A Novel Log-based Relevance Feedback Technique in Content-based Image Retrieval

Steven C. H. HOI
Singapore Management University, chhoi@smu.edu.sg

Michael R. LYU
Chinese University of Hong Kong

Citation

HOI, Steven C. H. and LYU, Michael R.. A Novel Log-based Relevance Feedback Technique in Content-based Image Retrieval. (2004). *MULTIMEDIA 2004: Proceedings of the 12th ACM International Conference on Multimedia, New York, October 10-16*. 24-31.
Available at: https://ink.library.smu.edu.sg/sis_research/2397

A Novel Log-based Relevance Feedback Technique in Content-based Image Retrieval

Chu-Hong Hoi and Michael R. Lyu
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong S.A.R.
{chhoi, lyu}@cse.cuhk.edu.hk

ABSTRACT

Relevance feedback has been proposed as an important technique to boost the retrieval performance in content-based image retrieval (CBIR). However, since there exists a semantic gap between low-level features and high-level semantic concepts in CBIR, typical relevance feedback techniques need to perform a lot of rounds of feedback for achieving satisfactory results. These procedures are time-consuming and may make the users bored in the retrieval tasks. For a long-term study purpose in CBIR, we notice that the users' feedback logs can be available and employed for helping the retrieval tasks in CBIR systems. In this paper, we propose a novel scheme to study the log-based relevance feedback (LRF) technique for improving retrieval performance and reducing the semantic gap in CBIR. In order to effectively incorporate the users' feedback logs, we propose a modified support vector machine (SVM) technique called soft label support vector machine (SLSVM) to construct the LRF algorithm in CBIR. We conduct extensive experiments to evaluate the performance of our proposed algorithm. Compared with the typical approach using query expansion (QEX) technique, we demonstrate that our proposed scheme can significantly improve the retrieval performance of semantic image retrieval from detailed experiments.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval—*Relevance Feedback, Query Reformation, Retrieval Model*

General Terms

Algorithms, Human Factors, Experimentation

Keywords

Content-based Image Retrieval, Relevance Feedback, Support Vector Machines, Users Logs

1. INTRODUCTION

Content-base image retrieval (CBIR) has attracted broad range of research interests from many computer communities in the past decade [27]. Although extensive research studies have been conducted for many years, finding a desired image from multimedia databases is still an open question. Many early-year studies on CBIR focused primarily on the feature analysis and similarity measure. However, because of the complexity of image understanding and the semantic gap between low-level features and high-level human perceptions, it is almost impossible to discriminate all the images by some kinds of similarity measure on the low-level features in real-world image databases. One of the feasible ways to bridge the semantic gap is to build the image indexing with textual descriptions. However, manual indexing on image databases is typically time-consuming, costly and subjective for different people, which is unlikely to be fully deployed in practical applications. While the automatic image indexing is still on a long way off [2, 7, 15, 17], relevance feedback, as an alternative feasible technique to mitigate the semantic gap issue, has been proposed and investigated in recent years [8, 9, 11, 21, 23, 22].

A variety of relevance feedback techniques from heuristic techniques to sophisticated learning techniques have been proposed and studied in [5, 10, 12, 16, 29, 32]. From the past research studies, relevance feedback has been shown as an effective scheme to improve the retrieval performance of CBIR [23] and has already been incorporated as a key part when designing a CBIR system. In general, relevance feedback mechanism has to solicit users for evaluating the relevance on the initial query results by the CBIR system. While a user has made the relevance judgements on the initial retrieval results, relevance feedback is engaged as a query refinement technique to improve the retrieval results. Since the learning task of relevance feedback is very tough, it typically needs to repeat many rounds of feedback in order to achieve satisfactory results. Therefore the learning task of relevance feedback can be a very time-consuming procedure.

Moreover, the relevance feedback procedure to specify the relevance of images is viewed as a tedious and boring step for the users. Hence, we hope that the CBIR system by relevance feedback can achieve satisfactory results within as few feedback steps as possible, at best down to only one step. In the past studies, some research efforts proposed to accelerate the relevance feedback by the active learning techniques [30]. However, traditional relevance feedback techniques have limited help when the relevant samples are scarce in the initial

query. Consequently, they cannot explore the feature spaces of the image databases well. From a long-term learning perspective, the feedback logs accumulated from the users could be served as an important resource to help the relevance feedback task in CBIR. In the past relevance feedback studies, less research efforts are devoted to study the log-based relevance feedback algorithms. To our knowledge, there is only one recent research study in [8] which is similar to our work in this paper. The authors suggested to learn a semantic space by learning the users' relevance feedback in image retrieval. Although they proposed to incorporate the user's feedbacks, they only considered the positive feedbacks in their proposed scheme which will lose the important negative information.

In this paper, we present a novel scheme to study the users' feedback logs for helping the learning task of relevance feedback in CBIR. The rest of this paper is organized as follows. We discuss the log-based relevance feedback problem and related techniques in Section 2. Section 3 provides a background review of SVM theory and algorithm. Section 4 proposes a modified SVM algorithm called soft label support vector machine (SLSVM) and formulates the log-based relevance feedback employing the SLSVM technique. Section 5 presents detailed experiments and performance comparison. Section 6 discusses problems and challenges for the log-based relevance feedback algorithms in CBIR and Section 7 gives our conclusions.

2. LOG-BASED RELEVANCE FEEDBACK

In a CBIR system, after a user submits a query by a given sample, the system will return a set of similar images to the user. The returned images may not be fully relevant to the user's targets. In order to learn the query concept of the user, relevance feedback is engaged as a query refinement technique for helping the retrieval task. The relevance feedback mechanism solicits the user to mark the relevance on the retrieved images and then refines the results by learning the feedbacks from the user. The relevance feedback procedure is repeated again and again until the targets are found. As stated previously, the semantic gap problem in CBIR is very challenging and regular learning techniques normally need a lot of rounds of feedback for finding satisfactory results. In order to reduce the learning difficulty, we suggest to seek the help from the user's feedback logs which can be available in CBIR systems from a long-term learning perspective.

To engage the users' feedback logs in CBIR, one may refer to the similar techniques studied in traditional information retrieval in which a lot of studies have described how to employ the users' logs to improve the retrieval performance [6, 1]. Some typical approaches are based on the query expansion (QEX) techniques [6, 34, 24, 33, 19]. Query expansion is viewed as a multiple-instance sampling technique [13] in which the returned samples of the next round are selected from the neighborhood from the positive samples of the previous rounds. Although query expansion showed successes in document retrieval, it may not be very effective to solve the problem in multimedia retrieval [20, 31].

In order to learn the users' logs effectively, we employ a popular yet powerful machine learning technique–support vector machines (SVM)–to attack the problem. SVM is a state-of-the-art classification technique with very good generalization performance. It has been shown with successful

applications in relevance feedback [9, 10, 11, 30]. Before introducing the SVM technique, we here briefly describe how to log the relevance feedback results from users and how to engage them in our proposed algorithm. In each relevance feedback session, a user will specify N_+ samples are relevant (positive) and N_- samples are irrelevant (negative). The information will be logged in our database after the end of the session. When retrieving the relevance feedback information from the logs, a relevance matrix (RM) is constructed to describe the relevance relationship between the images in the database. The column of the relevance matrix represents the image samples in the image database and the row represents the session number in the log database. For each given session, the relevance information of positive samples and negative samples are recorded as relevant $(+1)$, irrelevant (-1) or unknown (0). For example, suppose image i is marked as relevant and j is marked as irrelevant in a given session k, then the corresponding value in the matrix is $RM(k, i) = 1$ and $RM(k, j) = -1$. Therefore, relationship of two images i and j can be computed by the following modified correlation formula:

$$
\begin{aligned}
R_{ij} &= \sum_k \delta_k \cdot RM(k, i) \cdot RM(k, j) \quad (1) \\
\delta_k &= \begin{cases} 1 & \text{if} \quad RM(k, i) + RM(k, j) \geq 0 \,, \\ 0 & \text{if} \quad RM(k, i) + RM(k, j) < 0 \,. \end{cases}
\end{aligned}
$$

where R_{ij} represents the relevance relationship and the δ_k term is engaged to remove the element pair $(-1, -1)$ for the correlation formula. If R_{ij} is positive, it indicates that image i and image j are relevant otherwise they are irrelevant. Then for each given image sample, we can find a set of relevant samples and a set of irrelevant samples ranking by their relationship.

From the above descriptions, we can find that the relevant samples for each given sample may be with different relationship values, meaning that their relationship is with different confidence degrees. Typical SVM cannot well explore the relevance information with different confidence degrees. In order to utilize the advantages of SVM for solving the relevance feedback problem, we propose a modified support vector machine called soft label support vector machine in which the labels of the training data are with different confidence degrees. Before formulating our SLSVM, we first briefly introduce the SVMs and the advantages for solving the relevance feedback issues.

3. SUPPORT VECTOR MACHINES

As a state-of-the-art classification technique [3], SVM implements the principle of Structural Risk Minimization (SRM) based on the Vapnik-Chervonenkis (VC) theory [25, 26, 31]. SVM has very good generalization performance on many pattern classification problems. We briefly illustrate the basic theory and algorithm of SVM as follows.

In a general pattern recognition problem, suppose we are given l observations: $(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y}$, $i = 1, \cdots, l$, where $\mathcal{X} \subseteq R^n$ is the input space and $y_i \in \mathcal{Y}$ is the associated label given a trusted source. For a binary classification task, $\mathcal{Y} = \{+1, -1\}$. It is assumed these data are independently drawn and identically distributed (i.i.d.) from some unknown probability distribution $P(\mathbf{x}, y)$. A pattern recognition task is to learn the mapping $\mathbf{x}_i \longmapsto y_i$ with a learning machine, defined by a set of possible mappings

$\mathbf{x} \longmapsto f(\mathbf{x}, \alpha)$, where the functions $f(\mathbf{x}, \alpha)$ are given by some adjustable parameters α. A trained machine is obtained by choosing a set of parameters and the expectation of the test error for the trained machine is

$$R(\alpha) = \int \frac{1}{2} |y - f(\mathbf{x}, \alpha)| dP(\mathbf{x}, y) \qquad (2)$$

This quantity $R(\alpha)$ is called the expected risk or the risk. The VC theory provides a risk bound for the expected risk with respect to the empirical risk on trained data and a VC confidence interval. The SRM principle followed by SVM is to choose the learning machine which gives the lowest upper bound on the risk. We omit the detailed theory and illustrate the algorithm of SVM below [31].

Let us first begin with the simplest case, i.e. linear SVM on separable data. Suppose the training data $(\mathbf{x}_i, y_i), i = 1, \cdots, l, y_i \in \{+1, -1\}$ can be separated by some hyperplane ("separating hyperplane") $\mathbf{w} \cdot \mathbf{x} + b = 0$, which separates the positive from the negative samples. If we denote d_+ (d_-) the shortest distance from the separating hyperplane to the closest positive (negative) sample, "margin" of a separating hyperplane can be defined as $d_+ + d_-$. The strategy of SVM algorithm is to look for the optimal separating hyperplane with largest margin. It is evident that a separating hyperplane satisfies the constraints

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 , i = 1, \cdots, l . \qquad (3)$$

Hence, the SVM algorithm is to find the optimal hyperplane by maximizing the margin $d_+ + d_- = 2/\|w\|$, i.e. minimizing $\|\mathbf{w}\|^2$, subject to the constraints in Eq. 3.

For nonseparable data, nonlinear SVM with soft-margin separating hyperplane is suggested [31]. The input vectors are first mapped into a high-dimensional feature space $\mathcal{H}$ with a nonlinear mapping $\Phi : R^n \longmapsto \mathcal{F}$. Normally, a kernel function K is engaged to compute the dot products in $\mathcal{F}$ such that $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. The optimal separating hyperplane in the feature space can be obtained by solving the constrained optimization problem below

$$\min_{\mathbf{w},b,\xi,\rho} \quad \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum_i \xi_i \qquad (4)$$

$$\text{subject to} \quad y_i((\Phi(\mathbf{x}_i) \cdot \mathbf{w}) + b) \geq \rho - \xi_i \qquad (5)$$
$$\rho \geq 0 , \xi_i \geq 0 , i = 1, \ldots, l .$$

Here, the parameter $0 \leq \nu \leq 1$ is used to control the fraction of SVs and margin errors, and the positive slack variable ξ_i is introduced for counting the error in nonseparable cases [25]. From the above formulation, we can see that the training samples are treated fairly to impact the decision boundary.

4. LOG-BASED RELEVANCE FEEDBACK USING SLSVM

4.1 Soft Label Support Vector Machine

Regular SVM techniques assume that the labels of the data are from a trusted source. However, when some labels of data are incorrect, they may hardly impact on the predicted decision. In order to reduce the risk arising from the unconfident data, we incorporate the label confidence degree of data in the regular SVM formulation and suggest the Soft Label SVM as follows.

Let us consider the training data associated with labels of different confidence degrees $(\mathbf{x}_i, s_i) \in \mathcal{X} \times S, i = 1, \cdots, l$.

Here, $S = [-1, +1]$ and s_i is the soft label of the training data to indicate the degree of relevance or irrelevance belonging to the labelled class. Given a soft label set S, the corresponding "hard label" set Y can be obtained by applying a sign function on S

$$Y = sign(S) = \{+1, -1\} . \qquad (6)$$

We modify the objective function in Eq. 4 as

$$\min_{\mathbf{w},b,\xi,\rho} \quad \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum_i y_i s_i \xi_i \qquad (7)$$

$$\text{subject to} \quad y_i((\Phi(\mathbf{x}_i) \cdot \mathbf{w}) + b) \geq y_i s_i \rho - \xi_i ,$$
$$\rho \geq 0 , \xi_i \geq 0 , i = 1, \ldots, l .$$

From this objective function, we can see the margin error ξ_i will be larger if it was constrained with a smaller label value s_i when facing a non-separable case. If the label value is larger, the allowed margin error should be smaller when facing a non-separable case. Let us introduce Lagrange Multiplier technique to derive the dual problem

$$L(\mathbf{w}, \xi, b, \rho, \alpha, \beta, \delta) = \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum_i y_i s_i \xi_i$$
$$- \sum_i \left(\alpha_i(y_i(\Phi(\mathbf{x}_i) \cdot \mathbf{w} + b) - y_i s_i \rho + \xi_i) - \beta_i \xi_i \right) - \delta\rho . \quad (8)$$

Then, we take the partial derivative of L with respect to $\mathbf{w}$, ξ_i, b and ρ, respectively.

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x}_i) = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x}_i) ;$$

$$\frac{\partial L}{\partial \xi_i} = y_i s_i \frac{1}{l} - \alpha_i - \beta_i = 0 \Rightarrow 0 \leq \alpha_i \leq y_i s_i \frac{1}{l} ;$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{l} \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 ;$$

$$\frac{\partial L}{\partial \rho} = -\nu + \sum_{i=1}^{l} \alpha_i y_i s_i - \delta = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i s_i - \delta = \nu .$$

The dual of the primal optimization problem is turned into

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

$$\text{subject to} \quad \sum_i \alpha_i y_i = 0 , \qquad (9)$$

$$0 \leq \alpha_i \leq y_i s_i \frac{1}{l}, \quad i = 1, \ldots, l , \quad (10)$$

$$\sum_i \alpha_i y_i s_i \geq \nu . \qquad (11)$$

Then, the decision function can be derived as the form below

$$f(\mathbf{x}) = sign(\sum_i \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b) . \qquad (12)$$

From the dual optimization function above, we can find the important difference compared with the regular SVM. In the regular SVM, the constraint of the dual variables α_i is

$$0 \leq \alpha_i \leq \frac{1}{l} , \qquad (13)$$

while in the Soft Label SVM, the constraint becomes

$$0 \leq \alpha_i \leq y_i s_i \frac{1}{l} . \qquad (14)$$

This indicates the Support Vector (SV) with a large label will have a larger impact on the decision boundary than the SV with a smaller label. Hence, the probability of misclassification on the future data will be reduced.

4.2 LRF Algorithm by SLSVM

Implementation of LRF algorithm with SLSVM is similar to building a relevance feedback algorithm using SVM. The difference is that the labels of training samples in SLSVM are associated with different confidence degrees. The key step for constructing the LRF algorithm is the training sample selection problem. When a user begins a query session, the CBIR system first returns top N samples to the user as the initial query results by a given query sample. The user then specify the relevance of the N initial samples in which N_+ samples are marked as "positive" and N_- samples are "negative". Normally, there are few positive samples in the initial retrieval and insufficient training samples will degrade the performance of relevance feedback algorithms. To attack the problem, we can solicit the help from the users' feedback logs in the CBIR system.

In order to look for more training samples, we employ the N_+ initial positive samples as the seeds. For each positive seed (e.g. the ith image), we can compute the relationship value R_{ij} between this seed (the ith image) and each image sample j in the database according to Eq. 1. Then, the soft label value s_j^i of each image j with respect to the positive seed i can be obtained by normalizing the relationship value as follows

$$s_j^i = \begin{cases} R_{ij}/\max_j R_{ij} & \text{if} \quad R_{ij} > 0 \,, \\ -R_{ij}/\min_j R_{ij} & \text{if} \quad R_{ij} < 0 \,, \\ 0 & \text{if} \quad R_{ij} = 0 \,. \end{cases} \quad (15)$$

By considering the N_+ positive seeds, we can calculate a final soft label pair for each image sample j of the database $(s_j^+, s_j^-) = (\max_i s_j^i, \min_i s_j^i)$. Based on the soft label pairs, we can choose N_+' additive positive training samples through ranking images with s_j^+ and choose N_-' additive negative training samples by ranking them with s_j^-. Here, we constrain with $N_+ + N_+' = N_- + N_-'$ to avoid imbalanced dataset issues.

Based on the above selection strategy, we can collect $(N_+ + N_+')$ positive training sample and $(N_- + N_-')$ negative training samples, each of them is associated with a soft label for representing the confidence degree. Obviously, the soft labels of the initial positive sample and negative should be $+1$ and -1 respectively. Hence, these training samples are engaged to train an SLSVM classifier. With the trained SLSVM classifier, we can construct its evaluation function for the LRF algorithm according to the decision function in Eq. 12. Therefore, the images in the dataset can be evaluated and ranked by the constructed evaluation function which is similar to a regular SVM algorithm [11].

5. EXPERIMENTAL RESULTS

5.1 Datasets

To perform empirical evaluation of our proposed algorithm, we choose the real-world images from the COREL image CDs. There are two sets of data collected in our experiments: 20-Category (20-Cat) and 50-Category (50-Cat). The 20-Cat dataset contains 20 categories and the 50-Cat one contains 50 categories. Each category in the datasets consists exactly 100 images selected from the COREL image CDs. The categories are with different semantic meanings, such as *antique, antelope, aviation, balloon, botany, butterfly, car, cat, dog, firework, horse* and *lizard*, etc. The motivation to select the semantic categories are explained in two aspects. First, it can evaluate whether the approach can retrieve the relevant images not only visually relevant but also with similar semantic meaning. Second, the approach can help us evaluate the performance automatically, which can reduce the subjective errors in manual evaluations by different people.

5.2 Image Representation

Image representation is an important step to evaluate relevance feedback algorithms in CBIR. Three different features are chosen in our experiments to represent the images: color, edge and texture.

Color feature is widely adopted in CBIR for its simplification and effectiveness. The color feature engaged in our experiments is color moment since it is closer to human perception naturally, and many previous research studies have showed the effectiveness of color moment applied in CBIR. For the employed color moment, we extract 3 moments: color mean, color variance and color skewness in each color channel (H, S, and V), respectively. Thus, 9-dimension color moment is adopted as the color feature in our experiments.

Edge feature could be very effective in CBIR when the contour lines of images are evident. The edge feature in our experiments is edge direction histogram [14]. The images in the datasets are first translated to gray images. Then Canny edge detector is applied to obtain the edge images. From the edge images, the edge direction histogram can then computed. The edge direction histogram is quantized into 18 bins of 20 degree each, hence an 18-dimension edge direction histogram is employed to represent the edge feature.

Texture feature is proved as an important cue for image retrieval. A variety of texture analysis methods have been studied in the past years. In our experiments, we employ the wavelet-based texture technique [28, 18]. The original color images are transformed to gray images. Then we perform the Discrete Wavelet Transformation (DWT) on the gray images employing a Daubechies-4 wavelet filter [28]. Each wavelet decomposition on a gray 2D-image results in four subimages with a $0.5 * 0.5$ scaled-down image of the input image and the wavelets in three orientations: horizontal, vertical and diagonal. The scaled-down image is fed into the DWT operation to produce the next four subimages. In total, we perform 3-level decompositions and obtain 10 subimages in different scales and orientations. One of the 10 subimages is a subsampled average image of the original image which is discarded since it contains less useful texture information. For other 9 subimages, we compute the entropy of each subimage respectively. Therefore, we obtain a 9-dimention wavelet-based texture feature to describe the texture information for each image.

5.3 Experimental Setup

In order to evaluate the log-based relevance feedback algorithm, we have developed a CBIR system with relevance feedback mechanism [9, 10]. The collection of user logs is an important work for empirical evaluations of our proposed technique. In our experiments, we solicit 10 researchers to
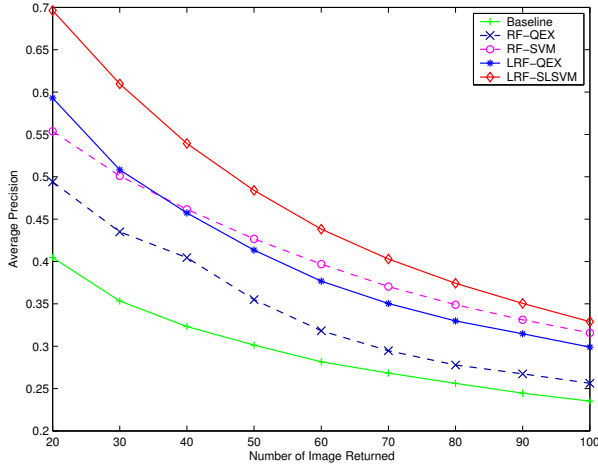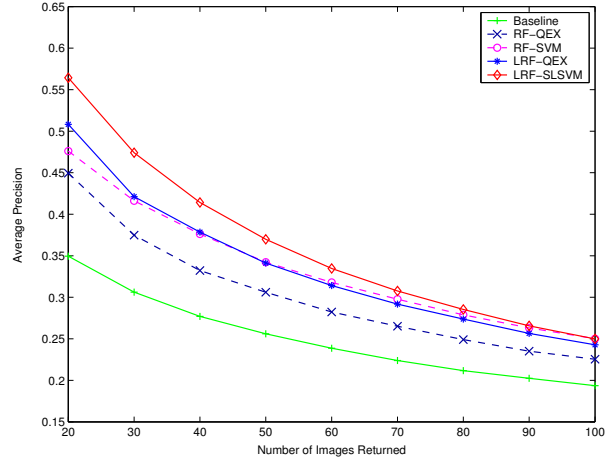
(a) 20-Category (b) 50-Category

Figure 1: Average precision for the compared algorithms on two datasets

serve as the users for helping evaluate our developed CBIR system and create the log information. The users are requested to perform the query-by-example (QBE) execution on the CBIR system and provide feedback on the retrieval results.

For the purpose of evaluation, we denote a log session (LS) as a basic log unit. Each log session is corresponding to a user's QBE execution and feedback, in which 20 images are reviewed by the user in such a log session. In order to cover the logs over the database, for each QBE execution of the user, a query sample is randomly seeded from the image database to serve as the query concept. Given the randomly seeded sample, the CBIR system returns top 20 ranking images by employing the Euclidean distance measure on the image database. For the 20 returned images, the user gives feedback information (positive or negative) on the images by judging whether they are relevant to the desired query concept. The feedback information is logged as a log session accumulated in a log database. Although the users may give noisy information to the logs, we assume the noise in the log information collected from the users is small. We will also evaluate the performance of noisy logs by injecting further noise into the collected logs.

In our experiments, each user is solicited to perform 10 times of QBE execution and feedback, hence we collect 100 log sessions from 10 users in total. At the first sight, someone may say the number of log sessions is not large enough to evaluate the performance. However, for 100 log sessions, the users actually need to review 2000 images. Compared with our two testing datasets (a dataset with 2000 images and another dataset with 5000 images), the collected 100 log sessions is enough for evaluating the performance. We can collect more logs in real-world CBIR applications; nonetheless, we hope to evaluate whether our proposed technique can work well even with limited log sessions.

5.4 Performance Comparison

The major performance measure metric employed in our experiments is based on the *Average Precision*, which is defined as the average ratio of the number of relevant images of the returned images over the number of total returned images. In our experiments, we compared our proposed log-based relevance feedback algorithm using SLSVM (LRF-SLSVM) over the traditional methods.

The traditional Relevance feedback method for our comparison is the query expansion technique (RF-QEX). As stated before, query expansion is considered as a multiple-instance sampling technique widely used for relevance feedback in information retrieval. We here briefly describe its implementation for CBIR in our experiments similar to the approach in [30]. To return the next round images in relevance feedback, we retrieve the images by looking for 5 nearest-neighboring images around the query sample by employing the Euclidean distance measure. After finding the 5 nearest samples, we recursively look for 5 nearest samples around the finding samples. In total, 20 images are returned to the user in each round. Although the QEX technique is simple yet effective for relevance feedback applications, it may not be the best choice for relevance feedback in CBIR applications for learning in a high dimensional feature space since it is simply based on the Euclidean distance measure.

We also extend the traditional query expansion to the log-based version called the log-based relevance feedback by query expansion (LRF-QEX). Similar to the approach of previous QEX approach, 5 nearest-neighborhood samples are retrieved recursively. However, we incorporate the log information to weight the distance measure. From Eq. 1, we compute the relationship values between the query sample and the samples in the database. Similar to the soft label strategy, we employ the result in Eq. 15 to weight the Euclidean distance. If the soft label s_j^i is positive, it is obvious that a larger label should correspond to a smaller distance weight (simply taking $1/s_j^i$). If the images is with $s_j^i = 0$, its distance is weighted with a constant C, such that $1/C = min_j\{s_j^i | s_j^i > 0\}$. If the soft label s_j^i is negative, its weight is equal to $C - s_j^i$.

Besides the query expansion and the log-based query expansion techniques, we also evaluate the performance of the traditional SVM-based relevance feedback (RF-SVM) techniques [11, 25]. Several problems for SVM-based relevance
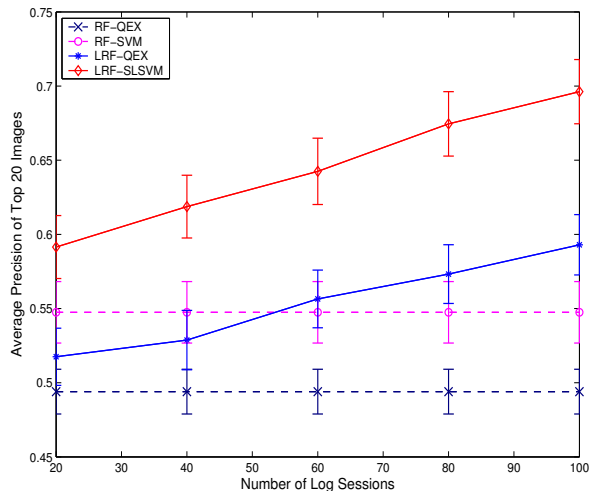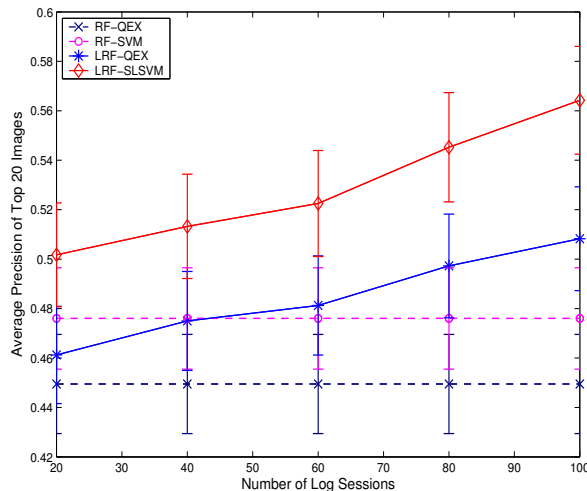
(a) 20-Category   (b) 50-Category

Figure 2: Average precision of the compared algorithms for different number of log sessions on two datasets.

feedback techniques should be addressed here. The kernel function and different parameter settings may impact largely on the performance. In order to enable objective evaluations, we choose a set of same kernel function and parameters for the traditional SVM-based relevance feedback and our suggested log-based technique using SLSVM. For the experiments, we implement the proposed SLSVM algorithm by modifying the regular SVM algorithm based on the public libsvm library available at [4].

The first case of performance evaluation is to measure the Average Precision of top k returned images for different approaches. For each method, we repeat 200-round executions to evaluate the precision and take the average over the 200 results of precision. For each execution round, we randomly pick a query sample from the database and then run the initial query and the LRF algorithms to refine the retrieval results in which the retrieval precision is computed as the retrieval performance. For each approach of the relevance feedback algorithms, the selected 200 evaluation rounds are identical in order to avoid an unfair comparison for different approaches.

The first set of experimental results on the 20-Category dataset is shown on Fig. 1 (a). In this comparison, the retrieval performance of the LRF-SLSVM method employing 100 log sessions is measured and evaluated. In the figure, the "Baseline" result are the average precision of the initial query by randomly generated query ID. From the figure, we can see that log-based relevance feedback algorithm can effectively improve the retrieval precision. With a lower rate percentage of logs, our LRF-SLSVM scheme can significantly improve the retrieval performance and outperform other approaches. For example, with limited 100 log sessions, our proposed algorithm outperform the typical RF-SVM approach by over 25%. Also our LRF-SLSVM method surpasses the log-based technique using query expansion by over 15%.

The second set of experimental results on the 50-Category dataset is shown on Fig. 1 (b). From the figure, we also can find the comparison results similar to those of the 20-

Category dataset. Although the difference is slightly reduced, our LRF-SLSVM technique still improve over 18% on the regular RF-SVM method and over 11% on the LRF-QEX technique.

In order to further evaluate the impact of the number of log sessions on our proposed algorithm, we evaluate two log-based relevance feedback algorithms with different number of log sessions. The experimental results on 20-Category and 50-Category are shown in Fig. 5.3 and Fig. 5.3, respectively. From these two figures, we can observe that the retrieval performance of log-based relevance feedback will be improved while the number of log sessions is increased. We also observe some interesting and surprising results that even with a very low number of log sessions, i.e., with 20 log sessions, our proposed method improve over 8% on the typical relevance feedback technique using SVM in 20-Category dataset and over 5% in 50-Category dataset. However, in real-world CBIR applications, thousands of log sessions can be available, hence more promising results for practical applications of our approach can be expected.

Moreover, we understand noisy logs have important impact on the performance of log-based algorithms. In order to evaluate our algorithm for noisy logs, we inject about 15% noise into the 100 log sessions. Then, we evaluate the relevance feedback algorithms employing SVM and SLSVM with or without noisy logs. From Table 1, we can see that our SLSVM algorithm is more robust than the regular SVM algorithm although it also suffers from the noisy logs.

Finally, we also notice that the kernel plays an important role for the SLSVM training. In order to evaluate the impact of kernels in our proposed algorithm, we conduct experiments to compare the performance between several popular kernels: Linear kernel (Linear), Polynomial (Polynomial-2: Degree 2 Polynomial and Polynomial-4: Degree 4 Polynomial), Radial Basis Function (RBF) and Sigmoid kernel (Sigmoid). In these experiments, parameters for each individual kernel are determined by a validation set (40 runs) and are tested with 160 rounds of executions of each one. From the experimental results shown in Table 2, we can ob-

Table 1: Retrieval performance of proposed algorithms on noisy logs

Methods	20-Category			50-Category		
	Top 20	Top 50	Top 100	Top 20	Top 50	Top 100
Baseline	0.405 ± 0.019	0.301 ± 0.015	0.235 ± 0.012	0.350 ± 0.018	0.256 ± 0.014	0.194 ± 0.010
SVM without logs	0.554 ± 0.021	0.427 ± 0.018	0.315 ± 0.014	0.476 ± 0.021	0.342 ± 0.016	$\mathbf{0.250 \pm 0.012}$
SVM with logs	0.608 ± 0.021	0.413 ± 0.017	0.295 ± 0.014	0.512 ± 0.021	0.330 ± 0.016	0.222 ± 0.011
SLSVM with logs	$\mathbf{0.653 \pm 0.021}$	$\mathbf{0.448 \pm 0.018}$	$\mathbf{0.317 \pm 0.014}$	$\mathbf{0.543 \pm 0.021}$	$\mathbf{0.356 \pm 0.017}$	0.242 ± 0.012

Table 2: Retrieval performance of different kernels on the 20-Cat and 50-Cat datasets

Kernels	20-Category			50-Category		
	TOP 20	TOP 50	TOP 100	TOP 20	TOP 50	TOP 100
Linear	0.573 ± 0.026	0.379 ± 0.020	0.253 ± 0.014	0.370 ± 0.023	0.208 ± 0.014	0.126 ± 0.009
Polynomial-2	0.604 ± 0.025	0.387 ± 0.020	0.257 ± 0.014	0.381 ± 0.022	0.210 ± 0.014	0.130 ± 0.009
Polynomial-4	0.612 ± 0.025	0.397 ± 0.020	0.262 ± 0.014	0.383 ± 0.023	0.212 ± 0.014	0.133 ± 0.010
RBF	$\mathbf{0.700 \pm 0.022}$	$\mathbf{0.483 \pm 0.019}$	$\mathbf{0.334 \pm 0.014}$	$\mathbf{0.574 \pm 0.022}$	$\mathbf{0.388 \pm 0.018}$	$\mathbf{0.267 \pm 0.013}$
Sigmoid	0.597 ± 0.025	0.359 ± 0.020	0.224 ± 0.014	0.422 ± 0.022	0.212 ± 0.014	0.120 ± 0.009

serve that the RBF kernel obviously outperforms others in all cases, which shows that RBF kernel is more suitable to learn the feature space in our problem. The performance of the Sigmoid kernel is similar to the Polynomial kernels in which the Degree-4 Polynomial is slightly better than the Degree-2 Polynomial one. The remaining Linear kernel obtains the worst performance in most cases.

6. DISCUSSIONS

In the experiments, we have evaluated the retrieval performance of our proposed log-based relevance feedback algorithm over the traditional approaches. Although we have shown the effectiveness of our proposed methods, some further work can be extended in the future.

First, dimension of the logs should be considered when the number of collected log sessions are very large. Thus, we may need to study some dimension reduction techniques for investigation in the future work. From some previous study on this problem, singular value decomposition (SVD) is regarded as a popular technique for the dimension reduction [8]. We could incorporate the method to our proposed scheme in the future work.

Second, we understand the noisy logs should be well considered in real-world CBIR applications and how to filter the noise from the logs is an important future work. Some previous work showed that incorporating the SVD technique can help reduce the impact of noise [8]. Also, it is encouraged to study more advanced filtering techniques for removing the noise from the logs.

Furthermore, when dimension of the log sessions and number of images in datasets are very large, complexity of our proposed algorithm should be addressed and the efficiency of our proposed algorithm may be reduced. Thus, we have to study the efficiency problem and solve it with more sophisticated technique for speeding up the retrieval task. Moreover, we need to enlarge our database and conduct more sophisticated evaluations on varied cases. Nevertheless, our current experimental results produce enough evidences in demonstrating the effectiveness of our proposed algorithm.

7. CONCLUSIONS

In this paper, we suggest to study the users' feedback logs for improving the relevance feedback algorithms. We pro-

pose a novel scheme for log-based relevance feedback learning in CBIR using modified SVM techniques. Contributions of our work can be summarized as follows:

First, we propose and formulate a modified support vector machine called soft label support vector machine for solving the log-based relevance feedback task in CBIR. The proposed technique is generic and may also be applied to other fields.

Second, we conduct detailed empirical evaluations in our experiments, in which we study the users' feedback logs and show how our approaches improve the retrieval performance of relevance feedback in CBIR. From the experimental results, we show that the log-based relevance feedback employing SLSVM is promising and effective in CBIR.

Finally, we discuss several important problems for formulating the log-based relevance feedback algorithms. We show that the noisy logs problem will degrade the performance of log-based algorithms and should be considered in real-world applications. We also indicate that kernel function plays an important role on the performance of our proposed algorithm, in which we show that the RBF kernel is more effective than other kernels.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] P. Anick. Using terminological feedback for web search refinement: a log-based study. In *Proceedings of the 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, pages 88–95, 2003.

[2] D. Blei and M. I. Jordan. Modeling annotated data. In *Proceedings of the 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, pages 127–134, 2003.

[3] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[4] 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`.

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

[6] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Probabilistic query expansion using query logs. In *Proceedings of the 11th Intl. World Wide Web Conf. (WWW'2002)*, 2002.

[7] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the 7th European Conference on Computer Vision (ECCV'02)*, pages 97–112, 2002.

[8] X. He, O. King, W.-Y. Ma, M. Li, and H. J. Zhang. Learning a semantic space from user's relevance feedback for image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):39–48, Jan. 2003.

[9] C.-H. Hoi and M. R. Lyu. Biased support vector machine for relevance feedback in image retrieval. In *Proceedings of Intl. Joint Conf. on Neural Networks (IJCNN'04)*, Budapest, Hungary, 2004.

[10] C.-H. Hoi and M. R. Lyu. Group-based relevance feeedback with support vector machine ensembles. In *Proceedings of the 17th Intl. Conf. on Pattern Recognition (ICPR'04)*, Cambridge, UK, 2004.

[11] P. Hong, Q. Tian, and T. Huang. Incorporate support vector machines to content-based image retrieval with relevant feedback. In *Proceedings of IEEE Intl. Conf. on Image Processing (ICIP'00)*, volume 3, pages 750–753, Vancouver, BC, Canada, 2000.

[12] T. S. Huang and X. S. Zhou. Image retrieval by relevance feedback: from heuristic weight adjustment to optimal learning methods. In *Proceedings of IEEE Intl. Conf. on Image Processing (ICIP'01)*, volume 3, pages 2–5, Thessaloniki, Greece, Oct. 2001.

[13] Y. Ishikawa, R. Subramanya, and C. Faloutsos. MindReader: Querying databases through multiple examples. In *Proceedings of the 24th Int. Conf. Very Large Data Bases (VLDB'98)*, pages 218–227, 1998.

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

[15] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, pages 119–126, 2003.

[16] I. King and J. Zhong. Integrated probability function and its application to content-based image retrieval by relevance feedback. *Pattern Recognition*, 36(9):2177–2186, 2003.

[17] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In *Advances in Neural Information Processing Systems (NIPS'03)*, 2003.

[18] B. Manjunath, P. Wu, S. Newsam, and H. Shin. A texture descriptor for browsing and similarity retrieval. *Signal Processing Image Communication*, 2001.

[19] K. Porkaew, K. Chakrabarti, and S. Mehrotra. Query refinement for multimedia retrieval and its evaluation techniques in mars. In *Proceedings of ACM Intl. Conf. on Multimedia (MM'99)*, Orlando, Florida, USA, 1999.

[20] K. Porkaew, M. Ortega, and S. Mehrotra. Query reformulation for content based multimedia retrieval in MARS. In *Proceedings of ICMCS*, volume 2, pages 747–751, 1999.

[21] Y. Rui and T. Huang. Optimizing learning in image retrieval. In *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'00)*, volume 1, pages 236–243, South Carolina, USA, 2000.

[22] Y. Rui and T. S. Huang. A novel relevance feedback technique in image retrieval. In *Proceedings of ACM Intl. Conf. on Multimedia (MM'99)*, pages 67–70, Orlando, Florida, USA, 1999.

[23] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, Sept. 1998.

[24] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 44(4):288–287, 1990.

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

[26] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

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

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

[29] K. Tieu and P. Viola. Boosting image retrieval. In *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'00)*, South Carolina, USA, 2000.

[30] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *Proceedings of the 9th ACM Intl. Conf. on Multimedia*, pages 107–118. ACM Press, 2001.

[31] V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[32] N. Vasconcelos and A. Lippman. Bayesian relevance feedback for content-based image retrieval. In *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries (CVPR'00)*, South Carolina, USA, 2000.

[33] L. Wu, C. Faloutsos, K. P. Sycara, and T. R. Payne. FALCON: Feedback adaptive loop for content-based retrieval. *The VLDB Journal*, pages 297–306, 2000.

[34] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In *Proceedings of the 19th Intl. ACM SIGIR Conf. (SIGIR'96)*, pages 4–11, 1996.