

SVM-based Relevance Feedback Using Random Subspace Method

Dacheng Tao and Xiaoou Tang

Department of Information Engineering, The Chinese University of Hong Kong

[dctao2, xtang}@ie.cuhk.edu.hk](mailto:{dctao2, xtang}@ie.cuhk.edu.hk)

Abstract

Relevance feedback (RF) schemes based on Support Vector Machine (SVM) have been widely used in content-based image retrieval (CBIR). However, the performance of SVM based RF is often poor when the number of labeled feedback samples is small. In order to solve this problem, we propose a RF algorithm using the Random Subspace Method. The algorithm can overcome the classifier unstable and over-fitting problems that are common to the SVM based RF. Through extensive experiments on 17, 800 images, the proposed algorithm is shown to outperform existing algorithms significantly.

1. Introduction

Relevance feedback (RF) [1] is an important tool to improve the performance of content-based image retrieval (CBIR) [2]. Many RF methods have been developed in recent years. One approach [1] adjusts the weights of various features to adapt to the user's perception. Another approach [3] estimates the density of the positive feedback samples. Discriminant learning has also been used as a feature selection method for RF [4]. These methods all have some limitations. The method in [1] is only heuristic based. The density estimation method [3] loses information contained in negative samples. The discriminant learning [4] often suffers from the matrix singular problem.

Recently, classification-based RF [5-7] becomes a popular technique in CBIR and the Support Vector Machine (SVM) based RF (SVMRF) has shown promising results owing to its good generalization ability. However, when the amount of positive feedbacks is small, the performance of SVMRF becomes poor. This is mainly due to the following reasons. First, SVM classifier is unstable for small size training set, i.e. the optimal hyper-plane of SVM is sensitive to the training samples when the size of the training set is small. Moreover, the size of the training set is much smaller than the dimension of the feature vector in RF, thus may cause the over fitting problem. Because of the existence of noise, some features can only discriminant the marked positive and negative feedbacks but cannot discriminant other relevant or irrelevant images in the database. Therefore, the learned

SVM classifier cannot work well for the remaining images in the database.

In order to overcome these problems, we aggregate several SVMs using the Random Subspace Method (RSM), one of the Classifier Committee Learning (CCL) [8-10] methods, for RF in CBIR. Since each classifier has its own unique ability to classify relevant and irrelevant samples, the CCL can pool a number of weak classifiers to improve the recognition performance. CCL works well when the original classifier is not stable, which is exactly the case for the SVM based RF.

2. SVM in CBIR RF

To better understand the proposed SVM based RF scheme, we first give a brief review of SVM. SVM [11] is an effective binary classification algorithm. Consider a linearly separable problem:

$$\{(x_i, y_i)\}_{i=1}^N \text{ and } y_i = \{+1, -1\}, \quad (1)$$

where x_i is an n-dimension vector and y_i is the label of the class that the vector belongs to. SVM separates the two classes of points by a hyper-plane,

$$w^T x + b = 0, \quad (2)$$

where x is an input, w is the weight vector, and b is the bias.

SVM finds parameters w and b for the optimal hyper-plane to maximize the geometric margin $2/\|w\|$, subject to

$$y_i (w^T x_i + b) \geq +1. \quad (3)$$

The solution is a Wolfe dual problem with α_i :

$$Q(\alpha) = \sum_{i=1}^m \alpha_i - \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) / 2, \quad (4)$$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^m \alpha_i y_i = 0$.

In the dual format, the data points only appear in the inner product. To get a potentially better representation of the data, the data points are mapped into the Hilbert Inner Product space through a replacement:

$$x_i \cdot x_j \rightarrow \phi(x_i) \cdot \phi(x_j) = K(x_i, x_j), \quad (5)$$

where $K(\cdot)$ is a kernel function. We then get the kernel version of the Wolfe dual problem:

$$Q(\alpha) = \sum_{i=1}^m \alpha_i - \sum_{i,j=1}^m \alpha_i \alpha_j d_i d_j K(x_i, x_j) / 2. \quad (6)$$

Thus for a given kernel function, the SVM classifier is

$$F(\mathbf{x}) = \text{sgn}(f(\mathbf{x})), \quad (7)$$

where $f(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$ is the output of SVM.

$f(\mathbf{x})$ can be used to measure the similarity [5,6] between a given pattern and the query.

3. Ensemble SVMs by RSM for CBIR RF

RSM [9] benefits from the bootstrapping and aggregation. Bootstrapping is based on random sampling with replacement on the training features. Because each weak classifier generated by bootstrapping is unique in terms of classification capability, the aggregated classifier may give better results than any individual classifiers. So RSM can be used to build a strong classifier with a set of weak classifiers, which are trained on different randomly sampled features.

For SVM based RF, the constructed single SVM classifier is unstable. Moreover, over fitting happens if the training set is relatively small compared to the high dimensional feature vector. In order to avoid these problems, we can use RSM to sample the training feature to reduce the discrepancy between the training data size and the feature vector length. Using such a random sampling method, we construct a multiple number of SVMs free of the over fitting problem. We then combine these SVMs to construct a more powerful classifier solving the unstable problem by the majority voting rule (MVR). The RSM based SVM (RSMSVM) algorithm is described in Table 1.

Table 1: Algorithm of RSMSVM.

Input: feature set \mathbf{F} , weak classifier I (SVM), integer T (number of generated classifiers), \mathbf{x} is the test sample.

1. For $i=1$ to T {
2. $\mathbf{F}_i =$ bootstrap feature from \mathbf{F} .
3. $C_i = I(\mathbf{F}_i)$
4. }
5. $C^*(\mathbf{x}) = \arg \max_{y \in Y} \sum_{i: C_i(\mathbf{x})=y} 1$.

Output: classifier C^* .

For CBIR, the similarity measure is required. For a given pattern, we first utilize the MVR to recognize it as query relevant or irrelevant. Then we measure the dissimilarity between the pattern and the query using the output of the individual SVM classifier, which gives the same label as the MVR with the highest confidence (the absolute value of the output of the SVM).

We now compare the computational complexity of the SVM and RSMSVM algorithms at the training and testing stages. From [11], the computational complexity to train a SVM is $O(SVM) = O(N_s^3 + N_s^2 L + N_s N_f L)$, where N_s is the

number of support vectors, N_f is feature dimension, and L is the size of the training set. From the formula of the output of SVM, the number of the support vectors N_s determines the computational complexity in the testing stage. We denote the computational complexity for a multiple and addition of two real values as \otimes and \oplus , respectively. Then the computation complexities of SVM and RSMSVM are given in Table 2. Because the size of the training set is very small compared to the database, the overall computational complexity is mainly determined by the testing stage. Because $N_f \approx T_f N'_f$ and $N_s \approx N'_s$, where T_f is the number of RSM classifiers, N'_f and N'_s are the number of features and support vectors of a RSM classifier, respectively. The computational complexities of SVM and RSMSVM are similar.

Table 2: Comparison of computational complexity.

	Training	Testing
SVM	$O(SVM)$	$N_s \cdot N_f \cdot (\otimes + \oplus)$
RSMSVM	$T_f \cdot O(SVM)$	$T_f \cdot N'_s \cdot N'_f \cdot (\otimes + \oplus)$

4. Experimental Results

We compare our new algorithm with existing algorithms through experiments on 17, 800 images of 90 concepts from the Corel Photo Gallery. Three main features, color, texture, and shape, are used to represent the corresponding image. For color, we select the color histogram [12] in HSV color space. Hue, Saturation, and Value are quantized into 8, 8, and 4 bins respectively. Texture is extracted by 3-level pyramidal wavelet transform (PWT) with Haar wavelet of the Y component in YCrCb color space. The mean value and standard deviation are calculated for the sub-bands at each decomposition level. Edge histogram [13] is also calculated on Y component in YCrCb. Edges are grouped into four classes, which are horizontal, 45 diagonal, vertical, and 135 diagonal. We combine the color, texture, and shape features into a feature vector (284 dim), and then we normalize it to a normal distribution.

The experiments are simulated by the computer automatically. First, 300 queries are randomly selected from the data, and then RF is automatically done by computer as: the first 5 query relevant and irrelevant images are marked as positive and negative feedbacks in the top 48 images, respectively. *Precision* and *Standard Deviation* (SD) are used to evaluate the performance of a RF algorithm. Precision is the percentage of relevant images in the top N retrieved images. The precision curve is the averaged precision values of the 300 queries, and SD curve is the SD values of the 300 queries' precision. The precision curve evaluates the effectiveness of a given

algorithm and SD curve evaluates the robustness of the algorithm. In the precision and SD curves, the maximum feedback iteration is 6, with 0 feedback referring to the retrieval based on Euclidean distance measure without RF.

We compare the proposed algorithm with the original SVM based RF [5], the constrained similarity measure SVM (CSM) based RF [7], and the biased discriminant analysis (BDA) based RF [4], which are all current state of the art. We chose the Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = e^{-\rho \|\mathbf{x} - \mathbf{y}\|^2}$ with $\rho=1$ (the default value in the OSU-SVM [14] MatLab™ toolbox) for all algorithms. The performances of all the SVM algorithms are stable over a range of ρ values.

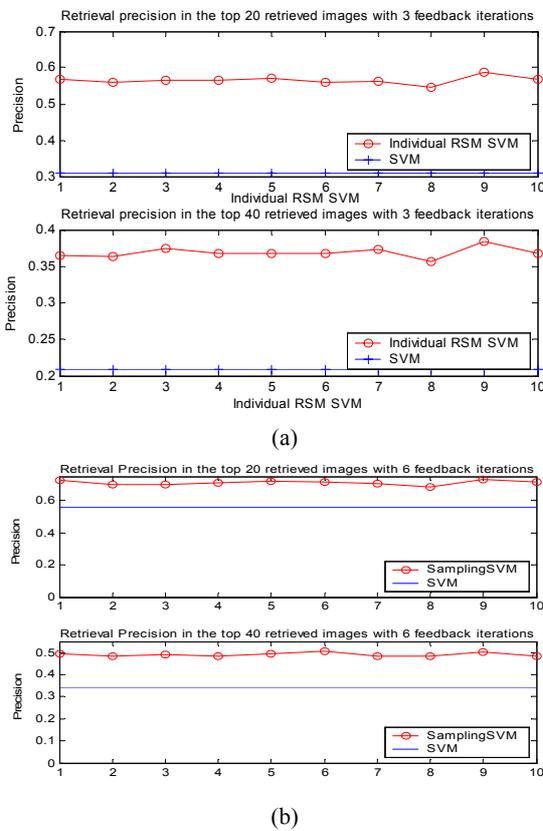


Figure 1. Compare the original SVM method with each individual RSM generated SVM classifier.

Experimental results are summarized in Fig. 1 and 2. In Fig. 1, we compare the original SVM method with each individual RSM generated SVM classifier. We can see that even before multiple classifier integration, each RSM generated classifier with short feature vector is already better than the original classifier. In Fig. 2 (a-d), we test the aggregation performances using different number of weak SVM classifiers. The results show that 5 weak SVMs is already enough for CBIR RF. Further increasing the number of classifiers to be aggregated does not improve the performance. Finally, we compare our

algorithm with existing methods in Fig. 2 (e-h). For RSMSVM, we randomly sample 30 features from the original feature vector. When we compare the RSMSVM with SVM, CSM, and BDA, we select 5 weak SVMs, and the computational complexity of RSMSVM is similar to SVM and CSM based RF. The results clearly show the superiority of the new algorithm over existing state-of-the-art algorithms.

5. Conclusion

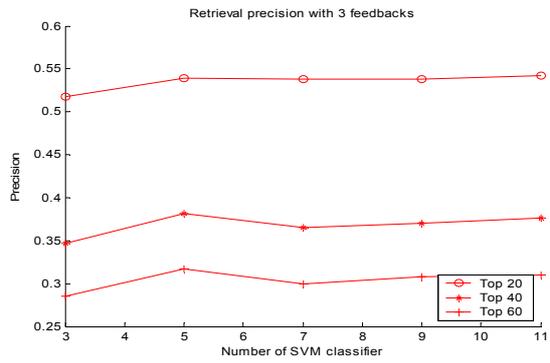
In this paper, we constructed a SVM based RF model using the random subspace method. The proposed algorithm can overcome the classifier unstable and over fitting problems effectively. The new algorithm shows significant improvement on both the effectiveness and robustness of the relevance feedback without increasing the computational complexity. This is demonstrated through extensive experiments on the Corel Photo database.

6. Acknowledgement

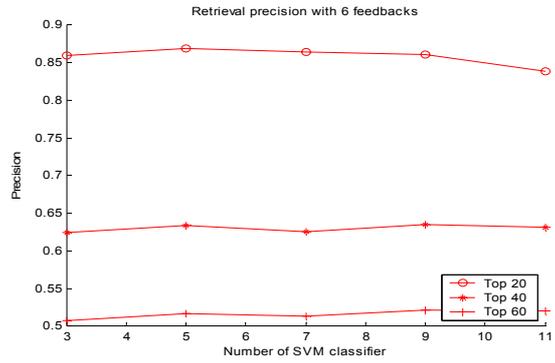
The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong SAR. (Project no. AoE/E-01/99).

7. References

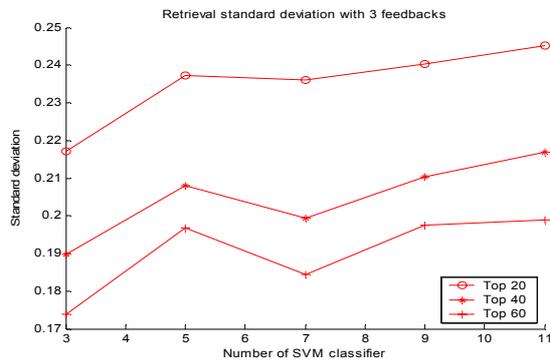
- [1] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," In Proc. IEEE ICIP, 1997.
- [2] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Trans. on PAMI, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.
- [3] Y. Chen, X. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," In Proc. IEEE ICIP, 2001.
- [4] X. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using biasmap," In Proc. IEEE CVPR, 2001.
- [5] L. Zhang, F. Lin, and B. Zhang, "Support vector machine learning for image retrieval," In Proc. IEEE ICIP, 2001.
- [6] P. Hong, Q. Tian, and T. S. Huang, "Incorporate support vector machines to content-based image retrieval with relevant feedback," In Proc. IEEE ICIP, 2000.
- [7] G. Guo, A. K. Jain, W. Ma, and H. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," IEEE Trans. on NN, vol. 12, no. 4, pp.811-820, 2002.
- [8] L. Breiman, "Bagging Predictors," Int. J. on Machine Learning, no. 24, pp 123-140, 1996.
- [9] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE Trans. On PAMI. vol. 20, no. 8, pp. 832-844, Aug. 1998.
- [10] J. Kittler, M. Hatef, P.W. Duin, and J. Matas, "On Combining Classifiers," IEEE Trans. On PAMI. Vol. 20, pp. 226-239, Mar. 1998.
- [11] V. Vapnik: The Nature of Statistical Learning Theory, Springer-Verlag, New York (1995).
- [12] D. Bahler and L. Navarro, "Methods for combining heterogeneous sets of classifiers," 17th Natl. Conf. on AAAI 2000.
- [13] B.S Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," IEEE Trans. on CSVT, pp. 703 -715, Jun 2001.
- [14] http://www.eleceng.ohio-state.edu/~maj/osu_svm/



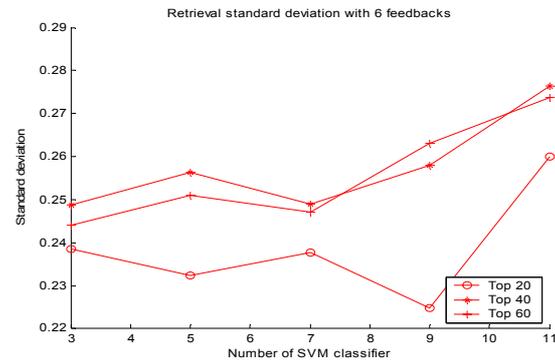
(a)



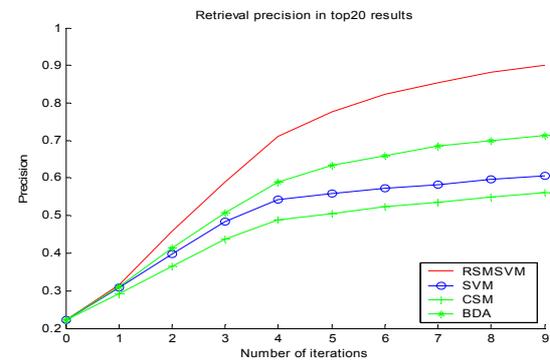
(b)



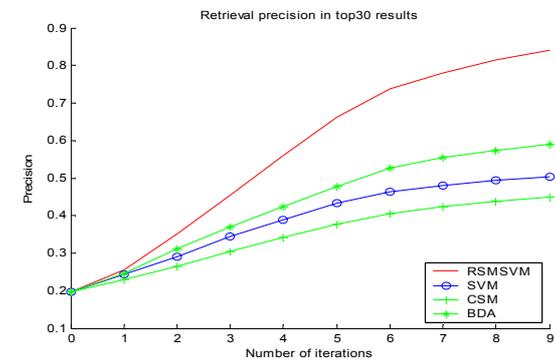
(c)



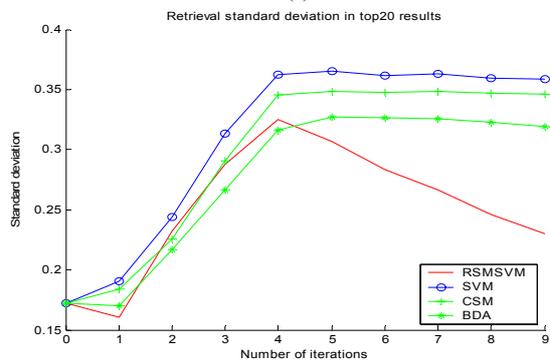
(d)



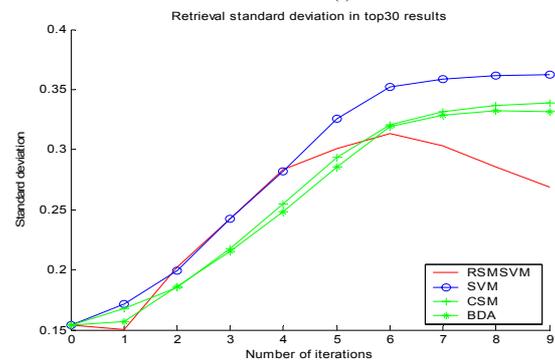
(e)



(f)



(g)



(h)

Figure 2. (a-d) test the aggregation performance with different number of weak SVM classifiers. (a,b) show precision curves. (c,d) are the corresponding SD curves. (e,f) compare the precision curves for different RF algorithms. (g,h) are the corresponding SD curves.