In Proc. IEEE Int'l Conf. on Image Processing 2001, Thessaloniki, Greece

ONE-CLASS SVM FOR LEARNING IN IMAGE RETRIEVAL

Yunqiang Chen, Xiang Zhou, and Thomas S. Huang
{chenyq, xzhou2,huang}@ifp.uiuc.edu

405 N. Mathews Ave., Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61820

ABSTRACT

Relevance feedback schemes using linear/quadratic estimators have been applied in content-based image retrieval to significantly improve retrieval performance. One major difficulty in relevance feedback is to estimate the support of target images in high dimensional feature space with a relatively small number of training samples. In this paper, we develop a novel scheme based on one-class SVM, which fits a tight hyper-sphere in the non-linearly transformed feature space to include most of the target images based on the positive examples. The use of kernel provides us an elegant way to deal with non-linearity in the distribution of the target images, while the regularization term in SVM provides good generalization ability. To validate the efficacy of the proposed approach, we test it on both synthesized data and real-world images. Promising results are achieved in both cases.

1. INTRODUCTION

1.1 Background

Content-based image retrieval has received much interest in the last decades due to the large digital storage and easy access of images on computers and through the World Wide Web [1]. A major challenge in CBIR system comes from the dynamic interpretation of images by different users at different times, thus adaptive real-time learning and/or classification is required. The computer can only detect the low-level features, e.g., texture, color histogram and edge features while the user's demand may be very high-level concepts. To bridge this large gap between humans and computers, computers have to be able to learn which features best describe the pictures in user's mind on-line.

Relevance feedback and on-line learning techniques have been shown to provide dramatic performance boost in CBIR systems [3][5][9][10]. The strategy is to ask the user to give some feedbacks on the results returned in the previous query round and try to refine the search strategy and come up with a better result-set based on these feedbacks. Majority of the work uses relevance feedback to learn the relative importance of different features, with some tries to learn a feature weighting scheme either with [5] or without [3][6] considering correlations among feature components; while others either use a probabilistic scheme [4], or Self-Organizing Maps [15], or boosting technique [13], etc., to do so. Many of the algorithms are heuristics-based, which are fast and robust, but relying on the condition that one can find the right parameters [4][5][14]. Discriminant analysis on the examples given by the user is applied for dimensionality reduction, assuming two classes, before the Expectation-Maximization (EM) algorithm is used in a transductive learning framework [10]. This scheme has the potential difficulty in computational expenses, especially when the database is large. Recently there is also preliminary attempt to incorporate Support Vector Machine (SVM) into relevance feedback process [8].

1.2 One Class or Two Classes

A typical problem with CBIR system with relevance feedback is the relatively small number of training samples and the high dimension of the feature space. The system can only present the user with a few dozen of images to label (relevant or irrelevant). The interesting images to the user are only a very small portion of the large image database, in which most images remain unlabeled. Much work regards the problem as a strict two-class classification problem, with equal treatments on both positive and negative examples. It is reasonable to assume positive examples to cluster in certain way, but negative examples usually do not cluster since they can belong to any class. It is almost impossible to estimate the real distribution of negative images in the database based on the relevant feedback. An illustration of the undesirable result reached by two-class SVM is given in Figure 1.

1.3 Prior Knowledge (Assumptions)

Different assumptions about the distribution of the target images have been proposed in the literature.

Gaussian assumption is the most common and convenient one [14]. It assumes the target pictures are distributed in the feature space as a single mode Gaussian. But because of the large gap between the high level concepts and low level features, it is hard to justify this assumption. Some go to another extreme. They assume each returned positive image as a mode and try to get the target images from the nearest neighbor of each relevant image. This method will be too sensitive to the training images and it does not have good generalization capability.

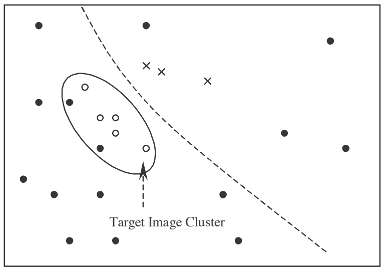


Figure 1 Decision boundary of a two-class SVM: The circles are the positive images, the crosses are the negative ones, and the black dots are the unlabeled images. The decision boundary (the dashed line) will classify many non-target images as positive.

1.4 The Proposed Scheme

In this paper, we try to fit a tight hyper-sphere in the feature space to include most positive training samples. This task is formulated into an energy optimization problem. It is then changed to a dual form and kernels are introduced. Because the distribution of the target images cannot be decided, we should not restrict ourselves to a special kind of distribution assumption. In this scheme, the non-linearity can be represent implicitly in the kernel evaluations. The computational cost will not change much from the non-kernel counterpart. As long as the assumption about the distribution can be represented in a kernel form, our algorithm can be used with high efficiency for online image retrieval.

The rest of the paper is organized as follows. Section 2 describe our One-class SVM in details and give some discussion on different kernels. Section 3 presents the experimental results on the real database. The conclusions and future work are listed in section 4.

2. ONE-CLASS SVM

In this paper, we try to estimate the distribution of the target images in the feature space without over-fitting to the user feedbacks. Because of the good generalization ability of SVM [7], we try to estimate the support that can include most of the relevant images with some regularization to single out outliers. The algorithm is named One-class SVM [11] since only positive examples are used in training and testing.

2.1 One class SVM

We first introduce terminology and notation conventions. We consider training data

$$X_1, X_2, \cdots X_l \in \aleph$$

where $l \in N$ is the number of observations. Let Φ be a feature map $\aleph \to F$.

Our strategy is to map the data into the feature space and then try to use a hyper-sphere to describe the data in feature space and put most of the data into the hyper-sphere. This can be formulated into an optimization problem. We want the ball to be as small as possible while at the same time, including most of the training data. We only consider the positive points and get the objective function in the following form (primal form):

$$\min_{R\in\Re,\zeta\in\Re^l,c\in F} R^2 + \frac{1}{vl}\sum_i \zeta_i$$

$$s.t. \quad \|\Phi(X_i) - c\|^2 \le R^2 + \zeta_i, \quad \zeta_i \ge 0 \ for\ i \in [l]$$

The trade off between the radius of the hyper-sphere and the number of training samples that it can hold is set by the parameter $v \in [0,1]$. When v is small, we try to put more data into the "ball". When v is larger, we try to squeeze the size of the "ball".

We can solve this optimization with Lagrangian multipliers:

$$L(R,\zeta,c,\alpha,\beta) = R^2 + \sum_{i=1}^{l} \alpha_i \left[\|\Phi(X_i) - c\|^2 - R^2 - \zeta_i \right] + \frac{1}{vl}\sum_{i=1}^{l}\zeta_i - \sum_{i=1}^{l}\beta_i\zeta_i$$

$$\frac{\partial L}{\partial R} = 2R(1 - \sum \alpha_i) = 0 \quad \Rightarrow \quad \sum \alpha_i = 1 \qquad (1)$$

$$\frac{\partial L}{\partial \zeta_i} = \frac{1}{vl} - \alpha_i - \beta_i = 0 \quad \Rightarrow \quad 0 \le \alpha_i \le \frac{1}{vl} \qquad (2)$$

$$\frac{\partial L}{\partial c} = -\sum 2\alpha_i \left(\Phi(X_i) - c\right) = 0 \qquad (3)$$
$$\Rightarrow \quad c = \sum \alpha_i \Phi(X_i)$$

The equation (1) and (2) turn out to be the constraints while equation (3) tell us the c (center of the ball) can be expressed as the linear combination of $\Phi(X)$, which make it possible to express the dual form with kernel functions.

$$\min_{\alpha} \sum_{i,j} \alpha_i \alpha_j k(X_i, X_j) - \sum_i \alpha_i k(X_i, X_i)$$

$$s.t. \quad 0 \le \alpha_i \le \frac{1}{vl}, \quad \sum_i \alpha_i = 1$$

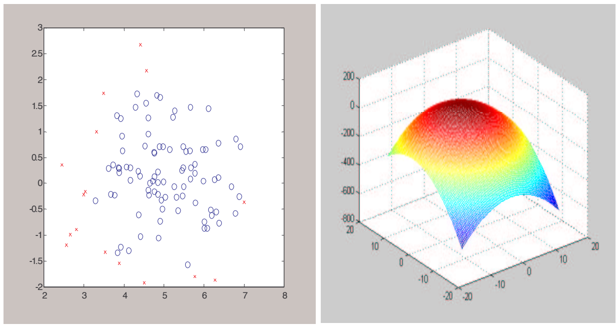
The optimal α's can be got after solving this dual problem by the QP optimization methods. We can rank all the images in the database by the following decision function:

$$f(X) = R^2 - \sum_{i,j} \alpha_i \alpha_j k(X_i, X_j) + 2\sum_i \alpha_i k(X_i, X) - k(X, X)$$

The images with higher scores are more likely to be the target images.

2.2 Linear Case: LOC-SVM (One-Class SVM)

First we try the linear case. For linear case, the algorithm just tries to fit a hyper-sphere to cover the training points with outlier detection. A synthetic training data set is generated to illustrate our algorithm. The training sample is randomly generated according to $x = N(\mu, \sigma)$, where $\mu = (5, 0)^T$, $\sigma = I$. One can see that in Figure 2(b), the learning machine catches the distribution without over-fitting. The decision function evaluates the largest value at [5.0, 0.15] which is very close to the true center. It tries to put a circle in the 2D dimension to include most positive samples while leave some out as outlier. The parameter ν can be tuned to control the amount of outliers.



(a) (b)

Figure 2: (a) shows the training points we generated. The dots are the samples that have positive evaluation from the decision function after training. The crosses are the samples that are detected as outliers and have negative evaluation from the decision function. (b) is the decision value for all the points in the 2D feature space. It takes the largest value at [5.0, 0.15]

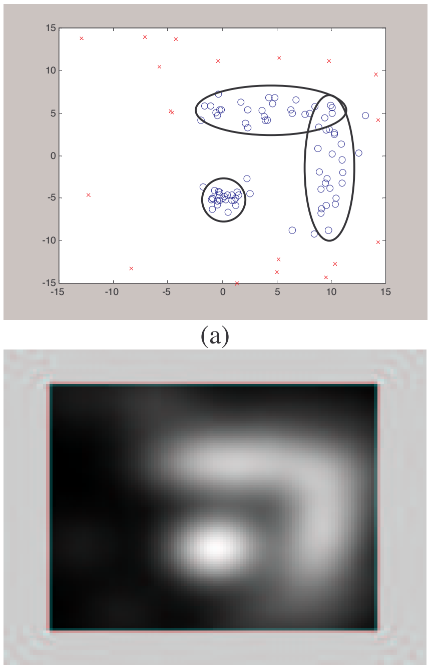
But in feature space, we cannot assume that the target images are clustered in spherical shape—images can have complicated non-linear distributions. However by using kernel-based form of this algorithm, non-linear distribution can be dealt with in the same framework.

2.3 Non-linear Case Using Kernel (KOC-SVM)

In this section, we discuss the use of reproducing kernel to deal with non-linear, multi-mode distributions using KOC-SVM (Kernel One-class SVM). A good choice is the Gaussian kernel in the following form:

$$K(X, Y) = e^{-\|X-Y\|^2 / 2\sigma^2}$$

To test KOC-SVM's ability to capture non-linearity such as a multi-mode distribution, training data are jointly sampled from three Gaussian modes. We also generate some sparse outliers from a uniform distribution over the feature space. After training, the decision function in the feature space is shown in Figure 3 (b). It is apparent that KOC-SVM captures the multimode fairly well. It is important to note that this learning machine has the capability of removing outliers in an intelligent way, unlike the way a non-parametric Parzen window density estimator works.



(a)

(b)

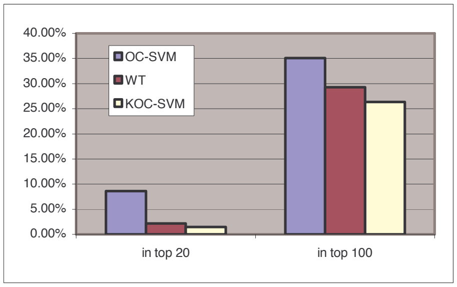
Figure 3: (a) shows all the training set we generated from three Gaussians. The dots are the samples which have positive evaluation from the decision function. The crosses are the samples which are detected as outliers and get negative decision function values. (b) decision values for all the points in the 2D feature space. White means high value while black means low value.

3. EXPERIMENT RESULTS

Finally, tests on the real-world images are conducted. We constructed a fully labeled image database. It has five classes each with 100 images. These are: airplanes, cars, horses, eagles and stained glasses. 10 images are randomly drawn as training samples and the learned decision function is applied to rank all the 500 images in the database. The hit rates in the first 20 and 100 images are used as the performance measures. Specifically, for each class 100 tests are performed with only 10 randomly drawn training samples and the average error rate is recorded as shown in the following table. Tested against is the WT (Whitening Transform), which is the relevance feedback technique reported in [3][6] (Note that in [6], a two-level structure is adopted to better deal with singularity issues. Here we use one-level and use regularization terms to bring the covariance matrix out of singularity.) LOC-SVM is the worst for apparent reasons that it lacks flexibility in modeling variations in distributions. WT is better since it can model multivariate Gaussian distribution. KOC-SVM gives the best results due to its capability in nonlinear modeling.

Table 1: Averaged Error rate for image retrieval using LOC-SVM (One-class SVM), WT (Whitening Transform), and KOC-SVM (Kernel One-class SVM), all with 10 training samples.

Average Error Rate	LOC-SVM	WT	KOC-SVM
in top 20	8.63%	2.20%	1.47%
in top 100	35.12%	29.28%	26.38%



4. CONCLUSION AND FUTURE WORK

In this paper, statistical learning method is used to attack the problems in content-based image retrieval. We developed a common framework to deal with the problem of training with small samples. Kernel machines provide us a way to deal with non-linearity in an elegant way. Promising results are presented.

Some more research should be done in how to choose appropriate kernel for CBIR. Also, it is desirable to have a systematic scheme for tuning the parameters such as the spread of the Gaussian and the strength of the regularization term.

ACKNOWLEDGEMENTS

This work was supported in part by National Science Foundation Grant CDA 96-24396, IIS-00-85980 and EIA 99-75019.

REFERENCES

[1] M. Flickner. et al., "Query by image and video content: The qbic system", IEEE Computers. 1995

[2] G. Rätsch, B. Schölkopf, S. Mika, and K. R. Müller. "SVM and boosting: One class", Technical Report 119, GMD FIRST, Berlin, November 2000.

[3] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "MindReader Query databases through multiple examples", in Proc. Of the 24th VLDB Conf. (New York), 1998

[4] J. Peng, B. Bhanu, and S.Qing, "Probabilistic feature relevance learning for content-based image retrieval", Computer Vision and Image Understanding, 75:150-164, 1999

[5] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval", IEEE Tran on Circuits and Systems for Video Technology, Vol 8, No. 5, Sept., 644-655, 1998,

[6] Y. Rui, T. S. Huang, "Optimizing learning in image retrieval", Proc. IEEE CVPR, 2000, 236-243

[7] V. Vapnik, "The Nature of Statistical Learning Theory", Springer Verlag, New York, 1995

[8] Q. Tian, P. Hong, T. S. Huang, "Update Relevant Image Weights for Content-Based Image Retrieval Using Support Vector Machines", IEEE International Conference on Multimedia and Expo (ICME'2000), Hilton New York & Towers, New York, NY, July 30 - Aug. 2, 2000.

[9] N. Vasconcelos, A. Lippman, "Bayesian relevance feedback for content-based image retrieval", in Proc.IEEE Workshop on Content-based Access of Image and Video Libraries, CVPR'00, Hilton Head Island, SC, 2000

[10] Y. Wu, Q. Tian, T. S. Huang, "Discriminant EM Algorithm with Application to Image Retrieval", IEEE Conf. Computer Vision and Pattern Recognition (CVPR)'2000, Hilton Head Island, South Carolina, June 13-15, 2000.

[11] B. Scholkopf, J. C. Platt, J. T. Shawe, A. J. Smola, R. C. Williamson, "Estimation the support of a high-dimensional Distribution", Technical Report MSR-TR-99-87, Microsoft Research

[12] I. J. Cox, M. Miller, T. Minka, P. Yianilos, "An Optimized Interaction Strategy for Bayesian Relevance Feedback", IEEE Conf. Computer Vision and Pattern Recognition (CVPR'98)

[13] K. Tieu and P. Viola, "Boosting Image Retrieval", IEEE Conf Computer Vision and Pattern Recognition (CVPR'00), Hilton Head, South Carolina

[14] C. Nastar, M. Mitschke and C. Meilhac "Efficient Query Refinement for Image Retrieval", IEEE Conf. Computer Vision and Pattern Recognition CVPR'98, Santa Barbara, CA, June 1998.

[15] J. Laaksonen, M. Koskela, and E. Oja. "PicSOM: Self-Organizing Maps for Content-Based Image Retrieval", Proc. of IJCNN'99. Washington, DC. July 1999