Random Sampling Based SVM for Relevance Feedback Image Retrieval

Dacheng Tao and Xiaoou Tang
Department of Information Engineering
The Chinese University of Hong Kong
{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. However, the performance of SVM based RF is often poor when the number of labeled positive feedback samples is small. This is mainly due to three reasons: 1. SVM classifier is unstable on small size training set; 2. SVM's optimal hyper-plane may be biased when the positive feedback samples are much less than the negative feedback samples; 3. overfitting due to that the feature dimension is much higher than the size of the training set. In this paper, we try to use random sampling techniques to overcome these problems. To address the first two problems, we propose an asymmetric bagging based SVM. For the third problem, we combine the random subspace method (RSM) and SVM for RF. Finally, by integrating bagging and RSM, we solve all the three problems and further improve the RF performance.

1. Introduction

Relevance feedback (RF) [1] is an important tool to improve the performance of content-based image retrieval (CBIR) [2]. In a RF process, the user first labels a number of relevant retrieval results as positive feedbacks and some irrelevant retrieval results as negative feedbacks. Then the system refines all retrieval results based on these feedbacks. The two steps are carried out iteratively to improve the performance of image retrieval system by gradually learning the user's perception.

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 examples. Discriminant learning has also been used as a feature selection method for RF [4]. These methods all have certain limitations. The method in [1] is only heuristic based. The density estimation method in [3] loses information contained in negative samples. The discriminant learning in [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 number 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. In SVM RF, the optimal hyper-plane is determined by the feedbacks. However, more often than not the users would only label a few images and cannot label each feedback accurately all the time. Hence the performance of the system may be poor with the inexactly labeled samples.

Second, in the RF process there are usually much more negative feedback samples than positive ones. Because of the imbalance of the training samples for the two classes, SVM's optimal hyper-plane will be biased toward the negative feedback samples. Consequently, SVMRF may mistake many query irrelevant images as relevant.

Finally, in RF, the size of the training set is much smaller than the dimension of the feature vector, thus may cause the over fitting problem. Because of the existence of noise, some features can only discriminant the positive and negative feedbacks but cannot discriminant the relevant or irrelevant images in the database. So the learned SVM classifier cannot work well for the remaining images in the database.

In order to overcome these problems, we design several new algorithms to improve the SVM based RF for CBIR. The key idea comes from the Classifier Committee Learning (CCL) [8-10]. 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. We use bagging and random subspace method to improve the SVM since they are especially effective when the original classifier is not very stable.

2. SVM in CBIR RF

SVM [11,12] is a very effective binary classification algorithm. Consider a linearly separable binary classification problem:

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

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,

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

where $\mathbf{x}$ is an input vector, $\mathbf{w}$ is an adaptive weight vector, and b is a bias. SVM finds the parameters $\mathbf{w}$ and b for the optimal hyper-plane to maximize the geometric margin $2/\|\mathbf{w}\|$, subject to

$$y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq +1 . \qquad (3)$$

The solution can be found through a Wolfe dual problem with Lagrangian multiplie α_i :

$$Q(\alpha) = \sum_{i=1}^{m}\alpha_i - \sum_{i,j=1}^{m}\alpha_i\alpha_j y_i y_j(\mathbf{x}_i \cdot \mathbf{x}_j)\Big/2 , \qquad (4)$$

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:

$$\mathbf{x}_i \cdot \mathbf{x}_j \rightarrow \phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) , \qquad (5)$$

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

$$Q(\alpha) = \sum_{i=1}^{m}\alpha_i - \sum_{i,j=1}^{m}\alpha_i\alpha_j d_i d_j K(\mathbf{x}_i \cdot \mathbf{x}_j)\Big/2 . \qquad (6)$$

Thus for a given kernel function, the SVM classifier is given by

$$F(\mathbf{x}) = \operatorname{sgn}\left(f(\mathbf{x})\right), \qquad (7)$$

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

In general, when $|f(\mathbf{x})|$ for a given pattern is high, the corresponding prediction confidence will be high. On the contrary, a low $|f(\mathbf{x})|$ of a given pattern means the pattern is close to the decision boundary and its corresponding prediction confidence will be low. Consequently, the output of SVM, $f(\mathbf{x})$ has been used to measure the dissimilarity [5,6] between a given pattern and the query image, in traditional SVM based CBIR RF.

3. CCL for SVMs

To address the three problems of SVM RF described in the introduction, we propose three algorithms in this section.

3.1. Asymmetric Bagging SVM

Bagging [8] strategy incorporates the benefits of bootstrapping and aggregation. Multiple classifiers can be generated by training on multiple sets of samples that are produced by bootstrapping, i.e. random sampling with replacement on the training samples. Aggregation of the generated classifiers can then be implemented by majority voting rule (MVR) [10].

Experimental and theoretical results have shown that bagging can improve a good but unstable classifier significantly [8]. This is exactly the case of the first problem of SVM based RF. However, directly using Bagging in SVM RF is not appropriate since we have only a very small number of positive feedback samples. To overcome this problem we develop a novel asymmetric Bagging strategy. The bootstrapping is executed only on the negative feedbacks, since there are far more negative feedbacks than the positive feedbacks. This way each generated classifier will be trained on a balanced number of positive and negative samples, thus solving the second problem as well. The Asymmetric Bagging SVM (ABSVM) algorithm is described in Table 1.

Table 1: Algorithm of Asymmetric Bagging SVM.

```
Input: positive training set $\mathbf{S}^+$, negative training set $\mathbf{S}^-$,
weak classifier $I$ (SVM), integer $T$ (number of
generated classifiers), $\mathbf{x}$ is the test sample.
1.  For $i = 1$ to $T$ {
2.      $\mathbf{S}_i^-$ = bootstrap sample from $\mathbf{S}^-$, with $|\mathbf{S}_i^-| = |\mathbf{S}^+|$.
3.      $C_i = I(\mathbf{S}_i^-, \mathbf{S}^+)$
4.  }
5.  $C^*(\mathbf{x}) = aggregation\{C_i(\mathbf{x}, \mathbf{S}_i^-, \mathbf{S}^+), 1 \leq i \leq T\}$.
Output: classifier $C^*$.
```

In ABSVM, the aggregation is implemented by Majority Voting Rule (MVR). The asymmetric Bagging strategy solves the classifier unstable problem and the training set unbalance problem. However, it cannot solve the small sample size problem. We will solve it by the Random Subspace Method (RSM) in the next section.

3.2. Random Subspace Method SVM

Similar to Bagging, RSM [9] also benefits from the bootstrapping and aggregation. However, unlike Bagging that bootstrapping training samples, RSM performs the bootstrapping in the feature space.

For SVM based RF, over fitting happens when the training set is relatively small compared to the high dimensionality of the feature vector. In order to avoid over fitting, we sample a small subset of features 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 over fitting problem. We then combine these SVMs to construct a more powerful classifier. Thus the over fitting problem is solved. The RSM based SVM (RSVM) algorithm is described in Table 2.

Table 2: Algorithm of RSM SVM.

```
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. $C^*(\mathbf{x}) = aggregation\{C_i(\mathbf{x}, \mathbf{F}_i), 1 \le i \le T\}$.
Output: classifier $C^*$.
```

3.3. Asymmetric Bagging RSM SVM

Since the asymmetric Bagging method can overcome the first two problems of SVMRF and the RSM can overcome the third problem of the SVMRF, we should be able to integrate the two methods to solve all the three problems together. So we propose an Asymmetric Bagging RSM SVM (ABRSVM) to combine the two. The algorithm is described in Table 3.

In order to explain why Bagging RSM strategy works, we derive the proof following a similar discussion on Bagging in [8].

Let $(y, \mathbf{x})$ be a data sample in the training set L with feature vector F, where y is the class label of the sample $\mathbf{x}$. L is drawn from the probability distribution P. Suppose $\varphi(\mathbf{x}, L, F)$ is the simple predictor (classifier) constructed by the Bagging RSM strategy, and the aggregated predictor is $\varphi_A(\mathbf{x}, P) = E_F E_L \varphi(\mathbf{x}, L, F)$.

Let random variables $(Y, \mathbf{X})$ be drawn from the distribution P independent of the training set L. The average predictor error, estimated by $\varphi(\mathbf{x}, L, F)$, is $e_a = E_F E_L E_{Y,\mathbf{X}}\left(Y - \varphi(\mathbf{X}, L, F)\right)^2$. The corresponding error estimated by the aggregated predictor is

$$e_A = E_{Y,\mathbf{X}}\left(Y - \varphi_A(\mathbf{X}, P)\right)^2. \tag{8}$$

Using the inequality $\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N}\left(z_{ij}\right)^2 \ge \left(\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N} z_{ij}\right)^2$, we have:

$$E_F E_L \varphi^2(\mathbf{X}, L, F) \ge \left(E_F E_L \varphi(\mathbf{X}, L, F)\right)^2 \tag{9}$$

$$E_{Y,X} E_F E_L \varphi^2(\mathbf{X}, L, F) \ge E_{Y,X} \varphi_A^2(\mathbf{X}, P) \tag{10}$$

Thus,

$$\begin{aligned} e_a &= E_{Y,X} Y^2 - 2E_{Y,X} Y \varphi_A + E_{Y,X} E_F E_L \varphi^2(\mathbf{X}, L, F) \\ &\ge E_{Y,X}\left(Y - \varphi_A\right)^2 = e_A \end{aligned}. \tag{11}$$

Therefore, the predicted error of the aggregated method is reduced. From the inequality, we can see that the more diverse is the $\varphi(\mathbf{x}, L, F)$, the more accurate is the aggregated predictor. In CBIR RF, the SVM classifier is unstable both for the training features and the training samples. Consequently, the Bagging RSM strategy can improve the performance.

Here we made an assumption that the average performance of all the individual classifier $\varphi(\mathbf{x}, L, F)$, trained on a subset of feature and training set replica is similar to a classifier, which use the full feature set and the whole subset training set. This can be true when the size of feature and training data subset is adequate to approximate the full set distribution. Even when this is not true, the drop of accuracy for each simple classifier may be well compensated in the aggregation process.

Table 3: Algorithm of Asymmetric Bagging RSM SVM.

```
Input: positive training set $\mathbf{S}^+$, negative training set $\mathbf{S}^-$, feature set $\mathbf{F}$, weak classifier $I$ (SVM), integer $T_s$ (number of Bagging classifiers), integer $T_f$ (number of RSM classifiers), $\mathbf{x}$ is the test sample.
1. For $j = 1$ to $T_s$ {
2.     $\mathbf{S}_j^-$ = bootstrap sample from $\mathbf{S}^-$.
3.         for $i = 1$ to $T_f$ {
4.             $\mathbf{F}_i$ = bootstrap sample from $\mathbf{F}$.
5.             $C_{i,j} = I\left(\mathbf{F}_i, \mathbf{S}_j^-, \mathbf{S}^+\right)$.
6.         }
7. }
8. $C^*(\mathbf{x}) = aggregation\left\{\begin{matrix} C_{ij}(\mathbf{x}, \mathbf{F}_i, \mathbf{S}_j^-, \mathbf{S}^+) \\ 1 \le i \le T_f, 1 \le j \le T_s \end{matrix}\right\}$
Output: classifier $C^*$.
```

Since Bagging RSM strategy can generate more diversified classifiers than using Bagging or RSM alone, it should outperform the two. In order to achieve maximum diversity, we choose to combine all generated classifiers in parallel as shown in Figure 1. This is better than combining Bagging or RSM first then combing RSM or Bagging.

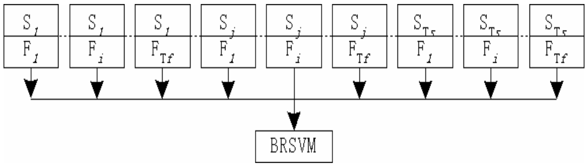

Figure 1. Aggregation structure of the ABRSVM.

For a given test sample, we first recognize it by all $T_f \cdot T_s$ weak SVM classifiers:

$$\left\{ C_{ij} = C\left(F_i, S_j\right) | 1 \le i \le T_f, 1 \le j \le T_s \right\}. \tag{12}$$

Then, an aggregation rule is used to integrate all the results from the weak classifiers for final classification of the sample as relevant or irrelevant.

3.4 Dissimilarity Measure

For a given sample, we first use the MVR to recognize it as query relevant or irrelevant. Then we measure the dissimilarity between the sample and the query as the output of the individual SVM classifier, which gives the same label as the MVR and produces the highest confidence value (the absolute value of the decision function of the SVM classifier).

4. Image Retrieval System

To evaluate the performance of the proposed algorithms, we develop the following general CBIR system with RF. In the system, we can use any RF algorithm for the block "Relevance Feedback Model".

In Figure 2, when a query image is input, the low-level features are extracted. Then, all images in the database are sorted based on a similarity metric (here we use Euclidean distance). The user labels some top images as positive and negative feedbacks. Using these feedbacks, a RF model is trained based on a SVM. Then the similarity metric is updated based on the RF model. All images are resorted by the updated similarity metric. The RF procedure will be iteratively executed until the user is satisfied with the outcome.

In our retrieval system, three main features, color, texture, and shape are extracted to represent the image. For color feature, we use the color histogram [13] in HSV color space. Here, the color histogram is quantized into 256 levels. Hue, Saturation and Value are quantized into 8, 8, and 4 bins respectively. Texture is extracted from Y component in the YCrCb space by pyramid wavelet transform (PWT) with Haar wavelet. The mean value and standard deviation are calculated for each sub-band at each decomposition level. The feature length is $2 \times 4 \times 3$. For shape feature, edge histogram [14] is calculated on Y component in the YCrCb color space. Edges are grouped into four categories, horizontal, 45 diagonal, vertical, and 135 diagonal. We combine the color, texture, and shape features into a feature vector, and then we normalize each feature to a normal distribution.

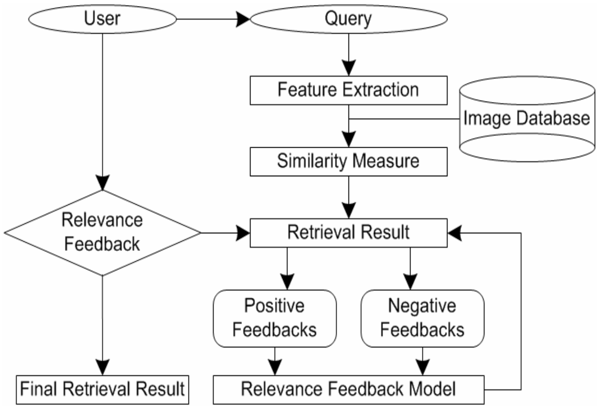

Figure 2. Flowchart of the image retrieval system.

5. Experimental Results

In this section, we compare the new algorithms with existing algorithms through experiments on 17, 800 images of 90 concepts from the Corel Photo Gallery. The experiments are simulated by a computer automatically. First, 300 queries are randomly selected from the data, and then RF is automatically done by the computer: all query relevant images (i.e. images of the same concept as the query) are marked as positive feedbacks in the top 40 images and all the other images are marked as negative feedbacks. In general, we have about 5 images as positive feedbacks. The procedure is close to the real circumstances, because the user typically would not like to click on the negative feedbacks. Thus requiring the user to mark only the positive feedbacks in top 40 images is reasonable.

In this paper, 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 values. 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 0 feedback refers to the retrieval based on Euclidean distance measure without RF.

We compare all the proposed algorithms with the original SVM based RF [5] and the constrained similarity measure SVM (CSVM) based RF [7]. 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 [15] MatLab$^{\text{TM}}$ toolbox) for all the algorithms. The performances of all the SVM algorithms are stable over a range of ρ .

5.1. Performance of Asymmetric Bagging SVM

Figure 3 shows the precision and SD values when using different number of SVMs in ABSVM. The results show that the number of SVMs will not affect the performance of the asymmetric Bagging method.

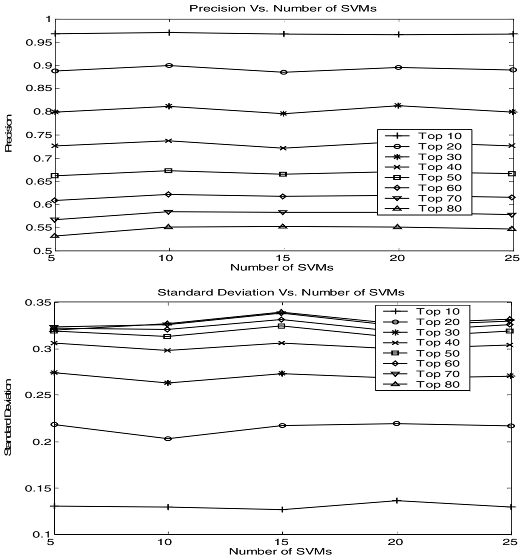

Figure 3. Asymmetric Bagging SVM based RF.

The second experiment compares the performances of the ABSVM using 5 weak SVM classifiers and the standard SVM and CSVM based RF. The experimental results are shown in Figure 5.

From these experiments, we can see that 5 weak SVMs are enough for ABSVM, and ABSVM clearly outperforms the SVM and CSVM. The precision curve of ABSVM is higher than that of SVM and CSVM, and SD curve is lower than that of SVM and CSVM.

5.2. Performance of RSM SVM

Figure 4 shows the precision and SD values when using different number of SVMs of RSVM. The results show that the number of SVMs does not affect the performance of RSVM.

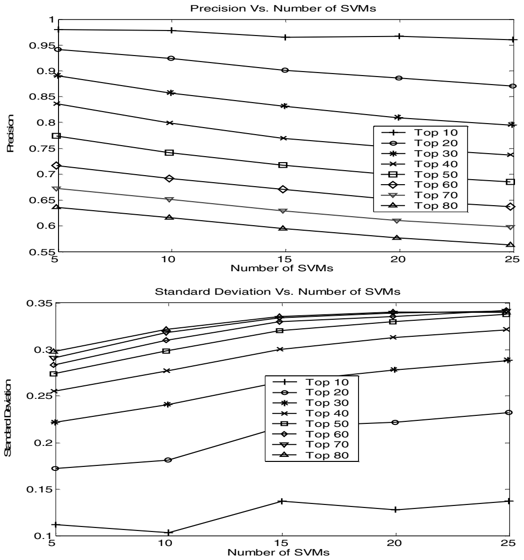

Figure 4. RSM SVM based RF.

The second experiment compares the RSVM using 5 weak SVM classifiers with the standard SVM and CSVM based RF. The experimental results are also shown in Figure 5.

The results demonstrate that 5 weak SVMs are enough for the RSVM, and RSVM outperforms the SVM and CSVM.

5.3. Performance of Asymmetric Bagging RSM SVM

This experiment evaluates the performance of the proposed ABRSVM, ABSVM, and RSVM based RF. In this experiment, we chose $T_s = 5$ for ABSVM, $T_f = 5$ for RSVM, and $T_s = T_f = 5$ for ABRSVM.

The results in Figure 5 show that the ABRSVM gives the best performance followed by RSVM then ABSVM. They all outperform SVM and CSVM.

5.4. Computational Complexity

To verify the efficiency of the proposed algorithms, we record the computational time when conducting the experiments. The ratio for the time used by different methods are SVM: CSVM: ABSVM: RSVM: ABRSVM = 25: 25: 11: 3: 5. This is show that the new SVM based algorithms are much more efficient than existing SVM based algorithms.

6. Conclusion

In this paper, we design a new Asymmetric Bagging Random Subspace Method for SVM based RF. The proposed algorithm can address the classifier unstable problem, the unbalanced training set problem, and the small sample size problem effectively. Extensive experiments on a Corel Photo database with 17, 800 images show that the new algorithm can improve the performance of relevance feedback significantly.

7. 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).

8. 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.W.M. 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, July 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, no. 3, pp. 226-239, Mar. 1998.

[11] Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag, New York (1995).

[12] J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Int. J. on. DMKD 2, pp.121-167, 1998.

[13] M.J. Swain and D.H. Ballard, "Color indexing," IJCV, 7(1):11--32, 1991.

[14] B.S Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," IEEE Trans. on CSVT, Vol. 11, pp. 703 -715, Jun 2001.

[15].http://www.eleceng.ohio-state.edu/~maj/osu_svm/

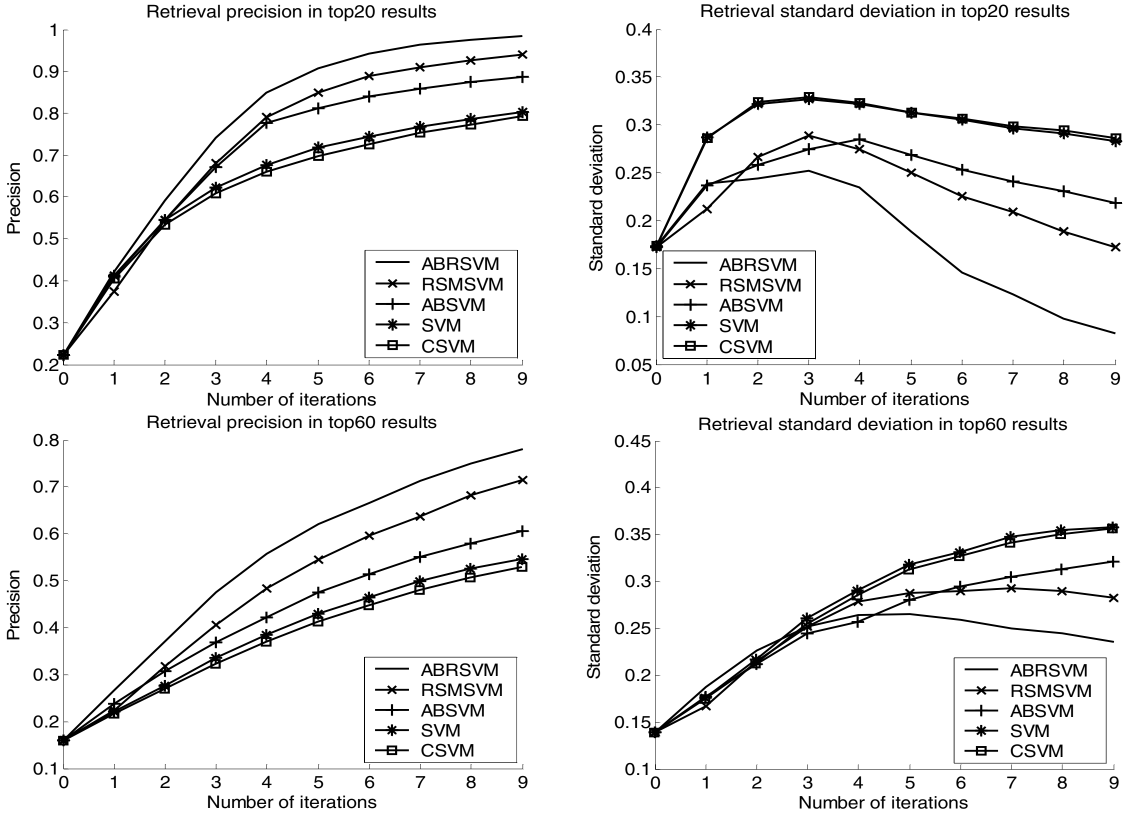

Figure 5. Performance of all proposed algorithms compared to existing algorithms. The algorithms are evaluated over 9 iterations.