TELKOMNIKA, Vol. 11, No. 7, July 2013, pp. 3634 ~ 3640
e-ISSN: 2087-278X

Improving Relevance Feedback in Image Retrieval by Incorporating Unlabelled Images

Guizhi Li
Computer Center, Beijing Information and Science and Technology University, Beijing, China
e-mail: lgz403@sina.com

Abstract

In content-base image retrieval, relevance feedback (RF) schemes based on support vector machine (SVM) have been widely used to narrow the semantic gap between low-level visual features and high-level human perception. However, the performance of image retrieval with SVM active learning is known to be poor when the training data is insufficient. In this paper, the problem is solved by incorporating the unlabelled images into the learning process. We proposed a semi-supervised active learning algorithm which uses not only labeled training samples but also unlabeled ones to build better models. In relevance feedback, active learning algorithm is often used to reduce the cost of labeling by selecting only the most informative data. In addition, we introduced a semi-supervised approach which employed Nearest-Neighbor technique to label the unlabeled sample with a certain degree of uncertainty in its class information. Using these samples, Fuzzy support vector machine (FSVM) which takes into account the fuzzy nature of some training samples during its training is trained. We compared our method with standard active SVM on a database of 10,000 images, the experiment results show that the efficiency of SVM active learning can be improved by incorporating unlabelled images, and thus improve the overall retrieval performance.

Keywords*: relevance feedback, image retrieval, FSVM, semi-supervised learning*

1. Introduction

Content-based image retrieval (CBIR) has become one of the most active research areas in multimedia signal processing. It is well known that the performance of content-based image retrieval systems is mainly limited by the semantic gap between low-level features and high-level concepts. Retrieving images from a large database according to the semantic meanings is a challenging problem, due to the lack of mature artificial intelligence mapping semantic meanings to low-level descriptors.

Relevance feedback techniques have been widely used in CBIR to bridge the semantic gap between low-level features and high-level semantics [1]. Relevance feedback is an interactive technique in the procedure of CBIR, which is originally used in the traditional text-based information retrieval systems. A relevance feedback-based approach allows the user to label the returned images as relevant or irrelevant. Such labeled examples are further used to refine retrieval results by short-term learning or long-term learning techniques.

Many relevance feedback algorithms have been adopted in CBIR systems and demonstrated considerable performance improvement [1-3]. A popular relevance feedback method in CBIR is centered on SVM. However, conventional SVM use only labeled data for training. Labeled instances are often difficult, expensive, or time consuming to obtain, as they require the efforts of experienced human annotators. Therefore, the performance of relevance feedback methods is often constrained by insufficient training samples. To deal with this problem, some works have been done to incorporate the unlabeled data to improve the learning performance [4]. SVM-based active learning has been proposed to carefully select the unseen images that are closest to the SVM decision hyperplane as the most informative images for user feedback [5]. Semi-supervised learning aims to employ unlabeled data to enhance the learning process and improve the retrieval performance [6]. Both semi-supervised learning and active learning can take advantage of the unlabeled data. It is quite natural to combine them to form a more effective method [7].

Received January 22, 2013; Revised April 6, 2013; Accepted April 18, 2013

In this paper, we propose a relevance feedback technique based on FSVM using semi-supervised active learning algorithm for content-based images. SVM active learning is used to alleviate the burden of labeling by selecting only the most informative data. A semi-supervised approach has been developed which uses Nearest-Neighbor technique to label the unlabeled data with a certain degree of uncertainty in its class information. FSVM is trained based on these automatically labeled samples and user labeled samples.

The rest of this paper is organized as follows. The proposed FSVM-based relevance feedback scheme incorporating unlabelled images is presented in section 2. Our experimental results are given in Section 3. Finally, Section 4 gives some concluding remarks of this study.

2. Proposed Method

2.1. Overview of Our Proposed Framework

In this paper, we propose a unified framework by fusing both unsupervised learning and active learning FSVM 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.

Step 1: Given a query image, the system performs the K-NN search using the Euclidean distance for similarity matching. The top n most similar images are returned to the user for feedback.
Step 2: The user labels the n images as either relevant or irrelevant.
Step 3: Train an initial SVM classifier Based on the n labeled images.
Step 4: The SVM active learning is employed by selecting m unlabeled images that are closest to the current SVM separating hyperplane for the user to label.
Step 5: Add the m images to the labeled training set.
Step 6: Use Nearest-neighbor technique to select unlabeled images, assign a label and evaluate the relevance membership of each label in its class.
Step 7: Train a FSVM using a hybrid of the user labeled and automatic labeled images.
Step 8: Repeat Steps 4–Step7 until the user is satisfied with the retrieval results.

2.2. SVM and FSVM

SVM, proposed by Vapnik [8], is known as an excellent tool for classification and regression problems with a good generalization performance. Its formulation embodies the Structural Risk Minimization principle. By use of kernel function mapping technique, SVM can achieve good ability of classification generalization through small data learning. In the case of pattern recognition, it can be divided into linearly separable case, linearly non-separable case and non-linear case.

In the linearly separable case, consider the following binary classification task. Let $\Omega = \{(x_i, y_i) | i = 1, 2, \cdots N\} \subset R^m \times \{-1, +1\}$ be a set of training examples, where $x_i \in R^m, y_i \in \{-1, +1\}$, m being the dimension of the input space. The goal is to find a decision function $g(x) = \mathrm{sgn}(f(x))$ that accurately predicts the labels of unseen data (x, y), and minimizes the classification error. If $f(x)$ is a linear function:

$$f(x) = w \cdot x + b, \text{ for } w \in R^m, b \in R \tag{1}$$

Then this gives a classification rule whose decision boundary $\{x | f(x) = 0\}$ is a hyperplane separating the class "+1" and class "-1" from each other. The problem of learning from data can be formulated as finding a set of parameters (w, b) such that $\mathrm{sgn}(w \cdot x_i + b) = y_i$ for all $i \in [1, N]$. And the margin between two classes is:

$$\rho(w, b) = 2 / \|w\| \tag{2}$$

The optimal separating hyperplane is given by maximizing the margin. Thus the problem of classification becomes the following optimization problem:

Minimize $\Phi(w) = \frac{1}{2}\|w\|^2$ (3)

Subject to $y_i((w \cdot x_i) + b) \geq 1, i \in [1, N]$ (4)

This constrained optimization problem is solved by introducing Lagrange Multiplier $\alpha_i \geq 0$ and a Lagrangian. The decision function of the SVM is obtained:

$$f(x) = sgn(\sum_{i=1}^{N} \alpha_i \ y_i (x \cdot x_i) + b \ ) \quad (5)$$

Where x_i represents the support vectors (SVs).

For linearly non-separable cases, one can introduce slack variables $\xi \geq 0$. The constraint of (4) is modified to:

$$y_i((w \cdot x_i) + b) \geq 1 - \xi_i, i \in [1, N] \quad (6)$$

The generalized optimal separation hyperplane is determined by the vector w, which minimizes the following function:

$$\Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i \quad (7)$$

subject to the constraint of (6).

In the non-linear case where a linear boundary is inappropriate, SVMs can map input vector into a high dimensional feature space. By choosing a non-linear mapping, the SVMs construct an optimal separation hyperplane in this higher dimensional space. $K(x, y)$ is kernel function performing non-linear mapping into feature space. And the kernel version of classification function is given by:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \quad (8)$$

FSVM is an extension of SVM that takes into account the different significance of the training samples. For FSVM, each training sample is associated with a fuzzy membership value $\{u_i\}_{i=1}^{n} \in [0,1]$.The membership value u_i reflects the fidelity of the data; in other words, how confident we are about the actual class information of the data. The higher its value, the more confident we are about its class label. The optimization problem of the FSVM is formulated as follows [9]:

Minimize $\Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} u_i \xi_i$ (9)

Subject to $y_i((w \cdot x_i) + b) \geq 1 - \xi_i, i \in [1, N]$ (10)

It is noted that the error term ξ_i is scaled by the membership value u_i.The fuzzy membership values are used to weigh the soft penalty term in the cost function of SVM. The weighted soft penalty term reflects the relative fidelity of the training samples during training. Important samples with larger membership values will have more impact in the FSVM training than those with smaller values.

2.3. Nearest-Neighbor Based Unlabeled Image Selection

Nearest-neighbor techniques are effective in all applications where it is difficult to produce a high-level generalization of a class of objects. Relevance learning in content based image retrieval may well fit into this definition, as it is difficult to provide a general model that can be adapted to represent different concepts of similarity [10]. Nearest-neighbor produces a probability density model and attempts to find the basic formation of each class by modeling the data. In this work, we propose to exploit Nearest-neighbor generative model to select the unlabeled images for labeling that have a high probability of belonging to each class. This will enlarge the training data set.

Let us recall that the nearest neighbor (NN) classifier is derived from the local estimation of densities in the neighborhood of the test pattern. Such a local density can be written as:

$$p_{NN}(x) = \frac{1/n}{V(\|x - NN(x)\|)} \tag{11}$$

Where n is the number of training patterns, x is the test image, NN denotes the nearest neighbor of x, and V is the volume of the minimal hypersphere centered in x, that contains $NN(x)$ (i.e., the radius of the hypersphere is equal to $\|x - NN(x)\|$).Thus we can compute the local density of relevant images as:

$$p_{NN}^{+}(x) = \frac{1/n}{V(\|x - NN^{+}(x)\|)} \tag{12}$$

Where NN^{+} is the nearest relevant image of x .Analogously the local density of non-relevant images can be computed as:

$$p_{NN}^{-}(x) = \frac{1/n}{V(\|x - NN^{-}(x)\|)} \tag{13}$$

Where NN^{-} is the nearest irrelevant image of x .

These densities can be used to estimate fuzzy membership of an unlabeled image in positive class as:

$$u_{NN}^{+}(x) = \frac{p_{NN}^{+}}{p_{NN}^{+} + p_{nn}^{-}} = \frac{\|x - NN^{-}(x)\|}{\|x - NN^{+}(x)\| + \|x - NN^{-}(x)\|} \tag{14}$$

Analogously the degree of relevance of an unlabeled image in negative class is:

$$u_{NN}^{-}(x) = \frac{p_{NN}^{-}}{p_{NN}^{-} + p_{nn}^{+}} = \frac{\|x - NN^{+}(x)\|}{\|x - NN^{-}(x)\| + \|x - NN^{+}(x)\|} \tag{15}$$

We Select $n1$ unlabeled samples with high probability according to equation (14).These samples are labeled as relevant and added to the training data set. We select $n2$ unlabeled samples with high probability according to equation (15). These samples are labeled as irrelevant and added to the training data set. In addition, $n1$ and $n2$ values are set according to the distribution of the positive and negative samples in the labeled data.

2.4. Fuzzy Membership Estimation of the Selected Unlabeled Images

We employ a fuzzy membership function to estimate the class membership of the automatic labeled images. The fuzzy information is then integrated into the FSVM for active learning. Equation (14) and equation (15) give the fuzzy membership of the selected unlabeled image in each class. Further, the fuzzy membership function of the unlabeled samples will also depend on a measure of how well the assigned label agrees with the label determined by the trained SVM. A Sigmoid function is employed to estimate the probability. The degree of relevance of the automatic labeled image in each class is:

$$u^{+}_{SVM}(x) == \frac{1}{1+\exp(-A \times d_{SVM}(x))} \tag{16}$$

$$u^{-}_{SVM}(x) == \frac{1}{1+\exp(A \times d_{SVM}(x))} \tag{17}$$

where A is a positive constant that can be estimated according to training data, $d_{SVM}(x)$ denote the distance from the automatic labeled image x to the decision boundary of SVM. We combine these two measures to compute the final predicted weight of x as:

$$u^{+(-)}(x) = u^{+(-)}_{NN}(x) u^{+(-)}_{SVM}(x) \tag{18}$$

3. Experiment Results and Analysis

The image database used in the experiment contains 10,000 color images of 100 different categories obtained from the Corel Gallery product. In our system, we use two types of visual features: color and texture. Color histogram, color moments, and color auto-correlogram are used as the representation for color features. Gabor wavelet and wavelet moments are used as the texture features representation [11-12].

In the experiments, we select 100 query images, one from each category, and evaluate the retrieval quality for a sequence of iterations starting with these initial queries. We perform 5 feedback iterations in addition to the initial query. At each iteration, active SVM selects 10 images to ask the user to label as relevant or irrelevant, then semi-supervised learning is used to select $n1$ (which is set to 10) and $n2$ (which is computed based on $n1$) unlabeled samples which are respectively labeled as relevant and irrelevant and added to the training data set. After the relevance feedback iterations have finished, the top-k most relevant images were retrieved by SVM to evaluate the retrieval performance. All the measurements are averaged over 100 queries. Precision and recall curve is used to measure the retrieval performance. Precision is defined as the number of retrieved relevant images over the number of total retrieved images. Recall is defined as the number of retrieved relevant images over the total number of relevant images in the database.

We have evaluated the performance of the proposed semi-supervised FSVM (SSFSVM) algorithm and compared with regular active SVM. Figure 1 illustrates the precision-recall graphs of the two methods after the last iteration. Each line is plotted with 10 points, each of which shows precision and recall as the number of retrieved images increases from 10 to 100. When the value is 100, since the number of retrieved images is equal to the number of relevant images, the value of precision and recall are the same. From the graph, we can observe that the proposed method outperforms the regular active SVM method in both cases. Our method provides higher recall rate at the same precision level, and higher precision rate at the same recall level.

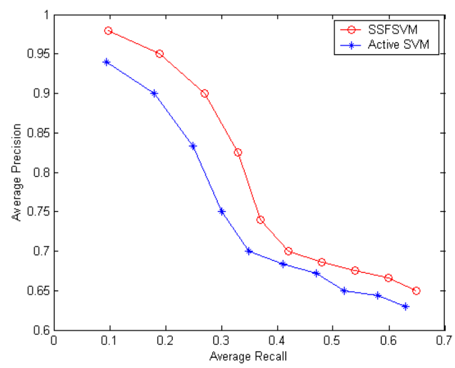

Figure 1. Average precision-recall graphs

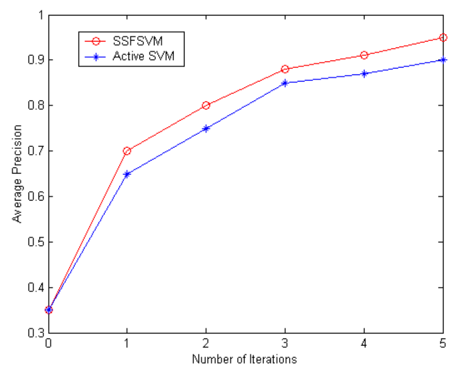

Figure 2. Precision graphs over the number of iterations

Figure 2 shows the precision graphs of the two approaches over the different iterations. From the graph, we can observe that they produce the same precision for the initial query. Then the precision of our method increase at each iteration and outperform those of the active SVM approach. The improvement mainly lies in the use of unlabeled data for effective learning.

Figure 3 illustrates the precision over the different number of retrieved images. According to Figure 3, the precision decreases with the increase of the number of retrieved images. This is because with the increase of the number of retrieved images, more and more images that are irrelevant to the query are involved in the retrieved images, which have negative impact on the precision. The graphs show that the retrieval performance of the proposed method is superior to the active SVM approach.

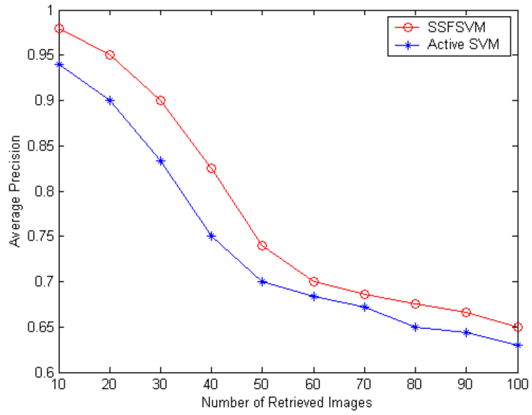

Figure 3. Precision graphs over the number of retrieved images

4. Conclusion

This paper presents a new content-based image retrieval method with relevance feedback technique using semi-supervised active learning algorithm. The approach uses a large amount of unlabeled data to alleviate the small sample problem in SVM based RF. A simple active learning is used to alleviate the burden of labeling by selecting only the most informative data. In addition, a semi-supervised approach has been developed which uses Nearest-Neighbor technique to label the unlabeled data with a certain degree of uncertainty in its class information. Then these automatically labeled samples with user labeled data are used to train FSVM. The experiment results show that our method has a better performance and demonstrate that it is an effective method to improve semantic image retrieval performance.

References

[1] Chandramouli K, Kliegr T et al. *Query Refinement and User Relevance Feedback for Contextualized Image Retrieval*. In Proceeding of Visual Information Engineering. 2008: 453-458.

[2] Rahman MM, Bhattacharya P, Desai BC. A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching Techniques with Relevance Feedback. *IEEE Transactions on Information Technology in Biomedicine*. 2007; 11(1): 58-69.

[3] Liu RJ, Wang YH et al. SVM-Based Active Feedback in Image Retrieval Using Clustering and Unlabeled Data. *Pattern Recognition*. 2008; 41: 2645-2655.

[4] Wang L, Chan KL, Zhang Z. *Bootstrapping SVM Active Learning by Incorporating Unlabelled Images for Image Retrieval*. Proceedings of IEEE International Conference of Computer Vision and Pattern Recognition. 2003: 629-634.

[5] Gosselin P, Cord M. Active Learning Methods for Interactive Image Retrieval. *IEEE Transactions on Image Process*. 2008; 17: 1200-1211.

[6] Lu K, Zhao JD, Cai D. An Algorithm for Semi-supervised Learning in Image Retrieval. *Pattern Recognition*. 2006; 39: 717-720.

[7] Hoi SCH, Lyu MR. *A Semi-supervised Active Learning Framework for Image Retrieval*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005; 2: 302-309.

[8] Vapnik VN. Statistical learning theory. New York: Wiley. 1998.
[9] Lin CF, Wang SD. Fuzzy Support Vector Vachines. *IEEE Transaction on Neural Networks*. 2002; 13(2): 464-471.
[10] Giacinto G. *A Nearest-Neighbor Approach to Relevance Feedback in Content Based Image Retrieval*. In Proceedings of the 6th ACM International Conference on Image and Video Retrieval. 2007: 456-463.
[11] Santosh NR, Anil BG. A Novel Approach For Face Recognition Using Fusion Of Local Gabor Patterns. *International Journal of Electrical and Computer Engineering.* 2012; 2(3): 345-352.
[12] Hendra K, NFN W, Adi S. Gabor-based Face Recognition With Illumination Variation Using Subspace-Linear Discriminant Analysis. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2012; 10(1): 119-128.