Statistical Machine Learning for Bridging the Semantic Gap in Image Retrieval

HOI, Chu Hong (Steven)

Supervised by

Prof. Michael R. LYU

November 2005

Abstract of thesis entitled:

Statistical Machine Learning for Bridging the Semantic Gap in Image Retrieval

Submitted by HOI, Chu Hong (Steven)

With the explosive growth of multimedia data, more and more research attentions have been devoted to visual information retrieval. Image retrieval, particularly content-based image retrieval (CBIR), has been actively studied in multimedia information retrieval community in the past decade. One of the most challenging difficulties in CBIR is the semantic gap between low-level visual features and high-level semantic concepts. This thesis investigates statistical machine learning techniques for attacking the semantic gap problem in image retrieval. In this thesis, a unified learning framework, integrating supervised learning, unsupervised learning, semi-supervised learning, active learning and distance metric learning, is proposed to bridge the semantic gap in image retrieval from several novel perspectives.

The first novelty in the proposed framework is the semi-supervised active learning (SSAL) scheme for online learning with users' relevance feedback in image retrieval. Different from traditional active learning approaches, which are usually supervised, our semi-supervised solution can be much more effective in finding the informative unlabeled data for narrowing down the semantic gap in image retrieval.

The second originality is the log-based relevance feedback (LRF) in that users' log data are engaged in the online learning tasks for image retrieval. Different from traditional relevance feedback, our LRF scheme exploits the users' log data as a critical resource for bridging

the semantic gap from the long-term learning view. To explore the users' log data effectively, we propose a novel Soft-Label Support Vector Machines (SLSVM) to formulate an effective LRF algorithm, which smoothly combines users' log data into the online learning task under a solid learning framework of regularization theory.

The third significant contribution is the distance metric learning (DML) scheme for the offline learning with the log data of users' relevance feedback. Different from traditional image retrieval, which is normally based on some fixed distance metric, we propose a collaborative image retrieval scheme in which a flexible distance metric is learned offline by exploiting users' log data of relevance feedback using a novel regularized distance metric learning (RDML) algorithm, which is much more effective and robust than traditional metric learning algorithms.

The last but not the least contribution is the Structural Similarity Measure (SSM) scheme by exploring unsupervised learning techniques for learning with unlabeled data to measure the similarity between images that goes beyond traditional distance measure for image retrieval. In this scheme, we investigate spectral clustering to perform unsupervised learning that can discover semantic relationship of images from the unlabeled data. Meanwhile, a marginalized kernel is extended to measure the similarity of images based on the grouping results of the unsupervised learning with spectral clustering.

As a conclusive note, although our application domain is in image retrieval, the theory, methodology and technology explored in this thesis can be generally applied and extended to other research areas in computer science such as information retrieval and data mining.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Content-based Image Retrieval

Along with the rapid growth of digital devices for image and video creation, storage and transmission, huge amounts of images and videos are produced everyday. Motivated by the enormous demand of information retrieval on image and video databases, more and more research efforts have been devoted to visual information retrieval in recent years [1, 2, 3, 4, 5]. A typical approach is to employ the textual descriptions or keywords for indexing and retrieving images [6, 7, 8]. However, a lot of limitations of the text-based methods make them far from working in real-world applications. A key task of text-based approaches is to annotate and index the images with keywords. Traditional techniques need amount of labors for annotating the images. This is quite challenging and almost prohibited in the practical applications. Although there are some promising advances of image annotation in recent research [9, 10, 11, 12], fully automatic and practical annotation techniques are still on a long way off. Moreover, textual descriptions may have limited capacity to capture the content of images. Like an old saying "An image worths a thousand of words", describing the rich content of an image by only a few keywords is almost impossible. Hence, fully text-based approach is not practical enough for current multimedia applications.

To overcome the shortcomings of text-based retrieval mechanisms, content-based image retrieval (CBIR) has been suggested as an alternative approach of text-based query for content accessing in multimedia databases in the early 1990's [13, 14, 15]. In recent years, CBIR has become one of the most active research areas in visual information retrieval [16, 4]. In CBIR, a set of low-level features (such as color, texture and shape, etc.) is first extracted to represent the visual content of images. The images in the database are indexed by these extracted visual features. Based on the indexing scheme by visual descriptions of low-level features, visual queries can be formulated and similarity of images can be measured by using an appropriate distance function defined in the CBIR system [17, 18, 19].

In the past decades, considerable amount of research efforts have been put in image retrieval, particular for feature representation and similarity measure. However, due to the high complexity of image understanding, it is almost impossible to discriminate images simply by distance measurements on low-level features. One well-know problem is the semantic gap between low-level features and high-level human semantic concepts [20, 15]. Before the mature of automatic image annotation [10], an alternative solution to narrow the semantic gap is relevance feedback which has been shown as a powerful tool to boost the retrieval performance in CBIR.

1.2 Relevance Feedback

To attack the semantic gap issue in CBIR, a variety of relevance feedback techniques, ranging from heuristic methods to sophisticated learning algorithms, have been suggested and studied In the past decade [21, 22, 23, 24, 25, 26]. Relevance feedback has already been considered a key component when designing a CBIR system. In general, any relevance feedback mechanism requires users' relevance judgements for the results returned by a CBIR system in response to a user query. Given the relevance judgments for the initially retrieved results, relevance

feedback is then engaged as a query refinement method to improve the retrieval accuracy of the CBIR system. Because of the diverse information needs from users and meanwhile the limited information exposed in solicited relevance judgments, very often multiple rounds of feedback are required to achieve satisfactory results. Given it is a tedious job to make relevance judgments for images in relevance feedback, it is advantageous if a CBIR system can achieve satisfactory results within a few feedback cycles. Although some research studies have suggested employing active learning techniques to speed up the relevance feedback procedure [27], traditional techniques for relevance feedback may not be able to tackle the problem well when only few (even none!) of the initially retrieved results are actually relevant.

1.3 Motivation of This Work

Recently, there have been a few studies on utilizing log data of users' relevance feedback to improve the retrieval accuracy of CBIR systems [28, 29, 30]. They assume that a CBIR system is able to collect and store large numbers of relevance judgments from users. In those studies, log data have been used as a supplementary resource to either improve the similarity measurement of images or automatically augment the pool of examples judged in users' feedback. In this work, we suggest treating the log-based relevance feedback as a multiple-modality learning task. We hypothesize that two images tend to be similar in their visual content when they have been judged similarly by a large number of users. Thus, in addition to low-level visual features, each image can also be represented by the related relevance judgments in log data. Then, the imposed question is how to exploit the two types of image representations for more effective relevance feedback in CBIR. One natural solution to a learning problem with two orthogonal representations is the co-training algorithms [31], which have shown to be effective for a number of learning tasks, such as web page classification [32] and natural language processing [33]. The main idea of co-training

algorithms for image retrieval is to effectively explore the correlation between two different image representations, namely the relevance prediction of images based on the two representation has to be consistent. To this end, we propose a novel learning technique, named "**Coupled Support Vector Machine**", that is able to effectively enforce the consistency in the relevance prediction between the low-level feature representation and the representation based on collaborative relevance judgments from log data.

1.4 Related Work

Traditional relevance feedback originates from document information retrieval [34, 35]. It is a bit surprising that relevance feedback has been received much more research attention from the image retrieval community in the past decade [21, 36, 15, 23, 37]. Most of the past research studies focused on studying various algorithms and theories for traditional relevance feedback schemes. Due to the difficulty of the learning task, it is almost impossible to bridge the semantic gap between low-level visual features and high-level semantic concepts by learning low-level image features only.

Exploiting log data of users' feedback has become a promising direction for reducing the semantic gap [30]. Although there are some research work for studying user logs in traditional text information retrieval [38, 39], there is little research done to image retrieval. Some related work in this area has been recently reported by Zhou and Zhang et al. [28], He and King et al. [40], Hoi and Lyu [30], He and Ma, et al. [41], amongst others. Different from previous work, our work in this paper is based on a novel Coupled support vector machine which can integrate the log information of users' feedback into traditional relevance feedback with learning on the low-level visual features of image content.

1.5 Organization of This Work

The rest of this paper is organized as follows. In Chapter 2, we first introduce the motivation and background of log-based relevance feedback. Then we propose the Coupled Support Vector Machine for attacking the log-based relevance feedback problem with the motivation of improving a simple solution using the linear combination of two support vector machines. In Chapter 3, we discuss our experiment methodology and present our experimental evaluations of our proposed algorithms. Chapter 4 sets out our conclusion and discusses some directions in our future work.

□ **End of chapter.**

Chapter 2

Log-based Relevance Feedback

2.1 Overview and Problem Formulation

Relevance feedback has become a necessary and important component when designing a CBIR system. In general, when a user submits his/her query target, the CBIR system will return a set of similar images to the user. The images returned initially may not be fully relevant to the query target. In order to learn the user's query concept, relevance feedback is employed as a query refinement method to help the retrieval task. A relevance feedback mechanism asks the user to judge the relevance of retrieved images and then refines the retrieval results by learning from the relevance judgments. The relevance feedback procedure is repeated until all the targets are found. Due to the semantic gap between low-level features and high-level concepts, regular feedback techniques normally require a number of rounds of feedback to achieve satisfactory results. To assist the learning task in relevance feedback, we propose to incorporate log data of users' feedback into relevance feedback for the retrieval task [30]. Fig. 2.1 shows a proposed architecture model of our log-based relevance feedback scheme for CBIR.

We first describe the organization of log data for users' relevance

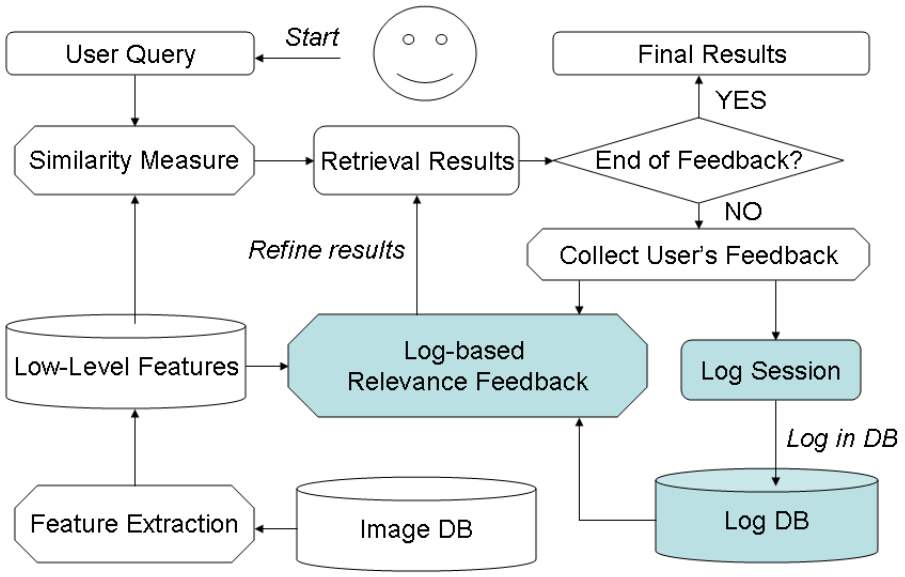

Figure 2.1: The proposed architecture of log-based relevance feedback scheme for CBIR

feedback. When a user launches a query in a CBIR system, he/she may choose to begin a relevance feedback learning procedure if he/she does not find the desired targets within the initially retrieved images. To quantify the log information, each round of relevance feedback can be viewed as a unit of user log session. For each user log session, suppose N_l images are returned to be judged by users, which are marked as either relevant or irrelevant. The relevant and irrelevant images are respectively recorded in the log database as "+1" (positive) and "−1" (negative). To formally describe the log data, a relevance matrix is constructed. Each user log session is represented as a row in the relevance matrix: an element is set to be "+1" when the corresponding image is judged as relevant, and "−1" when the corresponding image is judged as irrelevant; zeros are assigned to elements whose correspond images have not been judged in the log session.

More formally, let $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N\}$ be the collection of images in image retrieval, where N is the number of images in the image database. Let the first N_l images be the samples labeled by users,

and $\mathcal{S}_l$ be the set of labeled images, i.e. $\mathcal{S}_l = \{(\mathbf{z}_i, y_i)\}_{i=1}^{N_l}$, where $y_i \in \{-1, +1\}$ is the label of image $\mathbf{z}_i$. Let $N' = N - N_l$ be the number of unlabeled images, and $\mathcal{S}'$ be the collection of unlabeled images. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)$ represent the low-level information of image content. Let $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ be the relevance matrix for log data, in which each column corresponds to an image in the image database and each row represents a user log session in the log database. Each element $\mathbf{r}_{i,j}$ is the relevance judgement for the i-th image by the j-th user log session ("+1" and "−1" for being relevant and irrelevant, and "0" for not judged). Based on this representation, each image is represented by a vector of relevance judgments $\mathbf{r}_i$, whose dimension M is the total number of user log sessions collected.

Given the above notation, the log-based relevance feedback problem is that, given a query $\mathbf{q}$ and initial labeled collection $\mathcal{S}_l$, how can we combine log data $\mathbf{R}$ with low-level visual information $\mathbf{X}$ to learn users' information needs more efficiently? A natural solution to a learning problem with two orthogonal representations is the co-training algorithm. It explores the correlation between the low-level feature representation and the representation based on collaborative relevance judgments by enforcing their relevance prediction to be consistent. In this paper we formulate the co-training algorithm for image retrieval into a Coupled Support Vector Machine. To better motivate the Coupled SVM method, we will first present a simple approach for log-based relevance feedback, which linearly combines the relevance scores that are computed based on each of the two representations independently.

2.2 Solution I: Linear Combination of Two SVMs

2.2.1 Motivation and Overview

Given the large number of image features and user log session, support vector machines [42] is the ideal choice for log-based relevance feedback problem, which has enjoyed excellent performance for learning in high

dimensional space. For the log-based relevance feedback problem, one straightforward solution is to first learn a different SVM for each of the two representations, and then linearly combine their predictions. In the subsequent parts, we first briefly review the background of SVMs and then give the details of the solution by the linear combination of two SVMs.

2.2.2 Support Vector Machine

As a state-of-the-art discriminative learning technique, SVM has been successfully been applied to a large number of pattern recognition problems, drawing on its superior generalization performance. It has a sound theoretical foundation based on Structural Risk Minimization instead of Empirical Risk Minimization [43]. Let us briefly introduce the basic concept of SVM.

Suppose we are given a set of labeled training data $(\mathbf{x}_1, y_1)$,. . .,$(\mathbf{x}_l, y_l)$ in a binary classification task, where $\mathbf{x}_i$ are the data vectors in some input space $\mathcal{X} \subseteq R^n$, l is the number of training data instances, and $y_i \in \{+1, -1\}$ are the class labels. In the simplest situation, the learning goal of SVM is to find a separating hyperplane that separates the training data with a maximal margin. The primal form of SVM in a linear kernel setting can be expressed as:

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & \forall_{i=1}^{l} : y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i \ , \xi_i \geq 0 \ . \end{aligned}$$

The optimization of SVM is usually solved in a dual form as follows:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j \\ \text{subject to} \quad & \sum_{i=1}^{l}\alpha_i y_i = 0 \ , 0 \leq \alpha_i \leq C \ . \end{aligned}$$

In general, we can project the training data from the original data space $\mathcal{X}$ to a higher dimensional feature space $\mathcal{F}$ by a kernel function

K. The kernel function K, which satisfies a Mercer's condition [43], can be represented as $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, where "$\cdot$" denotes an inner product. $\Phi(\cdot)$ is a mapping function given by $\Phi : \mathcal{X} \longmapsto \mathcal{F}$. Therefore, the decision boundary of SVM with a kernel setting can be represented as: $f(x) = \mathbf{w} \cdot \Phi(\mathbf{x})$, where $\mathbf{w} = \sum_{i=1}^{l} \alpha_i \Phi(\mathbf{x}_i)$.

2.2.3 Linear Combination Approach

Now let us show how to solve the log-based relevance feedback by the linear combination of two SVMs. As we know, in a regular SVM based relevance feedback algorithm [27], only the low-level features of image content is considered. Typically, a vector $\mathbf{w}$ is introduced as the weights of image features, such that the magnitudes of $\mathbf{w}^T\mathbf{X}$ represent the relevance degrees of images to the given query $\mathbf{q}$. Formally, learning the optimal $\mathbf{w}$ by SVM can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{w}, b_w, \xi} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C_w \sum_{k=1}^{N_l} \xi_k \\ \text{subject to} \quad & \forall_{k=1}^{N_l} : \; y_k(\mathbf{w}^T\mathbf{x}_k + b_w) \geq 1 - \xi_k \; , \xi_k \geq 0 \; . \end{aligned} \tag{2.1}$$

Similarly, for the log information, we can also introduce a vector $\mathbf{u}$ as the weights assigned to different user log sessions such that the magnitudes of $\mathbf{u}^T\mathbf{R}$ represent the relevance degrees of images to the given query q. Hence, a maximum margin based approach can also be formulated as follows:

$$\begin{aligned} \min_{\mathbf{u}, b_u, \eta} \quad & \frac{1}{2}\|\mathbf{u}\|^2 + C_u \sum_{k=1}^{N_l} \eta_k \\ \text{subject to} \quad & \forall_{k=1}^{N_l} : \; y_k(\mathbf{u}^T\mathbf{r}_k + b_u) \geq 1 - \eta_k \; , \eta_k \geq 0 \; . \end{aligned} \tag{2.2}$$

The above two SVMs can be solved efficiently by available techniques. Once the two optimal weighting vectors $\mathbf{w}$ and $\mathbf{u}$ are acquired, the sum of their results becomes the solution to the log-based relevance feedback task.

2.3 Solution II: Coupled Support Vector Machine

2.3.1 Motivation and Overview

The straightforward solution by linear combination of two SVMs provides a simple solution to the log-based relevance feedback problem. But, it treats each image representation independently and thus unable to explore the correlation between the two representations. As a consequence, the relevance prediction based on the two representations can be inconsistent. In this section, we present the Coupled Support Vector Machine. It formulates the consistency requirement into a maximum margin based problem. Although we limit ourselves to the case of two representation, one can easily generalize the proposed Coupled SVM to the learning problems in which data points are associated with multiple representations. In the following, we first discuss the formulation of Coupled SVM, followed by the optimization strategy for the Coupled SVM.

2.3.2 Formulation

As described in Section 2.1, images are associated with two representations: the low-level feature presentation $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)$, and the representation based on log data $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$. In unify the two image representations, we can put the two objective functions above together, meanwhile force the relevance prediction based on the two representations to be consistent. More precisely, this idea can be

formulated into the following optimization:

$$
\begin{aligned}
\textit{Minimize over} \quad & (\mathbf{w}, \mathbf{u}, b_w, b_u, \xi, \xi', \eta, \eta', \mathbf{Y}') : \\
& \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2}\|\mathbf{u}\|^2 + C_w \sum_{i=1}^{N_l} \xi_i \; + \; C_u \sum_{i=1}^{N_l} \eta_i \\
& \qquad + \rho C_w \sum_{j=1}^{N'} \xi'_j + \rho C_u \sum_{j=1}^{N'} \eta'_j \qquad (2.3) \\
\text{subject to} \quad & \forall_{i=1}^{N_l} : \; y_i(\mathbf{w}^T \mathbf{x}_i + b_w) \geq 1 - \xi_i \; , \xi_i \geq 0 \\
& \forall_{i=1}^{N_l} : \; y_i(\mathbf{u}^T \mathbf{r}_i + b_u) \geq 1 - \eta_i \; , \eta_i \geq 0 \\
& \forall_{j=1}^{N'} : \; y'_j(\mathbf{w}^T \mathbf{x}'_j + b_w) \geq 1 - \xi'_j \; , \xi'_j \geq 0 \\
& \forall_{j=1}^{N'} : \; y'_j(\mathbf{u}^T \mathbf{r}'_j + b_u) \geq 1 - \eta'_j \; , \eta'_j \geq 0
\end{aligned}
$$

where N_l is the number of labeled images in relevance feedback, N' is the number of unlabeled images, and $\mathbf{Y}' = (y'_1, y'_2, ..., y'_{N'})^T \in -1, 1^{|\mathcal{S}'|}$ is the label (e.g., relevance) vector for the unlabeled images. The first four terms in the above formulism are the combination of (2.1) and (2.2), which include the maximum margins for the classifiers of the two representations and the classification errors for the labeled images. The most interesting part of (2.3) is the last two terms, which can be viewed as the classification errors for the unlabeled images. They distinguish the Coupled SVM approach from the linear combination approach, which does not incorporate the unlabeled images in its learning procedure. The importance of unlabeled data has been realized in the study of semi-supervised learning (e.g.,[44, 45]) , in which both labeled and unlabeled data are used to reduce the generalization errors of learned classifiers. More important, through the introduction of unlabeled images, we are able to enforce the relevance prediction based on the two representations to be consistent. This is realized through the third and fourth constraints, which require the relevance prediction based on image features and log data to be close to $\mathbf{Y}'$. Thus, through the unlabeled images, information is exchanged between the two representations until their relevance predictions are consistent. Given that only a few images have been judged in users' relevance feedback, the

sum of classification errors for unlabeled images can be substantially larger than the sum of classification errors for labeled images. Hence, a parameter ρ is introduced in (2.3) to avoid the dominance of unlabeled data in the learning task. When $\rho = 0$, the last two terms in the objective function in (2.3) becomes zeros and the unlabeled data are ignored in the learning procedure. Finally, it is straightforward to extend the above framework to nonlinear kernels as described in [43].

2.3.3 Alternating Optimization

Finding the optimal solution to the Coupled SVM is not an easy task. This is because the third and fourth constraints involve the product between variable y'_i and feature weights (i.e., $\mathbf{w}$ and $\mathbf{u}$), and thus are no longer linear constraints. Furthermore, they are not even convex constraints. In this section, we presentation a optimization strategy based on the Alternating Optimization (AO) [46] to tackle the problem.

First, we fix the parameters $\mathbf{Y}'$ and try to find the $(\mathbf{u}, b_u)$ and $(\mathbf{w}, b_w)$ that optimize the objective function in (2.3). This amounts to solving the following two independent optimization problems:

$$\begin{aligned} \min_{\mathbf{w},b_w,\xi,\xi',\mathbf{Y}'} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C_w \sum_{i=1}^{N_l} \xi_i + \rho C_w \sum_{j=1}^{N'} \xi'_j \\ \text{subject to} \quad & \forall_{i=1}^{N_l} : \; y_i(\mathbf{w}^T\mathbf{x}_i + b_w) \geq 1 - \xi_i \; , \xi_i \geq 0 \\ & \forall_{j=1}^{N'} : \; y'_j(\mathbf{w}^T\mathbf{x}'_j + b_w) \geq 1 - \xi'_j \; , \xi'_j \geq 0 \; , \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} \min_{\mathbf{u},b_u,\eta,\eta',\mathbf{Y}'} \quad & \frac{1}{2}\|\mathbf{u}\|^2 + C_u \sum_{i=1}^{N_l} \eta_i + \rho C_u \sum_{j=1}^{N'} \eta'_j \\ \text{subject to} \quad & \forall_{i=1}^{N_l} : \; y_i(\mathbf{u}^T\mathbf{r}_i + b_u) \geq 1 - \eta_i \; , \eta_i \geq 0 \\ & \forall_{j=1}^{N'} : \; y'_j(\mathbf{u}^T\mathbf{r}'_j + b_u) \geq 1 - \eta'_j \; , \eta'_j \geq 0 \; . \end{aligned} \tag{2.5}$$

To solve these two problems, we can simply apply the technique used in regular SVMs. For example, to solve the optimization in Eq. (2.4), we introduce non-negative Lagrange multipliers $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \ldots, \alpha_{N_l+N'})$

to enforce the constraints. For the convenience of discussion, let us denote $\hat{\mathbf{Y}}^T = \{\hat{y}_1, \ldots, \hat{y}_{N_l+N'}\}$, where $\hat{y}_i = y_i$ for $i = 1, \ldots, N_l$, and $\hat{y}_{N_l+j} = y'_j$ for $j = 1, \ldots, N'$. It is not difficult to derive the dual form for Eq. (2.4), which is

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - \boldsymbol{\alpha}^T\mathbf{1} \\ \text{subject to} \quad & \boldsymbol{\alpha}^T\hat{\mathbf{Y}} = 0 \\ & \forall_{i=1}^{N_l} : \ 0 \le \alpha_i \le C_w \\ & \forall_{i=N_l+1}^{N_l+N'} : 0 \le \alpha_i \le \rho C_w \end{aligned}$$

where $\mathbf{Q} = [Q_{i,j}]_{(N_l+N')\times(N_l+N')}$ is a positive semidefinite matrix, with $Q_{i,j} \equiv y_i y_j \mathbf{x}_i^T \mathbf{x}_j$. If a kernel is enabled, then $Q_{ij} \equiv \hat{y}_i \hat{y}_j K(\mathbf{x}_i, \mathbf{x}_j)$, and $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$. Note that, when $\rho = 0$, according to the second constraint in the above formulism, all α_is for unlabeled images (corresponding to index $i = N_l+1, ..., N$) have to be zeros, and therefore the unlabeled images are discarded in this setting.

Then, we fix $(\mathbf{w}, b_w)$ and $(\mathbf{u}, b_u)$ and turn to finding the optimal $\mathbf{Y}'$ that fits the data. Given the first four terms in (2.3) do not involve variables $\mathbf{Y}'$ and can be ignored, we have the optimization problem in (2.3) simplified as follows:

$$\begin{aligned} \min_{\xi', \eta', \mathbf{Y}'} \quad & C_w \sum_{j=1}^{N'} \xi'_j + C_u \sum_{j=1}^{N'} \eta'_j \\ \text{subject to} \quad & \forall_{j=1}^{N'} : \ y'_j(\mathbf{w}^T\mathbf{x}'_j + b_w) \ge 1 - \xi'_j \ , \xi'_j \ge 0 \\ & \forall_{j=1}^{N'} : \ y'_j(\mathbf{u}^T\mathbf{r}'_j + b_u) \ge 1 - \eta'_j \ , \eta'_j \ge 0 \ . \end{aligned}$$

If we substitute the slack variables into the objective function, the above optimization problem can be further simplified as follows:

$$\min_{\mathbf{Y}'} \quad \sum_{j=1}^{N'} \left\{ \begin{array}{l} C_w \max(0, 1 - y'_j(\mathbf{w}^T\mathbf{x}_j + b_w)) + \\ C_u \max(0, 1 - y'_j(\mathbf{u}^T\mathbf{r}_j + b_u)) \end{array} \right\}$$

Given that variables $\{y'_j\}_{j=1}^{N'}$ are independent from each other in the above objective function, it can be expanded into a set of univariate

optimization problems and each problem only involves a single variable y'_j. Since each y'_j can only take the value of $+1$ or -1 , an approach based on the exhaustive search can be employed to solve the above optimization problem.

To optimize the objective function in (2.3), we can alternate the above two steps. In particular, we can first randomly choose a set of labels for the unlabeled data, and then launch the alternating optimization procedure beginning with a small value of ρ in order to avoid a predominance of unlabeled data. This is similar to the strategy used in transductive SVM [47]. During the iteration of the above two-step procedure, we slowly increase the value of ρ until it achieves a setting threshold.

It is worth mentioning that although in the above we only describe the Coupled SVM for data with two representation, it can be easily extended to the case when data have multiple representations.

2.4 A Practical Algorithm for Coupled SVM

In the previous algorithm, we have presented a general algorithm for Coupled SVM. In this section, we will discuss one practical consideration in implementing Coupled SVM that could be critical to its performance. In the previous section, we assume that all unlabeled images are used for training Coupled SVM. However, this may not be the best strategy. Instead, we can select a subset of unlabeled images and train the Coupled SVM based on labeled images and selected unlabeled images. The reason for doing so is twofold: First, it is time consuming for training a Coupled SVM model when all unlabeled images are used. By selecting a subset of unlabeled images, we are able to substantially reduce the training cost. Second, the more important reason is because the assumption that both representations will make similar relevance prediction may not hold for *every* image. These two representations of images share very different characteristics and reflect different respects of image content. For example, there could be images

Algorithm LRF-CSVM:

Input:

– **q**: a query sample provided by a user

– $\mathcal{S}_l$: set of N_l labeled samples: $[(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_{N_l}, y_{N_l})]$

Parameters:

– C_w, C_u, ρ: regularization parameters in the optimization (1)

– Δ: a threshold value to control the degree of error

Output:

– $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{N_r}\}$: a set of N_r images most relevant the query **q**.

BEGIN

// 1. Select N' unlabeled samples for the learning task

($\mathbf{w}$, b_w, ξ) = SVM_QP($[(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_{N_l}, y_{N_l})]$, C_w) ;

($\mathbf{u}$, b_u, η) = SVM_QP($[(\mathbf{r}_1, y_1), \cdots, (\mathbf{r}_{N_l}, y_{N_l})]$, C_u) ;

FOR i=1 TO N DO

dist($\mathbf{z}_i$) = SVM_Dist($\mathbf{x}_i$, $\mathbf{w}$, b_w) + SVM_Dist($\mathbf{r}_i$, $\mathbf{u}$, b_u) ;

ENDFOR

$\mathcal{S}'$ = Add_Unlabeled_Samples_with_Max_Dist($N'/2$, dist[]) ;

$\mathcal{S}'$ = Add_Unlabeled_Samples_with_Min_Dist($N'/2$, dist[]) ;

// 2. Train the Coupled Support Vector Machine

$\rho^* = 10^{-4}$

WHILE ($\rho^* < \rho$) DO

($\mathbf{w}$, b_w, ξ, ξ') = SVM_QP($[(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_{N_l}, y_{N_l})]$, $[(\mathbf{x}'_1, y'_1), \cdots, (\mathbf{x}'_{N'}, y'_{N'})]$, C_w, $\rho^* C_w$) ;

($\mathbf{u}$, b_u, η, η') = SVM_QP($[(\mathbf{r}_1, y_1), \cdots, (\mathbf{r}_{N_l}, y_{N_l})]$, $[(\mathbf{r}'_1, y'_1), \cdots, (\mathbf{r}'_{N'}, y'_{N'})]$, C_u, $\rho^* C_u$) ;

WHILE ($\exists i$: ($\xi'_i > 0$) AND ($\eta'_i > 0$) AND ($\xi'_i + \eta'_i > \Delta$)) DO

FOR i=1 TO N' DO

IF (($\xi'_i > 0$) AND ($\eta'_i > 0$) AND ($\xi'_i + \eta'_i > \Delta$)) THEN

$y'_i = -y'_i$;

ENDIF

($\mathbf{w}$, b_w, ξ, ξ') = SVM_QP($[(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_{N_l}, y_{N_l})]$, $[(\mathbf{x}'_1, y'_1), \cdots, (\mathbf{x}'_{N'}, y'_{N'})]$, C_w, $\rho^* C_w$) ;

($\mathbf{u}$, b_u, η, η') = SVM_QP($[(\mathbf{r}_1, y_1), \cdots, (\mathbf{r}_{N_l}, y_{N_l})]$, $[(\mathbf{r}'_1, y'_1), \cdots, (\mathbf{r}'_{N'}, y'_{N'})]$, C_u, $\rho^* C_u$) ;

ENDWHILE

$\rho^* = min(2 * \rho^*, \rho)$;

ENDWHILE

// 3. Retrieve the results by the Coupled SVM

FOR i=1 TO N DO

dist($\mathbf{z}_i$) = CSVM_Dist($\mathbf{x}_i$, $\mathbf{r}_i$, $\mathbf{w}$, b_w, $\mathbf{u}$, b_u) ;

ENDFOR

$\{z_1, z_2, \ldots, z_{N_r}\}$ = Select_Samples_with_Max_CSVM_Dist(N_r,dist[]) ;

END

Figure 2.2: Algorithm for Log-based Relevance Feedback by Coupled SVM

that do not have any relevance judgments in log data and therefore it is impossible to make relevance prediction based on the representation of collaborative relevance judgments. One possible strategy is to choose the unlabeled images that are closest to the decision boundary of both SVMs. This is similar to the active learning approaches that have been successfully employed in the study of image retrieval [27, 48]. Unfortunately, this approach failed to achieve promising improvements in our empirical study. One possible explanation is that images closest to decision boundaries of SVMs could also be the ones for which both representations are *unlikely* to reach agreement on their relevance prediction. Thus, using those unlabeled images tend not to improve the performance of regular SVMs.

In contrast to the selective sampling strategy used in active learning, in the implementation of Coupled SVM, we choose the unlabeled images that are similar to the labeled ones. In particular, the similarity of images will be measured by both the low-level image features and the collaborative relevance judgments in log data as in [30]. Based on this selection strategy, it is likely for the two representations to make consistent prediction of relevance for the selected unlabeled images. Also, the unlabeled images that are similar to the labeled images can be viewed as variants of labeled ones that are slightly "corrupted" by noises. Thus, the introduction of similar unlabeled images is, to some degree, equivalent to introducing small noises into the training data for SVMs, which often improve the robustness of classification models. Thus, for practical implementation of Coupled SVM, we first train two SVM classifiers based on the two representations of images. Then, for each unlabeled image, we compute the sum of its distance to the decision boundaries of the two SVMs. Images with the largest distance will be chosen to form the pool of unlabled data for Coupled SVM. Fig. 2.2 shows the details of the algorithm for the log-based relevance feedback problem by the Coupled SVM (LRF-CSVM). It consists of three main steps, i.e. choosing the unlabeled data, training the Coupled SVM, and retrieving the results by the Coupled SVM. In the training procedure,

a parameter Δ is introduced for controling the error degree of label correction such that most of labeled images are correctly predicted by both representations.

□ **End of chapter.**

Chapter 3

Experimental Results

3.1 Overview of Experiments

In our experiment, we perform extensive comparison to evaluate the effectiveness of the proposed Coupled SVM for relevance feedback. In particular, the following questions will be addressed through the empirical studies. The first question is whether the log-based relevance feedback techniques can achieve better retrieval performance than the regular relevance feedback technique, and how significant is the improvement if they can. The second is whether the Coupled SVM will outperform the linear combination approach for log-based relevance feedback problem? The third question is regarding to the robustness of Coupled SVM. In particular, we will evaluate the performance of Coupled SVM in response to small amount of log data.

3.2 Datasets

To perform empirical evaluation of our proposed algorithm, we choose real-world images from the COREL image CDs. There are two sets of data collected in our experiment: 20-Category and 50-Category. The 20-Category dataset contains 20 categories and the 50-Category one contains 50 categories. Each category in the datasets consists exactly 100 images selected from the COREL image CDs. The categories rep-

resent different semantic meanings, such as *antique*, *antelope*, *aviation*, *balloon*, *botany*, *butterfly*, *car*, *cat*, *dog*, *firework*, *horse* and *lizard*, etc.

The motivations for selecting the semantic categories are twofold. First, it enables us to evaluate whether the approach can retrieve the images that are not only visually relevant but also have similar semantic meaning. Second, the approach can help us evaluate the performance automatically, which can reduce the subjective errors arising from manual evaluations by different people. In particular, two images are deemed relevant as long as they come from the same semantic category.

3.3 Image Representation

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

The color feature is widely adopted in CBIR for its simplicity and effectiveness. The color feature engaged in our experiment is color moment since it is naturally closer to human perception, 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-dimensional color moment is adopted as the color feature in our experiment.

The edge feature can be very effective in CBIR when the contour lines of images are evident. The edge feature in our experiment is the edge direction histogram [49]. The images in the datasets are first translated to gray images. Then a 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 employed to represent the edge feature.

The texture feature is known to be an important cue for image retrieval. A variety of texture analysis methods have been studied in past years. In our experiment, we employ the wavelet-based texture technique [50, 51]. 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 [51]. 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; this is discarded since it contains less useful texture information. For the other 9 subimages, we compute the entropy of each subimage respectively. Therefore, we obtain a 9-dimensional wavelet-based texture feature to describe the texture information for each image.

3.4 Log Data Collection of Users' Feedback

Log data collection of users' feedback is an important step toward performance evaluation of a log-based relevance feedback algorithm for CBIR. Instead of producing simulated log data by computers, we collect the feedback logs from real-world users. The main reason is that the users' feedback log data collected from real-world users typically contain more or less noise that is difficult to be simulated. In order to collect the log data, we have developed a CBIR system powered with a relevance feedback mechanism [37, 22]. In our CBIR system, users can judge the relevance of images simply by ticking out the relevant images. Fig. 3.1 shows the GUI of our CBIR system to collect users' feedback data.

In our experiment, the log data of relevance feedback are collected from users on both the 20-Category and 50-Category dataset. The

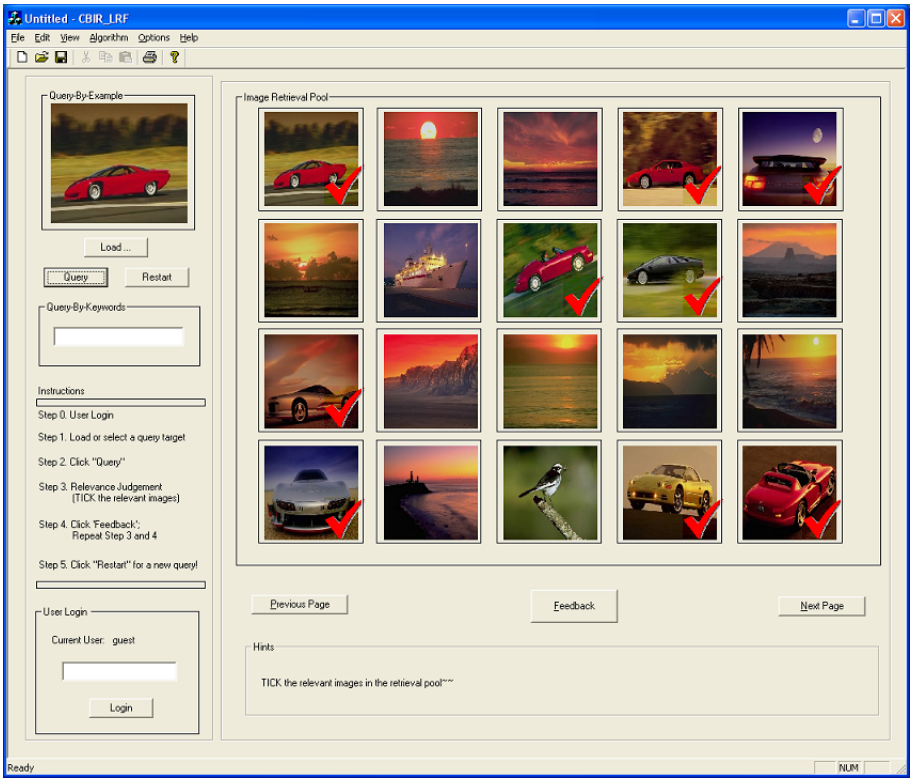

Figure 3.1: The GUI of our CBIR system to collect users' feedback

reason to evaluate the performance on the two datasets is that we want to evaluate algorithms on the datasets with different diversity. The 50-Category dataset is more diverse in visual content than the 20-Category dataset since the number of categories is larger than the 20-Category one. Hence, the log information may be less helpful for the 50-Category dataset which is more diverse than the 20-Category one. It is interesting for us to observe it empirically.

More specifically, the way to organize and collect the log data from users is articulated as follows. For each participant user, he or she first specifies a query example and submits it to the CBIR system. The CBIR system returns 20 initial similar images to the user according the measurement of low-level visual features of image content. The user then employs the relevance feedback tool to improve the retrieval performance. For the given 20 images, he/she marks positive (relevant)

or negative (irrelevant) labels on the images according to his/her query target. When a relevance feedback round is finished, the information of users' feedback will be logged into a log database. Each relevance feedback round corresponds to a log session unit of users' feedback in the log database. Since different people may have different subjectivity, a certain amount of noise is inevitable to appear in the collected log data. The noise problem is not further discussed in this paper, although it may also be a critical factor for the performance evaluation of the log-based relevance feedback algorithms.

In total, we respectively collect 150 log sessions for each of the two datasets from users in the experiment. Although the number of log sessions is not very large, they are enough to evaluate the effectiveness of our algorithm. In reality, many more log sessions can be collected in a real-world CBIR application from a long-term learning perspective; however, we hope to demonstrate that our proposed algorithm can work well even with limited log sessions.

3.5 The Effectiveness of Our Log-based Relevance Feedback Algorithms

We have developed the log-based relevance feedback algorithms by Coupled SVM (LRF-CSVM) in our experiment. We implement the Coupled SVM algorithm by modifying the LIBSVM library [52]. In order to evaluate the effectiveness of our method, we compare it to regular relevance feedback algorithm by SVM (RF-SVM), which learns a classification model based on the labeled images acquired from user relevance feedback. We also compare the performance of the Coupled SVM to the linear combination approach (LRF-2SVMs), which trains a different SVM approach for each image representation and linear combine their prediction. A Gaussian RBF kernel [52] for all learning schemes. The performance metric used in the experiment is *Average Precision*, which is defined as the percentage of the relevant samples among all returned images. 200 queries are randomly selected from the image

collection. As the referenced approach for image retrieval, a Euclidean distance is used to compute the difference between a query image and images in collections. The top 20 images with the least distance are returned for evaluation. The relevance of a returned image to the given query is determined by if they belong to the same category. Based on a query **q** and 20 labeled images, the three different relevance feedback schemes will be applied and their performance will be evaluated by the accuracy of the images retrieved by these three methods. The experimental results are obtained by taking an average of retrieval accuracy over the 200 queries.

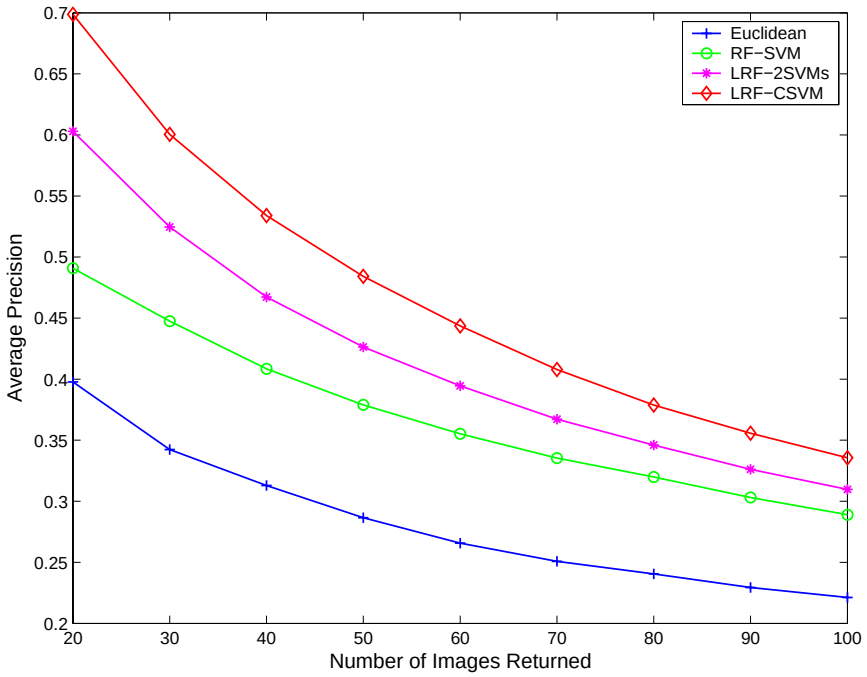

Figure 3.2: Performance comparison on 20-Category dataset

Fig. **??** and Fig. 3.3 illustrate the visual comparison of the experimental results on the two datasets. In the figures, the results based on Euclidean distance is given as a reference. The results of RF-SVM, represented by the curves with circles, is used as the baseline for performance comparison. Both figures evidently show that the log-based relevance feedback techniques, both LRF-2SVM and LRF-CSVM, can

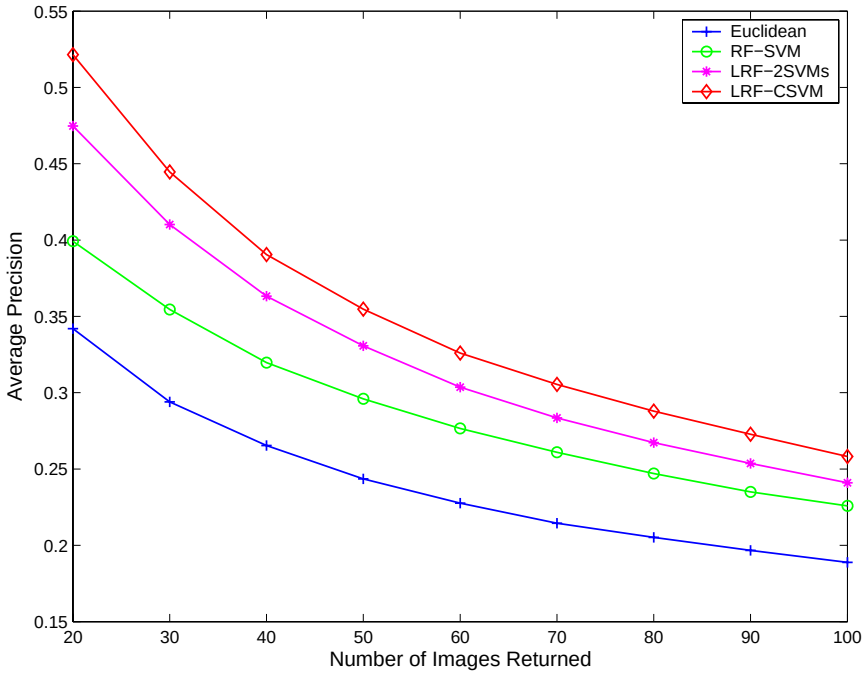

Figure 3.3: Performance comparison on 50-Category dataset

substantially improve the retrieval performance when compared with the regular relevance feedback scheme. Moreover, according to both figures, we clearly observe that the log-based relevance feedback by the Coupled SVM substantially outperform the log-based relevance feedback by a simple combination of two SVMs.

To examine the quantitative amount of improvement, let us look into more detailed experimental results in Table 3.1 and Table 3.2. These two tables show the average precision of the CBIR using Euclidean distance and three feedback techniques for the top returned images. They also include the relative improvement of the two log-based relevance feedback techniques (i.e., LRF-2SVM and LRF-CSVM) compared to the regular relevance feedback approach (RF-SVM). As indicated in Table 3.1 and 3.2, the overall trend is that the log-based feedback techniques make considerable improvement compared to the regular feedback technnique when the number of retrieved images is large, and the improvement starts to decrease when the number of

Table 3.1: Quantitative evaluation for different approaches on the 20-Category dataset

#TOP	Euclidean	RF-SVM	LRF-2SVMs	LRF-CSVM
20	0.398	0.491	0.603 (+22.9%)	0.699 (+42.4%)
30	0.342	0.448	0.525 (+17.2%)	0.600 (+34.2%)
40	0.313	0.409	0.467 (+14.4%)	0.534 (+30.7%)
50	0.287	0.379	0.426 (+12.5%)	0.484 (+27.8%)
60	0.266	0.355	0.394 (+11.0%)	0.444 (+24.9%)
70	0.251	0.335	0.367 (+9.5%)	0.408 (+21.6%)
80	0.241	0.320	0.346 (+8.2%)	0.379 (+18.4%)
90	0.229	0.303	0.326 (+7.6%)	0.356 (+17.4%)
100	0.221	0.289	0.310 (+7.2%)	0.336 (+16.1%)
MAP	0.283	0.370	0.418 (+12.3%)	0.471 (+25.9%)

Table 3.2: Quantitative evaluation for different approaches on the 50-Category dataset

#TOP	Euclidean	RF-SVM	LRF-2SVMs	LRF-CSVM
20	0.342	0.399	0.475 (+18.9%)	0.522 (+30.6%)
30	0.294	0.355	0.410 (+15.7%)	0.445 (+25.4%)
40	0.265	0.320	0.363 (+13.6%)	0.391 (+22.1%)
50	0.244	0.296	0.331 (+11.7%)	0.355 (+19.8%)
60	0.228	0.277	0.304 (+9.8%)	0.326 (+17.9%)
70	0.215	0.261	0.283 (+8.6%)	0.305 (+17.1%)
80	0.205	0.247	0.267 (+8.2%)	0.288 (+16.5%)
90	0.197	0.235	0.254 (+7.9%)	0.273 (+16.1%)
100	0.189	0.226	0.241 (+6.7%)	0.258 (+14.4%)
MAP	0.242	0.291	0.325 (+11.2%)	0.351 (+20.0%)

retrieved images is small. In other words, most of the improvement gained by the incorporation of log data happens to the top retrieved images. This is also true when we compare the two log-based feedback techniques. For example, on the 20-Category dataset, for the top 20 returned images, the log-based feedback technique by the Coupled SVM achieves 42.4% improvement compared to the regular relevance feedback approach. It substantially outperforms the log-based feedback technique by the linear combination, which only achieves relatively 22.9% improvement. In contrast, for the top 100 retrieved images, both approaches achieves less than 20% improvement compared to the regular feedback technique. On average, for the 20-category dataset, compared to regular relevance feedback, the Coupled SVM based approach achieves 25.9% improvement in average precision compared, while the LRF-2SVMs approach only obtains 12.3% improvement. Similarly, on the 50-Category dataset, the Coupled SVM approach achieves 20.0% improvement on average, while the linear combination approach only obtains 11.2% improvement. From Table 3.1 and 3.2, we also observe that the amount of improvement on the 50-Category dataset is less than that on the 20-Category dataset. This is because images in the 50-Category dataset is more diverse than images in the 20-Category, which makes it difficult for a learning algorithm to improve the retrieval accuracy. Nevertheless, the overall improvements by the log-based relevance feedback algorithms for both datasets are still very promising.

3.6 Performance Evaluation by Different Amount of Log Data

To further verify the performance of our proposed algorithms, we are interested to check how our algorithms perform with different amount of log data and whether the log-based relevance feedback scheme can achieve improvement under very limited log data. To this end, we generate several of log data with different sizes by randomly sampling user log sessions from the 150 log sessions that are collected from real-world

users on each dataset. Table 3.3 and Table 3.5 show the retrieval accuracy of the top 20 retrieved images for log-based feedback algorithms using different amount of log data. For the purpose of comparison, we also include the Euclidean distance as the reference point and the regular relevance feedback technique by SVM as the baseline model. According to Table 3.3 and 3.5, on average, we can observe the Coupled SVM approach is more effective than the linear combination when the amount of log data is small. For example, with only 90 user log sessions, the retrieval accuracy of the Coupled SVM is 63.7%, which is substantially higher than that for the linear combination approach using 150 log sessions. Furthermore, we can see that even with limited amount of log data, the Coupled SVM algorithm achieves reasonably good results. For example, with only 30 log sessions, the Coupled SVM approach achieve about 12% improvement on the 20-Category dataset and 7.4% improvement on the 50-Category dataset. One may find that the improvement on the 50-Category dataset is smaller than the 20-Category one. The reason is that the 50-Category dataset contains more categories and more diverse content which requires larger amount of users' log data less to achieve the same amount of improve as the 20-Category dataset. However, this problem can be better from a long-term learning purpose in which one can reasonably assume a CBIR system typically can collect large amount of feedback log data from users. Thus, we conclude that our algorithm for log-based relevance feedback works even with very limited amount of log data.

3.7 Evaluation of Time Efficiency

Although the above experimental results show that the Coupled Support Vector Machine is effective for log-based relevance feedback, one problem for applying the Coupled SVM to log-based relevance feedback, i.e., the time efficiency problem, must be addressed. The Coupled SVM is based on the co-training scheme which usually takes more time cost compared with a regular support vector machine algorithm.

Table 3.3: Performance evaluation via different amount of log data on the 20-Category dataset. The results are average precision on top-20 returned images.

#Log Sessions	Euclidean	RF-SVM	LRF-2SVMs	LRF-CSVM
30	0.398	0.491	0.524 (+6.8%)	0.550 (+12.1%)
60	0.398	0.491	0.555 (+13.0%)	0.579 (+17.9%)
90	0.398	0.491	0.580 (+18.2%)	0.637 (+29.9%)
120	0.398	0.491	0.598 (+21.9%)	0.671 (+36.7%)
150	0.398	0.491	0.603 (+22.9%)	0.699 (+42.4%)
MAP	0.398	0.491	0.572 (+16.6%)	0.627 (+27.8%)

Table 3.4: Performance evaluation via different amount of log data on the 50-Category dataset. The results are average precision on top-20 returned images.

#Log Sessions	Euclidean	RF-SVM	LRF-2SVMs	LRF-CSVM
30	0.342	0.399	0.425 (+6.5%)	0.429 (+7.4%)
60	0.342	0.399	0.439 (+9.9%)	0.446 (+11.6%)
90	0.342	0.399	0.441 (+10.5%)	0.453 (+13.5%)
120	0.342	0.399	0.456 (+14.2%)	0.489 (+22.4%)
150	0.342	0.399	0.475 (+18.9%)	0.522 (+30.6%)
MAP	0.398	0.399	0.447 (+12.0%)	0.468 (+17.1%)

Table 3.5: Time cost of algorithms over 200 executions on the 20-Category and 50-Category datasets (seconds).

Dataset	RF-SVM	LRF-2SVMs	LRF-CSVM
20-Category	1.50	3.25	36.59
50-Category	1.47	3.12	35.52

Hence, it is necessary to evaluate whether the Coupled SVM approach is efficient enough for a practical relevance feedback problem. To evaluate the time efficiency, we measure the training time cost of the algorithms in our scheme. Table shows the experimental results of time cost evaluated on 200 executions. We can see that although extra time cost has to be paid in the Coupled SVM scheme, the resulting time efficiency is still acceptable. On average, each feedback execution by Coupled SVM on both datasets took less than 0.2 second which is fast enough for a practical relevance feedback problem. But the time efficiency could not be ignored when applying for large scale problems which will be studied in our future work.

3.8 Discussions and Limitations

3.8.1 Log-based Relevance Feedback vs. Relevance Feedback

Although we have demonstrated the effectiveness of our log-based relevance feedback algorithms in the empirical study, several important issues and limitations of our algorithm are worth discussing here. First, one may challenge whether the log-based relevance feedback is feasible in real-world CBIR applications. The answer is positive since the log data of users' relevance feedback are one of the most important resources for reducing the semantic gap between low-level features and high-level concepts. The goal of log-based relevance feedback is to narrow down the semantic gap by learning with users' log data as well as low-level image content. Second, one may ask will the log-based relevance feedback technique fails to improve the performance in some

cases. This is possible due to the unpredicted nature and the diversity in users' information needs and the related feedback. Hence, we must emphasize the log-based relevance feedback algorithm is not to replace the traditional relevance feedback, but served as an important tool to boost the retrieval performance before launching a regular relevance feedback procedure.

3.8.2 Log-based Relevance Feedback via Coupled SVM

The promising experimental results show the Coupled SVM is effective, but several issues and limitations of the solution by Coupled SVM need to be addressed. First, the selection strategy of unlabeled data for the Coupled SVM is important in an image retrieval environment. In this paper, we employ a heuristic strategy, which chooses unlabeled images closest to the positive labeled images for half the samples, and those closest to the negative labeled images for the other half. In the future work, we will examine other more principled approach for selecting examples. Second, the choice of parameter ρ is also important for the algorithm. We have not yet found a theoretic justification for selecting an optimal parameter. Finally, the time efficiency of log-based relevance feedback is important when applying to large databases in real-world application. This should be studied in our future work.

3.8.3 The Coupled SVM for Multi-Modality Learning

It is worth making some comments on the Coupled SVM for learning on multiple-modality problems. In addition to data with two representations, the Coupled SVM can be easily generalized to learn the data with multiple representations. However, there are several open problems to be solved. First, the current approach for the optimization of the Coupled SVM is based on the Alternating Optimization technique, which may not be able to guarantee the optimal solution globally. It is interesting to seek other optimization techniques for tackling the problem. Moreover, whether existing a better formulation of the Coupled

SVM is worth discussing both theoretically and empirically.

□ **End of chapter.**

Chapter 4

Conclusions and Future Work

4.1 Conclusions

A log-based relevance feedback scheme for content-based image retrieval was studied in this article by integrating the log data of users' feedback with low-level image content. We suggest to represent the image by two types of information, i.e., one is based on low-level image content, another is based on log data of users' feedback. In order to learn with these two representations effectively, a co-training algorithm, formulated mathematically as **"Coupled Support Vector Machine"**, is suggested to solve the log-based relevance feedback problem. Extensive empirical results have shown the proposed scheme is effective. Moreover, the Coupled Support Vector Machine, is not only effective for the log-based relevance feedback problem, but also can be a promising solution for a general learning problem in which data are given by multiple representations.

4.2 Future Work

In our future work, we are going to evaluate and improve our log-based relevance feedback scheme in the following directions. First of all, to

further evaluate the performance of our proposed algorithms, we plan to test our algorithm on larger datasets and diverse environments. For the Coupled Support Vector Machine, although it is used for solving the log-based relevance feedback in this work, the general idea can be employed to solve other related problems. We will consider to apply it for other multiple modalities problem, e.g. Web image retrieval problems. Moreover, some problems, such as the convergence issue and sample selection strategies, need to be further studied in our future work. Finally, we notice the time efficiency is a disadvantage of Coupled Support Vector Machine and can be critical when applying for larger datasets. To improve the computational performance, one possible solution is to introduce probability constrains on unlabeled data which results in a quadratic programming problem of the original optimization that can be solved efficiently. We will study this possible solution in our future work.

□ **End of chapter.**

Bibliography

[1] A. Gupta and R. Jain, "Visual information retrieval," *Communications of the ACM*, vol. 40, no. 5, pp. 70–79, 1997.

[2] D. Aigrain, H. Zhang, and D. Petkovic, "Content-based representation and retrieval of visual media: A state of the art review," *Multimedia Tools and Applications*, vol. 3, pp. 178–202, 1996.

[3] Y. Rui, T. S. Huang, and S.-F Chang, "Image retrieval: Current techniques, promising directions, and open issues," *J. Visual Comm. and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.

[4] 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*, vol. 22, no. 12, pp. 1349–1380, 2000.

[5] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee, "Shape measures for content-based image retrieval: A comparison," *Information Processing Management*, vol. 33, no. 3, pp. 319–337, 1997.

[6] S. Chang and A. Hsu, "Image informations systems: where do we go from here?," *IEEE Trans on Knowledge and Data Engineering*, vol. 4, no. 5, pp. 431–442, 1992.

[7] H. Tamura and N. Yokoya, "Image database systems: A survey," *Pattern Recognition*, vol. 17, no. 1, pp. 29–43, 1984.

[8] E. Riloff and L. Hollaar, "Text databases and information retrieval," *ACM Comput. Surv.*, vol. 28, no. 1, pp. 133–135, 1996.

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

[10] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *Proc. the 7th European Conf. on Computer Vision*, 2002, pp. 97–112.

[11] 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)*, 2003, pp. 119–126.

[12] 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.

[13] W. Niblack, R. Barber, and et al., "The QBIC project: Querying images by content, using color, texture, and shape," in *Storage and Retrieval for Image and Video Databases (SPIE)*, 1993, pp. 173–187.

[14] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Past, present, and future," in *International Symposium on Multimedia Information Processing*, 1997.

[15] 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*, vol. 8, no. 5, pp. 644–655, Sept. 1998.

[16] Essam A. El-Kwae and Mansur R. Kabuka, "A robust framework for content-based retrieval by spatial similarity in image databases," *ACM Trans. Inf. Syst.*, vol. 17, no. 2, pp. 174–198, 1999.

[17] M. Stricker and M. Orengo, "Similarity of color images," in *Proc. of SPIE Storage and Retrieval for Image and Video Databases*, 1995, vol. 2420, pp. 381–392.

[18] S. Santini and R. Jain, "Similarity measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 871–883, 1999.

[19] G. D. Guo, A.K. Jain, W.Y. Ma, and H.J. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 811–820, 2002.

[20] C. H. Hoi and Michael R. Lyu, "Web image learning for searching semantic concepts in image databases," in *Poster Proc. 13th Int. World Wide Web Conference (WWW 2004)*, New York, US, 2004.

[21] 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)*, Santa Barbara, CA, USA, 1998, pp. 553–558.

[22] C. H. Hoi and Michael R. Lyu, "Group-based relevance feedback with support vector machine ensembles," in *Proceedings 17th International Conference on Pattern Recognition (ICPR'04)*, Cambridge, UK, 2004, pp. 874–877.

[23] T. S. Huang and X. S. Zhou, "Image retrieval by relevance feedback: from heuristic weight adjustment to optimal learning methods," in *Proceedings of IEEE International Conference on Image Processing (ICIP'01)*, Thessaloniki, Greece, Oct. 2001.

[24] 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.

[25] K. Tieu and P. Viola, "Boosting image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, South Carolina, USA, 2000.

[26] 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.

[27] Simon Tong and Edward Chang, "Support vector machine active learning for image retrieval," in *Proceedings of the ninth ACM international conference on Multimedia (MM'01)*, 2001, pp. 107–118.

[28] Xiang-Dong Zhou, Liang Zhang, Li Liu, Qi Zhang, and Bai-Le Shi, "A relevance feedback method in image retrieval by analyzing feedback log file," in *Proc. International Conference on Machine Learning and Cybernetics*, Beijing, 2002, vol. 3, pp. 1641–1646.

[29] X. He, W.-Y. Ma, and H.-J. Zhang, "Learning an image manifold for retrieval," in *Proceedings of ACM MM 2004*, 2004.

[30] C. H. Hoi and Michael R. Lyu, "A novel log-based relevance feedback technique in content-based image retrieval," in *Proc. ACM Multimedia (ACM-MM'04)*, New York, USA, 2004, pp. 24–31.

[31] Avrim Blum and Tom Mitchell, "Combining labeled and unlabeled data with co-training," in *COLT' 98: Proceedings of the eleventh annual conference on Computational learning theory*, New York, NY, USA, 1998, pp. 92–100, ACM Press.

[32] Kamal Nigam and Rayid Ghani, "Analyzing the effectiveness and applicability of co-training," in *CIKM '00: Proceedings of the ninth international conference on Information and knowledge management*, New York, NY, USA, 2000, pp. 86–93, ACM Press.

[33] Hang Li and Cong Li, "Word translation disambiguation using bilingual bootstrapping," *Comput. Linguist.*, vol. 30, no. 1, pp. 1–22, 2004.

[34] J. Rocchio, "Relevance feedback information retrieval," in *The Smart retrieval system: experiments in automatic document processing*, G. Salton, Ed., 1971.

[35] G. Salton and C. Buckley, "Improving retrieval performance by relevance feedback," *Journal of the American Society for Information Science*, vol. 44, no. 4, pp. 288–287, 1990.

[36] Y. Rui, T.S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in mars," in *Proceedings of IEEE International Conference on Image Processing (ICIP'97)*, Washington, DC, USA, Oct. 1997, pp. 815–818.

[37] C. H. Hoi and Michael R. Lyu, "Biased support vector machine for relevance feedback in image retrieval," in *Proceedings International Joint Conference on Neural Networks (IJCNN'04)*, Budapest, Hungary, 2004, pp. 3189–3194.

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

[39] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma, "Probabilistic query expansion using query logs," in *Proc. of the eleventh international conference on World Wide Web (WWW'02)*, 2002.

[40] Xiaofei 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*, vol. 13, no. 1, pp. 39–48, Jan. 2003.

[41] X. He, W.-Y. Ma, and H.-J. Zhang, "Learning an image manifold for retrieval," in *Proc. ACM Multimedia*, New York, US, 2004, pp. 17–23.

[42] C.J.C. Burges, "A Tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

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

[44] T. Joachims, "Transductive Inference for Text Classification using Support Vector Machines," in *Proceedings of The Sixteenth International Conference on Machine Learning (ICML 99)*, 1999.

[45] Lei Wang, Kap Luk Chan, and Zhihua Zhang, "Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'03)*, 2003, vol. 1, pp. 629–634.

[46] James C. Bezdek and Richard J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel Sci. Comput.*, vol. 11, no. 4, pp. 351–368, 2003.

[47] Thorsten Joachims, "Transductive inference for text classification using support vector machines," in *Proc. 16th Int. Conf. on Machine Learning (ICML'99)*.

[48] Simon Tong and Daphne Koller, "Support vector machine active learning with applications to text classification," *Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2001.

[49] A. K. Jain and A. Vailaya, "Shape-based retrieval: a case study with trademark image database," *Pattern Recognition*, , no. 9, pp. 1369–1390, 1998.

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

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

[52] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.