is randomly chosen from the unlabeled images. This strategy reflects what happens in the regular passive learning setting. Precision is used as the basic evaluation measure. When top *N* images are considered and there are *R* relevant images, the precision within top *N* images is defined to be $P(N) = R / N$. *N* is also called scope in the following. The average $P(10)$ vs. number of iterations graph is used for the comparison and shown in Figure 1. From the figure we can see that the two active learning strategies, i.e. the FS and SM strategy are consistently better than the passive learning strategy, i.e. the RD strategy after the second iteration. After 11 iterations the average $P(10)$ of FS is higher than that of RD by 15%. On the other hand, the performance of FS and SM are almost the same.

As discussed in [7], it is more convenient and practical for a user to label several images with fewer rounds. With the proposed active learning scheme, such multiple requests selection could be performed by choosing a group of images with largest information values instead of only the most informative one. More specific, during each feedback iteration, 5 images with smallest $| P(C \mid I) - 0.5|$ (closest to the separating hyperplane) were chosen as new requests when FS (SM) strategy was used. The number of iterations is limited to be 5. As a result, the final training set contained 25 images. The aforementioned three strategies were re-evaluated with the new feedback setting. The results are shown in Figure 2. The superiority of active strategies over the passive strategy is clearer here. After 5 iterations of feedback, the average $P(10)$ of FS is higher than that of RD by 22%. In addition, the performance difference between two active learning strategies is still negligible.

The effect of different relevance feedback settings on the active learning scheme has also been evaluated. The performance of single request setting after 11 (SR_11) and 3 (SR_3) iterations and that of multiple requests setting after 3 (MR_3) iterations are compared with FS being the selection strategy. The average precision vs. scope graph is used for the evaluation and shown in Figure 3. The performance of SR_11 is a higher than that of MR_3 which uses the same number of training images, i.e. 15. This observation could be explained by the truth that with the single request setting the active learner has more control and freedom to adapt. On the other hand, MR_3 is better than SR_3. Furthermore, the performance

difference between MR_3 and SR_3 is almost double of that between SR_11 and MR_3, which suggests that it is more beneficial to use multiple requests with fewer rounds of feedback.

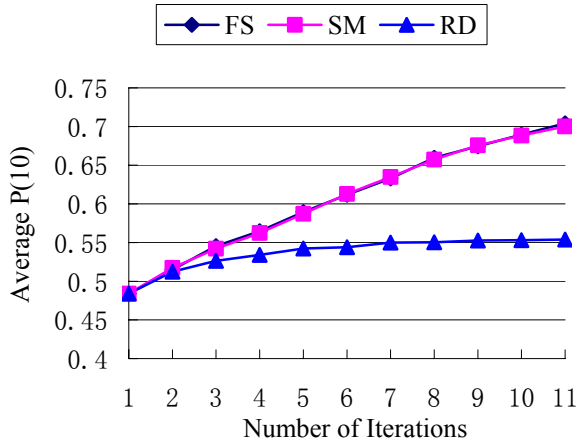

Figure 1. Accuracy comparison of different selection strategies under a single request setting. FS, SM and RD denote fitting sigmoid, simple margin and random selecting strategy respectively.

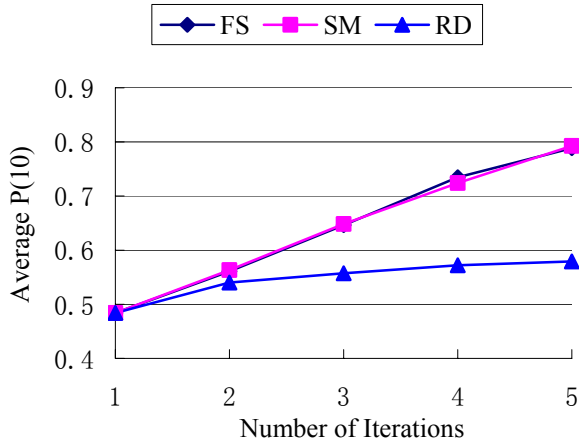

Figure 2. Accuracy comparison of different selection strategies under a multiple requests setting. FS, SM and RD denote fitting sigmoid, simple margin and random selecting strategy respectively.

5. CONCLUSION

We have introduced a new criterion for performing active learning in content-based image retrieval. To apply the criterion with SVMs, two techniques are utilized which results in two active selection strategies. Experimental results on large scale image database demonstrate that the active selection strategies are much more effective than a passive strategy for the purpose of query concept learning in content-based image retrieval.

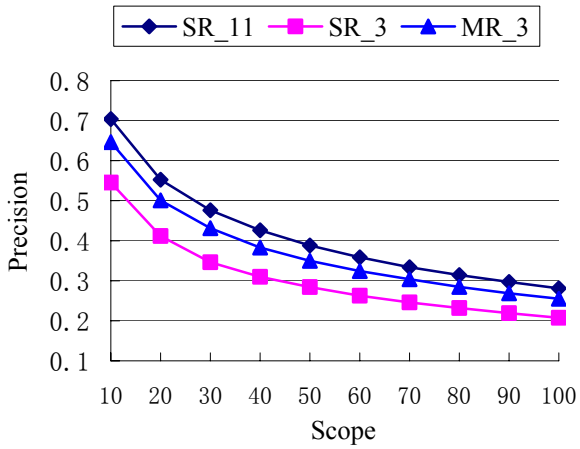

Figure 3. Comparison of single request setting after 11 (SR_11) and 3 (SR_3) iterations and multiple requests setting after 3 iterations (MR_3) using fitting sigmoid selection strategy.

6. ACKNOWLEDGMENTS

Feng Jing and Bo Zhang are supported in part by NSF Grant CDA 96-24396.

7. REFERENCES

[1] Chapelle, O., Haffner, P., and Vapnik, V., "SVMs for Histogram-based Image Classification". IEEE Transaction on Neural Networks, 10(5), Sep. 1999, pp. 1055-1065.

[2] Cristianini, N., Shawe-Taylor, J., "An Introduction to Support Vector Machines." Cambridge University Press, Cambridge, UK, 2000.

[3] Huang, J., et al, "Image indexing using color correlograms". In Proc. IEEE Comp. Soc. Conf. Comp. Vis. and Patt. Rec., pages 762--768, 1997.

[4] Jing, F., Li, M., Zhang, H.J., Zhang, B., "Support Vector Machines for Region-Based Image Retrieval", Proc. IEEE International Conference on Multimedia & Expo, 2003.

[5] Platt, J., "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods", in Advances in Large Margin Classifiers, MIT Press, 2000. pp. 61-74.

[6] Salton, G., "Automatic text processing", Addison-Wesley, 1989.

[7] Tong, S. and Chang, E. "Support vector machine active leaning for image retrieval," ACM Multimedia 2001, Ottawa, Canada.