A Semi-Supervised Active Learning Framework for Image Retrieval

Steven C. H. Hoi and Michael R. Lyu
Dept. of Computer Sci. & Eng. and Shun Hing Inst. of Advanced Eng.
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong S.A.R.
{chhoi, lyu}@cse.cuhk.edu.hk

Abstract

Although recent studies have shown that unlabeled data are beneficial to boosting the image retrieval performance, very few approaches for image retrieval can learn with labeled and unlabeled data effectively. This paper proposes a novel semi-supervised active learning framework comprising a fusion of semi-supervised learning and support vector machines. We provide theoretical analysis of the active learning framework and present a simple yet effective active learning algorithm for image retrieval. Experiments are conducted on real-world color images to compare with traditional methods. The promising experimental results show that our proposed scheme significantly outperforms the previous approaches.

1. Introduction

Image retrieval has attracted more and more research interests from several computer communities as the volumes of image data have grown rapidly in recent years. Content-based image retrieval (CBIR) is one of the most important and challenging research topics in this field [12]. It is well known that the main difficulty in CBIR is to bridge the semantic gap between low-level features and high-level semantic concepts. One feasible way to address this problem is through learning from the user's relevance feedback [9].

Many relevance feedback algorithms have been proposed for image retrieval in past years. Support vector machines (SVM) based approaches represent the state-of-the-art technique in image retrieval [14, 2]. One key step in relevance feedback is to prompt users to label the images. This is a burdensome task for users. Normally there are very few user-labeled images available at first. This is typically called *insufficient training data* problem in machine learning. Like many other supervised learning techniques, SVM inevitably suffers the problem although it enjoys excellent generalization performance. It is imperative to find a solution for solving the *insufficient training data* problem confronted by SVM based relevance feedback methods.

Although some recent studies have noticed that unlabeled images can be useful for the learning tasks in image retrieval [18], few schemes proposed for image retrieval that can exploit both labeled and unlabeled data effectively. One recent approach is to apply a transductive learning technique, i.e., Transductive SVMs (TSVM) [16]. Although TSVM showed positive results in some text classification tasks [5], finding the exact solution is NP-hard. Moreover, some study challenged that TSVM might not be so helpful from unlabeled data in theory and in practice [17].

Recently the idea of learning from both labeled and unlabeled data, i.e., semi-supervised learning (SSL), has attracted much attention in machine learning [11]. Some promising techniques have been proposed and often show a measure of improvement in typical classification tasks [19]. Although these semi-supervised learning techniques show some advantage for very few labeled data, they may not always outperform traditional supervised learning, e.g. SVMs, when the amount of labeled data is increased. Moreover, there is still no a clear answer about the generalization performance of these semi-supervised learning techniques.

In order to exploit the advantages of both the emerging semi-supervised learning techniques and the regular SVMs, we present a novel semi-supervised active learning framework by comprising a fusion of the two. The proposed framework is general, but the engaged semi-supervised learning technique in this paper is based on the Gaussian fields and harmonic functions approach proposed by Zhu et al. [19].

The rest of this paper is organized as follows. Section 2 gives the background of techniques related to this work, including SVMs, active learning, and semi-supervised learning. Section 3 describes our proposed framework, methodology, and algorithm. Section 4 presents the experimental results of our performance evaluation. Section 5 gives some related work and Section 6 sets out our conclusion.

2. Background

2.1. SVM and Active Learning

Support vector machine, the representative of large margin classifiers, enjoys a sound theoretical foundation based on Structural Risk Minimization [15]. It has achieved many successes in various empirical applications thanks to its superior generalization performance. Here introduces its basic concept and the version space concept for SVM based active learning.

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

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

Typically one can project the data from the original data space $\mathcal{X}$ to a higher dimensional feature space $\mathcal{F}$ via a Mercer kernel K [15], which can be represented as $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, where $\Phi(\cdot)$ is a mapping function $\Phi : \mathcal{X} \longmapsto \mathcal{F}$ and "$\cdot$" denotes an inner product. Therefore, the decision boundary of SVM with the kernel setting can be represented as: $f(x) = \mathbf{w} \cdot \Phi(\mathbf{x})$, where $\mathbf{w} = \sum_{i=1}^{l} \alpha_i \Phi(\mathbf{x}_i)$.

At this point, the concept of version space must be addressed, as it is critical to SVM based active learning. If there exists a set of hyperplanes that can linearly separate the training data in the feature space, this set of hyperplanes or hypotheses is called the version space [7]. This can be defined as $\mathcal{V} = \{f \in \mathcal{H} | y_i f(\mathbf{x}_i) > 0, i = 1, \ldots, l\}$ or $\mathcal{V} = \{\mathbf{w} \in \mathcal{W} | \|\mathbf{w}\| = 1, y_i(\mathbf{w} \cdot \Phi(\mathbf{x}_i)) > 0, i = 1, \ldots, l\}$ [14], where $\mathcal{H}$ is the set of possible hypotheses and $\mathcal{W}$ is the parameter space for the unit vectors $\mathbf{w}$. Note that there is a duality between $\mathcal{F}$ and $\mathcal{W}$, i.e., points in the feature space correspond to hyperplanes in the parameter space and vice versa [15].

Active learning, known as pool-based active learning, is an interactive learning technique designed to reduce the labour cost of labeling in which the learning algorithm can freely assign the unlabeled data instances to the training set. The basic idea is to select the most informative data instances for labeling by the users in the next learning round. In other words, the strategy of active learning is to select an optimal set of unlabeled data instances that minimizes the expected risk of the next round. Among the various active learning techniques, SVM based active learning is one of the most promising methods currently available [10, 14]. The key idea of the SVM based approach is to reduce the version space of SVM as much as possible so as to minimize the expected risk associated with the unseen data.

2.2. Semi-Supervised Learning

Semi-supervised learning, namely learning with labeled and unlabeled data, has attracted considerable research attention recently [11, 1]. One of the promising and competitive approaches is SSL by Gaussian fields and harmonic functions proposed by Zhu et al. [19]. The method belongs to the category of graph-based methods; this is a major family of semi-supervised learning techniques [1]. The basic idea of this technique is to construct a weighted graph with both labeled and unlabeled data and then formulating the learning problem as a Gaussian random field on the graph. The mean of the field is a harmonic function that can be efficiently computed via matrix methods. This model enjoys many beneficial properties compared with other approaches. For example, and importantly for the work described in this paper, class priors and the predictions of external classifiers by supervised learning can be consistently combined with the harmonic learning model to improve the overall performance. We introduce the basic concept and offer the major results of their work as follows.

Suppose there are l labeled data instances $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)$ and u unlabeled data instances $\mathbf{x}_{l+1}, \ldots, \mathbf{x}_{l+u}$, where $l \ll u$ typically. Let us assume the labels are binary in the first instance. We can construct a graph $G = (V, E)$, where the vertex set $V = L \cup U$, L and U are the sets of labeled and unlabeled data, respectively. The task of semi-supervised learning is to assign labels to the unlabeled set U. In order to formulate the model, an $n \times n$ weighted matrix W is constructed on the edges of the graph that encodes the similarity between the instances. For example, given any two data instances $\mathbf{x}_i$, $\mathbf{x}_j \in R^n$, the weight of the edge between these two instances can be given as $w_{ij} = exp(-\sum_{d=1}^{n} \frac{(x_{id} - x_{jd})^2}{\sigma_d^2})$, where x_{id} is the d-th component of the data vector $\mathbf{x}_i$, and σ_d is length scale parameter of each dimension.

The learning strategy is to find an optimal real-valued function $g : V \longmapsto \mathcal{R}$ on the graph G and then to employ the function g for assigning the labels. The function g on the graph is constrained to take the values $g(\mathbf{x_i}) = g_l(\mathbf{x_i}) = y_i$ on the labeled data for $i = 1, \ldots, l$. Many semi-supervised learning techniques assume a default principle that data points located closely normally share similar label information. Based on this principle, one can define a quadratic energy function:

$$E(g) = \frac{1}{2}\sum_{i,j} w_{ij}(g(\mathbf{x}_i) - g(\mathbf{x}_j))^2 \ . \qquad (1)$$

In order to provide a probability distribution on the function g, a Gaussian field $p_\beta(g) = \frac{e^{-\beta E(g)}}{Z_\beta}$ is applied, where β is an inverse temperature parameter, and Z_β is the partition function with $Z_\beta = \int_{g|L=g_l} exp(-\beta E(g))dg$. Then the function g can be solved by minimizing the energy function as follows:

$$g = argmin_{g|L=g_l} E(g) \ . \tag{2}$$

According to graph theory, the resulting function g enjoys the *harmonic* property, i.e., it satisfies $\Delta g = 0$ on unlabeled data U, and is same to g_l on the labeled data L. Here, Δ is the *combinatorial Laplacian* which is given by $\Delta = D - W$, where $D = diag(d_i)$ is the diagonal matrix containing the entries $d_i = \sum_j w_{ij}$ and W is the weight matrix.

In order to represent the harmonic solution in terms of matrix operations, let $P = D^{-1}W$, and then split the matrices W, D and P into 4 blocks similar to the following example:

$$W = \begin{bmatrix} W_{ll} & W_{lu} \\ W_{ul} & W_{uu} \end{bmatrix} . \tag{3}$$

Let $g = \begin{bmatrix} g_l \\ g_u \end{bmatrix}$, where g_u is the values of g for the unlabeled data; then the harmonic solution to the function g_u can be represented as follows:

$$g_u = (D_{uu} - W_{uu})^{-1} W_{ul} g_l = (I - P_{uu})^{-1} P_{ul} g_l \ . \tag{4}$$

When this harmonic function g_u is solved, for each unlabeled data point $\mathbf{x}_i$, $sign(g(\mathbf{x}_i)) = 1$ if $g(\mathbf{x}_i) \geq 0.5$ and $sign(g(\mathbf{x}_i)) = 0$ otherwise.

3. Semi-Supervised Active Learning

3.1. Overview of Our Proposed Framework

In the discussion above, we introduced state-of-the-art methodologies in supervised and semi-supervised learning. The goal of our work in this paper is to combine these two into a unified semi-supervised active learning framework for image retrieval. We employ a proportion of unlabeled images in the learning tasks in order to attack the problems of there being insufficient training data. We describe our proposed framework as follows.

Our strategy in this framework is that we first employ SVM to learn a rough decision boundary based on the labeled data instances. As this is a supervised learning method, the learning procedure can be done quickly. The unlabeled images can then be given by a rough real-valued labels by computing their distances from the SVM decision boundary. In the second phase, we employ the Gaussian fields and harmonic functions-based Semi-Supervised Learning (SSL) technique to smooth the labels so as to improve the classification performance.

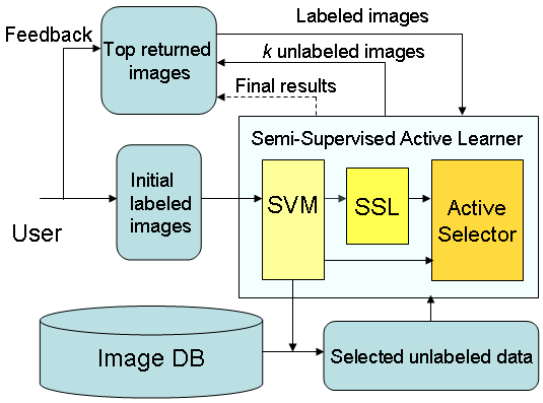


Figure 1. The architecture of our proposed framework

The architecture model of our proposed framework is shown in Fig. 1. First of all, a user provides a set of initial labeled images. The number of labeled images is very few but must include at least one positive and one negative instance. In the initial round, in order to increase the number of positive samples, we combine the SVM and SSL approach to train a classifier on both labeled and unlabeled data, and return to the user the most relevant results. In the subsequent feedback rounds, an active learner is launched by applying a semi-supervised active learning algorithm based on SVM and SSL by Gaussian fields and harmonic functions. Then, in the final round, the learning system trains a classifier combining SVM and the SSL on both the labeled and unlabeled data, and returns the most relevant results. It is important to indicate that only a small portion of unlabeled data is selected for the learning task. The motivation and selection strategy are discussed in the subsequent section.

3.2. Formulation and Theoretical Analysis

We first formulate the semi-supervised classifier by fusing the SVM and SSL based on Gaussian fields and harmonic functions, and then analyze the associated active learning technique. Suppose an SVM classifier f is trained on the given labeled data. Let $d_{SVM}(\mathbf{x}_i)$ denote the distance from a data instance $\mathbf{x}_i$ to the decision boundary of SVM. In order to normalize the distance metric to probability label metric within $[0, 1]$, a Sigmoid function is employed for fitting the probability similar to [8]:

$$f_u(\mathbf{x}_i) = \frac{1}{1 + \exp(-A \times d_{SVM}(\mathbf{x}_i))}, \tag{5}$$

where A is a positive constant that can be estimated according to training data.

Now let us combine the SVM prediction results into the harmonic energy minimization function. Following the suggestion in [19], we attach a "dongle" node for each unlabeled data point. Each dongle node $\mathbf{x}_i$ is assigned a value based on $g_u(\mathbf{x}_i)$. The transition probability from $\mathbf{x}_i$ to its dongle is given by ρ, while all other transitions from the node $\mathbf{x}_i$ have a probability of $1-\rho$. Here, ρ is introduced as a coupling factor to fuse two kinds of information. Based on this modified graph, we can solve the harmonic energy minimization problem in the usual way. The harmonic solution on this enhanced graph can be shown in the following form:

$$g_u = (I - (1-\rho)P_{uu})^{-1}((1-\rho)P_{ul}g_l + \rho f_u)\ . \quad (6)$$

Then, the fused harmonic solution above can be offered as the final relevance evaluation F, namely $F = g_u$.

Based on the fusion of SVM and SSL, let us now focus on the analysis of the involved active learning. We propose to perform active learning through optimally selecting the unlabeled data to minimize the risk on both the SVM classifier and the harmonic energy minimization functions. Let D be the given dataset, D_l^i and D_u^i be the labeled and unlabeled data sets respectively in the i-th active learning round, and D_k be the set of the k selected unlabeled instances. For simplicity, let D^i denote the learning data in the i-th round including D_l^i and D_u^i. The goal of semi-supervised active learning is to choose an optimal subset D_k^* that can minimize the risk of the fused relevance evaluation function F, namely

$$\begin{aligned} D_k^* &= arg\min_{D_k \subset D_u^i} \mathcal{R}(F^{D^{i+1}}) \\ &= arg\min_{D_k \subset D_u^i} \mathcal{R}(F^{(D_l^i + D_k)\vee(D_u^i - D_k)})\ . \end{aligned} \quad (7)$$

In theory, solving this minimization problem leads directly to the optimal selection set. Unfortunately, this is often a retraining problem that is computationally intensive and so impractical for image retrieval. Hence, we have to make some approximation for the optimization in practice. First, the risk can be approximated by decomposing it into two components: SVM and harmonic functions. Taking consideration on the coupling factor, we can obtain the first approximated optimal subset $\widehat{D}_k^*$ as follows

$$\widehat{D}_k^* = arg\min_{D_k \subset D_u^i} \left\{ \rho\mathcal{R}(f^{D^{i+1}}) + (1-\rho)\mathcal{R}(g^{D^{i+1}}) \right\}\ . \quad (8)$$

It is still intractable in practice, so we have to make a further approximation by reducing k to 1, namely to select only one unlabeled instance each time and repeat the operation for selecting k targets in each round. This can be represented as

$$\mathbf{x}_k^* = arg\min_{\mathbf{x}_k \in D_u^i} \left\{ \rho\mathcal{R}(f^{+\mathbf{x}_k}) + (1-\rho)\mathcal{R}(g^{+\mathbf{x}_k}) \right\} \quad (9)$$

where $\mathcal{R}(f^{+(\mathbf{x}_k, y_k)})$ represents the risk of SVM classifier after adding a new labeled instance $(\mathbf{x}_k, y_k)$. However, the retraining problem still exists. Thus, we have to eliminate the components that require retraining after adding new data or finding efficient ways for the retraining problem. We attack this difficulty in two ways.

For dealing with the risk of SVM, no very efficient way is available for the retraining problem. Fortunately, many previous studies have shown that there are some approximated approaches that can solve the problem very effectively [10, 14]. One of the most popular and effective approaches is to choose the instances closest to the decision boundary. This reduces the version space as much as possible so as to reduce the overall risk greatly. Theoretical support for this heuristic yet effective strategy has already been published in previous work [10, 14, 16].

On the other hand on, for the risk of harmonic functions, there is an efficient retraining way available in [20]. We here offer the main results as follows. Let $p^*(y_k|D_l)$ be the unknown true label distribution for node $\mathbf{x}_k$ given the labeled data. As the graph is based on Gaussian field model, it is reasonable to assume $p^*(y_k = 1|D_l) \approx g_k$, where g_k is the probability to reach label "1" in a random walk on the graph. Then the approximated risk of harmonic function $\widehat{\mathcal{R}}(g^{+\mathbf{x}_k})$ can be given as

$$\widehat{\mathcal{R}}(g^{+\mathbf{x}_k}) = (1-g_k)\mathcal{R}(g^{+(\mathbf{x}_k,0)}) + g_k\mathcal{R}(g^{+(\mathbf{x}_k,1)})\ , \quad (10)$$

where the estimated risk $\mathcal{R}(g^{+(\mathbf{x}_k,y_k)})$ can be computed based on harmonic concepts as follows

$$\mathcal{R}(g^{+(\mathbf{x}_k,y_k)}) = \sum_{k=1}^{m} \min(g_k^{+(\mathbf{x}_k,y_k)}, 1 - g_k^{+(\mathbf{x}_k,y_k)})\ . \quad (11)$$

The remaining question now is how to compute the harmonic function $g^{+(\mathbf{x}_k,y_k)}$ after adding the labeled data $(\mathbf{x}_k, y_k)$. The work in [20] has shown this can be computed efficiently by the following equation:

$$g_u^{+(\mathbf{x}_k,y_k)} = g_u + (y_k - g_k)\frac{(\Delta_{uu}^{-1})_{\cdot k}}{(\Delta_{uu}^{-1})_{kk}} \quad (12)$$

where $(\Delta_{uu}^{-1})_{\cdot k}$ is the k-th column of the inverse Laplacian on unlabeled data, and $(\Delta_{uu}^{-1})_{kk}$ is the k-th diagonal element of the matrix Δ_{uu}^{-1}. Both of them are available when computing the harmonic function g in the i-round semi-supervised learning.

3.3. A Practical Active Learning Algorithm

Although we can implement the algorithm by following the framework and formulation proposed previously, there are some practical problems that must be considered when designing an active learning algorithm for image retrieval.

SVM-SSAL Algorithm:
For the initial or last feedback rounds:
1) Learn an SVM on the available labeled images ;
2) Select u unlabeled samples with maximal d_{SVM} ;
3) Perform SSL enhanced by SVM results in Eq. (6) ;
4) Output top-k images with maximal g_u from Eq. (6) .
For each of other feedback rounds:
1) Learn an SVM on the available labeled images ;
2) Select u unlabeled samples with minimal $|d_{SVM}|$;
3) Perform SSL by Eq. (4) ;
4) Compute the risk by Eq. (9) ;
5) Output top-k images with minimal risk from Eq. (9) .

Figure 2. The summary of SVM-SSAL algorithm

As we know, active learning in image retrieval is an interactive procedure; hence, the response time is an important issue for the learning system. It is important to balance the classification performance and speed of response when designing and running the algorithm. In general, the computational cost of semi-supervised learning is proportional to the amount of unlabeled data. It is not practical to engage all of unlabeled images for the learning task due to the fast response requirement.

Hence, it is critical to choose the most valuable unlabeled data in the learning task. Our proposed strategy for selecting the unlabled data is based on the SVM and active learning theory. Fig. 2 summarizes our suggested SVM-SSAL algorithm.

In above algorithm, unlabeled data are chosen to the learning task by two ways. First, in the first and last relevance feedback rounds, we select the unlabeled data with large SVM distances; in other feedback rounds, we choose unlabeled data closest to the decision boundary of SVM, when they are required for the semi-supervised active learning phase. Our strategy is based on two assumptions discussed previously: (1) an instance with a larger SVM distance will be more relevant than a smaller one; and (2) instances closest to the decision boundary of SVM will be more informative for active learning. Other minor ways of enhancing the performance in our suggested algorithm will be discussed in the next section.

4. Experimental Results

4.1. Datasets

To conduct empirical evaluation of our proposed framework, we pick real-world images from the COREL image CDs. There are two image datasets used 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 have different semantic meanings, such as *antique*, *balloon*, *car*, and *lizard* et al. The rationale for the selection of the semantic categories is as follows. First, it enables us evaluate whether the approach can not only retrieve images that are visually similar but also the images that are relevant semantically after learning with users' feedback. Secondly, the approach can enables us to evaluate the performance automatically, reducing subjective errors arising from manual evaluations by different people.

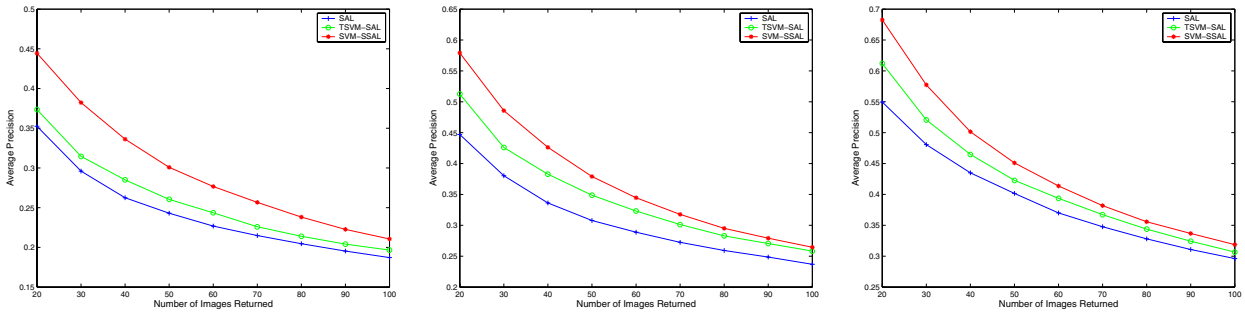
4.2. Image Representation

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

Color features are widely adopted in CBIR on account of their simplicity. The color feature employed in our experiments is color moment since it is close to natural human perception; many previous research studies have shown the effectiveness of color moment applied in CBIR. For the employed color moment, we extract 3 moments: color mean, color variance and color skewness in each color channel (H, S, and V), respectively. Thus, a 9-dimensional color moment is adopted as the color feature in our experiments.

Edge features can be very effective in CBIR when the contour lines of images are evident. The edge feature used in our experiments is the edge direction histogram [3]. The images in the datasets are first translated to gray images. Then a Canny edge detector is applied to obtain the edge images. From the edge images, the edge direction histogram can then computed. The edge direction histogram is quantized into 18 bins of 20 degrees each; hence an 18-dimensional edge direction histogram is employed to represent the edge feature.

Texture feature is proven to be an important cue for image retrieval. In our experiments, we employ the wavelet-based texture technique [13, 6]. The original color images are first transformed to gray images. Then we perform the Discrete Wavelet Transformation (DWT) on the gray images employing a Daubechies-4 wavelet filter [13]. 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 decomposition and obtain 10 subimages in different scales and orientations. One of the 10 subimages is a subsampled average image of the original image; this is discarded, since it contains less useful texture information. For the other 9 subimages, we compute the entropy of each subimage separately. Therefore, we obtain a



(a) Round 1 (b) Round 2 (c) Round 3

Figure 3. Experimental results on the 20-Cat dataset.

9-dimensional wavelet-based texture feature to describe the texture information for each image.

4.3. Performance Evaluation

We compare our proposed semi-supervised active learning (SVM-SSAL) algorithm with two previous well-known active learning algorithms: SVM active learning (SAL) proposed by Tong et al. [14], and Transductive SVM based active learning (TSVM-SAL) proposed by Wang et al. [16]. A lot of previous studies have shown the SVM active learning approaches to be effective and beneficial for image retrieval.

For all three compared schemes, we employ the same experimental settings to enable objective evaluation. The software used for training SVM and TSVM is SVM^{light} [4]. The kernel function employed for the SVMs is Gaussian RBF, $K(\mathbf{x}_i, \mathbf{x}_j) = exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$; γ is fixed according to the training samples in the experiments [16]. The regularization parameter, C, is fixed at 100.

Initially 10 images are presented to the user for assignment of labels; the user must assign at least one positive and one negative label. After obtaining the labeled images from the user, we perform the active learning algorithms by employing the three schemes separately. For each scheme, three rounds of relevance feedback are conducted in total and 10 images are presented to users in each round. In order to evaluate the performance from each round, we output the retrieval results (the top 100 most relevant images) after each round in each scheme. Because of intensive computation load it would incur, it is impractical (and unnecessary) to incorporate all the unlabeled data in our learning scheme. Therefore, in our experiments, 5% of the unlabeled images from the image dataset are engaged for our active learning algorithm. For transductive SVM learning, we incorporate 2.5% of the images since we found the performance is worse when engaging more images in this approach. We conduct the experimental evaluation automatically; the user's relevance feedbacks are simulated automatically based on the ground truth data. The evaluation metric in our experiments is based on *Average Precision* that is defined as the correct retrieved images over the total returned images. In total, we perform 100 executions on each of the two datasets and obtain the comparison results shown in Fig. 3 and Fig. 4.

Table 1. Average precision of top-20 images for different initial labeled images on 20-Cat dataset

#Labels	SAL	TSVM-SAL	SVM-SSAL
5	0.267	0.300 (+12.36%)	0.349 (+30.52%)
10	0.353	0.374 (+5.96%)	0.445 (+26.10%)
15	0.439	0.460 (+4.90%)	0.517 (+17.90%)
20	0.507	0.529 (+4.44%)	0.557 (+9.87%)
25	0.531	0.548 (+3.20%)	0.582 (+9.60%)
30	0.576	0.586 (+1.74%)	0.616 (+7.04%)
MAP	0.445	0.466 (+5.43%)	0.511 (+16.84%)

From the experimental results, we observe that enhancing SVM active learning by Tranductive SVM can only have limited improvement on the retrieval performance. In contrast, our proposed semi-supervised active learning algorithm significantly improves the learning performance. For example, in the 20-Cat dataset, at the first round, our proposed algorithm yielded 26% improvement over the regular SVM active learning on the top 20 returned images and an average improvement of 21% on the returned images. Our algorithm also outperformed the TSVM-SAL approach 19% improvement on the top 20 returned images and an average improvement of 14% on the returned images. Similar improvements can also be observed in the second and third round. In the 50-Cat dataset, we also measured a significant improvement. On average, our algorithm outperformed the traditional SAL scheme at least 25% and the TSVM-SAL method by at least 15%.

Moreover, we are also interested to evaluate the retrieval performance under different numbers of initial labeled images. Table 1 and Table 2 show the average precision of top-20 returned images for learning with different numbers of initial labeled images on 20-Cat and 50-Cat respectively. From the table of 20-Cat dataset, we can observe our proposed algorithm outperformed the regular SVM approach by about 16% improvement, while the transductive SVM approach only yielded an 5% improvement. Similarly, on the 50-Cat dataset, our algorithm also achieved a significant improvement compared with the transductive SVM approach. It is interesting to note that we recorded some negative improvement with the transductive SVM approach when engaging larger amount of unlabeled data. From Table 1 and Table 2, we also found the improvement of our algorithm is decreased when the number of initial labeled images is increased. This phenomenon supports our assumption that SVM can learn better when engaging more labeled data; hence, there is less scope for improvement. Nevertheless, our algorithm still yields considerable improvement compared with SVM even for many labeled images.

4.4 Discussion

We studied the effect of employing more unlabeled data in the transductive SVM model. The results showed this model to be sensitive to the amount of unlabeled data resulted in an improvement that was smaller or even negative in some situations. In contrast, for our SVM-SSAL algorithm, the improvements are always increased when engaging more unlabeled images. One of the reasons for this is that the prediction results of SVM are very compatible with the harmonic functions and can then be fused consistently. In the experiments, we achieve the promising results above simply by setting the coupling factor ρ to 0.5. Even greater improvements can be obtained when tuning better coupling factors.

Some questions remain for further study of our proposed framework. First of all, albeit that our semi-supervised active learning algorithm is effective, it still involves some greedy approximations. We would like find an effective algorithm that can reduce or eliminate these. Secondly, although we have shown that fusing the prediction results of SVM with the harmonic function is compatible using the graph perspective and the empirical experimental results, no theoretical connection between two kinds of classifiers is available yet. In our current approach, we fuse them by means of a coupling factor ρ. Finding the best way of optimizing this coupling factor is also an open question. Moreover, the response time is critical for relevance feedback in image retrieval, learning with the unlabeled data is a tradeoff between performance and efficiency. Fortunately, the engaged semi-supervised learning algorithm in our framework is quite efficient. We will study the analysis of computation cost and quantitative evaluation of efficiency in our future work.

Table 2. Average precision of top-20 images for different initial labeled images on 50-Cat dataset

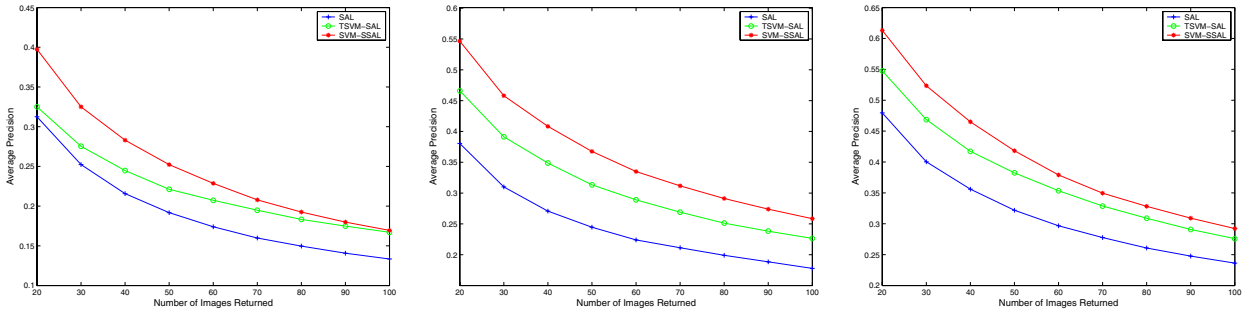
#Labels	SAL	TSVM-SAL	SVM-SSAL
5	0.250	0.255 (+2.20%)	0.314 (+25.85%)
10	0.313	0.325 (+4.00%)	0.398 (+27.20%)
15	0.383	0.391 (+2.22%)	0.457 (+19.35%)
20	0.426	0.453 (+6.22%)	0.501 (+17.61%)
25	0.477	0.498 (+4.41%)	0.531 (+11.44%)
30	0.500	0.529 (+5.70%)	0.549 (+9.80%)
MAP	0.391	0.408 (+4.13%)	0.458 (+18.54%)

5. Related Work

The work in this paper is based on several significant reports of recent years. The most important work related to our paper is the semi-supervised learning by Gaussian fields and harmonic functions approach proposed by Zhu et al. [19]. They proposed an elegant semi-supervised learning technique and also presented an active learning scheme based on Gaussian fields and harmonic functions [20]. Another important related work is the SVM active learning scheme. Many studies have been done for SVM active learning [14, 10]. Also a noteworthy work is the transdutive SVM approach [16] that provides some theoretical analysis of SVM active learning. Recently, although some work has been done to incorporate unlabeled data in image retrieval [18], there are few schemes that can work very effectively. To our knowledge, our work is the first to combine supervised learning (e.g. SVM) and semi-supervised learning for active learning in image retrieval.

6. Conclusion

In this paper we have proposed a novel semi-supervised active learning framework for image retrieval. The suggested active learning scheme is based on the fusion of two different kinds of learning techniques, namely one is supervised another is semi-supervised. In the proposed algorithm, support vector machines and semi-supervised learning with Gaussian fields and harmonic functions are fused together. We have analyzed the motivation and approach of the proposed approximated active learning algorithm. To evaluate the performance of our suggested algorithm, detailed experiments have been conducted and significant improvements have been demonstrated over some other leading approaches. We believe the suggested semi-supervised active learning framework will be a significant tool for



(a) Round 1 (b) Round 2 (c) Round 3

Figure 4. Experimental results on the 50-Cat dataset.

learning in image retrieval and its associated ideas will also be applicable in other fields.

Acknowledgment

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK4182/03E), and partially supported by Shun Hing Institute of Advanced Engineering (SHIAE).

References

[1] A. Blum and S. Chawla. Learning from labeled and unlabeled data using graph mincuts. In *Proc. 18th ICML*, pages 19–26. Morgan Kaufmann, San Francisco, CA, 2001.

[2] C.-H. Hoi and M. R. Lyu. Group-based relevance feeedback with support vector machine ensembles. In *Proc. 17th ICPR*, volume 3, pages 874–877, Cambridge, UK, 2004.

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

[4] T. Joachims. Making large-scale svm learning practical. In *Advances in Kernel Methods - Support Vector Machines*. MIT Press, Cambridge, MA, 1999.

[5] T. Joachims. Transductive inference for text classification using support vector machines. In *Proc. 16th ICML*, pages 200–209, Bled, SL, 1999.

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

[7] T. Mitchell. Generalization as search. *Aritificial Intelligence*, 28:203–226, 1982.

[8] J. Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. *Advances in Large Margin Classifiers*, pages 61–74, 2000.

[9] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Trans. on CSVT*, 8(5):644–655, Sept. 1998.

[10] G. Schohn and D. Cohn. Less is more: Active learning with support vector machines. In *Proc. 17th ICML*, pages 839–846. Morgan Kaufmann, San Francisco, CA, 2000.

[11] M. Seeger. Learning with labeled and unlabeled data (Technical Report). University of Edinburgh, 2001.

[12] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. on PAMI*, 22(12):1349–1380, 2000.

[13] J. Smith and S.-F. Chang. Automated image retrieval using color and texture. *IEEE Trans. on PAMI*, Nov. 1996.

[14] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *Proc. 9th ACM Multimedia Conference*, pages 107–118, 2001.

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

[16] L. Wang, K. L. Chan, and Z. Zhang. Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval. In *Proc. CVPR*, volume 1, pages 629–634, 2003.

[17] T. Zhang and F. Oles. A probability analysis on the value of unlabeled data for classification problems. In *Proc. 17th ICML*, pages 1191–1198. San Francisco, CA, 2000.

[18] Z.-H. Zhou, K.-J. Chen, and Y. Jiang. Exploiting unlabeled data in content-based image retrieval. In *Proc. 15th ECML*, Pisa, Italy, 2004.

[19] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proc. 20th ICML*, 2003.

[20] X. Zhu, J. Lafferty, and Z. Ghahramani. Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In *ICML 2003 workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, 2003.