

SUPPORT VECTOR MACHINE LEARNING FOR IMAGE RETRIEVAL

Lei Zhang, Fuzong Lin, Bo Zhang

State Key Laboratory of Intelligent Technology and Systems,
Department of Computer Science and Technology, Tsinghua University, Beijing 100084

ABSTRACT

In this paper, a novel method of relevance feedback is presented based on Support Vector Machine learning in the content-based image retrieval system. A SVM classifier can be learned from training data of relevance images and irrelevance images marked by users. Using the classifier, the system can retrieve more images relevant to the query in the database efficiently. Experiments were carried out on a large-size database of 9918 images. It shows that the interactive learning and retrieval process can find correct images increasingly. It also shows the generalization ability of SVM under the condition of limited training samples.

1. INTRODUCTION

With advances in the multimedia technologies and the advent of the Internet, Content-Based Image Retrieval (CBIR) has been an active research topic since the early 1990's. Most of the early researches have been focused on low-level vision alone. However, after years of research, the retrieval accuracy is still far from users' expectations. It is mainly because of the large gap between high-level concepts and low-level features.

Motivated by the limitations of the low-level based approach, an interactive learning mechanism was appeared in recent years [1,2,3,4]. The basic idea is to build a model according to the relevance information, fed back by users to indicate which images he or she thinks are relevant to the query, and to do retrieval again for better result.

In [1,2,3], the query point movement method is used to improve the estimate of the "ideal query point" by moving it towards good examples point and away from bad example points. On the other hand, the re-weighting method is also used to change the distance metric to make relevant images closer. This method tries to approximate the semantic concepts by mapping images to a new feature space. In [4], a semantic network is represented by a set of keywords having links to the images in the database. Weights associated to each individual link can be updated

by feedback techniques. In this way, keywords based technology and content-based image retrieval technology are combined to improve the performance.

From another viewpoint, CBIR can also be considered as a search problem in the feature space to find more images similar to the query image. With the feedback technique, the available information is not only the feature space itself, but also the relevance information given by users. Therefore, we can build a classifier to separate two classes of relevance images and irrelevance images. Using the classifier model, we can retrieve much more images relevant to the query efficiently in the feature space.

In this paper, we propose a novel learning approach based on Support Vector Machine (SVM) [5,6]. Based on SVM, A classifier can be learned from training data of relevance images and irrelevance images marked by users. Then the model can be used to find more relevance images in the whole database. Compared with other learning algorithms, the SVM approach is considered a good candidate because of its high generalization performance without the need to add *a priori* knowledge, even when the dimension of the input space is very high.

This paper is organized as follows. In section 2, we will provide a brief overview to SVM and then present the proposed learning algorithm. In Section 3, we will describe the experimental environment and provide the result of the proposed algorithm. Thereafter, we will give concluding remarks in Section 4.

2. THE PROPOSED METHOD

To better understand the proposed method, we given in this section a very brief introduction to SVM [5,6] and then present the novel learning method.

2.1 Support vector machine

The followings will describe the principle of SVM in linear separable case.

Given a set of linear separable training samples $(\mathbf{x}_i, y_i)_{1 \leq i \leq N}$, $\mathbf{x}_i \in R^d$, $y_i \in \{-1, 1\}$ is the class label which \mathbf{x}_i belongs to. The general form of linear classification function is $g(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$, which corresponds to a separating hyperplane $\mathbf{w} \cdot \mathbf{x} + b = 0$.

We can normalize $g(\mathbf{x})$ to satisfy $|g(\mathbf{x})| \geq 1$ for all \mathbf{x}_i , so that the distance from the closest point to the hyperplane is $1/\|\mathbf{w}\|$.

Among the separating hyperplanes, the one for which the distance to the closest point is maximal is called *optimal separating hyperplane* (OSH). Since the distance to the closest point is $1/\|\mathbf{w}\|$, finding the OSH amounts to minimizing $\|\mathbf{w}\|$ and the objective function is:

$$\min \phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} (\mathbf{w} \cdot \mathbf{w})$$

Subject to: (1)

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, N$$

If we denote by $(\alpha_1, \dots, \alpha_N)$ the N non-negative Lagrange multipliers associated with constraints in (1), we can uniquely construct the OSH by solving a constrained quadratic programming problem. The solution \mathbf{w} has an expansion $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$ in terms of a subset of training patterns, called support vectors, which lie on the margin. The classification function can thus be written as

$$f(\mathbf{x}) = \text{sign} \left(\sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \right) \quad (2)$$

When the data is not linearly separable, on the one hand, SVM introduces slack variables and a penalty factor such that the objective function can be modified as

$$\phi(\mathbf{w}, \xi) = \frac{1}{2} (\mathbf{w} \cdot \mathbf{w}) + C \left(\sum_1^N \xi_i \right) \quad (3)$$

On the other hand, the input data can be mapped through some nonlinear mapping into a high-dimensional feature space in which the optimal separating hyperplane is constructed. Thus the dot production can be represented by $k(\mathbf{x}, \mathbf{y}) := (\phi(\mathbf{x}) \cdot \phi(\mathbf{y}))$ when the kernel k satisfy Mercer's condition [6]. Finally, we obtain the classification function

$$f(\mathbf{x}) = \text{sign} \left(\sum_i \alpha_i y_i \cdot k(\mathbf{x}_i, \mathbf{x}) + b \right) \quad (4)$$

Because SVM can be analyzed theoretically using concepts from the statistical learning theory, it has particular advantages when applied to problems with limited training samples in the high-dimensional space. Consequently, it can achieve good performance when applied to real problems.

2.2 The learning algorithm in image retrieval

During the process of relevance feedback, users can mark an image as either relevance or irrelevance. Considering top N_{RT} images in the result as training data, we can carry out two classes learning algorithm by SVM, and construct a classifier suitable to represent concepts of user's query. Thereafter, other images can be classified into either relevance class or irrelevance class according to the

distance from each image to the separating hyperplane. Sorting images according to their distance to the hyperplane, we can thus obtain a better result. The process is described below.

1. Retrieve by a traditional method.
2. Mark top N_{RT} images into two classes: relevance set I^+ and irrelevance set I^o .
3. Prepare for SVM the training data (\mathbf{x}_i, y_i) ,

$$\mathbf{x}_i \in I^+ \cup I^o, y_i = \begin{cases} +1, & \text{if } \mathbf{x}_i \in I^+ \\ -1, & \text{if } \mathbf{x}_i \in I^o \end{cases}$$

4. Construct classification function using SVM algorithm.

$$f(\mathbf{x}) = \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b$$

Note: In order to output the similarity distance to the query, we ignored the function $\text{sign}(\cdot)$ in the classifier $f(\mathbf{x})$.

5. Calculate the score for each image I_i in the database.

$$\text{score}(I_i) = f(\mathbf{x}_i)$$

6. Sort all images by score and return new result.

Obviously, in the first learning iteration, both marked positive samples and unmarked irrelevance samples are all close to the query. Such samples are very suitable to construct the SVM classifier because support vectors are just those who lie on the separating margin while other samples far away from the hyperplane will contribute nothing to the classifier. In the following iterations, more relevance samples fed back by users can be used to refine the classifier. Although training samples are limited compared to the testing images, they provide satisfactory information to separate two classes in the feature space.

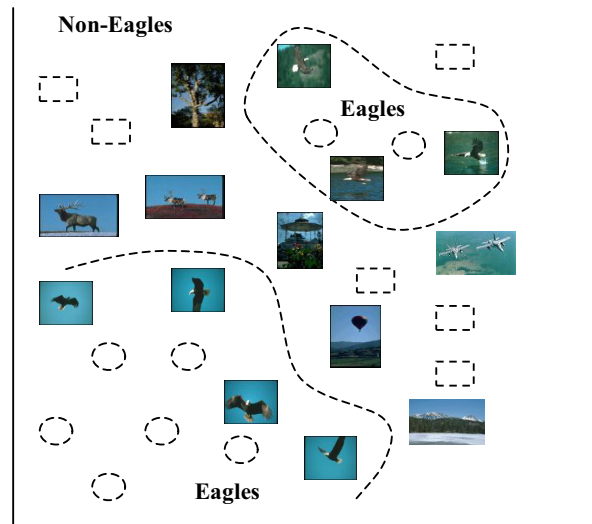


Figure 1. Geometrical interpretation of how the SVM separates the "eagle" and "non-eagle" classes

Figure 1 gives a geometrical interpretation of how to implement relevance feedback using SVM classifier when the query is “eagle”. Users mark the result returned by last iteration and indicate which images are relevant to “eagle”. The system can thus construct a SVM classifier using training data of “eagle” class and “non-eagle” class. The classifier corresponds to a separating hyperplane in the feature space (see Figure 1). Applying this model to classify each image in the database, we can retrieve more images relevant to “eagle”. Circles in Figure 1 imply new “eagle” images found by classifier after learning.

3. EXPERIMENTAL RESULTS

3.1 Performance measures

Let $\{Q_1, \dots, Q_q\}$ be the set of query images. For the i -th query Q_i , let $I_1^{(i)}, \dots, I_{a_i}^{(i)}$ are correct answers and $rank(I_j^{(i)})$ is the rank of $I_j^{(i)}$ in the result. We use three performance measures [7]:

$$(1) Avg-r = \frac{1}{q} \sum_{i=1}^q \frac{1}{a_i} \sum_{j=1}^{a_i} rank(I_j^{(i)}).$$

$$(2) Avg-p = \frac{1}{q} \sum_{i=1}^q \frac{1}{a_i} \sum_{j=1}^{a_i} \frac{j}{rank(I_j^{(i)})}.$$

(3) **Recall vs. Scope:** For query Q_i and scope $S(S>0)$:
 $recall\ r = |\{I_j^{(i)} \mid rank(I_j^{(i)}) \leq S\}| / a_i.$

3.2 Experimental environment

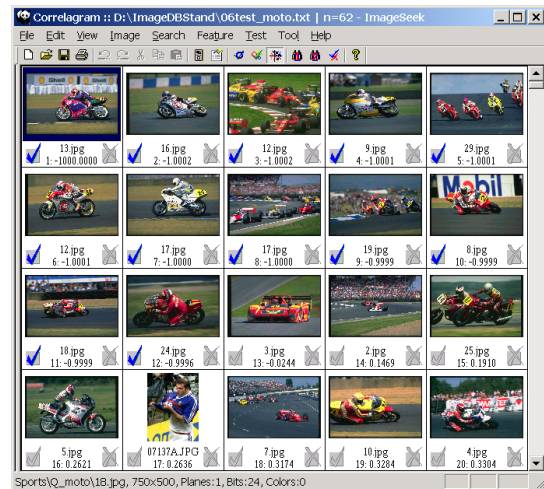
Database: The database consists of 9918 images, collected from Corel Photo CD, web site <http://www.yestart.com/pic/>, <http://202.102.233.12/pic/main.asp>, <http://home.gz.cninfo.net/suca/>, <ftp://ftp.igd.edu.cn>. The database is quite heterogeneous, including peoples, natural scenes, animals, plants, buildings, indoor scenes and sports.

Featurebase: We adopt auto-correlogram [7] as the feature for each image. We consider the RGB color space with quantization into $4*4*4=64$ colors. Then we use the distance set $D=\{1,3,5,7\}$ for computing the auto-correlogram. The dimension of the feature is 256. In general, the proposed method in this paper can use any other features suitable for image retrieval.

Query set: We use 6 query sets, see Table 1.

Table 1: Query set

	1	2	3	4	5	6
Query set	Eagle	Sunset	Rose	Tiger	Horse racing	Motor racing
Correct answers	56	75	14	23	26	62



Note: The top-left image is the query image. Images marked with \checkmark are positive samples, other “racing” images without \checkmark are retrieved after one iteration of SVM learning.

Figure 2. An instance of SVM learning in ImageSeek

For each query, five iterations were carried out to study the learning behavior on the ImageSeek system we developed for content-based image retrieval. The main user interface and an instance of learning are shown in Figure 2. The user is able to select multiple images among top N_{RT} images in the result and give feedback to the ImageSeek system. By learning from the feedback information, the system can retrieve again and get better result.

3.4 Results

In SVM, a kernel function is used to represent the dot production in the high-dimensional feature space. There are currently no techniques available to “learn” the form of the kernel. In this paper, we choose $K_{Gaussian}(x, y) = e^{-\rho \|x-y\|_2^2}$ as the kernel in the experiments because $K_{Gaussian}$ gives better performance than other kernels such as the polynomial kernel and the sigmoid kernel.

For the parameter ρ in the Gaussian kernel, an experiment is conducted to select the appropriate parameter. We set $\rho = 0.01, 0.1, 0.5, 1.0, 2.0, 10.0$, respectively, and set $N_{RT} = 100$ to allow users marking top 100 images in the result. We select another parameter in equal (3) $C=1000.0$. To better represent the performance of different parameter, we draw out the learning curve of *Recall* for five iterations in Figure 3. Obviously, when $\rho = 0.5$, the SVM learning achieved the best performance. Following experiments are thus conducted using $\rho = 0.5$.

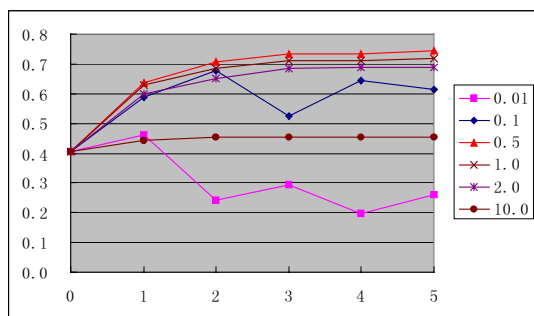


Figure 3. Learning curves of *Recall* (*Scope* = 100) using different parameter ρ of $K_{Gaussian}$

In addition to verify the effectiveness of the proposed method, we have compared it with the learning technique used in MARS [1] under the same environment. The result is shown in Table 2. We also drew out learning curves of *Recall* for five iterations in Figure 4.

Table 2. Comparison between SVM learning in ImageSeek and Re-weight learning in MARS
Test condition: $N_{RT}=100$, *Scope* = 100

Iterations	0	1	2	3	4	5
<i>Recall</i>	0.407	0.544	0.523	0.529	0.524	0.524
<i>Avg-r</i>	1056	1241	1200	1165	1165	1155
<i>Avg-p</i>	0.234	0.423	0.366	0.411	0.390	0.404

(a) Results in MARS

Iterations	0	1	2	3	4	5
<i>Recall</i>	0.407	0.637	0.706	0.733	0.733	0.743
<i>Avg-r</i>	1056	1508	1035	1392	867	1372
<i>Avg-p</i>	0.234	0.606	0.689	0.724	0.676	0.717

(b) Result in ImageSeek

As we can see from the results, our system based on SVM technique achieves better performance than MARS based on re-weighting technique. In our system, both *Recall* and *Avg-p* increase the most in the first iteration and keep continuous increase in later iterations. While in MARS, *Recall* and *Avg-p* only increase in the first iteration. Later iterations results in minor increase or decrease in *Recall* and *Avg-p*. This is the phenomenon of over learning because of too many training data.

Re-weighting techniques, like in MARS[1] and MindReader[3], use weighted or generalized Euclidean distance as the model to capture the distribution of relevant images [3]. Such model assumes that relevant images conform to the single Gaussian distribution, whose center and deviation are calculated by query point movement method and re-weighting method respectively. But in general case, relevant images conform to rather the mixture Gaussian distribution than the single Gaussian distribution. Obviously, re-weighting method cannot capture the distribution effectively and therefore has limited ability to improve the retrieval result.

Contrasting to re-weighting method, SVM does not make any assumption to training data but analyzes the

classification problem based on the statistical learning theory. Because of its high generalization ability, SVM method achieves better performance than re-weighting method.

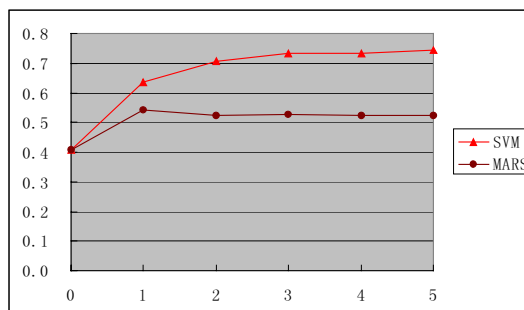


Figure 4. Learning curves of *Recall* (*Scope* = 100)

4. CONCLUSIONS

In this paper, we propose a novel learning method, which integrates support vector machine into the process of relevance feedback in image retrieval. A SVM classifier can be learned from relevance images fed back by users, and the classifier can retrieve more images relevant to the query effectively. Experiments were carried out on a large-size database of 9918 images. It shows the generalization ability of SVM under the condition of limited training samples. Both the recall rate and the precision rate are improved after several learning iterations.

5. REFERENCES

- [1] 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*, 8(5), pp. 644-655, September 1998.
- [2] Y. Rui, T. S. Huang, "A novel relevance feedback technique in image retrieval", *ACM Multimedia*, 1999.
- [3] Y. Ishikawa, R. Subramanya, and C. Faloutsos. "Mindreader: Query Databases Through Multiple Examples," In *Proc. of the 24th VLDB Conference*, pp. 218-227, 1998.
- [4] Y. Lu, C. Hu, X. Zhu, H. J. Zhang, and Q. Yang, "A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems", *ACM multimedia*, 2000.
- [5] C. Cortes, V. Vapnik, "Support-Vector Networks", *Machine Learning*, 20, pp. 273-297, 1995
- [6] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, 2(2), pp. 1-47, 1998.
- [7] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu and R. Zabih, "Image indexing using color correlograms", In *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 762-768, 1997.