Biased Support Vector Machine for Relevance Feedback in Image Retrieval

Chu-Hong Hoi, Chi-Hang Chan, Kaizhu Huang, Michael R. Lyu and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong SAR
E-mail: {chhoi,chchan,kzhuang,lyu,king}@cse.cuhk.edu.hk

Abstract— **Recently, Support Vector Machines (SVMs) have been engaged on relevance feedback tasks in content-based image retrieval. Typical approaches by SVMs treat the relevance feedback as a strict binary classification problem. However, these approaches do not consider an important issue of relevance feedback, i.e. the imbalanced dataset problem, in which the negative instances largely outnumber the positive instances. For solving this problem, we propose a novel technique to formulate the relevance feedback based on a modified SVM called Biased Support Vector Machine (Biased SVM or BSVM). Mathematical formulation and explanations are provided for showing the advantages. Experiments are conducted to evaluate the performance of our algorithms, in which promising results demonstrate the effectiveness of our techniques.**

I. Introduction

Content-based image retrieval (CBIR) has been widely investigated in the past decade [18]. Different from traditional image retrieval approaches based on keywords annotation, CBIR systems employ the visual content of images, such as color, shape, and texture features, to index the images [15]. At the early stage of CBIR research, major efforts focused on the feature identification and expression for the best representation of the content of images [18]. However, early CBIR systems with heuristic feature selections and rigid distance weighting did not achieve satisfactory performance. Later, researchers noticed and recognized the difficulties in CBIR, i.e. the semantic gap problem between low-level features and high-level concepts, and the subjectivity of human perception [15].

Relevance feedback was introduced to attack the semantic gap and the subjectivity of human perception problems in CBIR [15]. It has been shown as a powerful tool to improve the retrieval performance of CBIR systems [15], [6]. Recently, classification techniques are introduced to attack relevance feedback tasks [9], [24], [5], [22], in which SVM-based techniques are considered as the most promising techniques. However, previous studies on relevance feedback by SVMs normally treat the problem as a strict binary classification problem without noticing an important issue of relevance feedback, i.e. the imbalanced dataset problem, in which the negative instances significantly overnumber the positive ones [10]. This imbalanced dataset problem may cause the positive instances to be overwhelmed by the negative instances. In order to attack this problem, we propose a modified Support Vector Machine [17], [21], [13] called Biased Support Vector Machine (Biased SVM or BSVM) which can better model the relevance feedback problem and reduce the performance degradation caused by the imbalanced dataset problem.

The rest of the paper is organized as follows. In Section II, we review some related research efforts on relevance feedback and address their disadvantages. Then we provide a brief introduction for two-class SVM and one-class SVM in Section III. In Section IV, we present and formulate our proposed Biased SVM algorithm. We then formulate the relevance feedback technique employing Biased SVM and show the benefits compared with the conventional techniques in Section V. Experiments, performance evaluations, and discussions are given in Section VI. Finally, Section VII concludes our work.

II. Related Work

In the past years, relevance feedback techniques have evolved from early heuristic weighting adjustment techniques to various machine learning techniques recently [14], [15], [8], [5], [6]. In [8], Self-organizing Map (SOM) was proposed to construct the relevance feedback algorithm. Besides the SOM, many popular machine learning techniques were also suggested, such as Decision Tree [9], Artificial Neural Network [12], and Bayesian learning [3], etc. Moreover, many state-of-the-art classification techniques were proposed to attack the relevance feedback, such as Nearest-Neighbor classifiers [25], Bayesian classifiers [20] and Support Vector Machines [2], [5], [22], etc. Among them, SVM-based techniques are the most promising and effective techniques to solving the relevance feedback tasks.

Typical relevance feedback approaches by SVMs are based on strict binary classifications [5], [22] or one-class classifications [2]. However, the strict binary classifications do not consider the imbalanced dataset problem in relevance feedback. The one-class technique seems to avoid the imbalanced dataset problem, but the relevance feedback work cannot be done properly without the help of negative information [26]. In order to fuse the negative information, we propose the Biased Support Vector Machine derived from one-class SVM to construct the relevance feedback technique in CBIR.

III. Support Vector Machines

We here briefly introduce the basic concepts of regular two-class SVM [23] and one-class SVM (1-SVM) [17], [21], [13].

0-7803-8359-1/04/$20.00 ©2004 IEEE

SVMs implement the principle of structural risk minimization by minimizing Vapnik-Chervonenkis dimensions [23]. On pattern classification problems, SVMs provide very good generalization performance in empirical applications.

Let us discuss SVMs in a binary classification case. Generally speaking, a binary classification problem can be formalized as a task to estimate a function $f : \mathbb{R}^m \rightarrow \{-1,+1\}$ based on independent identically distributed (i.i.d.) data $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_l, y_l) \in \mathcal{X} \times Y, \quad Y = \{-1,+1\}$ [16]. Here, the training instances are vectors in some space $\mathcal{X} \subseteq \mathbb{R}^m$ and l is the number of training instances. The goal of the learning process is to find an optimal decision function f which can classify the unseen data $\mathbf{x}$ correctly. In theory, the goal is to find the optimal function f with the smallest risk

$$R[f] = \int_{\mathcal{X}} L\,(f, \mathbf{x}, y)\; dP(\mathbf{x}, y)\,, \tag{1}$$

where P is the probability measure for the generation of the training data and L is a loss function. In the simplest form, the goal of learning in SVMs is to find a hyperplane with the maximum margin (see Fig. 1). The vectors closest to the hyperplane are called support vectors.

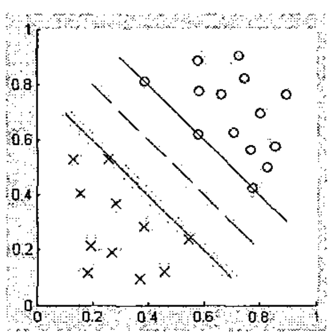


Fig. 1. The linear separating hyperplane of SVMs for separable data: The circles and crosses are called positive instances and negative instances, respectively. The circles and the crosses on the two solid lines are called support vectors. The dashed line between the two solid lines is called the decision hyperplane.

More generally, the training data in the original space $\mathcal{X}$ can be projected to a higher dimensional feature space $\mathcal{F}$ which is spanned by a mapping function Φ. The mapping function corresponds to a Mercer kernel $k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}))$ which implicitly computes the dot product in $\mathcal{F}$. Hence, the goal of SVMs is to find the optimal separating hyperplane depicted by a vector $\mathbf{w}$ in the feature space $\mathcal{F}$ with the following form

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}). \tag{2}$$

The task for finding the optimal hyperplane turns out to be solving the primal optimization problem in the form of soft margin SVMs (also called ν-SVM [16]):

$$\min_{\mathbf{w}\in\mathcal{F}} \quad \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum \xi_i \tag{3}$$

$$s.t. \quad y_i(\mathbf{w} \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i \tag{4}$$

$$\xi_i \geq 0, \rho \geq 0, \quad i = 1, \ldots, l, \tag{5}$$

where ξ_i represent the margin errors for the non-separable training data. When the margin errors $\xi = 0$, one can show that the two classes are separated by a margin with $2\rho/\|\mathbf{w}\|$ from Eq. (4). By introducing the Lagrange multipliers, the optimization problem can be shown with the dual form as follows [16], [23]:

$$\max_{\alpha} \quad -\frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$s.t. \quad \sum_i \alpha_i y_i = 0\,, \tag{6}$$

$$0 \leq \alpha_i \leq \frac{1}{l}\,, \tag{7}$$

$$\sum_i \alpha_i \geq \nu, \quad i = 1, 2, \ldots, l\,. \tag{8}$$

1-SVMs are derived from classical SVMs for solving density estimation problems. In typical formulations of 1-SVMs, only positive instances are considered for estimating the density of the data. There are two kinds of different formulations of 1-SVMs in the literature [17], [21]. Here, we choose to illustrate the sphere-based approach with an explicit and good geometric property. Fig. 2 illustrates an example of 1-SVMs.

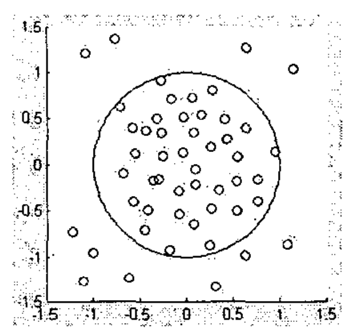


Fig. 2. The sphere hyperplane in 1-SVM for constructing the smallest soft sphere that contains most of the positive instances. The circles represent the positive instances.

The optimal decision function of the sphere-based approach of 1-SVMs can be found by solving the optimization problem as follows [17], [21]:

$$\min_{R\in\mathbb{R}, \mathbf{c}\in\mathcal{F}} \quad R^2 + \frac{1}{\nu l}\sum_i \xi_i \tag{9}$$

$$s.t. \quad \|\Phi(\mathbf{x}_i) - \mathbf{c}\|^2 \leq R^2 + \xi_i, \tag{10}$$

$$\xi_i \geq 0 \quad i = 1, \ldots, l\,. \tag{11}$$

Here, $\nu \in (0, 1]$ is a parameter to control the tradeoff between the radius of the hyper-sphere and the number of positive training instances.

IV. Biased Support Vector Machine

In order to incorporate the negative information, we propose the Biased Support Vector Machine derived from 1-SVMs for overcoming the imbalance problem of relevance feedback tasks. Our strategy is to describe the data by employing a pair of sphere hyperplanes in which the inner one captures most of the positive instances while the outer one pushes out the negative instances. Therefore, the goal of our problem is to find an optimal sphere hyperplane which not only can contain most of the positive data but also can push most of the negative data out of the sphere. The problem is visually illustrated in Fig. 3. The dashed sphere in the figure is the desired sphere-hyperplane in our goal. The task can be formulated as an optimization problem and the mathematical formulation of our technique are given as follows.

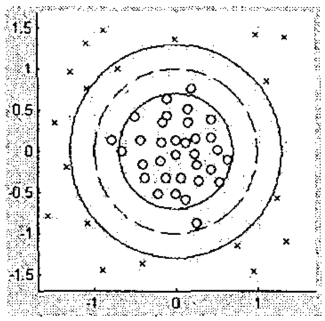


Fig. 3. The sphere hyperplane of BSVM. The circles and the crosses represent the positive instances and the negative instances, respectively. The dashed sphere is the decision hyperplane.

Let us consider the following training data:

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l) \in \mathbb{R}^m \times Y, \quad Y = \{-1, +1\} \qquad (12)$$

where l is the number of training instances and m is the dimension of the input space.

The objective function for finding the optimal sphere-hyperplane can be formulated below:

$$\min_{R\in\mathbb{R},\rho\in\mathbb{R},\mathbf{c}\in\mathcal{F}} \quad bR^2 - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i \qquad (13)$$

$$s.t. \quad y_i(||\Phi(\mathbf{x}_i) - \mathbf{c}||^2 - R^2) \leq -\rho + \xi_i \;, (14)$$

$$b > 0\,, \rho \geq 0\,, 0 \leq \nu \leq 1, \xi_i \geq 0\,, \qquad (15)$$

where ξ_i are the slack variables for margin errors, $\Phi(\mathbf{x}_i)$ is the mapping function, $\mathbf{c}$ is the center of the optimal sphere-hyperplane, and b is a parameter to control the bias.

The optimization task can be solved by introducing the Lagrange multipliers:

$$L(R,\xi,c,\alpha,\beta,\lambda) = bR^2 - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\beta_i\xi_i - \lambda\rho + \sum_{i=1}^{l}\alpha_i[y_i(||\Phi(\mathbf{x}_i) - \mathbf{c}||^2 - R^2) + \rho - \xi_i]\,. \qquad (16)$$

Let us take the partial derivative of L with respect to R, ξ_i, c and ρ, respectively. By setting their partial derivatives to zero, we obtain the following equations:

$$2R(b - \sum_{i=1}^{l} y_i\alpha_i) = 0 \;\Rightarrow\; \sum_{i=1}^{l} y_i\alpha_i = b\,; \qquad (17)$$

$$\frac{1}{\nu l} - \alpha_i - \beta_i = 0, \;\Rightarrow\; 0 \leq \alpha_i \leq \frac{1}{\nu l}\,; \qquad (18)$$

$$\sum_{i=1}^{l} 2\alpha_i y_i(\Phi(\mathbf{x}_i) - \mathbf{c}) = 0 \;\Rightarrow\; \mathbf{c} = \frac{1}{b}\sum_{i=1}^{l}\alpha_i y_i \Phi(\mathbf{x}_i) \qquad (19)$$

$$-1 + \sum_{i=1}^{l}\alpha_i - \lambda = 0 \;\Rightarrow\; \sum_{i=1}^{l}\alpha_i \geq 1\,. \qquad (20)$$

By substituting the above derived results to the objective function in Eq. (16), the dual of the primal optimization can be shown to take the form

$$\max_{\alpha} \quad \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}_i) - \frac{1}{b}\sum_{i,j}\alpha_i\alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \qquad (21)$$

$$s.t. \quad \sum_i \alpha_i y_i = b\,, \qquad (22)$$

$$0 \leq \alpha_i \leq \frac{1}{\nu l}\,, \qquad (23)$$

$$\sum_i \alpha_i \geq 1, \quad i = 1, 2, \ldots, l\,. \qquad (24)$$

This dual problem can be solved by Quadratic Programming (QP) techniques [11]. Then, the resulting decision function takes the form

$$f(x) \;=\; sgn(R^2 - ||\Phi(\mathbf{x}) - \mathbf{c}||^2)\,, \qquad (25)$$

where $\mathbf{c}$ can be obtained from Eq. (19) and R can be solved by support vectors. Based on the decision function, we can know the instances inside the sphere hyperplane will be predicted as positive, and negative otherwise.

V. Relevance Feedback Using BSVM

A. Advantages of BSVM in Relevance Feedback

From the above formulations, one may see that the optimization in Eq. (21) is similar to the one in the ν-SVM. Now, we show the mathematical differences compared with regular SVMs and the advantages of our BSVM from a geometric perspective for solving the relevance feedback problems.

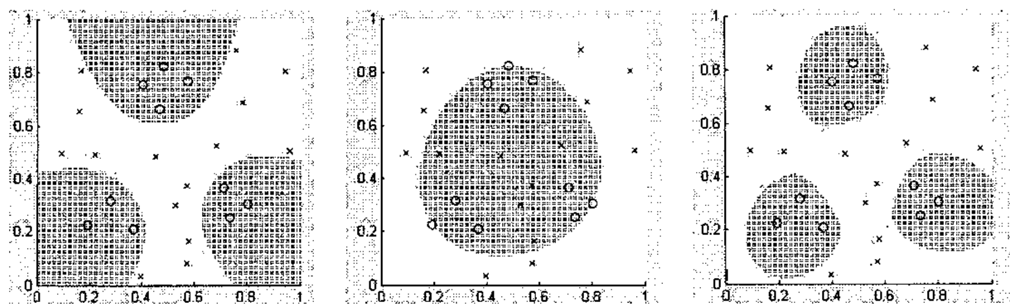
From the results of mathematic deduction in the optimization function, we see that BSVM is with the following constraint from Eq. (22)

$$\sum_i \alpha_i y_i = b\,. \qquad (26)$$

When replacing y_i with $+1$ for the positive class and -1 for the negative one, the constraint can be written as

$$\sum_{i\in S^+}\alpha_i - \sum_{i\in S^-}\alpha_i = b\,, \qquad (27)$$

where S^+ denotes the positive class and S^- denotes the negative one. However, in the regular SVMs (ν-SVM), the



(a) SVM (ν-SVM) (b) 1-SVM (c) BSVM

Fig. 4. Decision boundaries of three classification methods with the same kernels (RBF) and parameters (γ=0.1): (a) ν-SVM, (b) 1-SVM, (c) BSVM. The circles and crosses represent the positive and negative instances, respectively. The boundaries of the shadow regions represent the decision boundaries.

constraint is with the form

$$\sum_{i \in S^+} \alpha_i - \sum_{i \in S^-} \alpha_i = 0 \ . \qquad (28)$$

The difference indicates that the weight allocated to the positive support vectors in BSVM will be larger than the negative ones when setting a positive bias factor b. This can be useful for solving the imbalance dataset problem. However, ν-SVM treat the two classes without any bias, which is not effective enough to model the relevance feedback problem.

Moreover, we can also see the difference from the geometric perspective. Fig. 4 provides a comparison of the decision boundaries of regular SVM, 1-SVM and BSVM on the synthetic data with the same kernels (Radial Basis Function) and parameters (γ=0.1). We can see that the geometric property of BSVM is better than ν-SVM and 1-SVM. BSVM can describe the data in a cluster behavior by the sphere-based boundary and can flexibly control the weight of the positive class for the imbalanced dataset problem by setting a bias factor. Therefore, compared with the ν-SVM and 1-SVM, BSVM is more reasonable and more effective to model the relevance feedback tasks.

B. Relevance Feedback Algorithm By BSVM

From the above comparisons, we have shown the benefits of BSVM for solving relevance feedback issues. Here, we describe how to formulate the relevance feedback algorithm by employing the BSVM technique. Applying SVMs based techniques in relevance feedback is similar to the classification task. However, the relevance feedback needs to construct an evaluation function to produce the relevance value of the retrieval instances. From the decision function in Eq. (30), we build the evaluation function by substituting the derived result in Eq. (19)

$$\begin{aligned} f(\mathbf{x}) &= R^2 - ||\Phi(\mathbf{x}) - \mathbf{c}||^2 \\ &= R^2 - ||\Phi(\mathbf{x}) - \frac{1}{b}\sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x}_i)||^2 \ , \qquad (29) \end{aligned}$$

where the radius R can be solved by a set of support vectors. However, for the relevance evaluation purpose, constant values can be eliminated. Then, the evaluation function can be shown to take the following concise form

$$f(\mathbf{x}) \ = \ \frac{2}{b}\sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}) \ . \qquad (30)$$

Once the parameters α_i are solved in Eq. (21), the evaluation function can be constructed. Consequently, we can rank the images based on the scores of the evaluation function $f(\mathbf{x})$. The images with higher scores will be more likely to be chosen as the targets.

VI. Experiments

In the experiments, we compare the performance of three different algorithms for relevance feedback: ν-SVM, 1-SVM and our proposed BSVM. The experiments are evaluated both on a synthetic dataset as well as two real-world image datasets.

A. Datasets

1) A Synthetic Dataset: We generate a synthetic dataset to simulate the real-world image dataset. The dataset consists 40 categories, each of which contains 100 data points randomly generated by 7 Gaussians in a 40-dimensional space. The means and covariance matrices for the Gaussians in each category are randomly generated in the range of [0,10].

2) COREL Image Datasets: The real-world images are chosen from the COREL image CDs. We organize two datasets which contain various images with different semantic meanings, such as *antique*, *aviation*, *balloon*, *botany*, *butterfly*, *car* and *cat*, etc. One of the datasets is with 20 categories (20-Cat) and another is with 50 categories (50-Cat). Each category includes 100 images belonging to the same semantic class.

B. Image Representation

For the real-world image retrieval, the image representation is an important step for evaluating the relevance feedback algorithms. We extract three different features to represent the images: color, shape and texture.

The color feature engaged is the color moment since it is closer to human natural perception. We extract three moments: color mean, color variance, and color skewness in each color channel (H, S, and V), respectively. Thus, a 9-dimensional color moment is employed as the color feature in our experiments.

We employ the edge direction histogram as the shape feature in our experiments [7]. Canny edge detector is applied to obtain the edge images. From the edge images, the edge direction histogram can then be computed. The edge direction histogram is quantized into 18 bins of 20 degrees each, hence an 18-dimensional edge direction histogram is used to represent the edge feature.

We apply the wavelet-based texture feature for its effectiveness [19]. We perform the Discrete Wavelet Transformation (DWT) on the gray images employing a Daubechies-4 wavelet filter [19]. In total, we perform 3-level decompositions and obtain ten subimages in different scales and orientations. Then, we choose nine subimages with most of the texture information and compute the entropy of each subimage. Hence, a 9-dimensional wavelet-based texture feature is obtained to describe the texture information for each image.

C. Experimental Results

In the following, we present the experimental results by three algorithms on both the synthetic data and the real-world images. For the purpose of objective measure of performance, we assume that the query judgement is defined on the image categories [22]. The metric of evaluation is the *Average Precision* which is defined as the average ratio of the number of relevant images of the returned images over the total number of the returned images.

In the experiments, a category is first picked from the database randomly, and this category is assumed to be the user's query target. The system then improves retrieval results by relevance feedbacks. In each iteration of the relevance feedback process, 10 instances are picked from the database and labelled as either positive or negative based on the ground truth of the database. For the first iteration, two positive instances and eight negative instances are randomly picked, and all three methods are applied with the same set of initial data points. For the iterations afterward, each method selects 10 instances closest to the decision boundaries. In the retrieval process, the instances in the positive region are selected and ranked by their distances from the boundaries. The precision of each method is then recorded, and the whole process is repeated for 200 times to produce the average precision in each iteration for each method.

The algorithms implemented in our experiments are based on modifying the codes in the *libsvm* library [1]. We notice that the experimental settings are important to impact on the evaluation results. To enable an objective measure of performance without bias, we choose the same kernels and parameters for all the settings. All the kernels are based on Radial Basis Function (RBF) which outperforms other kernels in the experiments.

The first evaluation is on the synthetic dataset. Fig. 5 shows the evaluation results of top-30 returned results. We can observe that BSVM outperforms the other approaches. The 1-SVM achieves the worst performance without considering the negative information.

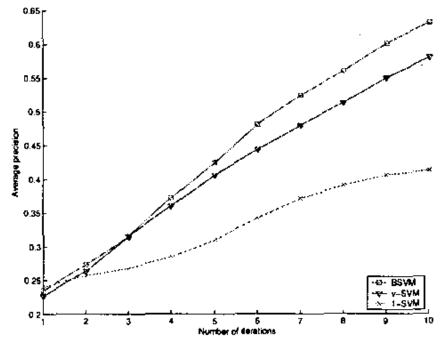


Fig. 5. Experimental results on the synthetic dataset

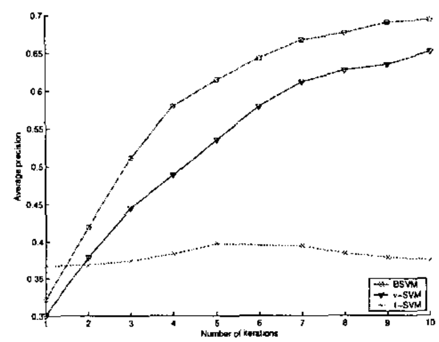


Fig. 6. Experimental results on the 20-Cat image dataset

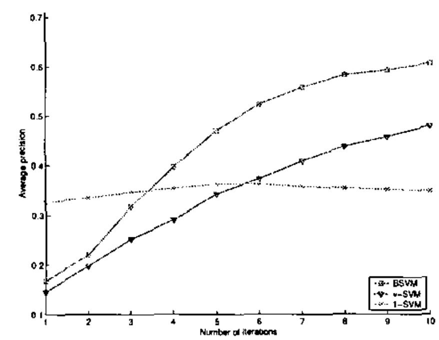


Fig. 7. Experimental results on the 50-Cat image dataset

The second evaluation is on the real-world datasets. Fig. 6 and Fig. 7 show the evaluation results on the 20-Cat dataset and 50-Cat dataset, respectively. From the results on the real-world datasets, we can see our proposed BSVM also outperforms the other approaches. However, we notice that the performance of 1-SVM in the beginning feedback steps

TABLE I

AVERAGE PRECISION AFTER 10 ITERATIONS

Methods	Top20@20-Cat	Top30@20-Cat	Top50@20-Cat
ν-SVM	0.656	0.648	0.608
1-SVM	0.401	0.396	0.346
BSVM	0.713	0.694	0.650
Methods	Top20@50-Cat	Top30@50-Cat	Top50@50-Cat
ν-SVM	0.487	0.480	0.465
1-SVM	0.376	0.358	0.344
BSVM	0.639	0.614	0.588

is better than that of other approaches. The reason is that 1-SVM can reach the enclosed positive region quickly, but it cannot be further improved without the help of the negative information in subsequent steps. In order to observe the detailed comparison of the three methods after 10-iterations, we list the retrieval results in Table I. From the results, we can also see the similar results matching the above comparisons.

D. Discussions

From the experimental results, we see that our proposed BSVM performs better than the regular SVM approaches. The typical approaches by SVMs (ν-SVM) without considering the bias in the retrieval tasks is not appropriate in solving the relevance feedback problem. We also see that regular one-class SVMs do not consider the negative information which cannot learn the feedback well. Furthermore, we know there are other methods to address the imbalanced dataset problem in literature [10], [4]. We may also consider to include them in our scheme in the future. Nevertheless, we have observed the promising results in demonstrating the effectiveness of our proposed BSVM technique for the relevance feedback problems.

VII. CONCLUSIONS

In this paper, we investigate SVM techniques for solving the relevance feedback problems in CBIR. We address the imbalanced dataset problem in relevance feedback and propose a novel relevance feedback technique with Biased Support Vector Machine. The advantages of our proposed techniques are explained and demonstrated. We perform the experiments both on synthetic data and real-world image datasets to evaluate the performance. The experimental results demonstrate that our Biased SVM based relevance feedback algorithm is effective and promising for improving the retrieval performance in CBIR.

ACKNOWLEDGMENT

The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK4182/03E and Project No. CUHK4351/02E).

REFERENCES

[1] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[2] Y. Chen, X. Zhou, and T. S. Huang. One-class svm for learning in image retrieval. In *Proc. IEEE International Conference on Image Processing (ICIP'01)*, Thessaloniki, Greece, 2001.

[3] I. J. Cox, M. Miller, T. Minka, and P. Yianilos. An optimized interaction strategy for bayesian relevance feedback. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, pages 553–558, Santa Barbara, CA, USA, 1998.

[4] C. Drummond and R. Holte. Class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. In *Workshop on International Conference on Machine Learning (ICML'2003)*, 2003.

[5] P. Hong, Q. Tian, and T. S. Huang. Incorporate support vector machines to content-based image retrieval with relevant feedback. In *Proc. IEEE International Conference on Image Processing (ICIP'00)*, Vancouver, BC, Canada, 2000.

[6] T. S. Huang and X. S. Zhou. Image retrieval by relevance feedback: from heuristic weight adjustment to optimal learning methods. In *Proc. IEEE International Conference on Image Processing (ICIP'01)*, Thessaloniki, Greece, October 2001.

[7] A. K. Jain and A. Vailaya. Shape-based retrieval: a case study with trademark image database. *Pattern Recognition*, (9):1369–1390, 1998.

[8] J. Laaksonen, M. Koskela, and E. Oja. Picsom: Self-organizing maps for content-based image retrieval. In *Proc. International Joint Conference on Neural Networks (IJCNN'99)*, Washington, DC, USA, 1999.

[9] S. MacArthur, C. Brodley, and C. Shyu. Relevance feedback decision trees in content-based image retrieval. In *Proc. IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 68–72, 2000.

[10] M. Maloof. Learning when data sets are imbalanced and when costs are unequal and unknown. In *Workshop on ICML'2003*, 2003.

[11] O. L. Mangarasian. *Nonlinear Programming*. McGraw Hill, New York, 1969.

[12] F. Qian, M. Li, W.-Y. Ma, F. Lin, and B. Zhang. Alternating feature spaces in relevance feedback. In *3rd Intl Workshop on Multimedia Information Retrieval (MIR2001)*, 2001.

[13] G. Ratsch, S. Mika, B. Scholkopf, and K.-R. Muller. Constructing boosting algorithms from svms: an application to one-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1184–1199, 2002.

[14] Y. Rui, T. S. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. In *Proc. IEEE International Conference on Image Processing (ICIP'97)*, pages 815–818, Washington, DC, USA, October 1997.

[15] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*.

[16] B. Scholkof, A. J. Smola, R. Williamson, and P. Bartlett. New support vector algorithms. *Neural Computation*, 12:1083–1121, 2000.

[17] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Constructing boosting algorithms from svms: an application to one-class classification. *Neural Computation*, 13(7):1443–1471, 2001.

[18] 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. Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[19] J. Smith and S.-F. Chang. Automated image retrieval using color and texture. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, November 1996.

[20] Z. Su, H. Zhang, and S. Ma. Relevant feedback using a bayesian classifier in content-based image retrieval. In *SPIE Electronic Imaging 2001*, January 2001.

[21] D. Tax and R. Duin. Data domain description by support vectors. In *Proc. European Symp. Artificial Neural Network (ESANN'99)*, pages 251–256, Bruges, Belgium, 1999.

[22] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *Proc. ACM Multimedia*, pages 107–118, 2001.

[23] V. N. Vapnik. Statistical learning theory. 1998.

[24] N. Vasconcelos and 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*, South Carolina, USA, 2000.

[25] P. Wu and B. S. Manjunath. Adaptive nearest neighbour search for relevance feedback in large image database. In *ACM Multimedia conference*, 2001.

[26] R. Yan, A. Hauptmann, and R. Jin. Negative pseudo-relevance feedback in content-based video retrieval. In *ACM Multimedia (MM'03)*, Berkeley, CA, USA, 2003.