Learning on Relevance Feedback in Content-based Image Retrieval

HOI, Chu-Hong

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Computer Science and Engineering

July 2004

Abstract of thesis entitled:

Learning on Relevance Feedback in Content-based Image Retrieval
Submitted by HOI, Chu-Hong
for the degree of Master of Philosophy
at The Chinese University of Hong Kong in July 2004

Relevance feedback has been shown as a powerful technique for interactive Content-based Image Retrieval (CBIR). Although a number of state-of-the-art techniques have been devoted to relevance feedback, such as Support Vector Machines (SVMs), existing techniques still have many drawbacks and limitations, including (1) without concerning the imbalanced dataset problem; (2) paying little attention to the insufficient training samples problem; (3) assuming training samples are simply from one positive class and one negative class; and (4) requiring many rounds of feedback to achieve satisfactory results. This thesis investigates the learning techniques on relevance feedback to address the above problems. It also proposes effective algorithms to improve them through different perspectives.

We first present a novel technique to attack the imbalance problem of relevance feedback in CBIR. Regular relevance feedback techniques typically treat fairly with relevant and irrelevant samples. However, in real-world relevance feedback tasks, the irrelevant samples often overnumber the relevant samples. Moreover, the relevant samples normally are clustered in the same way while the irrelevant samples are positioned in different ways. In order to deal with this imbalance problem, we suggest a modified SVM technique called Biased SVM to formulate the relevance feedback algorithm, which is verified from convincing experimental results.

Second, we indicate similarity measure metrics learned by regular SVM-based relevance feedback techniques will be skew when facing insufficient training samples. To overcome the problem, a unified learning scheme with the optimizing learning and the SVM-based boundary constraint is proposed in the thesis. Experimental results are reported to show the effectiveness of the suggested scheme.

Furthermore, traditional learning techniques for relevance feedback usually assume training samples are drawn from one positive class and one negative class. We argue that these approaches are not powerful enough. It is more practical and reasonable to consider the relevant samples coming from multiple positive classes and the irrelevant samples coming from one negative classes in real-world applications. Based on this relaxation, a novel group-based relevance feedback algorithm constructed with SVM ensembles is presented in this thesis.

In addition to regular relevance feedback, we also study log-based relevance feedback algorithms through a long-term learning perspective. In CBIR retrieval tasks, users are often required to repeat many rounds of relevance feedback in order to achieve satisfactory results. To overcome the problem in a long-term purpose, a novel scheme is proposed to learn the users' feedback logs engaging a modified SVM technique called Soft Label SVM in this thesis. The suggested scheme is validated by promising experimental results.

Finally, a simple yet meaningful application employing relevance feedback is presented. It suggests to learn Web images for searching semantic concepts in image databases by engaging a relevance feedback mechanism. Promising results are observed from the proposed application.

在基於內容的圖像檢索中學習相關反饋技術之研究

摘要

相關反饋技術已經被證實是一項強大的技術應用於可交互式的基於內容圖像檢索。儘管已經有不少先進的学习技術提出來構建相關反饋算法，比如最近的支持向量機器(SVM)等，目前的技術仍有許多缺陷和不足，其中包括以下几个方面：(1) 没有考虑到不平衡数据集的问题；(2) 较少考虑不足够训练样本的影響；(3) 通常假设训练样本只来自一个正类和一个负类； (4) 需要重復多次反饋學習才能得到比較理想的結果。針對這些問題，本論文研究基於內容圖像檢索中相關反饋的學習問題，並通過不同的學習角度提出有效的算法來改善這些問題。

首先，我們提出一個新的技術來克服基於內容圖像檢索中相關反饋學習的數據集不平衡問題。正規的相關反饋技術通常會平衡地處理相關和無關的訓練樣本。然而，在實際的相關反饋任務里，無關的訓練樣本數目通常要遠超出相關樣本的數目。而且，相關樣本通常會按同一方式群集在一起，而無關樣本則會以不同方式分佈。爲了處理相關反饋數據集的不平衡問題，我們提出一個改良的支持向量機技術 Biased SVM 來構造相關反饋算法，並且通過實驗核實其有效性。

其次，我們指出當面對不足夠訓練樣本的時候，利用正規的基於 SVM 相關反饋技術學習得到的相似性尺度會偏離正確的尺度。爲了改善這個問題，我們提出一個統一的學習方案，整合了最優化學習和基於 SVM 學習的約束。實驗結果表明我們提出的方案在不足夠訓練樣本下能有效地改善檢索的性能。

此外，傳統的相關反饋學習技術通常認爲訓練樣本只來自一個正類和一個負類，這種處理實際是不夠有效的。在現實的應用中，更實用和合理的做法是考慮相關樣本來自多個正類，而無關樣本來自多個負類。基於這種擴展，我們提出一種新的基於“組”的相關反饋算法，並且使用 SVM Ensembles 技術來構造該算法。

除了正規的相關反饋技術，出於長期的學習考慮，我們也研究基於用戶日誌的相關反饋算法。在基於內容圖像檢索的任務，用戶通常需要重復多次相關反饋學習才能獲得比較滿意的結果。爲了從長遠的目標來改善這個問題，我們提出一個新的方案，採用一項改進的 SVM 技術 Soft Label SVM 來學習用戶的日誌。頗佳的實驗結果證實我們提出的方案能有效地改善檢索的性能。

最後，我們給出一個簡單而有意義的例子，説明如何利用相關反饋技術到實際的應用中。在該例子里，我們提出通過使用網上圖像，採用相關反饋機制來學習從圖像數據庫里搜索語義概念的技術。從該應用的實驗里，我們觀察到不錯的實驗結果。

Acknowledgement

I would like to thank my supervisor, Prof. Michael R. Lyu, for providing the guidance and constructive criticism, and for giving me freedom to explore the interesting research areas over the past two years. Without his support, inspiration, and patience, my work in these two years would not be so much fruitful. I also would like to take this opportunity to thank my other committee members who are nice and supportive. Prof. Yiu-Sang Moon and Prof. Irwin King have contributed a lot to my research work from discussions and cooperations.

I would like to specially thank my fellow research partners, Wei Wang, Kaizhu Huang, Chi-Hang Chan, and Haiqin Yang, who have given me help or constructive comments on my research work. And I also like to show my gratitude to the professors and staffs in the department of Computer Science and Engineering, CUHK, for the provision of the high quality research environment and varied assistances. The chairman of department, Prof. Kwong-Sak Leung, and the head of graduate division, Prof. Kam-Wing Ng, have kindly provided much help on all kinds of fairs of my postgraduate study. Prof. J. C. Liu and Prof. T. T. Wong had taught me a lot of things through the TA work with them.

Moreover, I would like to thank my dear friends and fellow colleagues, Yingge Qu, Shi Lu, Zhiyong Liu, Charles Chiu, Xuelei Hu, John Tang, Kun Zhang, Yang Lu, Xinyu Chen, Edith Ngai, Pat Chan, GiGi Li, Xia Cai, Yangfan Zhou, and many others. I have had enjoyable and memorable time with them in the past two years.

This work is dedicated to my beloved parents and family for their persistent love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Content-based Image Retrieval

Along with the rapid development 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 accessing in image and video databases, more and more research efforts have been devoted to visual information retrieval in recent years [1, 2, 3, 4, 5, 6, 7]. One typical approach is to employ the textual descriptions or keywords for indexing and retrieving images [8, 9, 10]. However, many limitations of the text-based methods make them still far from working in real-world applications nowadays. A key task of text-based approaches is to annotate and index the images with keywords. Traditional techniques require much human labor to annotate the images. This is quite challenging and almost prohibited in practical applications. Although there are some promising advances of image annotation techniques in recent research work [11, 12, 13, 14, 15], fully automatic and practical annotation methods are still on a long way off. Moreover, textual descriptions may have limited capacity to represent 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 the text-based retrieval mechanism, Content-based Image Retrieval (CBIR) has been suggested as an alternative approach for content accessing in multimedia databases in the early 1990's [16, 17, 18]. In recent years, CBIR has become one of the most active research areas in the visual information retrieval [3, 4, 19, 20, 21]. In CBIR, low-level features (such as color, shape, and texture, etc.) are 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 with visual descriptions of low-level features, visual queries can be formulated and similarity of images can be measured by employing a distance function defined in the CBIR systems [22, 23, 24].

At the early stage of CBIR, major research efforts focused on the feature identification and expression for the best representation of the content of images [4, 25]. Fruitful research outputs had been contributed to varied effective feature representations on color [22, 26, 27], shape [5, 28, 29], and texture [25, 30, 31, 32], etc. At that period, researchers typically adopted the "*computer centric*" based mechanism to design the CBIR systems [18, 25]. In the "*computer centric*" based CBIR systems, users first are required to select the features they are interested and specify the weights of the features in accordance with their preferences. Then, the CBIR systems return the most similar images by ranking the similarity based on the selected features and their corresponding weights.

However, those early CBIR systems with heuristic feature selections and fixed weighting schemes did not achieve satisfactory performance. Later, researchers noticed and recognized the difficulties of CBIR [18], i.e., the semantic gap between high-level concepts and low-level features, and the subjectivity of human perception. The "*computer centric*" based CBIR systems are not powerful enough to capture the high-level concepts and overcome the subjectivity of human perception.

In order to surmount the limited capacity of the early "*computer centric*" based CBIR systems, human's interaction was suggested to be involved in the retrieval procedures for constructing the interactive mechanism in recent CBIR research work [17]. Under this circumstance, relevance feedback was introduced into CBIR as a powerful technique to attack these challenges [18, 33].

1.2 Relevance Feedback

Relevance feedback is originally from traditional text-based information retrieval [34], in which it has been proved as an effective technique to improve the retrieval performance. It is an interactive technique which refines the retrieval performance by engaging user's fed-back information. With the motivation to overcome the limitations of the early CBIR systems, relevance feedback was employed to build the interactive mechanism for refining user's query concepts in the retrieval tasks [35, 36, 37]. In order to acquire a user's feedback information, the CBIR system first displays the user a few image samples simply employing some kind of similarity measure metric. Based on the initial returned results, the user marks labels on the images to indicate which of them are "relevant" and which are "irrelevant". By incorporating with the fed-back information, the relevance feedback module is employed to refine the user's query concepts and returns a set of better results to the user. By engaging such a relevance feedback mechanism, the user can retrieve his/her desired images round-by-round through the interactions with the computers. Based on the assumption that high-level concepts can be captured by low-level features, the refinement scheme with relevance feedback will return a set of images which are considered as the desired targets of the user after several rounds of feedback.

In the past years, much research work has been devoted to relevance feedback in CBIR [33, 38, 39, 40, 41]. It has been shown as a powerful technique to narrow down the semantic gap and to overcome the subjectivity of human perception existed in CBIR [18, 42]. However, due to the difficulties of CBIR retrieval tasks, relevance feedback has not yet been a very sophisticated and successful technique in real-world applications. There is still a long way to complete it in the research societies. Recently, along with the promising advances in statistical learning theory and machine intelligence, more and more machine learning techniques are suggested to attack the relevance feedback problems [43, 44, 45, 46, 47, 48, 49]. These approaches have created a new research realm for exploring more effective relevance feedback algorithms and gradually make it to become a more sophisticated technique in empirical CBIR applications. Although a lot of efforts have been devoted to relevance feedback, existing techniques still have many drawbacks and limitations, including (1) without concerning the imbalanced dataset problem; (2) paying little attention to the insufficient training samples problem; (3) assuming training samples are simply from one positive class and one negative class; and (4) requiring many rounds of feedback to achieve satisfactory results.

The goal of this thesis is to investigate the learning issues on relevance feedback in the context of CBIR particularly in exploring the state-of-the-art machine learning techniques, such as Support Vector Machines (SVMs) [50]. The thesis suggests to study the learning problem on the relevance feedback tasks through different perspectives to overcome the drawbacks and limitations of traditional methods. The contributions and organization of this work are discussed in the following sections.

1.3 Contributions

This thesis focuses on studying the learning techniques for relevance feedback algorithms in CBIR. Our main contributions are contributed to the following aspects.

One of our major contributions is to investigate the learning technique for the imbalance problem of relevance feedback in CBIR. Regular relevance feedback normally assume the relevance feedback as a strict binary classification task without bias consideration [44, 43]. To overcome the imbalanced dataset problem of relevance feedback tasks, we propose a novel Biased SVM technique to construct the relevance feedback algorithm in CBIR. The proposed technique outperforms conventional techniques from promising experimental results. The related paper of this work can be found in [51].

The second important contribution in this thesis is a unified scheme for learning similarity measure metrics with relevance feedback by optimizing learning [52] and the SVM-based constraint [44]. The proposed scheme can reduce the impact of SVM-based learning scheme when facing insufficient training samples. Also, it significantly improves the retrieval performance of traditional optimizing leaning scheme by engaging the SVM-based constraint. Details of this work can be found in [53].

Furthermore, this thesis suggests that it is more reasonable and practical to assume that relevant samples are drawn from multiple positive classes and irrelevant samples are drawn from one negative class. Based on this relaxation different from conventional approaches [54, 55], we propose a novel group-based relevance feedback scheme by employing the SVM ensembles technique in CBIR. We show that our proposed scheme is more effective than traditional techniques. The related paper can be found in [56]

Besides of the contributions in regular relevance feedback, the thesis also contributes to log-based relevance feedback techniques [47]. In order to learn users' feedback logs in a long-term purpose, we propose a log-based relevance feedback algorithm by engaging a modified SVM technique called Soft Label SVM. We demonstrate promising experimental results based on our suggested algorithm. The contribution of this work can be found in [57].

Finally, the thesis also shows a novel application to demonstrate how to apply the relevance feedback technique to a meaningful application. We suggest to learn Web images for searching semantic concepts in image databases by engaging a relevance feedback mechanism [58]. To find more detailed contributions of my work, please refer to the list of publications in the appendix.

1.4 Organization of This Work

Chapter 2, "Background", surveys the research work on relevance feedback in CBIR and reviews a related machine learning technique, i.e. Support Vector Machine, which is an important and promising technique employed to formulate the relevance feedback algorithm in CBIR. It also discusses, at a high-level, advantages and disadvantages of state-of-the-art techniques for relevance feedback in CBIR.

Chapter 3, "Relevance Feedback with Biased SVM", presents a novel relevance feedback algorithm by employing a modified SVM technique called Biased SVM in CBIR. The proposed technique is motivated for attacking the imbalanced dataset problem of relevance feedback tasks in CBIR.

Chapter 4, "Optimizing Learning with SVM Constraint", proposes a novel scheme for learning on relevance feedback by unifying the optimizing learning and the SVM technique in CBIR. The advantages of the proposed unified scheme are addressed and its effectiveness is verified from detailed experiments.

Chapter 5, "Group-based Relevance Feedback", suggests a novel group-based relevance feedback scheme by engaging SVM ensembles technique in CBIR. Different from traditional approaches, the proposed group-based relevance feedback assumes that relevant samples are drawn from multiple positive classes and irrelevant samples are from one negative class. We develop a related group-based CBIR system to validate our proposed group-based relevance feedback technique in this chapter.

Chapter 6, "Log-based Relevance Feedback", studies the log-based relevance feedback technique from a long-term learning perspective. Different from regular relevance feedback, it presents a novel log-based relevance feedback algorithm employing Soft Label SVM to engage users' feedback logs in a long-term learning purpose.

Chapter 7, "Application: Web Image Learning", presents a simple yet meaningful application by engaging a relevance feedback mechanism for learning the semantic concepts in image databases with Web images. Promising results are provided to for a demonstration.

The thesis concludes in chapter 8 with discussions of future research directions on relevance feedback and the difficulties of learning tasks in CBIR.

□ **End of chapter.**

Chapter 2

Background

2.1 Relevance Feedback

Relevance feedback, originally proposed for text-based information retrieval, is a powerful technique to improve the retrieval performance [59]. It is the process of automatically altering an existing query employing information provided by users about the relevance of previous retrieved objects in order to approach the users' query targets [59, 34]. It is a bit surprising that although there are limited research focusses on relevance feedback in traditional text-based information retrieval, relevance feedback has been become one of the most active research topic in CBIR community. The major reason is that traditional relevance feedback methods for text-based information retrieval have poor performance when directly applied in CBIR systems [18].

In the past years, a variety of relevance feedback techniques have been proposed in CBIR [18, 33, 41, 60]. Relevance feedback techniques in CBIR have evolved from early heuristic weighting adjustments to the optimization methods and various machine learning techniques proposed recently [41, 43, 52]. Early relevance feedback approaches for CBIR can be traced back to the last decade [41]. These early algorithms for CBIR were introduced from the traditional text-based information retrieval field [34, 61, 59]. In the later years, relevance feedback became a hotspot in CBIR and more systematic formulations of relevance feedback were suggested for CBIR [39, 52]. In most recent years, various machine learning techniques are proposed to formulate the state-of-the-art relevance feedback algorithms in CBIR [41]. We introduce the main development trends of relevance feedback in CBIR and briefly review the representative approaches.

2.1.1 Heuristic Weighting Methods

Early approaches of relevance feedback in CBIR were inspired by the regular relevance feedback techniques in traditional document information retrieval fields [61, 62]. For example, the relevance feedback approach in [33] transformed the "term-frequency" and "inverse document frequency" based methods to the "positive samples" and "negative samples" based approaches. These early approaches employed heuristic methods for adjusting the involved parameters. They normally assume the user has an ideal query point and each of the axises (corresponding with different visual features) accounts different weights. Two representative methods including the Query Representation Modification and Query Reweighting were proposed as the prototypes of query refinement techniques in the early stage of CBIR [33, 40, 63].

Query Representation Modification

The mechanism allows users to modify the query point and refine their representation. Two methods were suggested for refining query representation: Query Point Movement (QPM) [33] and Query Expansion (QEX) [63]. Query Point Movement assumes there exists an ideal query and tries to estimate the query point in the feedback tasks. The idea of this technique was often found in traditional information retrieval fields [61]. One typical QPM formulation based the vector space model was given as follows [33]

$$Q' = \alpha\ Q + \beta \frac{1}{N_p} \sum_{i=1}^{N_p} \frac{R_i}{N_p} - \gamma \sum_{i=1}^{N_n} \frac{S_i}{N_n} \tag{2.1}$$

where Q is the original query point before feedback, Q' is the estimated query point after the feedback, R_i is the vector of relevant sample i and S_i is the vector of non-relevant sample i, N_p denotes the number of relevant samples and N_n denotes the number of non-relevant samples. Query expansion tries to add relevant information to query. It can be viewed as an multiple-instance sampling technique to retrieve the relevant samples. In each round, relevant samples are added to the query set.

In literature, QPM and QEX were suggested as the relevance feedback method in the MARS (Multimedia Analysis and Retrieval System) [33, 63]. The research work in [63] showed that the QEX approach had slightly improvement with the QPM method [63].

Query Reweighting

Query Reweighting (QRW) was another typical query refinement technique in the early stage of CBIR [33, 63, 64]. The QRW technique refined the query by changing the weighting of different components of the feature space. The QRW intuitively learn the distance function for different users by changing the weights of components of the representation. A typical QRW approach simply took the inverse of the standard deviation of the i-th feature values in the feature matrix as the weight of i-th feature of the feature space, i.e. $w_i = \frac{1}{\sigma_i}$. QRW was also suggested in the MARS and detailed discussion and improvement can be found in [40, 63].

2.1.2 Optimization Formulations

In the later stage, more research work tried to seek systematic formulations of relevance feedback for CBIR [39, 52]. MindReader proposed a scheme to formulate the previous query refinement approaches systematically [39]. Different from the heuristic approaches of previous works, MindReader formulated the feedback task as an optimization problem in which parameters are learned by minimizing the sum of overall distances from the query centroid to all relevant samples [39]. However, the MindReader only found a good representation for a single feature space and it lacked analyzing the working conditions in CBIR. Addressing these problems, a more vigorous approach called optimizing learning (OPL) scheme was proposed for overcoming the limitations of previous work. The OPL technique systematically formulate the relevance feedback as an optimizing problem and suggested the hierarchical learning rather than the flat model of the MindReader. Along with this research direction, more techniques can also be found in literature [65, 66]. However, we would not take more discussions on these approaches since most of these work focused more on efficiency problem.

2.1.3 Various Machine Learning Techniques

With the rapid advances of machine intelligence in recent years, sophisticated machine learning techniques have been suggested to formulate the relevance feedback tasks in CBIR. All kinds of machine learning methods had been proposed, such as Artificial Neural Network [67], Bayesian Learning [38, 68, 69, 70, 71], Decision Tree [46], Boosting techniques [45], Discriminant Analysis [41], etc. Unsupervised learning techniques, like SOM [60] and EM algorithms [72, 73], were also tried in the literature. Recently, SVMs are widely explored in the fields of machine learning and pattern classification [50, 74]. A lot of research work have tried to employ SVMs for attacking the relevance feedback task in CBIR. From the past research efforts, SVM has been demonstrated as one of the most promising and successful approaches for relevance feedback problems [43, 44, 56]. We briefly introduce the basic concept of SVM as follows.

As a popular machine learning technique, SVMs have sound theoretical foundations and excellent performance in pattern classification problems [50]. SVMs implement the principle of structural risk minimization by minimizing Vapnik-Chervonenkis dimensions [50, 74].

They can provide very good generalization performance in empirical applications. Fig. 2.1 gives an intuitive illustration of a linear separable case of SVM. From the example, we can see that the objective of SVM learning is to find the optimal hyperplane with a maximum margin. When constructing a relevance feedback algorithm, SVM typically is first employed to learn a decision boundary to classify the positive and negative samples. The decision boundary is then employed to formulate the distance function which is used to rank the samples in the database. A lot of past research efforts have shown the successes by employing SVMs techniques in relevance feedback [44, 56, 75].

2.2 Support Vector Machines

2.2.1 Setting of the Learning Problem

Suppose a generator produces random vectors $x \in \mathbb{R}^m$ drawn independently from a fixed but unknown probability distribution function $F(x)$. A supervisor is employed to response an output value y of every input vector x, according to a fixed and unknown conditional distribution function $F(y|x)$. The problem of a learning machine is to best approximate the supervisor's response by choosing a function from $f(x, \alpha)$, $\alpha \in \Lambda$, where Λ is a set of parameters [76].

For choosing the desired function, a training set of l independent and identically distributed (i.i.d.) observations are given

$$(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l) , \tag{2.2}$$

drawn according to the joint distribution function $F(x, y) = F(x)F(y|x)$. For a binary classification task, the output y only take two values y={0,1}. In order to figure out the best function to approximate the supervisor's response, one can measure the loss, $L(y, f(x, \alpha))$, between the response y of the supervisor for a given input x and the response $f(x, \alpha)$ of the learning machine. The expectation of the loss of a given function $f(x, \alpha)$ is given by the risk function as:

$$R(\alpha) = \int L(y, f(x, \alpha))dF(x, y) . \tag{2.3}$$

The goal of learning is to look for the function $f(x, \alpha_0)$ that minimizes the risk function $R(\alpha)$. In order to approach the goal, the Empirical Risk Minimization (ERM) inductive principle is proposed [50]. In ERM principle, the risk function $R(\alpha)$ in Eq.(2.3) is replaced by the empirical risk function

$$R_{emp}(\alpha) = \frac{1}{l}\sum_{i=1}^{l}(y_i - f(x_i, \alpha))^2 \tag{2.4}$$

which is constructed on the training set. Then, the ERM principle suggests to approximate the $f(x, \alpha_0)$ by the function $f(x, \alpha_l)$ that minimizes the empirical risk Eq.(2.4).

However, the ERM principle only minimizing the training error (i.e. the empirical risk) does not imply a small average test error (actual risk) $R(\alpha)$ over the general test examples drawn from the underlying distribution $P(x, y)$. To control the generalization ability of the learning machine, the VC (Vapnik-Chervonenkis) theory provides the bounds of test error (the actual risk) [77]. The bounds of the actual risk depend on the empirical risk and the capacity of the function class [76]. The motivation to minimize these bounds leads to the Structural Risk Minimization inductive principle [50].

2.2.2 Optimal Separating Hyperplane

Support Vector Machine (SVM) is a hyperplane learning algorithm that implements the SRM inductive principle. Let us consider the class of hyperplanes

$$(\mathbf{w} \cdot \mathbf{x}) + b = 0 \quad \mathbf{w} \in \mathbb{R}^m, b \in \mathbb{R} , \tag{2.5}$$

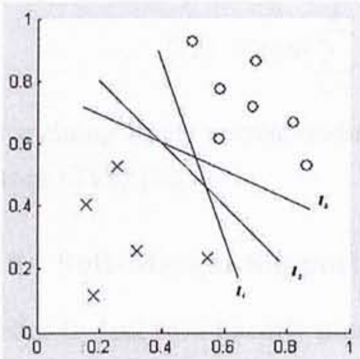
with the corresponding decision functions

$$f(x) = sgn((\mathbf{w} \cdot \mathbf{x}) + b) , \tag{2.6}$$

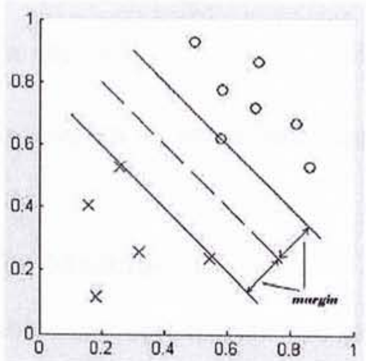
and figure out the learning algorithm to the given data

$$(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l) \;, \tag{2.7}$$

which can be separated by a hyperplane. Two facts stand for hyperplane learning algorithms. First, among the separating hyperplanes separating the data, there exists a unique hyperplane which yields the maximum margin between the classes. Second, the capacity of the function decreases when increasing the margin. The SVM learning is to construct such an optimal separating hyperplane, as shown in Fig. 2.1.



(a) Three different hyperplanes



(b) The optimal hyperplane

Figure 2.1: Illustration of linearly separable cases of SVM

In order to construct the optimal hyperplane, one has to solve the minimization problem:

$$\min_{\mathbf{w},b} \quad \frac{1}{2}\|\mathbf{w}\|^2 \tag{2.8}$$

$$s.t. \quad y_i((\mathbf{w}\cdot\mathbf{x}_i)+b) \geq 1, \quad i=1,\ldots,l \tag{2.9}$$

This optimization problem can be solved by introducing the Lagrangian multipliers:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\| - \sum_{i=1}^{l} \alpha_i (y_i \cdot ((\mathbf{x_i}\cdot\mathbf{w}) + b) - 1) \;. \tag{2.10}$$

By taking into account the Karush-Kuhn-Tucker (KKT) conditions [50], one can derive the dual problem of SVM as follows:

$$\min_{\mathbf{w},b} \quad \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (\mathbf{x_i} \cdot \mathbf{x_j}) \tag{2.11}$$

$$s.t. \quad \alpha_i \geq 0, i = 1, \ldots, l \quad and \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{2.12}$$

The dual variables α_i can be obtained after solving the above quadratic programming problem [78]. Then, the decision function, based on the optimal hyperplane, can be written as follows

$$f(x) = sgn(\sum_{i=1}^{l} y_i \alpha_i \cdot (\mathbf{x} \cdot \mathbf{x_i}) + b) \ . \tag{2.13}$$

The training items corresponding with nonzero α_i are called Support Vectors (SVs) [50].

2.2.3 Soft-Margin Support Vector Machine

In order to deal with linearly nonseparable cases, soft-margin and kernel techniques are introduced to formulate the learning algorithm. Two representative soft-margin SVMs are C-SVM [76] and ν-SVM [77]. We here discuss the later one.

For dealing with nonseparable cases, slack variables $\xi_i \geq 0$ are introduced to relax the constraints in Eq. (2.9). By taking with kernel, the optimization problem of ν-SVM can be written as

$$\min_{\mathbf{w}} \quad \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum \xi_i \tag{2.14}$$

$$s.t. \quad y_i(\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) \geq \rho - \xi_i \tag{2.15}$$

$$\xi_i \geq 0, \rho \geq 0, \quad i = 1, \ldots, l, \tag{2.16}$$

where ν is a parameter to control the bounds on the fraction of margin errors and the fraction of SVs, ρ is variable to be optimized, and Φ is a mapping function corresponding to a Mercer kernel $k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}))$. Detailed solutions to the optimization problem can be found in [77].

2.2.4 One-Class Support Vector Machine

One-class SVM (1-SVM) is derived from the classical SVM to solve density estimation problems. In typical formulations of 1-SVM, only positive instances are considered for estimating the density of the data. There are two kinds of different formulations of 1-SVMs in the literature [79, 80]. Here, we choose to illustrate the sphere-based approach with an explicit and good geometric property. Fig. 2.2 simply illustrates an example of 1-SVMs.

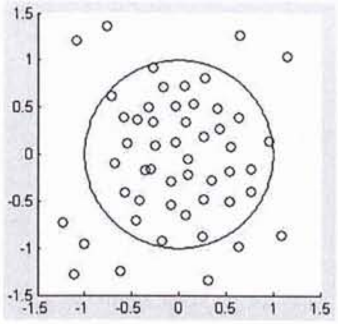


Figure 2.2: The sphere hyperplane in 1-SVM for constructing the smallest soft sphere that contains most of the positive instances. The circles represent the positive instances.

To construct the optimal hypersphere in 1-SVM, one can solve the following optimization problem [79, 80, 81]:

$$\min_{R,\mathbf{c}} \quad R^2 + \frac{1}{\nu l}\sum_i \xi_i \tag{2.17}$$

$$s.t. \quad \|\Phi(\mathbf{x}_i) - \mathbf{c}\|^2 \leq R^2 + \xi_i, \tag{2.18}$$

$$\xi_i \geq 0 \quad i = 1, \ldots, l\,. \tag{2.19}$$

Here, R is the radius of the hypersphere, $\mathbf{c}$ is the centroid of the hypersphere, and $\nu \in (0, 1]$ is a parameter to control the tradeoff between the radius R of the hypersphere and the fraction of positive training instances. The way to solve the optimization problem of one-class SVM can also employ the Lagrangian multipliers. The details can be found in [79].

□ **End of chapter.**

Chapter 3

Relevance Feedback with Biased SVM

3.1 Introduction

As an active research topic in computer science and engineering communities, CBIR has captured more and more attentions to study the relevance feedback algorithms in the past years [18, 41, 48]. Recently, along with the rapid development of statistical learning and machine intelligence, many machine learning techniques have been suggested to attack the relevance feedback in CBIR [43, 46, 71], in which Support Vector Machine (SVM) is one of the most effective and promising techniques [43, 44, 79, 80, 82]. With sound theoretical foundations, SVMs provide very good generalization performance in varied empirical pattern classification applications [50, 74]. They also have been proved for achieving successes in the relevance feedback tasks from past research efforts [43, 44, 75, 83].

However, previous studies on relevance feedback with SVMs usually regard the problem as a strict binary classification task without noticing an important feature of relevance feedback, i.e. the imbalanced dataset problem [84]. In real-world relevance feedback tasks, the negative samples often overnumber the positive samples. Moreover, the positive samples are often clustered in the same way while the negative samples are positioned in different ways. This imbalance problem may cause the positive samples to be overwhelmed by the negative samples if they are treated without any bias. In order to mitigate the impact of this problem, we propose a modified SVM technique called Biased Support Vector Machine (Biased SVM or BSVM) which can better model the relevance feedback and reduce the performance degradation caused by the imbalanced dataset problem.

The rest of this chapter is organized as follows. Section 3.2 presents and formulates the Biased SVM technique. Section 3.3 suggests to construct a relevance feedback algorithm employing our proposed BSVM and explains the benefits compared with conventional techniques. Experiments and performance evaluation are given in Section 3.4. Section 3.5 discusses the problems of our proposed algorithm. Finally, Section 3.6 concludes the work of this chapter.

3.2 Biased Support Vector Machine

Regular one-class SVM is originally for density estimation only dealing with the positive data [50]. In order to incorporate the negative information, we propose the Biased Support Vector Machine derived from 1-SVM to mitigate the imbalance problem of relevance feedback tasks. Our strategy is to describe the data by employing a pair of sphere hyperplanes in which the inner one captures most of the positive instances while the outer one pushes out the negative instances. Different from the regular two-class SVM techniques treating fairly with positive and negative data, this kind of approach focuses on the positive data and also includes the impact of negative information.

Hence, the goal of our problem is to find an optimal sphere hyperplane which not only can contain most of the positive data but also can push most of the negative data out of the sphere. The problem is visually illustrated in Fig. 3.1. The dashed sphere in the figure is the desired sphere-hyperplane in our goal. The task can be formulated as an optimization problem and the mathematical formulation of our technique is given as follows.

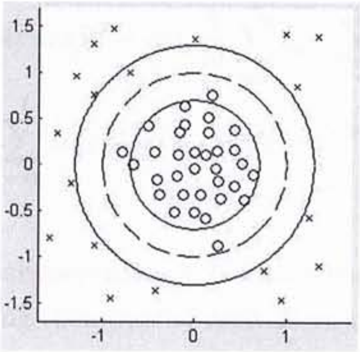


Figure 3.1: The sphere hyperplane of Biased SVM. The circles and the crosses represent the positive instances and the negative instances, respectively. The dashed sphere is the decision hyperplane.

Let us consider the following training data:

$$(x_1, y_1), \ldots, (x_l, y_l) \in \mathbb{R}^m \times Y, \quad Y = \{-1, +1\} \tag{3.1}$$

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

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

$$\min_{R\in\mathbb{R},\rho\in\mathbb{R},\mathbf{c}\in\mathcal{F}} \quad bR^2 - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i \tag{3.2}$$

$$s.t. \quad y_i(||\Phi(\mathbf{x}_i) - \mathbf{c}||^2 - R^2) \leq -\rho + \xi_i\ , \tag{3.3}$$

$$b > 0\ , \rho \geq 0\ , 0 \leq \nu \leq 1, \xi_i \geq 0\ , \tag{3.4}$$

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

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

$$L(R,\xi,c,\alpha,\beta,\lambda) = bR^2 - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\beta_i\xi_i - \lambda\rho + \sum_{i=1}^{l}\alpha_i[y_i(||\Phi(\mathbf{x}_i) - \mathbf{c}||^2 - R^2) + \rho - \xi_i]\,. \tag{3.5}$$

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

$$2R(b - \sum_{i=1}^{l} y_i\alpha_i) = 0 \;\Rightarrow\; \sum_{i=1}^{l} y_i\alpha_i = b\,; \tag{3.6}$$

$$\frac{1}{\nu l} - \alpha_i - \beta_i = 0, \;\Rightarrow\; 0 \le \alpha_i \le \frac{1}{\nu l}\,; \tag{3.7}$$

$$\sum_{i=1}^{l} 2\alpha_i y_i(\Phi(\mathbf{x}_i) - \mathbf{c}) = 0 \;\Rightarrow\; \mathbf{c} = \frac{1}{b}\sum_{i=1}^{l}\alpha_i y_i\Phi(\mathbf{x}_i) \tag{3.8}$$

$$-1 + \sum_{i=1}^{l}\alpha_i - \lambda = 0 \;\Rightarrow\; \sum_{i=1}^{l}\alpha_i \ge 1\,. \tag{3.9}$$

By substituting the above derived results to the objective function in Eq. (3.5), the dual of the primal optimization can be shown to take the form

$$\max_{\alpha} \quad \sum_i \alpha_i y_i k(\mathbf{x}_i,\mathbf{x}_i) - \frac{1}{b}\sum_{i,j}\alpha_i\alpha_j y_i y_j k(\mathbf{x}_i,\mathbf{x}_j) \tag{3.10}$$

$$s.t. \quad \sum_i \alpha_i y_i = b\,, \tag{3.11}$$

$$0 \le \alpha_i \le \frac{1}{\nu l}\,, \tag{3.12}$$

$$\sum_i \alpha_i \ge 1, \quad i = 1,2,\ldots,l\,. \tag{3.13}$$

This dual problem can be solved by Quadratic Programming (QP) techniques [78, 85]. Then, the resulting decision function takes the form

$$f(x) \;=\; sgn(R^2 - ||\Phi(\mathbf{x}) - \mathbf{c}||^2)\,, \tag{3.14}$$

where $\mathbf{c}$ can be obtained from Eq. (3.8) and R can be solved by support vectors. Based on the decision function, we can know the instances inside the sphere hyperplane will be predicted as positive, and negative otherwise.

3.3 Relevance Feedback Using Biased SVM

3.3.1 Advantages of BSVM in Relevance Feedback

From the above formulations, one may see that the optimization in Eq. (6.6) is similar to the one in the ν-SVM. Now, we explain mathematical differences compared with regular SVMs and the advantages of our BSVM from a geometric perspective for solving the relevance feedback problems.

From the results of mathematic deduction in the optimization function, we see that BSVM is with the following constraint from Eq. (3.11)

$$\sum_i \alpha_i y_i = b \, . \tag{3.15}$$

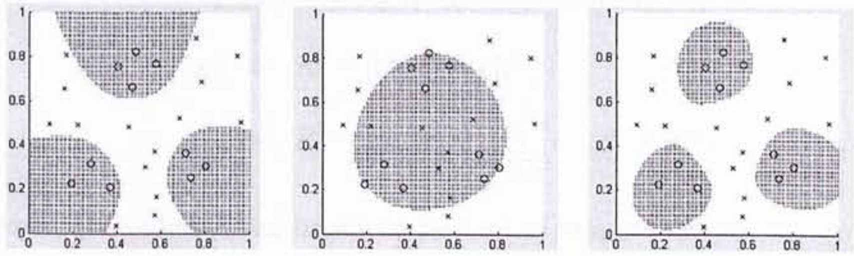
When replacing y_i with $+1$ for the positive class and -1 for the negative one, the constraint can be written as

$$\sum_{i \in S^+} \alpha_i - \sum_{i \in S^-} \alpha_i = b \, , \tag{3.16}$$

where S^+ denotes the positive class and S^- denotes the negative one. However, in the ν-SVM, the constraint is with the form

$$\sum_{i \in S^+} \alpha_i - \sum_{i \in S^-} \alpha_i = 0 \, . \tag{3.17}$$

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



(a) SVM (ν-SVM) (b) 1-SVM (c) BSVM

Figure 3.2: Decision boundaries of three classification methods with the same kernel (RBF) and parameters (γ=0.1).

Moreover, we can also see the difference from the geometric perspective. Fig. 3.2 provides a comparison of the decision boundaries of regular SVM, 1-SVM and BSVM on the synthetic data with the same kernels (Radial Basis Function) and parameters (γ=0.1). We can see that the geometric property of BSVM is better than ν-SVM and 1-SVM. BSVM can describe the data in a cluster behavior by the sphere-based boundary and can flexibly control the weight of the positive class for the imbalanced datasets problem by setting a bias factor. Therefore, compared with the ν-SVM and 1-SVM, BSVM is more reasonable and more effective to model the relevance feedback tasks.

3.3.2 Relevance Feedback Algorithm by BSVM

From the above comparisons, we have shown the benefits of BSVM for solving relevance feedback issues. Here, we describe how to formulate the relevance feedback algorithm by employing the BSVM technique. Applying SVMs based techniques in relevance feedback is similar to the classification task. However, the relevance feedback needs to construct an evaluation function to produce the relevance value of the retrieval instances. From the decision function in Eq. (3.19), we build the evaluation function by substituting the derived result in Eq. (3.8)

$$\begin{aligned} f(\mathbf{x}) &= R^2 - ||\Phi(\mathbf{x}) - \mathbf{c}||^2 \\ &= R^2 - ||\Phi(\mathbf{x}) - \frac{1}{b}\sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x}_i)||^2 , \end{aligned} \tag{3.18}$$

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

$$f(\mathbf{x}) = \frac{2}{b}\sum_{i} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}) . \tag{3.19}$$

Once the parameters α_i are solved in Eq. (3.10), the evaluation function can be constructed. Consequently, we can rank the images based on the scores of the evaluation function $f(\mathbf{x})$. The images with higher scores will be more likely to be chosen as the targets.

3.4 Experiments

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

3.4.1 Datasets

A Synthetic Dataset

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

COREL Image Datasets

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

3.4.2 Image Representation

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

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

We employ the edge direction histogram as the shape feature in our experiments [29]. Canny edge detector is applied to obtain the edge images. From the edge images, the edge direction histogram is then computed. The edge direction histogram is quantized into 18 bins of 20 degrees each, hence an 18-dimensional edge direction histogram is used to represent the edge feature.

We apply the wavelet-based texture feature for its effectiveness [25]. We perform the Discrete Wavelet Transformation (DWT) on the gray images employing a Daubechies-4 wavelet filter [25]. In total, we perform 3-level decompositions and obtain ten subimages in different scales and orientations. Then, we choose nine subimages with most of the texture information and compute the entropy of each subimage. Hence, a 9-dimensional wavelet-based texture feature is obtained to describe the texture information for each image.

The three kinds of extracted features are normalized and combined as a 36-dimensional feature vector in order to effectively capture the visual content of images. In fact, we observed that image retrieval based on the combined features outperforms the one with an individual feature separately in empirical experiments.

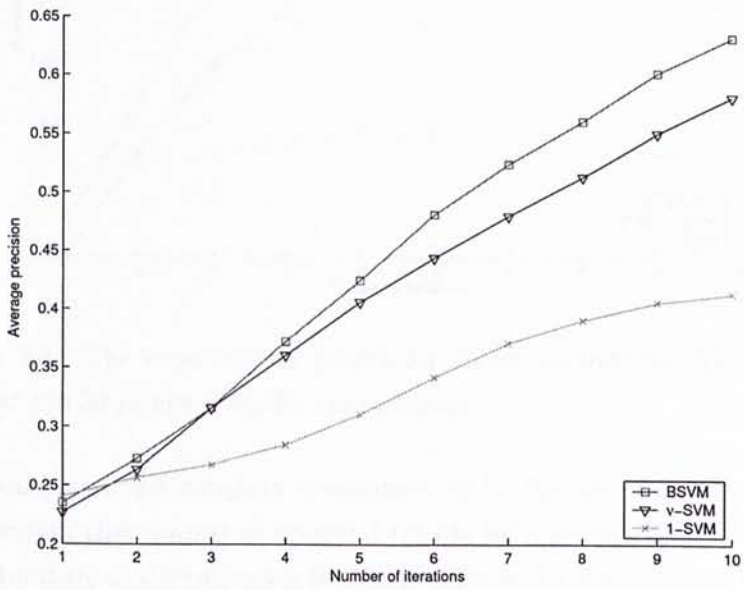


Figure 3.3: The experimental results for three methods on the synthetic dataset: top-30 returned results are evaluated.

3.4.3 Experimental Results

In the following, we present the experimental results by three algorithms on both the synthetic data and the real-world images. For the purpose of objective measure of performance, we assume that the query judgement is defined on the image categories [44]. The metric of evaluation is the *Average Precision* which is defined as the average ratio of the number of relevant images of the returned images over the total number of the returned images.

In the experiments, a category is first picked from the database

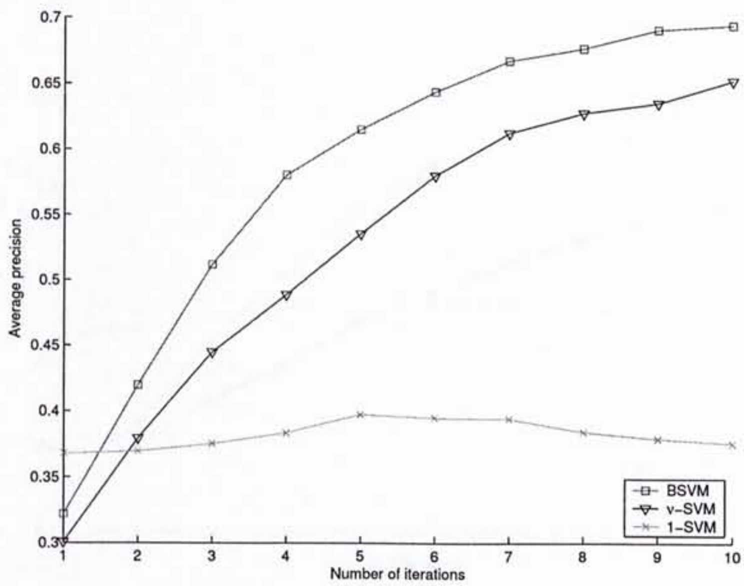


Figure 3.4: The experimental results for three methods on the 20-Cat dataset: top-30 returned results are evaluated.

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

The algorithms implemented in our experiments are based on modifying the codes in the *libsvm* library [86]. We notice that the experimental settings are important to impact on the evaluation results. To

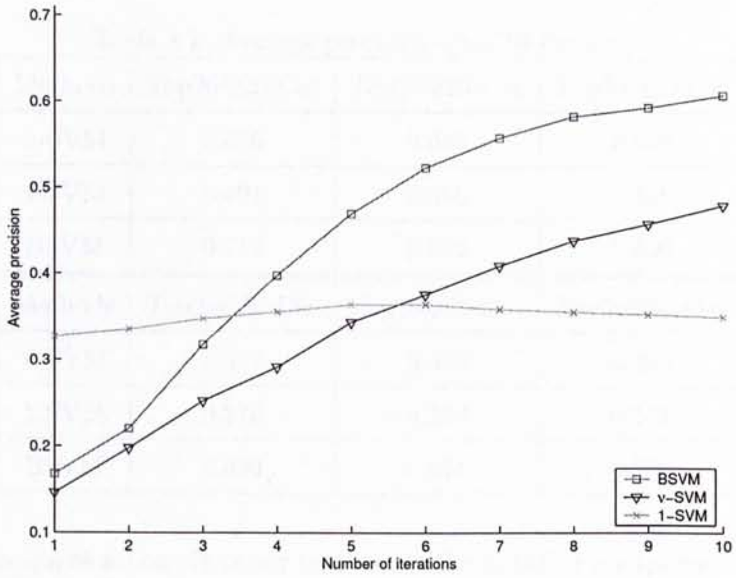


Figure 3.5: The experimental results for three methods on the 50-Cat dataset: top-30 returned results are evaluted.

enable an objective measure of performance without bias, we choose the same kernels and parameters for all the settings. All the kernels are based on Radial Basis Function (RBF) which outperforms other kernels in the experiments.

The first evaluation is on the synthetic dataset. Fig. 3.3 shows the evaluation results of top-30 returned results. We can observe that BSVM outperforms other approaches. The 1-SVM achieves the worst performance without considering the negative information.

The second evaluation is on the real-world datasets. Fig. 3.4 and Fig. 3.5 show the evaluation results on the 20-Cat dataset and 50-Cat dataset, respectively. From the results on the real-world datasets, we can see our proposed BSVM also outperforms other approaches. However, we notice that the performance of 1-SVM in the beginning feedback steps is better than that of other approaches. The reason is that 1-SVM can reach the enclosed positive region quickly, but it cannot be further improved without the help of the negative information

Table 3.1: Average precision after 10 iterations

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

in subsequent steps. In order to observe the detailed comparison of the three methods after 10-iterations, we list the retrieval results in Table 3.1. From the results, we can also see the similar results matching the above comparisons.

3.5 Discussions

We have observed that the proposed Biased SVM scheme performs better than the regular SVM approaches from the experimental results. The typical approaches by SVMs (ν-SVM) without considering the bias in the retrieval tasks is not appropriate in solving the relevance feedback problem. We also see that regular one-class SVMs do not consider the negative information which cannot learn the feedback well. Furthermore, we know there are other methods to address the imbalanced dataset problem in literature [84, 87]. We can also consider to include them in our scheme in the future. Nevertheless, we have observed the promising results in demonstrating the effectiveness of our proposed Biased SVM technique for the relevance feedback problems.

3.6 Summary

This chapter investigates the learning techniques to attack the imbalance problem of relevance feedback tasks in CBIR. We first pointed out the imbalanced dataset problem of relevance feedback, and then proposed a novel technique called Biased SVM to construct the relevance feedback algorithm in CBIR. The advantages of the proposed techniques are illustrated and justified. Experiments are conducted both on synthetic data and real-world image datasets for evaluating the retrieval performance of the proposed technique. The experimental results demonstrate that our proposed Biased SVM based relevance feedback algorithm is effective and promising to improve the performance of CBIR systems.

□ **End of chapter.**

Chapter 4

Optimizing Learning with SVM Constraint

4.1 Introduction

In a typical relevance feedback scheme, users are involved to provide the relevance judgements on the initial retrieval results from CBIR systems. Based on the relevance information fed-back from the users, relevance feedback is incorporated as a query refinement technique to learn a flexible similarity measure metric for approaching the users' query target. After a few rounds of feedback, the CBIR system returns the learning results which are considered as the most relevant and interested results of the users [88, 15].

Therefore, one of the most important problems of relevance feedback is to learn the similarity measure metric which can best represent the users' query targets [23]. In the literature, a variety of relevance feedback techniques have been proposed ranging from the simple heuristic approaches to various sophisticated learning techniques. Although a lot of advances have been achieved in the past research efforts [24, 52], learning the best similarity measure metric with relevance feedback is yet a challenging issue.

This chapter proposes a novel scheme to formulate and solve the learning task in relevance feedback. The optimizing learning technique is suggested to learn a flexible similarity measure metric with the constraint of the boundary learned by Support Vector Machine. The suggested scheme unifies two important techniques in relevance feedback learning: the optimizing learning and the SVM techniques. In the proposed scheme, SVM is first engaged to learn a coarse boundary by projecting the original data space to a high dimensional feature space. Constrained with the boundary of SVM, the optimizing technique is introduced to learn a flexible similarity measure metric. Finally, a best similarity measure scheme is obtained by fusing the learning results.

The key idea of our proposed scheme is that SVM can effectively classify the positive samples and negative samples by employing the kernel technique. But as a statistical learning algorithm, its performance will largely drop when facing insufficient training samples. The optimizing learning systematically formulate the learning task of similarity measure as a distance minimization problem. It is less sensitive to the insufficient training samples problem. However, it does not support kernel-based learning which can achieve better performance by learning in high dimensional feature spaces. Whereas, SVM has important advantages for solving the classification problem in high dimensional feature spaces and provides very good empirical performance. Hence, engaging the SVM technique to constrain the optimizing learning can importantly improve the retrieval performance of traditional approaches.

The rest of this chapter is organized as follows. Section 4.2 reviews the related work on relevance feedback and gives the motivation of our work. Section 4.3 presents our proposed scheme to employ the boundary of SVM learning for constraining the optimizing learning learning on learning a flexible similarity measure metric. Experiments and performance evaluation are presented in Section 4.4. Section 4.5 discusses several problems. Section 4.6 summaries our work.

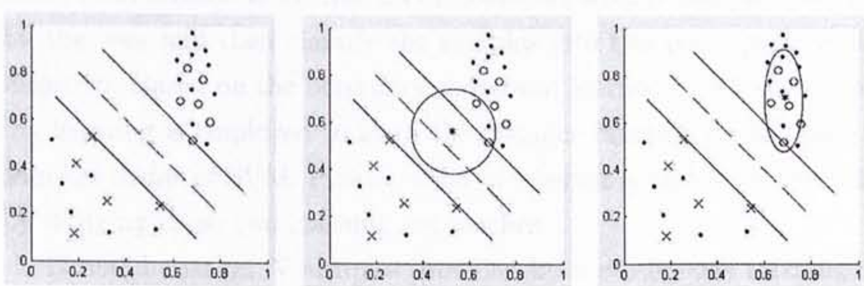
4.2 Related Work and Motivation

In the above discussion, we emphasize two important techniques: the OPL learning and the SVM. Although many successes have been demonstrated for these two techniques from past research efforts [52, 43], both of them have a lot of limitations.

First, although relevance feedback employing the OPL technique has the optimization claim, the solved optimal distance function may not be the best solution for the retrieval task. The so-called "optimal" is based on the assumption of minimization of total distances. Moreover, the OPL does not support kernel-based learning. SVM is a kernel-based technique which has excellent performance by learning in a projected high dimensional feature space.

However, limitations for SVMs should be addressed in formulating the relevance feedback. One of the big disadvantages of SVM based relevance feedback is the insufficient training samples problem. Although SVM is less sensitive to the insufficient training data problem among varied statistical algorithms, it still cannot avoid the performance degradation when facing insufficient training samples. With enough training samples, the decision boundary learned by SVM may deviate largely from the correct one. Another important issue of SVM-based relevance feedback learning is the ranking problem. Traditional SVM-based relevance feedback ranks the similarity of samples in the database by employing the distance from the boundary as the metric. However, there is no explicit and definite evidence to prove that the distance is the best metric for image similarity metric. Moreover, when the learned boundary deviates too much from the true distribution, the ranking method will perform poorly in practical applications.

To overcome these limitations, we suggest to unify the optimizing learning and the SVM technique to construct the relevance feedback algorithm in CBIR. In the training phase, we first employ SVM to learn the boundary for classifying the positive samples and negative samples. Then we engage the optimizing learning to compute the optimal similarity measure metric by minimizing the overall distance. At the ranking phase, the trained SVM classifier is employed to classify the samples in the database into two group: positive and negative. The positive samples are ranked by the optimizing learning metric while the negative samples are ranked by the distance from the boundary of SVM. Our approach is similar to the related work in literature [24]. The authors in [24] suggested to employ SVM to constrain the Euclidean distance for similarity measure. However, they suggested the straight Euclidean distance which is not effective enough for learning the similarity measure in practical applications. Moreover, their approach is a little bit heuristic and lacks a systematic formulation. Different from their approach, we formulate the similarity measure as an optimization problem and employ the SVM to constrain the similarity measure.



(a) SVM: the regular learning approach.

(b) SVM+EU: SVM with the straight Euclidean distance.

(c) SVM+OPL: SVM with the Generalized Ellipsoid distance by optimizing learning.

Figure 4.1: Illustration of methods learning with SVM. The circles, crosses and dotted points represent the positive, negative and unlabelled instances respectively. The star denotes the query or learned centroid.

4.3 Optimizing Learning with SVM Constraint

4.3.1 Problem Formulation and Notations

Before going detailed discussions, it is necessary to clearly state the objective of our task and introduce the basic idea of our proposed learning scheme. Table 4.1 gives a list of notations and their definitions for the following discussions.

In the context of CBIR, a user typically begins the query with a query sample $\vec{q}$ [4]. The CBIR system first returns the user a set of initial samples. Then, the user marks the relevance scores on the returned samples. Based on the feedback information, the task of relevance feedback algorithm is to learn the refinement technique for retrieving the best tentative list of samples to the users in the next round. The basic idea of our scheme is to train SVM classifiers with N samples provided by the user and then classify the samples into two part: positive and negative. Based on the boundary constraint learned by SVM, optimizing learning is employed to learn the distance function for the positive samples inside of SVM. Finally, a list of relevant samples are generated by unifying these two learning approaches.

Hence, based on N samples provided by the user, two learning objectives need be achieved for each sample $\vec{x}_n = [\vec{x}_{n1}, \ldots, \vec{x}_{nM}]$ in the database. First, the classification boundary of SVM and the corresponding distance function $\mathcal{D}_{SVM}(\vec{x}_n, \Theta)$. Secondly, a generalized distance function $\mathcal{D}_{OPL}(\vec{x}_n, \vec{q})$ need be learned by the optimizing learning constrained with the boundary of SVM.

4.3.2 Learning boundaries with SVM

In theory, SVM implements the SRM inductive principle [50] which can guarantee the bounds of the test error. Let us consider the SVM for binary classification problems. Generally speaking, a binary classification problem can be formulated as a learning task to estimate a function $f{:}\mathbb{R}^m \rightarrow \{-1, +1\}$ based on the given independent identically

Table 4.1: Notations and their definitions in the context

Notation	Definition
N	number of training sample vectors
M	number of feature components (e.g. color, shape and texture components)
S	set of support vectors for a SVM classifier
L_i	dimension of the i-th feature component
$\vec{x}_n = [\vec{x}_{n1}, \ldots, \vec{x}_{nM}]$	the n-th sample vector in the image database
$\vec{x}_{ni} = [x_{ni_1}, \ldots, x_{niL_i}]$	the i-th component of the n-th sample vector
$\vec{q} = [\vec{q_1}, \ldots, \vec{q_i}, \ldots, \vec{q_M}]$	the query vector
$\vec{q_i} = [q_{i_1}, \ldots, q_{iL_i}]$	the i-th feature component of the query vector
$\vec{u} = [u_1, \ldots, u_M]$	weights of feature components
$\vec{v} = [v_1, \ldots, v_N]$	goodness values of samples
$\mathbf{W_i} = [w_{jk}]$	real symmetric full distance matrix for distance functions
$\mathbf{C_i} = [c_{jk}]$	weighted covariance matrix of samples vectors
$K(\mathbf{x}, \mathbf{y})$	Mercer kernel function for SVM
Φ	a mapping function for SVM
$f()$	decision function for SVM
$\mathcal{D}_{SVM}()$	distance function of SVM
$\mathcal{D}_{OPL}()$	distance function of Optimizing Learning
$MaxDist$	maximum OPL distance inside the positive boundary
$\mathbf{Dis}()$	overall dissimilarity measure function

distributed (i.i.d.) data [89]

$$(x_1, y_1), \ldots, (x_N, y_N) \in \mathcal{X} \times Y, \quad Y = \{-1, +1\} . \tag{4.1}$$

Here, the training instances are vectors in some data space $\mathcal{X} \subseteq \mathbb{R}^m$ and N is the number of training instances. The goal of the SVM learning is to find an optimal decision function f which can classify the unseen data $\mathbf{x}$ correctly. The optimal boundary is corresponding to the hyperplane with a maximal margin.

In order to construct the optimal hyperplane of SVM, one can solve the following optimization problem based on the kernel technique and soft-margin concepts [50, 89]

$$\min_{\mathbf{w}\in\mathcal{F}} \quad \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{N}\sum \xi_i \tag{4.2}$$

$$s.t. \quad y_i(\mathbf{w}\cdot\Phi(\mathbf{x}_i)) \geq \rho - \xi_i \tag{4.3}$$

$$\xi_i \geq 0, \rho \geq 0, \quad i = 1, \ldots, N, \tag{4.4}$$

where ξ_i represent the margin errors for the non-separable training data. When the margin errors $\xi = 0$, one can show that the two classes are separated by a margin with $2\rho/\|\mathbf{w}\|$ from Eq. (4.3). By introducing the Lagrange multipliers, the optimization problem can be shown with the dual form as follows [74, 89].

$$\max_{\alpha} \quad -\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j})$$

$$s.t. \quad \sum_i \alpha_i y_i = 0\,, \quad 0 \leq \alpha_i \leq \frac{1}{N}\,, \tag{4.5}$$

$$\sum_i \alpha_i \geq \nu, \quad i = 1, 2, \ldots, N\,. \tag{4.6}$$

To evaluate the similarity measure, the distance from boundary is used as the distance function, namely

$$\mathcal{D}_{SVM}(\mathbf{x_n}, \Theta) = \sum_{\mathbf{x_i}\in S} \alpha_i y_i K(\mathbf{x_i}, \mathbf{x_n}) + b \tag{4.7}$$

The corresponding decision function for classification is given as follows

$$f(\mathbf{x_n}) = sign(\sum_{\mathbf{x_i}\in S} \alpha_i y_i K(\mathbf{x_i}, \mathbf{x_n}) + b)\,. \tag{4.8}$$

4.3.3 OPL for the Optimal Distance Function

In the traditional approaches of CBIR, similarity measure is formulated by simply taking the Euclidean distance as the distance function. Given an initial query vector $\mathbf{q}$, the simplest method to measure the distance d_n for a sample vector $\mathbf{x_n}$ is defined by the straight Euclidean distance as follows

$$d_n = (\mathbf{q} - \mathbf{x_n})^T(\mathbf{q} - \mathbf{x_n}) \ . \tag{4.9}$$

A more vigorous modification is based on the weighted Euclidean distance, namely

$$d_n = (\mathbf{q} - \mathbf{x}_n)^T\Lambda(\mathbf{q} - \mathbf{x_n}) \ , \tag{4.10}$$

where $\Lambda = [\frac{1}{\sigma_{kk}}]$ is a diagonal matrix. The k-th weight value in the diagonal matrix is simply taken by the inverse of standard deviation of the samples along with the k-th dimension. However, simply weighting approach with a flat model may not powerful enough to model the retrieval model. To introduce more sophisticated approach, Generalized Ellipsoid Distance is introduced to model the distance metric [39, 33]

$$d_n = (\mathbf{q} - \mathbf{x}_n)^T W(\mathbf{q} - \mathbf{x_n}) \ , \tag{4.11}$$

where W is a real symmetric full matrix. Moreover, hierarchical structures are proposed to formulate the distance function [33]. Each feature component is assigned with different value to indicate its importance. The weighting vector of feature components is denoted as $\mathbf{u} = [u_1, \ldots, u_M]$. Based on the hierarchical structure above, the distance function of a sample $\mathbf{x_n}$ can be written as follows

$$D(\mathbf{x_n}, \mathbf{q}) = \sum_{i=1}^{M} u_i(\mathbf{x_{ni}} - \mathbf{q_i})^T W_i(\mathbf{x_{ni}} - \mathbf{q_i}) \ . \tag{4.12}$$

The task now turns into the problem of finding the optimal distance matrices W_i. Actually, this task can be formulated as an optimization problem for minimizing the overall variance of distance values of all training samples [39].

Before giving the formulation of the optimization problem, a relevance vector of samples $\mathbf{v}$ need be addressed. For each positive training sample $\mathbf{x_i}$, a goodness value v_i is assigned to indicate its relevance degree. In the traditional approaches [39], the goodness value is assigned with users. However, weighting the relevance degrees by users is is a tedious job for the users and it also costs a lot of time during the relevance feedback. In order to alleviate this problem, we suggest to assign the goodness values of training samples by engaging the results of SVM learning. For each training sample $\mathbf{x_i}$, we employ the distance from the boundary of SVM to coarsely represent the relevance degree to the target. To formulate the concept, a sigmoid function of neural networks is suggested for the goodness value as follows

$$v(\mathbf{x_i}) = \frac{exp(D_{SVM}(\mathbf{x_i}, \Theta))}{1 + exp(D_{SVM}(\mathbf{x_i}, \Theta))} . \tag{4.13}$$

From the above formula, we can see that the larger the distance of SVM, the larger the goodness value.

Based on the given distance matrices and the goodness values, we can formulate the optimization problem as follows

$$\min_{W,\mathbf{q},\mathbf{u}} \quad \sum_{n=1}^{N}\sum_{i=1}^{M} v_n u_i (\mathbf{x_{ni}} - \mathbf{q_i})^T W_i (\mathbf{x_{ni}} - \mathbf{q_i}) \tag{4.14}$$

$$\text{s.t.} \quad \sum_{i=1}^{M} \frac{1}{u_i} = 1 \tag{4.15}$$

$$det(W_i) = 1 \quad i = 1, 2, \ldots, M . \tag{4.16}$$

This minimization problem can be solved by introducing the Lagrange multipliers technique [78]. We here provide the major conclusions [52, 39].

The optimal solution to the query point $\mathbf{q}^* = [\mathbf{q_i}^*]$ is given as follows

$$\mathbf{q_i}^* = \frac{X_i^T \mathbf{v}}{\sum_{n=1}^{N} v_n} , \tag{4.17}$$

where X_i is the matrix of N training sample vectors with i-th component. From this result, we can find that the ideal query point is the linear combination of the relevant samples by the goodness values.

The optimal solution to the matrix W_i can be solved with the following forms

$$W_i^* = \begin{cases} (det(C_i))^{\frac{1}{L_i}} C_i^{-1}, & N \geq L_i \\ diag(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_{L_i}^2}), & N < L_i \end{cases} \tag{4.18}$$

where C_i is the $(L_i \times L_i)$ weighting covariance matrix of X_i, namely as follows

$$C_{i_{st}} = \frac{\sum_{n=1}^{N} v_n(x_{ni_s} - q_{i_s})(x_{ni_t} - q_{i_t})}{\sum_{n=1}^{N} v_n} . \tag{4.19}$$

The above solution indicates that the distance matrix is equal to the weighted matrix with inverse standard deviations when the number of training samples is larger than the dimension of the corresponding feature component; otherwise, it can be solved with a more vigorous result.

The optimal solution for the feature weighting vector $\mathbf{u}^* = [u_i^*]$ is given as follows

$$u_i^* = \sum_{j=1}^{M} \sqrt{\frac{f_j}{f_i}} , \tag{4.20}$$

where $f_i = \sum_{n=1}^{N} v_n(\vec{x_{ni}} - \vec{q_i})^T W_i(\vec{x_{ni}} - \vec{q_i})$. This formula indicates that if the variance of the i-th component f_i is large, the corresponding weighting value u_i will be stressed.

4.3.4 Overall Similarity Measure with OPL and SVM

From the optimizing learning, we can obtain the optimal distance function (it may be a sub-optimal solution if the number of training samples is less than the dimension of the feature component). For each sample in the database $\mathbf{x_n}$, its OPL distance is denoted as $\mathcal{D}_{OPL}(\mathbf{x_n}, \mathbf{q}^*)$. Different from the previous approaches [33, 42, 39], we do not simply take the OPL distance as the similarity measure metric. Since SVM can learn a good boundary to classify the samples in a high dimensional feature space, we employ the boundary of SVM to constrain the

OPL distance. The samples inside the positive boundary are measured by the OPL distance, otherwise they are measured with the SVM distance. Based on this simple yet effective unification, we can formulate the overall similarity measure metric as follows

$$\mathbf{Dis}(\mathbf{x_n}) = \begin{cases} \mathcal{D}_{OPL}(\mathbf{x_n}, \mathbf{q}^*), & D_{SVM}(\mathbf{x_n}, \Theta) \geq 0 \\ MaxDis - \mathcal{D}_{SVM}(\mathbf{x_n}, \Theta), & \mathcal{D}_{SVM}(\mathbf{x_n}, \Theta) < 0 \end{cases} \tag{4.21}$$

Here, the $MaxDis$ is the maximum OPL distance of the samples inside the positive boundary of SVM which is given as follows

$$MaxDis = \max_{\forall x_i} D_{OPL}(\mathbf{x_i}, \mathbf{q}^*), \quad \text{if } D_{SVM}(\mathbf{x_i}, \Theta) \geq 0\,. \tag{4.22}$$

From the above formulation, we can see that our proposed similarity retrieval model not only can utilize the advantage of SVMs for kernel-based learning in the high dimensional feature space, but also can overcome disadvantages of SVMs by unifying the OPL distance metric.

4.4 Experiments

To verify the effectiveness of our proposed scheme, we perform the empirical evaluations on the real-world image databases. The experiments are performed on a PC with Windows 2000, 2.0G CPU and 512MB memory. The details of our experiments are described as follows.

4.4.1 Datasets

In order to evaluate the empirical performance, 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. Since there are many noisy categories in the COREL image CDs. We select the discriminative categories to form our datasets. In each category of our 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 motivations to select the semantic categories are explained in two aspects. Firstly, it can evaluate whether the approach can retrieve the relevant images not only visually relevant but also with similar semantic meaning. Secondly, the approach can help us evaluate the performance automatically, which can reduce the subjective errors in manual evaluations by different people.

4.4.2 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 [16, 25]. 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 [29]. 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 [25, 32]. 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 [25]. 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.

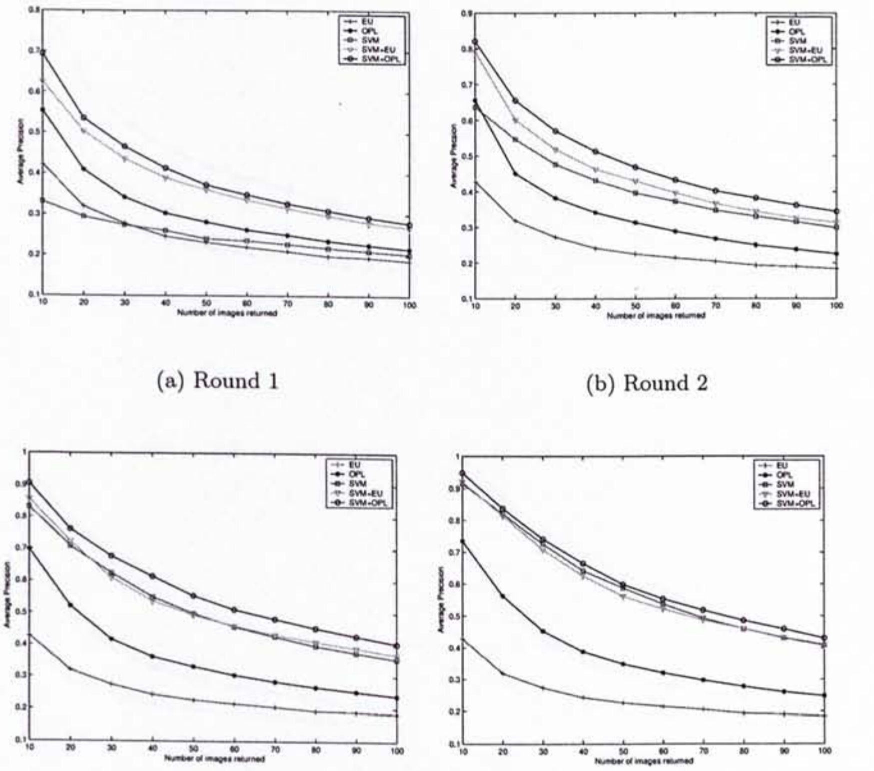
4.4.3 Performance Evaluation

To evaluate the performance of our proposed method, we compare our proposed method with traditional approaches. Four relevance feedback algorithms are compared: Optimizing Learning (OPL) [33], SVM [43], Euclidean with SVM (SVM+EU) [24] and our Optimal Learning with SVM (SVM+OPL). The evaluation metric used in our experiments is based on Average Retrieval Precision (ARP) which is defined by the correct retrieved images over the total returned images. Different from the work in [24], we do not consider the specific category in the experiment since it may lead to subjective results. Instead, we evaluate the performance on overall average performance on all the datasets.

Before discussing the experimental results, we would like to look at our experiment settings which are important for performance evaluations. The CBIR experiments sometimes are a bit subjective in comparisons [4]. In order to enable an objective performance measure without bias, we adopt the automatic approach to measure the precision. The images contained in a same categories are considered as relevant, otherwise they are non-relevant. The feedback procedures are also fully automatical. Moreover, we also notice that different parameter settings have important impacts on the retrieval performance. Hence, in the comparisons, we choose a same set of parameters for all algorithms and randomly pick same queries to evaluate their performance. The kernel of SVMs used in the experiments are all based on Radius Basis Function (RBF) which is considered as one of the most effective kernels for empirical applications [50, 74]. In the experiments, 100 executions of queries and feedbacks are conducted and averaged to produce to an ARP curve. For each execution, we first randomly pick an image sample from the database and take it as the query example. In each feedback round, 10 samples are returned to the user for marking the relevance to the target. In order to maximize the information from the user, the 10 returned samples are different from the previous samples returned to the user. Hence, the training samples are accumulated after each feedback round.

Fig 4.2 and Fig 4.3 show the experimental results on the 20-Cat and 50-Cat data sets, respectively.

In the figures, the EU curve represents the baseline of ARP with Euclidean distance retrieval without relevance feedback. These four figures show the retrieval results in four feedback rounds. From the results of the first round in both two datasets, we can found that SVM perform poorly without enough training samples in the initial round. The OPL, SVM+EU and SVM+OPL approaches performs better than SVM. And the Euclidean distance with SVM can perform better than typical SVM and the OPL approach, but our SVM+OPL achieves the best performance of all. In the second feedback round, the retrieval results are a bit different from the first round. The SVM approach performs better than the OPL approach in this round. The reason resides that the SVM can perform better if the number of training samples are increased since it can learn a better boundary in the high dimensional space. Even in this circumstance, our proposed scheme still can significantly improve the performance of regular SVM and



(a) Round 1

(b) Round 2

(c) Round 3

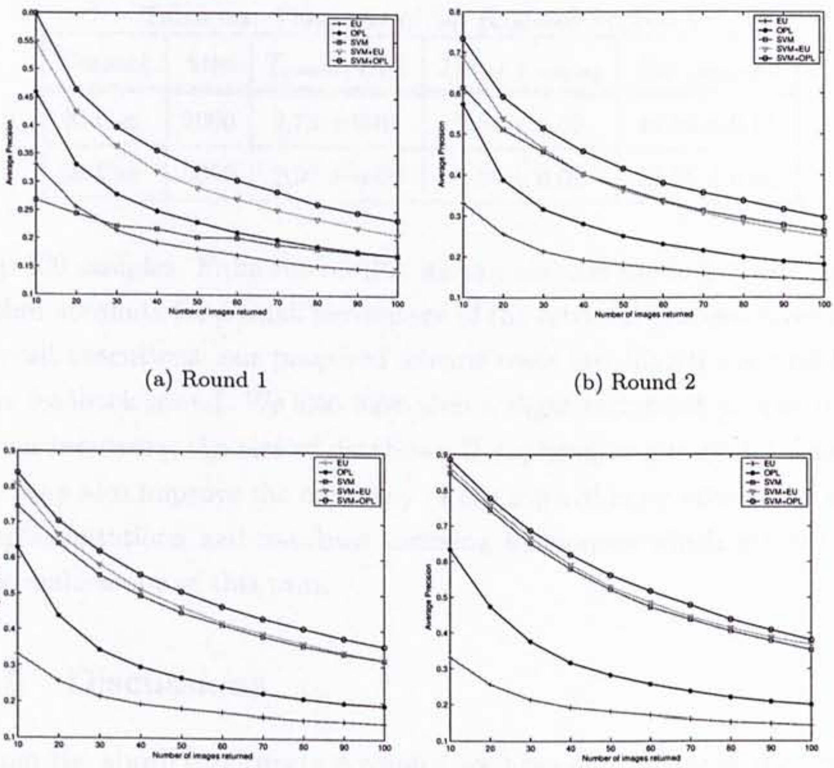
(d) Round 4

Figure 4.2: Experimental results on the 20-Cat image dataset

perform better than the heuristic Euclidean approach. The third and fourth feedback rounds also show the similar results.

4.4.4 Complexity and Time Cost Evaluation

Although we mainly focus on the retrieval performance of our proposed scheme, computational complexity and time cost may need be concerned when applying in large scale empirical applications. In our proposed scheme, we are interested to analyze the time cost of two major parts of our algorithm: the solution of OPL and the SVM training. The computational complexity for solving the parameters in OPL is $O(\sum_{i=1}^{M}(L_i^3 + 2NL_i^2))$ which was investigated in previous work [52].



(a) Round 1 (b) Round 2

(c) Round 3 (d) Round 4

Figure 4.3: Experimental results on the 50-Cat image dataset

The computational complexity of SVM training lies on the kernel and the size of training set for a convex optimization problem. Details of SVM training and analysis of computational complexity can be found in [50, 90].

We here provide the empirical analysis on the time cost of our propose scheme. The part of OPL algorithm is implemented in Matlab codes and the SVM training is based on the C++ implementation with LIBSVM [86]. Table 4.2 shows the time costs for 100 repeated executions with random query samples. The overall time for all executions includes solving the OPL, the SVM training, the computation of distances, computation of SVM distances and the retrieval ranking of

Table 4.2: Time cost of our proposed approach

Dataset	Size	$T_{solving\ OPL}$	$T_{SVM\ Training}$	$T_{all\ executions}$
20-Cat	2000	2.78 ± 0.05	5.81 ± 0.09	49.84 ± 0.11
50-Cat	5000	2.90 ± 0.05	6.39 ± 0.07	68.75 ± 0.08

top 100 samples. From the results, we can see that the cost of our algorithm accounts for a small percentage of the retrieval process. Even for overall executions, our proposed scheme costs less than 0.2 second for one feedback round. We also have seen a slight increment of time cost when increasing the size of database. If applying in a huge databases, we may also improve the efficiency of our algorithm by optimizing our implementations and database indexing techniques which are out of the main scope of this work.

4.5 Discussions

From the above experimental results, we have demonstrated the effectiveness of our proposed scheme. Nonetheless, our proposed scheme can still be improved in several aspects in the future work. One direction for improving our work is the feature weighting issue in our scheme. In the current proposed scheme, our feature weighting is based on the distance variances. We will try to investigate and combine some recently proposed weighting algorithms in Machine Learning community for improving our scheme [91, 92].

Moreover, we are interested to include the unlabelled data in our learning scheme which is also called the semi-supervised learning in Machine Learning field [93]. The unlabelled data may provide informative resource when the number of training is very small. Effective fusion of labelled and unlabelled data and the efficiency problems are the keys for the semi-supervised learning in CBIR. Whatever, this is still an open and challenging topic in Machine Learning and Information Retrieval communities [50].

4.6 Summary

In this chapter, a novel learning scheme is proposed for learning the similarity measure in CBIR with relevance feedback. We suggested an effective learning scheme by employing the SVM to constrain the optimizing learning for similarity measure in CBIR. Our proposed scheme not only can make use of the advantages of SVMs for excellent performance in kernel-based learning, but also can overcome the limitations of SVMs by unifying the optimizing learning techniques. Empirical evaluation on the real-world image datasets demonstrate our systematic formulation is more effective and more competitive than the traditional approaches.

□ **End of chapter.**

Chapter 5

Group-based Relevance Feedback

5.1 Introduction

In a typical relevance feedback mechanism of CBIR, users are solicited to make the relevance judgements on retrieved images from CBIR systems. In general, the retrieved images are marked as *relevant* or *irrelevant* to the user's query concepts. Relevant images are also called *positive* samples, and irrelevant ones are called *negative* samples. In orthodox relevance feedback techniques, the retrieval tasks by relevance feedback are simply regarded as a strict two-class (positive and negative) classification problems.

However, a simple two-class classification methodology may not be powerful enough to depict the user's high-level concepts [54]. For instance, a user would like to look for "cars" objects. The "cars" category may contain red cars, blue cars and white cars, etc. These "cars" images are simply assumed from one positive class in normal relevance feedback techniques. However, these approaches disregard much color information which is still useful to discriminate the "cars" category from other concepts.

In order to construct more powerful techniques, it is more reasonable to assume the relevant images are from multiple positive classes rather than one single positive class. Similarly, it may also consider the irrelevant images are from multiple negative classes. However, it is not practical for the multiple negative classes assumption as users are interested in the positive samples rather than the irrelevant samples. Moreover, the negative samples normally overnumbers the positive samples. Hence, classifying the negative samples into multiple negative classes will be a troublesome burden for users and trivial for the retrieval tasks.

Therefore, this chapter suggests to formulate the group-based relevance feedback technique which assumes relevant images are from multiple positive classes and irrelevant images are from one single negative class. However, under the assumption, the negative samples may largely overnumber the positive samples in each of the positive classes. In order to overcome the challenge, the SVM ensembles techniques are employed to construct our proposed group-based relevance feedback algorithm.

The rest of this chapter is organized as follows. Section 5.2 introduces the SVM ensembles and the advantages compared with the regular SVM. Section 5.3 formulates a group-based relevance feedback algorithm employing the SVM ensembles. Section 5.4 presents our experiments and performance evaluation. Section 5.5 discusses the limitation of our work. Section 5.6 concludes the work in this chapter.

5.2 SVM Ensembles

Although SVMs have been successfully applied in many empirical applications, they have a lot of limitations. First, the regular SVM is originally for binary classification problem. It may not achieve the best performance when applied in multi-class tasks. Moreover, the regular SVM treats fairly with the positive and negative instances. If instances in one of the two-class overnumber another ones, the performance of the regular SVMs may suffer dramatically. To overcome the drawbacks of the regular SVM, the SVM ensemble technique was proposed and have shown promising improvement over the regular SVMs [76, 94].

In general, an SVM ensemble is a collection of several SVM classifiers in which the decision to classify the test data is made by combining the decision functions of all individual classifiers. Suppose there is an SVM ensemble with n individual SVM classifiers, denoted as f_i $(i = 1, 2, \ldots, n)$, and a test data x, the classification result of data x is based on aggregating all the predicting results of n individual SVM classifiers. If all the n individual classifiers are all identical, the classification result is equivalent to each individual classifier. However, if the classifiers are different, the error of prediction can be reduced by combining the n classifiers. Therefore, an ensemble of several individual SVM classifiers is expected to outperform a single SVM classifier.

5.3 Group-based Relevance Feedback Using SVM Ensembles

5.3.1 (x+1)-class Assumption

Regular research efforts on relevance feedback simply consider relevance feedback as a two-class classification problem, in which the relevant instances are assumed from one positive class and the irrelevant ones are considered from another negative class. However, in practical applications, the training instances normally come from multiple positive and negative classes. To address this problem, Zhou et al. [55] suggested to represent the relevance feedback as a *(1+x)-class* classification problem (one positive class and multiple negative classes). Nakazato et al. [54] proposed to extend it as an *(x+y)-class* problem (multiple positive classes and multiple negative classes).

However, in relevance feedback tasks, users are more interested in the relevant instances rather than the irrelevant ones. Grouping the relevant instances are easier than classifying the irrelevant ones. Hence, asking the users to group the irrelevant instances is a troublesome and tedious job and it may cost much time for users. Therefore, it is more reasonable to represent the relevance feedback task as an *(x+1)-class* problem (multiple positive classes and one negative class).

5.3.2 Proposed Architecture

In order to deal with the *(x+1)-class* model, we suggest a novel group-based relevance feedback with the above suggested assumption. On the other hand, we know that the irrelevant instances in the single negative class may outnumber the relevant samples of the positive classes. To attack this problem, we employ the Support Vector Machine Ensembles technique to construct our group-based relevance feedback framework. Fig. 5.1 depicts the idea of our proposed architecture. For instance, in Fig. 5.1, there are two positive groups (PG-1 and PG-2) and a negative group (NG) provided by users. The negative group is partitioned into several parts based on some kind of sampling strategy. Each negative part is then combined with a positive group to train a SVM classifier. Then, a SVM ensemble combiner is employed to combine the classifiers for each positive group with some combination strategy. At the end, all positive groups are aggregated to construct the final decision function. The strategy to construct the SVM classifier in our scheme is the bagging technique [95, 96]. Although the boosting technique is also proposed for the SVM ensembles [97], it is not suitable to relevance feedback tasks for its huge computation load.

5.3.3 Strategy for SVM Combination and Group Aggregation

In our group-based relevance feedback scheme, after each SVM classifier has been constructed and trained, we need to combine the SVM classifiers for each SVM ensemble by choosing a proper combination strategy and then aggregate all ensembles to form the final group decision function by an appropriate strategy. In literature, several work has devoted to study the combination methods of SVM ensembles, such as Majority Voting [94, 98], Sum Rule [99], and other mixture mod-

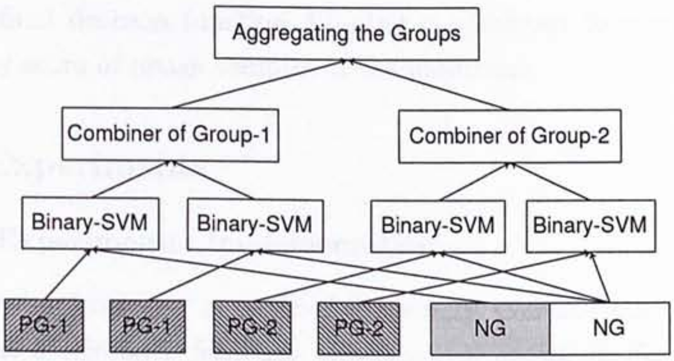


Figure 5.1: The model architecture of our proposed scheme based on SVM ensembles.

els [100], etc. However, previous work focusing on the classification problems are not appropriate for relevance feedback. By considering for practical applications, we adopt the Sum Rule method as a simple yet effective approach to combine the SVM classifiers. And the strategy for group aggregation is based on a linear combination with different weights. The detailed formulation of our group-based decision function is given as follows.

Let K_g be the number of positive groups in a group-based relevance feedback learning step. Let K_m be the number of SVM classifiers in the SVM ensemble of each positive group learning. Let us denote f_{ij} (j=1,2,...,K_m) as a decision function of the jth SVM classifier in the ith SVM ensemble and denote F_i (i=1,2,...,K_g) as a decision function of the ith SVM ensemble. Then for the ith ensemble, its decision function by the Sum Rule is given as

$$F_i(\mathbf{x}) = \sum_{j=1}^{K_m} f_{ij}(\mathbf{x}) . \tag{5.1}$$

Then let w_i be the weights of each positive group which is determined by the number of samples in each group. Hence, the final decision function $f_{GRF}(\mathbf{x})$ of the group-based algorithm can be obtained as

$$f_{GRF}(\mathbf{x}) = \sum_{i=1}^{K_g} w_i F_i(\mathbf{x}) = \sum_{i=1}^{K_g} \sum_{j=1}^{K_m} w_i f_{ij}(\mathbf{x}) . \tag{5.2}$$

The final decision function $f_{GRF}(\mathbf{x})$ is employed to compute the similarity score of image samples in the databases.

5.4 Experiments

5.4.1 Experimental Implementation

We have implemented a retrieval system to evaluate our proposed group-based relevance feedback algorithm, as shown in Fig. 5.2. In our relevance feedback mechanism, users drag and group the positive images which are considered as relevant from the retrieval pool in each round. The images remaining in the retrieval pool are considered as irrelevant (negative) in default. The positive images and negative images in the previous round will be accumulated to the next round learning. In our experiments, we compare the retrieval performance between our

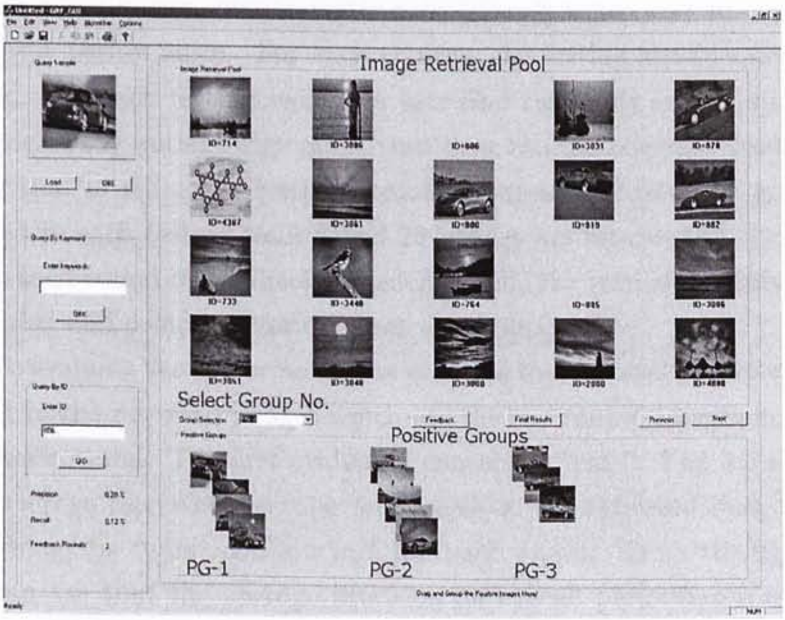


Figure 5.2: GUI of our group-based relevance feedback system.

proposed group-based relevance feedback with SVM ensemble (GRF-SVM.E) and traditional relevance feedback algorithm using SVMs (RF-

SVM). The test image dataset used in our experiments is selected from the COREL image data set. 50 categories of images are selected and each category contains 100 images.

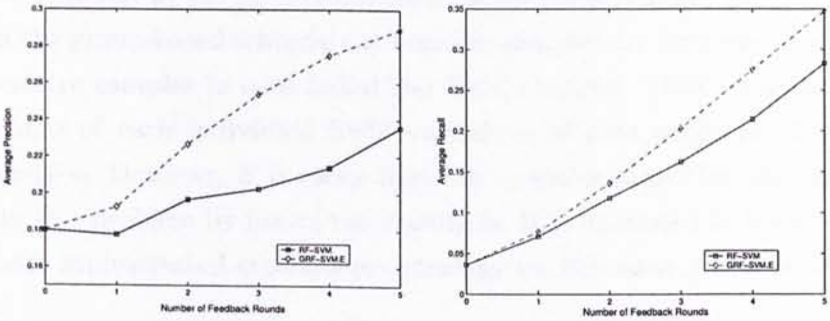
For image representation, three low-level features are extracted: color, shape and texture. Namely, a 36-dimensional low-level feature is engaged including 9-dimensional color moment, 18-dimensional edge direction histogram and 9-dimensional wavelet texture feature [30].

The kernel function for SVM used in our experiments is based on the Radial Basis Function (RBF). We notice that different parameters of the kernel function in SVM have large impact on the retrieval performance. To enable objective evaluation, the parameters are set to the same constant values for different algorithms respectively.

5.4.2 Performance Evaluation

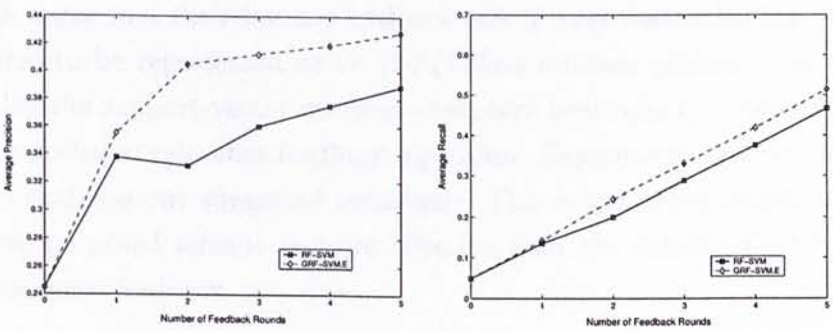
In the experiments, two semantic concepts are tested to evaluate the retrieval performance. For each concept, 10 testing sessions are engaged. For each testing session, a user first randomly selects a query sample as the initial query point, and then run the relevance feedback algorithm to refine the retrieval results. 5 rounds of feedback are executed in each testing session and 20 images are returned to the user after each round of feedback. In each round, the retrieved results are recorded and compared for different algorithms.

To evaluate the performance, we examine the retrieval precision and recall in the returned images which are the top ranked images in each feedback round. The first evaluated concept is "cars". Fig. 5.3 shows the average retrieval precision and recall on the retrieved images for searching the "cars" concept in 5 feedback rounds. From the figures, we can see that the average precision and recall performance of our group-based relevance feedback employing SVM ensembles is better than the regular SVM-based method. And we also observe the similar improvement for searching the "roses" concept from Fig. 5.4.



(a) Average Precision (b) Average Recall

Figure 5.3: Retrieval Performance for searching "cars"



(a) Average Precision (b) Average Recall

Figure 5.4: Retrieval Performance for searching "roses"

5.5 Discussions

Although we have already shown promising preliminary results, limitations of the proposed scheme should also be addressed. One of the important problems is the efficiency issue. Since the SVM ensembles technique is engaged, the relevance feedback algorithm is required to train a lot of SVM classifiers which costs a lot of time. However, since the number of training in relevance feedback tasks is typically rare, it still be acceptable for typical applications. Moreover, another problem encountered in the proposed scheme is the small sample size problem. In the group-based scheme, the negative samples are far more than the positive samples in each individual SVM classifier. Hence, the performance of each individual SVM may drop without sufficient training samples. However, it is lucky that the ensemble algorithm can alleviate this problem by fusing the classifiers. It is interested to investigate more sophisticated combination strategy for this issue in the future.

5.6 Summary

In this chapter, we propose a novel group-based relevance feedback scheme in the context of CBIR. Different from traditional approaches, we argue that the relevance feedback task is more reasonable and practical to be represented as an *(x+1)-class* learning problem. We employ the support vector machine ensembles technique to construct our group-based relevance feedback algorithm. Experiments are conducted to evaluate our suggested technique. The experimental results show our proposed scheme is more effective than the regular SVM-based relevance feedback.

□ **End of chapter.**

Chapter 6

Log-based Relevance Feedback

6.1 Introduction

Relevance feedback has been received a considerable amount of research from heuristic techniques to sophisticated learning techniques in the past years [56, 38, 41, 71, 45, 48]. It has been shown as an effective tool to improve the retrieval performance in CBIR [18] and has already been considered as a key part when designing a practical CBIR system. In general, relevance feedback mechanism has to solicit users for giving the relevance judgements on the initial query results by the CBIR system. While a user has made the relevance judgements on the retrieval results, relevance feedback is used as a query refinement technique to improve the retrieval results. As the learning task of relevance feedback is very tough, it generally needs to repeat many rounds of feedback in order to achieve satisfactory results. Therefore regular relevance feedback learning techniques are very time-consuming.

Moreover, the relevance feedback procedure to indicate the relevances of images is viewed as a tedious and boring step for 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 [44, 101, 102, 103, 104, 105]. 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 users could be served as important resources to help the relevance feedback task in CBIR. Although the selection of relevant samples by users is subjective for different people, it is reasonable to assume that the major semantic concept can be captured from the logs provided by users. In past relevance feedback research, less efforts are devoted to study on log-based relevance feedback algorithms. To our knowledge, there is only one recent research study in [47] which is similar to our work in this chapter. The authors suggested to learn a semantic space by learning the user's 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 chapter, a novel scheme is presented to study the user feedback logs in order to help the learning task of relevance feedback in CBIR. The rest of this chapter is organized as follows. Section 6.2 provides the background of log-based relevance feedback problem and discusses our motivation. Section 6.3 presents a modified SVM technique called Soft Label Support Vector Machine (SLSVM) and formulates a log-based relevance feedback algorithm employing the SLSVM technique. Section 6.4 describes detailed experiments and the performance evaluation. Section 6.5 discusses several problems for building the log-based relevance feedback algorithms in CBIR. Section 6.6 gives the conclusions.

6.2 Related Work and Motivation

As an interactive technique in CBIR, relevance feedback has to solicit users for relevance judgements on the retrieval results by CBIR systems. Due to the difficulties of the semantic gap and human perception subjectivity problems [18], the relevance feedback procedure normally has to repeat for many rounds in order to obtain satisfactory results. Hence, regular relevance feedback is a time-consuming and tedious job for users. In order to overcome the challenges, the user feedback histories can be logged and engaged to help the retrieval tasks through a long-term learning purpose.

It is evident that the user feedback logs can be collected in a CBIR system from a long-term perspective. How can we employ these resources to help the retrieval tasks? An intuitive way is to borrow the techniques in traditional text-based information retrieval, such as the query expansion (QEX) techniques [106, 107, 108]. Query expansion can be viewed as a multiple-instance sampling technique in which the returned samples of the next round are selected from the neighborhood from the relevant samples of the previous rounds [40, 109, 110, 111]. Although query expansion has already achieved successful applications in text-based information retrieval, it may not be very effective for solve the log-based retrieval tasks in CBIR.

Recently, kernel-based machine learning techniques have been shown with powerful performance in many pattern learning applications. In order to formulate an effective algorithm for engaging the user feedback logs, we suggest to solve the problem by employing a popular yet powerful kernel-based learning technique–Support Vector Machines (SVMs) which have shown with successful applications in relevance feedback tasks [51, 56, 43, 44]. However, regular SVM technique could not be directly applied to formulate the log-based relevance feedback problem. Hence, we propose a modified SVM technique called Soft Label Support Vector Machine to construct the log-based relevance feedback algorithm.

6.3 Log-based Relevance Feedback Using SLSVM

6.3.1 Problem Statement

We first describe the ways to log the feedback histories of users and how to employ them to help the regular relevance feedback tasks in CBIR systems. In a typical relevance feedback procedure, a CBIR system first returns N images from databases to a user based on some kinds of similarity measure. The user then specifies which images are relevant (positive) and which ones are irrelevant (negative). When the user submits his/her judgement results to the CBIR system, such a feedback session including N_+ positive samples and N_- negative samples (where $N_+ + N_- = N$) will be logged in the log database. The task for the log-based relevance feedback is to employ the accumulated feedback sessions in the log databases for improving relevance feedback in the retrieval tasks.

The first step toward the log-based relevance feedback is to access and effectively organize the feedback information of users from the log database. For solving this problem, a Relevance Matrix (RM) is constructed to express the relevance relationship between the image samples in the database. The column of the relevance matrix represents the ID of image samples in the database and the row represents the session number of feedback sessions in the log database. For each session in the log database, the relevance values 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:

$$R_{ij} = \sum_k \delta_k \cdot RM(k,i) \cdot RM(k,j) \tag{6.1}$$

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

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.

Hence, by engaging the log information, we can obtain a list of relevant samples and irrelevant samples associated with different relationship values in a relevance feedback procedure. The challenge of the log-based relevance feedback is how to formulate an effective algorithm to deal with the training samples associated with different confidence degrees in relevance or irrelevance. In order to solve this problem, we propose a modified SVM technique called Soft Label Support Vector Machine in which the training data are associated with soft labels of different confidence degrees.

6.3.2 Soft Label Support Vector Machine

Regular SVM techniques assume that the labels of the data are absolutely correct. However, when some labels of data are incorrect, they may hardly impact on the predicting decision. In order to reduce the errors arose from the unconfident data, we incorporate the label confidence degree of data in the regular SVM model and propose the Soft Label SVM as below. Assume the labels of the data are with different confidence degrees. Let us denote the training data,

$$(\mathbf{x_1}, s_1), ..., (\mathbf{x_l}, s_l) \in \mathcal{X} \times S, \quad S \subseteq [-1, +1]$$

where l is the number of training data for some training set $\mathcal{X} \subseteq \mathbb{R}^m$ in the m-dimension input space, $s_i \in [-1, 1]$ is the soft label of the training data which indicates the degree of relevance or irrelevance belonging to the class. Given a soft label set S, the corresponding "hard label" set Y can be obtained by applying a sign function on S

$$Y = sgn(S) = \{+1, -1\}. \tag{6.3}$$

The objective function of Eq. (2.4) becomes

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{x} \in \mathbb{R}^m, b \in \mathbb{R}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum_i y_i s_i \xi_i \\ \text{subject to} \quad & y_i((\Phi(\mathbf{x_i}) \cdot \mathbf{w}) + b) \geq y_i s_i \rho - \xi_i \,, \\ & \xi_i \geq 0, \quad i = 1, \ldots, l \,, \\ & 0 \leq \nu \leq 1, \quad \rho \geq 0 \,. \end{aligned} \tag{6.4}$$

From this objective function, we can see the margin errors ξ_i will be large if constrained with a small label value s_i and will be smaller for a larger one in non-separable cases.

Let us introduce Lagrange Multiplier technique to derive the dual problem

$$\begin{aligned} L(\mathbf{w}, \xi, b, \rho, \alpha, \beta, \delta) &= \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum_i y_i s_i \xi_i \\ - \sum_i (\alpha_i(y_i(\Phi(\mathbf{x_i}) \cdot \mathbf{w} + b) &- y_i s_i \rho + \xi_i) - \beta_i \xi_i) - \delta\rho \,. \end{aligned} \tag{6.5}$$

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

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x_i}) = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x_i}) \,; \\ \frac{\partial L}{\partial \xi_i} &= y_i s_i \frac{1}{l} - \alpha_i - \beta_i = 0 \Rightarrow 0 \leq \alpha_i \leq y_i s_i \frac{1}{l} \,; \\ \frac{\partial L}{\partial b} &= -\sum_{i=1}^{l} \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 \,; \\ \frac{\partial L}{\partial \rho} &= -\nu + \sum_{i=1}^{l} \alpha_i y_i s_i - \delta = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i s_i - \delta = \nu \,. \end{aligned}$$

The dual of the primal optimization problem is turned into

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x_i}, \mathbf{x_j})$$

$$\text{subject to} \quad \sum_i \alpha_i y_i = 0 \;, \tag{6.6}$$

$$0 \le \alpha_i \le y_i s_i \frac{1}{l}, \quad i = 1, 2, \ldots, l \;, \tag{6.7}$$

$$\sum_i \alpha_i y_i s_i \ge \nu \;. \tag{6.8}$$

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

$$f(\mathbf{x}) = sgn(\sum_i \alpha_i y_i k(\mathbf{x}, \mathbf{x_i}) + b) \;. \tag{6.9}$$

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

$$0 \le \alpha_i \le \frac{1}{l} \;, \tag{6.10}$$

while in the Soft Label SVM, the constraint becomes

$$0 \le \alpha_i \le y_i s_i \frac{1}{l} \;. \tag{6.11}$$

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.

6.3.3 LRF Algorithm by SLSVM

The implementation of LRF algorithm with SLSVM is similar to building a relevance feedback algorithm using SVM. The difference is that the training samples in SLSVM are with different confidence degrees. A key step for constructing the LRF algorithm is to select the training samples. 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 positive and N_- samples are negative. Normally, there are few positive samples in the initial retrieval. However, we can employ the user feedback logs to look for more positive samples by using the N_+ initial positive samples as the seeds. For each positive seed i, we compute the relationship value between seed i and the image x in the database: R_{ix}. The image with a large R_{ix} value will be selected as the potential positive sample with a specific confidence value. The negative training sample selections are the same as the positive ones. Hence, the soft label of a training sample x corresponding to the seed i is computed by the following formulation

$$\mathbf{s_x} = \begin{cases} R_{ix}/R_{max}, & R_{ix} \geq 0 \\ -R_{ix}/R_{min}, & R_{ix} < 0 \end{cases} \tag{6.12}$$

where R_{max} is the maximum of the relationship values and R_{min} is the minimum one.

Hence, based on the feedback and the logs, a proportion of training data can be selected by ranking their confidence values. After that, the training data are fed to the SLSVM classifier for the training purpose. For predicting the retrieval results, the relevance feedback is a bit different from regular classification problem in SVM. For the classification purpose, regular SVM output only with positive (+1) or negative (−1). In order to facilitate the evaluation of the retrieval results, we need to construct an evaluation function for the LRF algorithm based on SLSVM.

After training an SLSVM classifier, we can solve a set of optimal parameters α by the QP technique in Eq. 6.6. Similar to the derivation from regular SVM techniques in [89, 112], the evaluation function of SLSVM can be constructed by the decision function in Eq. (6.9) as follows:

$$f(\mathbf{x}) = \sum_i \alpha_i y_i k(\mathbf{x}, \mathbf{x_i}) + b\ , \tag{6.13}$$

where the parameter b can be solved by a set of the support vectors [89]. Based on the evaluation function, we can compute the relevance score of each image sample in the dataset and rank them by the relevance score. The image with a larger score is closer to the desired target.

6.4 Experimental Results

6.4.1 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. Secondly, the approach can help us evaluate the performance automatically, which can reduce the subjective errors in manual evaluations by different people.

6.4.2 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 [26]. 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 [22, 26, 27, 28]. 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 [5, 29]. 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 [25, 31, 32]. 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 [25]. 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.

6.4.3 Experimental Setup

In order to evaluate the log-based relevance feedback algorithm, we have developed a CBIR system with relevance feedback mechanism [51, 56]. The collection of user logs is an important work for empirical evaluations of our proposed technique. In our experiments, we solicit

10 researchers to 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 by 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.

In our experiments, each user is asked 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 application; nonetheless, we hope to evaluate whether our proposed technique can work well even with limited log sessions.

6.4.4 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 comparision 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 [44]. 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 to learn in a high dimensional feature space since it is simply based on the Euclidean distance.

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 weighting the distance measure. From Eq. 6.1, we compute the relationship values between the query sample and the samples in the database. Similar to the soft label strategy, we employ the form similar in Eq.6.12 to weight the Euclidean distance measure.

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. Several problems for SVM-based relevance 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 [86].

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 unfair an comparison for different approaches.

The first set of experimental results on the 20-Category dataset is shown on Fig. 6.1. 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 surpass 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. 6.2. 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. 6.3 and Fig. 6.4, 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.

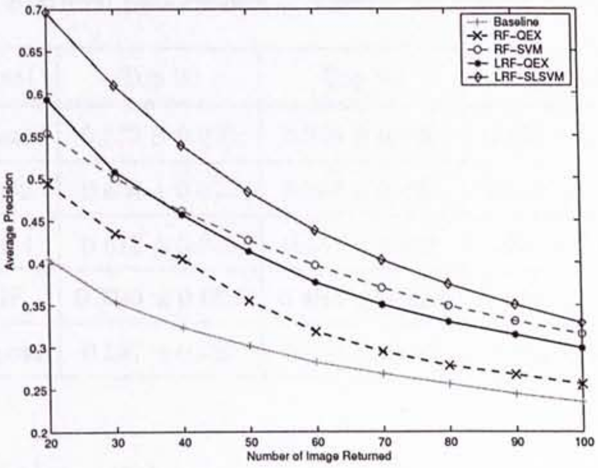


Figure 6.1: Average precision for the algorithms in the 20-Category dataset.

Moreover, 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 (Poly-2: Degree 2 Polynomial and Poly-4: Degree 4 Polynomial), Radial Basis Function (RBF) and Sigmoid kernel (Sigmoid). In these experiments, we choose the best parameters for each individ-

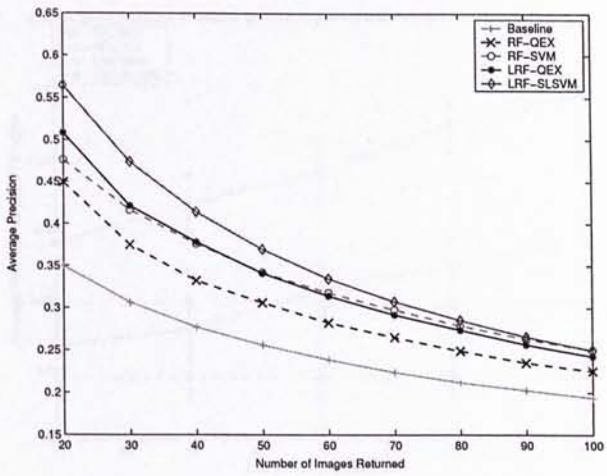


Figure 6.2: Average precision for the algorithms in the 50-Category dataset.

Table 6.1: Retrieval performance of different kernels on 20-Cat dataset

Kernel	Top 20	Top 50	Top 100
Linear	0.573 ± 0.026	0.379 ± 0.020	0.253 ± 0.014
Poly-2	0.604 ± 0.025	0.387 ± 0.020	0.257 ± 0.014
Poly-4	0.612 ± 0.025	0.397 ± 0.020	0.262 ± 0.014
RBF	$\mathbf{0.700 \pm 0.022}$	$\mathbf{0.483 \pm 0.019}$	$\mathbf{0.334 \pm 0.014}$
Sigmoid	0.597 ± 0.025	0.359 ± 0.020	0.224 ± 0.014

ual kernel and run 200 rounds for each one. From the experimental results shown in Table 6.1 and Table 6.2, we can observe 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.

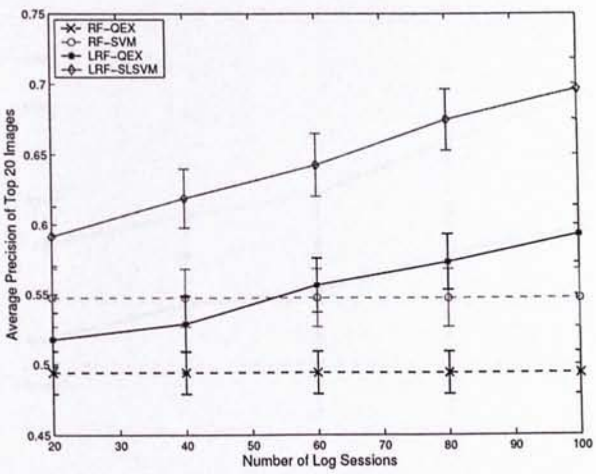


Figure 6.3: Average precision of algorithms for different number of log sessions in the 20-Category Dataset.

6.5 Discussions

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

First of all, we do not consider noisy cases from the users in our

Table 6.2: Retrieval performance of different kernels on 50-Cat dataset

Kernel	Top 20	Top 50	Top 100
Linear	0.370 ± 0.023	0.208 ± 0.014	0.126 ± 0.009
Poly-2	0.381 ± 0.022	0.210 ± 0.014	0.130 ± 0.009
Poly-4	0.383 ± 0.023	0.212 ± 0.014	0.133 ± 0.010
RBF	$\mathbf{0.574 \pm 0.022}$	$\mathbf{0.388 \pm 0.018}$	$\mathbf{0.267 \pm 0.013}$
Sigmoid	0.422 ± 0.022	0.212 ± 0.014	0.120 ± 0.009

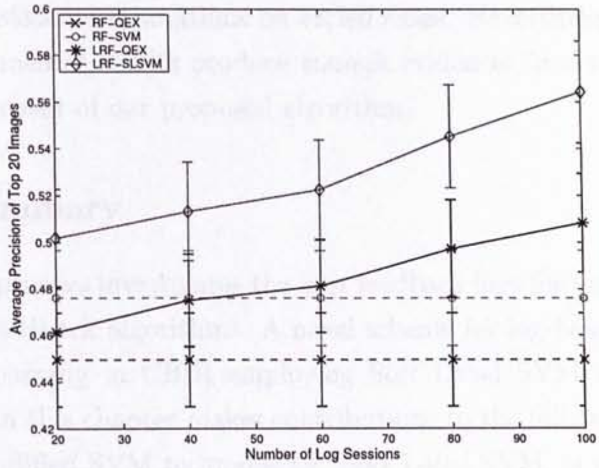


Figure 6.4: Average precision of algorithms for different number of log sessions in the 50-Category Dataset.

experiments although our experiments may have included the noisy data contributed by the users. In general, our proposed technique can reduce the impact of the noisy data better than the typical approaches. We may further investigate the advantages of our proposed technique in handling the noisy cases with different situations compared with traditional techniques.

Secondly, the dimension of the logs need to be considered when the number of feedback sessions in the log database are very large. Thus, we may need to study the 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 [47]. We can also incorporate the method to our proposed scheme in the future work.

Furthermore, when the dimension of the log sessions and the number of images are large, the complexity of our proposed algorithm should be addressed and the efficiency of the proposed algorithm may need be reduced. Thus, we will study the efficiency problem and improve it with more sophisticated technique to speed up the image retrieval task. Moreover, we may test our algorithm on larger databases and perform more sophisticated evaluations on varied cases. Nevertheless, our current experimental results produce enough evidences in demonstrating the effectiveness of our proposed algorithm.

6.6 Summary

In this chapter we investigates the user feedback logs for improving the relevance feedback algorithms. A novel scheme for log-based relevance feedback learning in CBIR employing Soft Label SVM is proposed. The work in this chapter makes contributions to the following aspects. First, a modified SVM technique,i.e. Soft Label SVM, is proposed for solving the log-based relevance feedback task in CBIR. The proposed technique is generic and may be useful for other applications. Second, detailed empirical evaluations are conducted to study the user feedback logs in our experiments, in which promising results show that our approach outperforms the regular technique. Finally, several important problems are addressed in formulating the log-based relevance feedback algorithms in CBIR. We indicate that the kernel function plays an important role on the performance of the proposed algorithm and report that the RBF kernel is more effective than other kernels in the experiments.

□ **End of chapter.**

Chapter 7

Application: Web Image Learning

7.1 Introduction

Along with the rapid development of multimedia devices and the internet, the amount of images have been dramatically increased in the past decade. Although Content-Based Image Retrieval (CBIR) has been studied for many years [4], searching semantic concepts in image databases is still a formidable task. Earlier approaches for CBIR are usually based on the Query-By-Example (QBE) strategy [4]. These approaches are inflexible since users may have difficulties in describing the query concepts. In general, searching by keywords is more easier to describe the query concepts than the QBE strategy. Recent research work begins to study the annotation techniques for attaching the textual labels to images [113]. However, fully automatic annotation techniques are yet a long way off.

In order to search the semantic concepts in images databases, we propose a scheme to engage Web images for learning the semantic concepts as the Web images associated with textual descriptions can serve as an important knowledge base. Our strategy is to search the semantic concepts by words from the Web and learn the returned Web images associated with the words. The Web images after filtering out the noisy images serve as the training set for learning in the image databases. The idea of our scheme is similar to a previous study which also proposed to engage the Web images for image recognitions [114]. With a different purpose, our interest is to investigate the issue for searching semantic concepts in image databases. Moreover, we propose to employ the Biased Support Vector Machine (SVM) techniques for attacking the learning tasks.

The rest of this chapter is organized as follows. Section 7.2 presents our proposed learning scheme and the associated techniques. Section 7.3 presents the experimental results. Section 7.4 discusses the limitation of our proposed scheme. Section 7.5 concludes our work.

7.2 A Learning Scheme for Searching Semantic Concepts

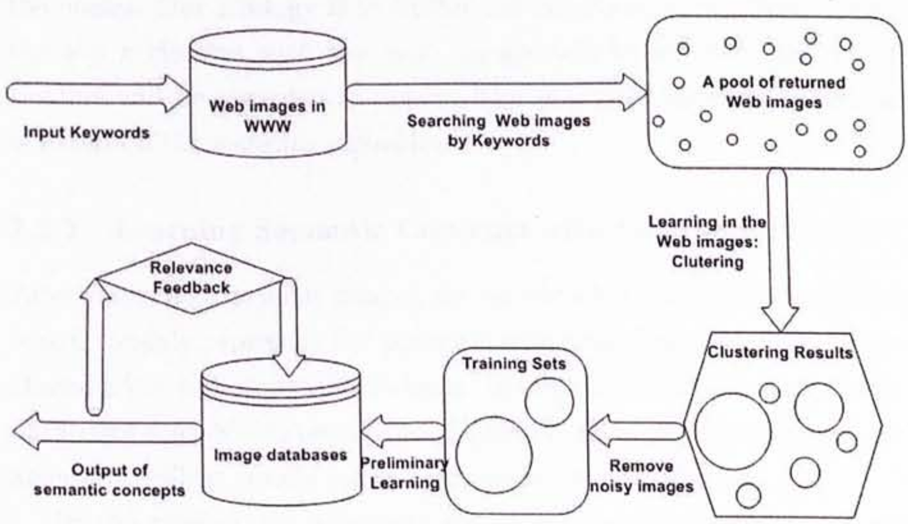


Figure 7.1: Overall architecture of our proposed scheme

Fig. 7.1 presents the overview of our proposed scheme for learning Web images to search the semantic concepts in image databases. We illustrate each step of our proposed system as follows.

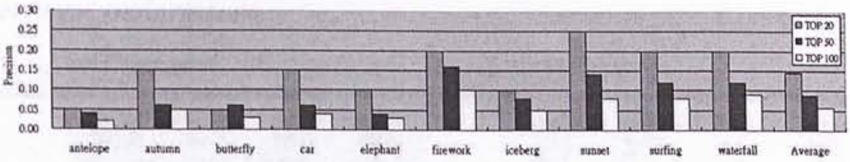


Figure 7.2: Experimental results on 10 testing semantic concepts

7.2.1 Searching and Clustering Web Images

In our proposed system, a user first keys in words to represent their desired semantic concepts. Then, the proposed system searches the images on the Web which are associated with the related words. In our approach, we solicit Web image search engines to do this job. From the Web, we collect a pool of images which have textual descriptions related the semantic concepts. However, the image pool may contain many noisy images. Thus, we employ clustering techniques to remove the noises. Our strategy is to cluster the images into k clusters. Then, the top p clusters with the most images will be selected, and other clusters will be regarded as noises. The engaged clustering technique is based on the k-means algorithm.

7.2.2 Learning Semantic Concepts with Relevance Feedback

After removing the noisy images, we can obtain a set of training images which roughly represent the semantic concepts. Then, we employ the Biased SVM techniques [51] to learn the semantic concepts in the image databases since SVMs provide good generalization performance and can achieve excellent results on pattern classifications problems [74].

In the preliminary searching round, as negative samples are not available, we simply employ the One-class SVMs (1-SVM) to learn the training set of images in the database. 1-SVM is derived from classical SVMs for solving density estimation problems. After learning by 1-SVMs, we can obtain the preliminary searching results. Then, we employ the relevance feedback with Biased SVM to improve the retrieval performance.

Figure 7.3: Top 40 Web images returned from Google by searching "firework"

7.3 Experimental Results

7.3.1 Dataset and Features

In our image database, we collect 20,000 images from the COREL image CDs which include 200 semantic categories, such as *antelope*, *car* and *sunset*, etc.

The image representation is an important step toward semantic learning in CBIR. We extract three features to represent the images: color, shape and texture. The color feature engaged is color moment, since it is closer to human natural perception. A 9-dimensional color moment is employed [4]. For the shape feature, edge direction histogram (EDH) is selected [4]. Canny edge detector is applied to obtain the edge images. The computed EDH from the edge image is quantized into 18 bins of 20 degrees each, hence an 18-dimensional EDH is used. For the texture feature, the wavelet-based texture is engaged. Dis-

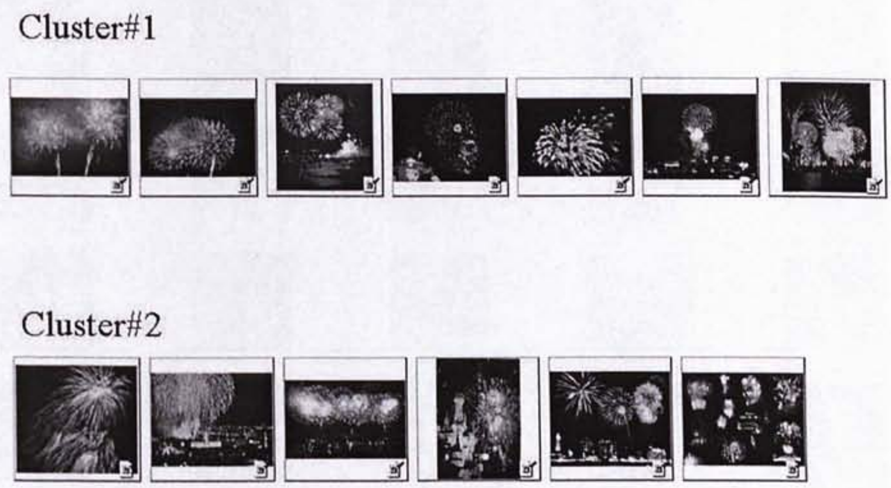


Figure 7.4: Clustering results after removing noisy Web images

crete Wavelet Transformation (DWT) on the gray images employing a Daubechies-4 wavelet filter. We perform 3-level DWT decomposition and obtain ten subimages, in which nine subimages are selected to compute the entropies. Thus, a 9-dimensional wavelet texture feature is obtained.

7.3.2 Performance Evaluation

To evaluate the performance of the proposed scheme in a large image database, we choose 10 semantic concepts, including *antelope*, *autumn*, *butterfly*, *car*, *elephant*, *firework*, *iceberg*, *sunset*, *surfing* and *waterfall*. To search Web images, we choose the Google Image Search Engine [1]. For each query semantic concept, top 40 returned imaged from Google were collected. For the clustering algorithm in our proposed scheme, we choose the parameters k=12 and p=2 in the k-means algorithm. The kernel function used in SVMs is based on the Radial Basis Function [74]. Fig. 7.2 shows the experimental results. We observe that the average retrieval precision on TOP 20, TOP 50, and TOP 100 results is over

[1]http://images.google.com/

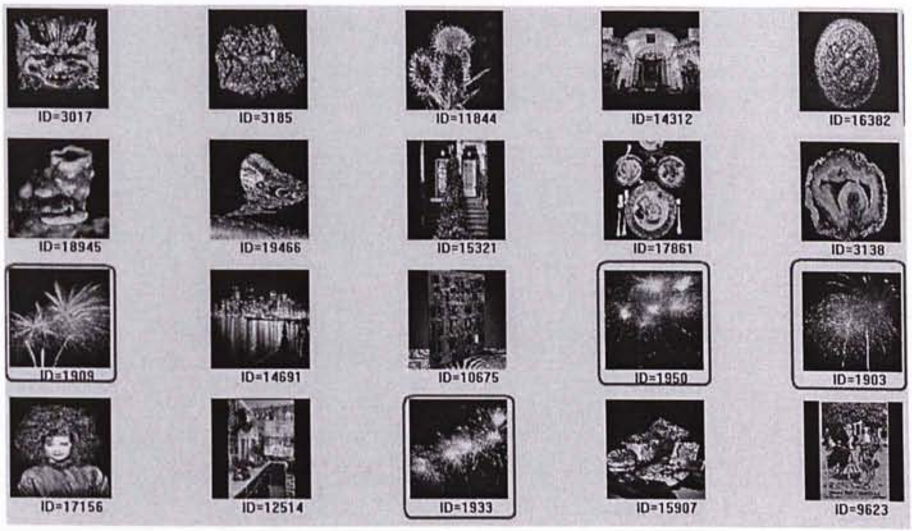


Figure 7.5: Preliminary learning results before relevance feedback

Table 7.1: Average retrieval precision by relevance feedback

Feedback Round	TOP 20	TOP 50	TOP 100
No Feedback	14.5%	8.8%	5.7%
1 Feedback	29.0%	15.2%	15.4%
2 Feedback	47.0%	26.4%	16.1%
3 Feedback	58.5%	32.2%	18.3%

14%, 8%, and 5%, respectively.

The preliminary searching results are further improved by relevance feedbacks using the Biased SVM [51]. In each feedback round, 50 images are presented to users for judging their relevance. Table 7.1 shows the retrieval performance improved by 3-round relevance feedbacks. We can see that the average precision in TOP 20, TOP 50 and TOP 100 after 3-round feedbacks can achieve 58%, 32% and 18%, respectively. Fig. 7.3, Fig. 7.4, Fig. 7.5, and Fig. 7.6 give an example to show the learning procedures for searching the semantic concept of "firework".

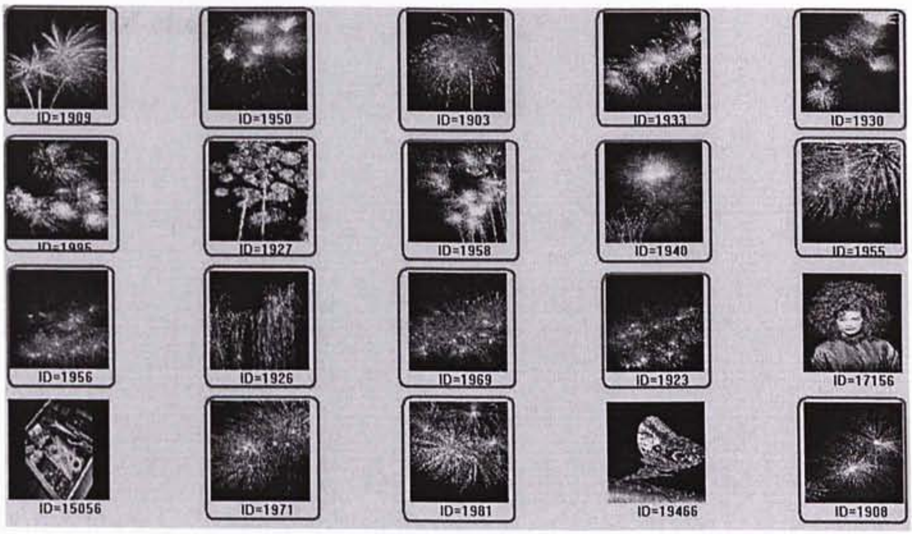


Figure 7.6: Learning results after 3-round relevance feedback

7.4 Discussions

Although we have demonstrated promising results from the above experiments, we also notice limitations of our scheme. One disadvantage is that the preliminary retrieval performance will be sensitive to the quality of the collected images from the web, which is one of our future research tasks for investigation. Moreover, we will engage more effective relevance feedback algorithms to help the retrieval tasks.

7.5 Summary

In this chapter, a novel scheme is proposed to learn Web images for searching semantic concepts in image databases. We suggest to employ the SVMs techniques to attack the learning tasks of searching semantic concepts in image databases. The promising results are presented to demonstrate the effectiveness of our scheme. Finally, we address the limitations of the suggested scheme and indicate some possible future work.

□ **End of chapter.**

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This thesis investigated the learning techniques for solving the relevance feedback tasks in CBIR. We addressed the disadvantages of traditional learning techniques on relevance feedback in CBIR. We also suggested effective relevance feedback algorithms to overcome the shortcomings of traditional techniques through different learning perspectives.

First of all, this thesis pointed out an important feature of relevance feedback learning, i.e. the imbalanced dataset problem. In order to attack the imbalance problem, we proposed a novel technique called Biased SVM to construct the relevance feedback algorithm in CBIR.

Secondly, this thesis indicated that typical relevance feedback with statistical learning algorithms have drawbacks when lacking sufficient training samples. In order to overcome the problem of SVM-based relevance feedback techniques, this thesis proposed an original scheme to unify the optimizing learning and the SVM techniques for formulating the relevance feedback algorithm.

Furthermore, the thesis argued that it is more reasonable and practical to assume the training samples in relevance feedback tasks are from multiple positive classes and one negative class in real-world CBIR applications. Based on the relaxation, a new group-based relevance feedback algorithm constructed with SVM ensembles techniques is proposed in the thesis.

In addition to regular relevance feedback techniques, this thesis also advised to study the log-based relevance feedback through a long-term learning perspective. In order to engage the user feedback logs well, a modified SVM technique called Soft Label SVM was proposed to formulate the log-based relevance feedback algorithm.

Finally, the thesis also presented an interesting and meaningful application to study Web image learning. We suggested to employ the Web images for learning the semantic concepts in image database by engaging a relevance feedback mechanism with the SVMs techniques.

8.2 Future Work

Although this thesis has devoted much effort to the learning techniques on relevance feedback in the context of CBIR, several important research directions still can be extended in future work. First of all, efficient algorithms can be studied to extend our suggested techniques. As this thesis mainly focused on the retrieval performance of algorithms from vision and learning perspectives with less consideration on the efficiency problem, it is encouraged to explore on the database performance issues and construct more efficient techniques. Recently, some international workshops and conferences have begun to gather researchers from computer vision, machine learning and database communities to study effective and efficient algorithms for real-world applications. We expect that more satisfactory results can be obtained in the future from this research direction and wish to provide contribution on the area in the future.

Furthermore, unsupervised learning techniques [50, 115] or semi-supervised techniques [116, 117, 118, 119, 93] in statistical learning area can be investigated to attack the relevance feedback in future work.

Since the learning techniques studied in this thesis are major based on supervised learning techniques which only learn with labelled data, it is reasonable to convince that the unlabelled data is informative to the relevance feedback tasks [120]. However, relevance feedback learning with unlabelled data may face risks if they are not engaged carefully. Moreover, the efficiency problem may also be a problem for unsupervised and semi-supervised learning techniques. Nonetheless, they still will be promising research areas to improve the retrieval performance of CBIR systems in the future.

☐ **End of chapter.**

Appendix A

List of Publications

Published or Accepted Papers

[1] **Chu-Hong Hoi** and Michael R. Lyu, A Novel Log-based Relevance Feedback Technique in Content-based Image Retrieval, in *Proceedings of ACM International Conference on Multimedia 2004 (MM'2004)*, New York, Oct, 2004

[2] **Chu-Hong Hoi** and Michael R. Lyu, Web Image Learning for Searching Semantic Concepts in Image Databases, in *Poster Proceedings of the 13th International World Wide Web Conference (WWW'2004)*, New York, USA, 17-22 May, 2004

[3] **Chu-Hong Hoi** and Michael R. Lyu, Group-based Relevance Feedback with Support Vector Machine Ensembles, in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'2004)*, Cambridge, UK, 23-26 August, 2004

[4] **Chu-Hong Hoi**, Chi-Hang Chan, Kaizhu Huang, Michael R Lyu and Irwin King, Biased Support Vector Machine for Relevance Feedback in Image Retrieval, in *Proceedings of International Joint Conference on Neural Networks (IJCNN'2004)*, Budapest, Hungary , 25-29 July, 2004

[5] **Chu-Hong Hoi** and Michael R. Lyu, Robust Face Recognition Using Minimax Probability Machine, in *Proceedings of The 2004 IEEE International Conference on Multimedia and Expo (ICME'2004)*, Taiwan, 27-30 June, 2004

[6] **Chu-Hong Hoi**, Wei Wang and Michael R. Lyu, A Novel Scheme for Video Similarity Detection, in *International Conference on Image and Video Retrieval (CIVR'2003)*, USA, pages 373-382, *Lecture Notes on Computer Science*, vol. 2728, Springer, 2003

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 Transactions 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] C. H. Hoi, W. Wang, and M. R. Lyu, "A novel scheme for video similarity detection," in *Proceedings of International Conference on Image and Video Retrieval (CIVR'03)*. 2003, vol. 2728 of *LNCS*, pp. 776–780, Springer.

[7] C. H. Hoi and M. R. Lyu, "Robust face recognition using minimax probability machine," in *Proceedings 2004 IEEE Interna-*

tional Conference on Multimedia and Expo (ICME'2004), Taipei, Taiwan, 2004.

[8] 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.

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

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

[11] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'03)*, 2003, pp. 127–134.

[12] 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 Conf. on Computer Vision*, 2002, pp. 97–112.

[13] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'03)*, 2003, pp. 119–126.

[14] 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, pp. 408–415.

[15] Y. Lu, H. Zhang, W. Liu, and C. Hu, "Joint semantics and feature based image retrieval using relevance feedback," *IEEE Transactions on Multimedia*, vol. 5, no. 3, pp. 339–347, 2003.

[16] 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.

[17] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Past, present, and future," *Journal of Visual Communication and Image Representation*, vol. 10, pp. 1–23, 1999.

[18] 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.

[19] E. A. El-Kwae and M. 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.

[20] Z. Jin, I. King, and X. Li, "Content-based image retrieval by relevance feedback," in *Advances in Visual Information Systems, 4th international Conference, VISUAL 2000*, Lyon, France, 2000, vol. 1929 of *Lecture Notes in Computer Science*, pp. 530–539, Springer.

[21] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "An effective region-based image retrieval framework," in *Proceedings of the tenth ACM international conference on Multimedia*, 2002, pp. 456–465.

[22] 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.

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

[24] 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.

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

[26] N. Sebe and M. S. Lew, "Robust color indexing," in *Proceedings of the seventh ACM international conference on Multimedia (MM'99)*, 1999, pp. 239–242.

[27] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[28] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recogoition Journal*, vol. 29, pp. 1233–1244, 1996.

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

[30] M. Mandal, T. Aboulnasr, and S. Panchanathan, "Image indexing using moments and wavelets," *IEEE Transactions on Consumer Electronics*, vol. 42, no. 3, pp. 557–565, 1996.

[31] A. K. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

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

[33] 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.

[34] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill Book Company, 1983.

[35] Y. Rui, T. S. Huang, and S. Mehrotra, "Relevance feedback techniques in interactive content-based image retrieval," in *Storage and Retrieval for Image and Video Databases (SPIE)*, 1998, pp. 25–36.

[36] Y. Lu, C. Hu, X. Zhu, H. Zhang, and Q. Yang, "A unified framework for semantics and feature based relevance feedback in image retrieval systems," in *Proceedings of the eighth ACM international conference on Multimedia (MM'00)*, 2000, pp. 31–37.

[37] Y. Wu and A. Zhang, "Adaptive pattern discovery for interactive multimedia retrieval," in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*, 2003, vol. 2, pp. 649–655.

[38] 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.

[39] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "MindReader: Querying databases through multiple examples," in *Proceedings of 24th International Conference on Very Large Data Bases (VLDB'98)*, 1998, pp. 218–227.

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

[41] 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.

[42] Y. Rui and T. S. Huang, "A novel relevance feedback technique in image retrieval," in *Proc. ACM Multimedia (MM'99)*, Orlando, Florida, USA, 1999, pp. 67–70.

[43] P. Hong, Q. Tian, and T. S. Huang, "Incorporate support vector machines to content-based image retrieval with relevant feedback," in *Proceedings of IEEE International Conference on Image Processing (ICIP'00)*, Vancouver, BC, Canada, 2000, vol. 3, pp. 750–753.

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

[45] K. Tieu and P. Viola, "Boosting image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, South Carolina, USA, 2000, pp. 228–235.

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

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

[48] I. King and J. Zhong, "Integrated probability function and its application to content-based image retrieval by relevance feedback," *Pattern Recognition*, vol. 36, no. 9, pp. 2177–2186, 2003.

[49] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "Relevance feedback in region-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 672–681, 2004.

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

[51] C. H. Hoi, C. H. Chan, K. Huang, M. R. Lyu, and I. King, "Biased support vector machine for relevance feedback in image retrieval," in *Proceedings International Joint Conference on Neural Networks (IJCNN'04)*, 2004.

[52] Y. Rui and T. S. Huang, "Optimizing learning in image retrieval," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, South Carolina, USA, 2000, pp. 236–243.

[53] C. H. Hoi and M. R. Lyu, "Optimizing learning with svm constraint for content-based image retrieval," *submitted to IEEE Trans. on Neural Networks*, 2004.

[54] M. Nakazato, C. Dagli, and T. S. Huang, "Evaluating group-based relevance feedback for content-based image retrieval," in *Proceedings of International Conference on Image Processing (ICIP'2003)*, Spain, 2003, vol. 2, pp. 599–602.

[55] X. Zhou and T. S. Huang, "Comparing discriminating transformations and svm for learning during multimedia retrieval," in *Proc. the 9th ACM Multimedia*, 2001, pp. 137–146.

[56] C. H. Hoi and M. R. Lyu, "Group-based relevance feedback using support vector machine ensembles," in *Proceedings 17th International Conference on Pattern Recognition (ICPR'04)*, Cambridge, UK, 2004.

[57] C. H. Hoi and M. R. Lyu, "A novel log-based relevance feedback in content-based image retrieval," in *Proceedings ACM Intl. Conf. on Multimedia (MM'04)*, New York, USA, 2004.

[58] C. H. Hoi and M. R. Lyu, "Web image learning for searching semantic concepts in image databases," in *Poster Proc. 13th International World Wide Web Conference (WWW'2004)*, 2004.

[59] C. Buckley and G. Salton, "Optimization of relevance feedback weights," in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, 1995, pp. 351–357.

[60] 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.

[61] J. Rocchio, *Relevance Feedback in Information Retrieval*, Prentice-Hall, Inc., 1971.

[62] J. Rocchio, "Relevance feedback information retrieval," in *The Smart retrieval system: experiments in automatic document processing*, G. Salton, Ed. 1971, pp. 313–323, Prentice-Hall.

[63] K. Porkaew, M. Ortega, and S. Mehrotra, "Query reformulation for content based multimedia retrieval in MARS," in *Proceedings of International Conference on Multimedia Computing and Systems (ICMCS'99)*, 1999, vol. 2, pp. 747–751.

[64] R. E. Schapire, Y. Singer, and A. Singhal, "Boosting and rocchio applied to text filtering," in *Proceedings of the Twenty-first Annual International ACM SIGIR Conference on Information Retrieval (SIGIR'98)*.

[65] T. Wang, Y. Rui, and S.-M. Hu, "Optimal adaptive learning for image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001, pp. 1140–1147.

[66] T. Wang, Y. Rui, S.-M. Hu, and J. Sun, "Adaptive tree similarity learning for image retrieval," *Multimedia Systems*, vol. 9, no. 2, pp. 131–143, 2003.

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

[68] N. Vasconcelos and A. Lippman, "Learning from user feedback in image retrieval systems," in *Advances in Neural Information Processing Systems (NIPS'99)*, 1999.

[69] I. J. Cox, M. L. Miller, T. P. Minka, T. Papathomas, and P. N. Yianilos, "The bayesian image retrieval system, pichunter: Theory, implementation and psychophysical experiments," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 20–37, January 2000.

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

[71] 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, pp. 63–67.

[72] Y. Wu, Q. Tian, and T. S. Huang, "Discriminant-em algorithm with application to image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, South Carolina, USA, 2000, pp. 222–227.

[73] K. Sia and I. King, "Relevance feedback based on parameter estimation of target distribution," in *Proceedings of International Joint Conference on Neural Networks (IJCNN'2002)*, 2002, pp. 1974–1979.

[74] 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.

[75] L. Zhang, F. Z. Lin, and B. Zhang, "Support vector machine learning for image retrieval," in *IEEE International Conference on Image Processing (ICIP'2001)*, 2001, vol. 2, pp. 721–724.

[76] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1999.

[77] P.-H. Chen, C.-J. Lin, and B. Scholkopf, "A tutorial on nu-support vector machines," in *http://www.csie.ntu.edu.tw/ cjlin/papers/nusvmtutorial.pdf*, 2003.

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

[79] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Constructing boosting algorithms from svms: an application to one-class classification," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

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

[81] G. Ratsch, S. Mika, B. Scholkopf, and K.-R. Muller, "Constructing boosting algorithms from svms: an application to one-class classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1184–1199, 2002.

[82] G. Ratsch, S. Mika, B. Scholkopf, and K.-R. Muller, "Constructing boosting algorithms from svms: an application to one-class classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1184–1199, 2002.

[83] Y. Chen, X. Zhou, and T. S. Huang, "One-class svm for learning in image retrieval," in *Proceedings of IEEE International Conference on Image Processing (ICIP'01)*, Thessaloniki, Greece, 2001, pp. 34–37.

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

[85] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.

[86] 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`.

[87] K. Huang, H. Yang, I. King, M. R. Lyu, and L. Chan, "Biased minimax probability machine for medical diagnosis," in *Proceedings the Eighth International Symposium on Artificial Intelligence and Mathematics (AMAI'2004)*, Fort Lauderdale, Florida, 2004.

[88] 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*, vol. 13, no. 1, 2003.

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

[90] C.-C. Chang and C.-J. Lin, "Training nu-support vector classifiers: Theory and algorithms," *Neural Computation*, vol. 13, no. 9, pp. 2119–2147, 2001.

[91] S. Canu and Y. Grandvalet, "Adaptive scaling for feature selection in svms," in *Advances in Neural Information Processing Systems (NIPS'03)*, 2003.

[92] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for svms," in *Advances in Neural Information Processing Systems (NIPS'01)*, pp. 668–674.

[93] O. Chapelle, J. Weston, and B. Schölkopf, "Cluster kernels for semi-supervised learning," in *Advances in Neural Information Processing Systems (NIPS'03)*, 2003.

[94] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, and S. Y. Bang, "Pattern classification using support vector machine ensemble," in *Proceeding of 16th International Conference on Pattern Recogntion (ICPR'02)*, 2002, pp. 160–163.

[95] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[96] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, and S. Y. Bang, "Support vector machine ensemble with bagging," in *Proceeding of International Workshop on Pattern Recognition with Support Vector Machine*, 2002, pp. 397–407.

[97] R. S. Y. Freund, "Experiments with a new boosting algorithm," in *Proceedings The Thirteen International Conference on Machine Learning (ICML'96)*, 1996, pp. 148–156.

[98] K. Goh, E. Chang, and K. T. Cheng, "Svm binary classifier ensembles for multi-class image classification," in *ACM International Conference on Information and Knowledgement Management (CIKM)*, 2001, pp. 395–402.

[99] R. Yan, Y. Liu, R. Jin, and A. Hauptmann, "On predicting rare classes with svm ensembles in scene classification," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2003)*, 2003, vol. 3, pp. 21–24.

[100] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[101] C. Zhang and T. Chen, "An active learning framework for content-based information retrieval," *IEEE Transactions on Multimedia*, vol. 4, no. 2, pp. 260–268, 2002.

[102] A. Dong and B. Bhanu, "Active concept learning for image retrieval in dynamic databases," in *Proc. 9th International Conference on Computer Vision (ICCV'03)*, Nice, France, 2003, pp. 90–95.

[103] T. Onod, H. Murata, and S. Yamada, "Relevance feedback with active learning for document retrieval," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN'03)*, 2003, pp. 1757–1762.

[104] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines," in *Proc. 17th International Conf. on Machine Learning (ICML'00)*, 2000, pp. 839–846.

[105] R. Yan, J. Yang, and A. Hauptmann, "Automatically labeling data using multi-class active learning," in *Proc. 9th International Conference on Computer Vision (ICCV'03)*, Nice, France, 2003, pp. 516–523.

[106] 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, pp. 325–332.

[107] P. Anick, "Using terminological feedback for web search refinement: a log-based study," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (SIGIR'03)*, 2003, pp. 88–95.

[108] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma, "Query expansion by mining user logs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 829–839, 2003.

[109] J. Xu and W. B. Croft, "Query expansion using local and global document analysis," in *Proc. the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96)*, 1996, pp. 4–11.

[110] 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.

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

[112] 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*, vol. 13, no. 7, pp. 1443–1471, 2001.

[113] K. Barnard and D. Forsyth, "Learning the semantics of words and pictures," in *Proc. International Conference on Computer Vision*, 2001, vol. 2, pp. 408–415.

[114] K. Yanai, "Web image mining toward generic image recognition," in *Poster Proc. 12th International World Wide Web Conference (WWW'2003)*, 2003.

[115] H. Cui, M.-Y. Kan, and T.-S. Chua, "Unsupervised learning of soft patterns for generating definitions from online news," in *Proc. 13th International World Wide Web Conference (WWW'04)*, 2004, pp. 90–99.

[116] A. Dong and B. Bhanu, "A new semi-supervised em algorithm for image retrieval," in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*, 2003, vol. 2, pp. 662–667.

[117] K. Bennett and A. Demiriz, "Semi-supervised support vector machines," in *Proc. Advances in Neural Information Processing Systems (NIPS'98)*, 1998, pp. 368–374.

[118] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proceedings The Twentieth Inter-*

national Conference on Machine Learning (ICML'2001), 2001, pp. 19–26.

[119] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *Proceedings The Twentieth International Conference on Machine Learning (ICML'2003)*, 2003, pp. 912–919.

[120] L. Wang, K. L. Chan, and Z. Zhang, "Bootstrapping svm active learning by incorporating unlabelled images for image retrieval," in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*, 2003, vol. 1, pp. 629–634.