Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP 2007)

A Pseudo-Labeling Framework for Content-based Image Retrieval

Kim-Hui Yap and Kui Wu
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
Email: {ekhyap, pg01537831}@ntu.edu.sg

Abstract-**This paper presents a new pseudo-label fuzzy support vector machine (PLFSVM)–based active learning framework in interactive content-based image retrieval (CBIR) systems. One of the main issues associated with relevance feedback in CBIR systems is the small sample problem where only a limited number of labeled samples are available for learning. This is because image labeling is time consuming and users are often reluctant to label too many images for feedback. Learning from insufficient training samples often constrains the retrieval performance. To address this problem, we propose a new algorithm based on the concept of pseudo-labeling. It incorporates carefully selected unlabeled images to enlarge the training data set and assigns proper pseudo-labels to them. Further, some fuzzy rules are utilized to automatically estimate class membership of the pseudo-labeled images. Fuzzy support vector machine (FSVM) is designed to take into account the fuzzy nature of some training samples during its training. In order to exploit the advantages of pseudo-labeling, active learning and the structure of FSVM, we develop a unified framework to perform content-based image retrieval. Experimental results based on a database of 10,000 images demonstrate the effectiveness of the proposed method.**

1. INTRODUCTION

Content-based image retrieval (CBIR) has been developed to retrieve a set of desired images from an image collection. It makes use of the visual contents such as color, texture, shape and spatial relationship that exist in the images. These low-level features, however, may not correspond to the users' dynamic and subjective interpretation of image contents under various circumstances. In view of this, relevance feedback has been introduced to bridge this gap. Relevance feedback is an interactive mechanism that involves user participation. Under this framework, the users provide their judgment on the relevance of the retrieved images. The systems then learn the user information needs based on these feedbacks. Many relevance feedback algorithms have been adopted in CBIR systems and demonstrated considerable performance improvement.

Despite the previous works on relevance feedback for CBIR systems, it is still a challenging task to develop effective and efficient interactive mechanisms to yield satisfactory retrieval performance. One key difficulty with relevance feedback is the lack of sufficient labeled images since users usually do not have the patience to label a large number of images. Therefore, the performance of relevance feedback methods is often constrained by the limited number of training samples. To deal with this problem, some works have been done to incorporate the unlabeled data to improve the learning performance. Discriminant Expectation Maximization (D-EM) algorithm has been introduced to incorporate the unlabeled samples to estimate the underlying probability distribution [1]. The results are promising, but the computational complexity can be significant for large databases. Transductive support vector machine (TSVM) for text classification has been proposed to tackle the problem by incorporating the unlabeled data [2]. It has also been applied for image retrieval [3]. The method proposes to incorporate unlabeled images to train an initial SVM, followed by standard active learning. It is, however, observed that the performance of this method may be unstable in some cases. Incorporating prior knowledge into the SVM has also been introduced to resolve the small sample problem [4]. All these proposed methods show some promising outcomes, however few can learn from the labeled and unlabeled data effectively.

In this paper, we develop a pseudo-label fuzzy support vector machine (PLFSVM) framework to perform content-based image retrieval. By exploiting the characteristics of the labeled images, unlabeled images are chosen carefully and assigned different pseudo-labels such as 'relevant' or 'irrelevant'. This process will enlarge the training data set. As these images are not labeled explicitly by the users, there is a potential imprecision embedded in their class information. In view of this, a fuzzy membership function is employed to estimate the class membership of the pseudo-labeled images. The fuzzy information is then integrated into the FSVM for active learning.

2. LEARNING FROM SMALL SAMPLES

A. Small Sample Problem

In interactive CBIR systems, it is not user friendly to let the users label too many images for feedback. This results in the small sample problem where learning from a small number of training samples may not produce good retrieval results, even

for powerful learning machine such as SVM. Therefore, it is imperative to find solutions to solve the small sample problem faced by relevance feedback.

Considering that obtaining a large number of labeled images is labor intensive while unlabeled images are readily available, we propose to augment the available labeled images by making use of the potential role of unlabeled images. It is worth noting that unlabeled images can degrade the performance if used improperly. Consequently, they should be carefully chosen so that they will be beneficial to the retrieval performance. Each selected, unlabeled image is assigned a pseudo-label of either 'relevant' or 'irrelevant' based on an algorithm to be explained in subsection 3-B. These pseudo-labeled images are fuzzy in nature since they are not explicitly labeled by the users. Therefore the potential imprecision embedded in their class information should be taken into consideration. We employ a fuzzy membership function to determine the degree of uncertainty for each pseudo-labeled image, hence putting into context the relative importance of these images. These pseudo-labeled samples are then combined with those labeled samples to train the FSVM.

B. Active Learning

SVM is an implementation of the method of structural risk minimization (SRM) [5]. It has been successfully utilized in many real-world applications. The basic idea of SVM for binary classification is to find an optimal separating hyperplane that maximizes the margin between two classes in a kernel-induced feature space. Despite the superior performance of SVM in solving classification problems, it is still limited to crisp classification where each training sample is classified into exactly one class or another. Nevertheless, there exist situations where the training samples do not fall neatly into discrete classes. They may belong to different classes with different degree of membership. To solve this problem, FSVM has been developed [6]. FSVM is an extended version of SVM that takes into consideration different significance of the training samples. It exhibits the following properties that motivate us to adopt it in our framework: integration of fuzzy data, strong theoretical foundation, and excellent generalization power.

We develop a unified PLFSVM framework that integrates the advantages of pseudo-labeling and FSVM. It exploits inexpensive unlabeled data to augment the small set of labeled data, hence potentially improves the retrieval performance. This is in contrast to most existing feedback approaches in CBIR systems that are concerned with the use of labeled data only. It is noted that the proposed PLFSVM differs from the traditional SVM in several ways. The PLFSVM is developed for resolving the small sample problem by incorporating pseudo-labeled images, while traditional SVM can only handle labeled images. Further, PLFSVM requires less user workload, thus making it more appealing for practical applications such as image retrieval over bandwidth-limited network. Lastly, PLFSVM can take relative significance of the training samples into consideration, and hence, is more general and flexible.

Active learning is designed to achieve maximal information gain or minimize uncertainty in decision making. It selects the most informative samples to query the users for labeling. SVM-based active learning aims to select samples that maximally reduce the version space of SVM [7]. It selects samples that are closest to the current SVM decision boundary as the most informative points. Samples that are farthest away from the boundary and on the positive side are considered as the most relevant images. The same selection strategy is adopted in this work. Integrating the merits of PLFSVM into active learning, we can achieve improved retrieval performance with less user labeling.

3. PSEUDO-LABEL FUZZY SUPPORT VECTOR MACHINE (PLFSVM)

A. Formulation of FSVM

We first provide a brief introduction on SVM. Let $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ be a set of n training samples, where $\mathbf{x}_i \in \Re^m$ is an m-dimensional sample in the input space, and $y_i \in \{-1,1\}$ is the class label of $\mathbf{x}_i$. SVM first transforms data in the original input space to higher dimensional feature space through a mapping function $\mathbf{z} = \varphi(\mathbf{x})$. It then finds the optimal separating hyperplane with minimal classification errors. The hyperplane can be represented as:

$$\mathbf{w}\ \mathbf{z} + b = 0 \tag{1}$$

where $\mathbf{w}$ is the normal vector of the hyperplane, and b is the bias which is a scalar. The optimal hyperplane can be obtained by solving the following constrained optimization problem [5]:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{subject to } y_i(\mathbf{w}\ \mathbf{z}_i + b) \quad 1-\xi_i, \quad \xi_i \quad 0, \quad i = 1,\mathrm{K}\ ,n \end{aligned} \tag{2}$$

where C is the regularization parameter controlling the tradeoff between margin maximization and classification error. Larger value of C produces narrow-margin hyperplane with less misclassification. ξ_i is called the slack variable that is related to classification errors in SVM. Misclassifications occur when $\xi_i > 1$. The optimization problem can be transformed into the following equivalent dual problem using the Lagrangian method:

$$\begin{aligned} &\text{maximize } \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \mathbf{z}_i\ \mathbf{z}_j \\ &\text{subject to } \sum_{i=1}^{n} y_i\alpha_i = 0, \quad 0 \le \alpha_i \le C, \quad i = 1,\mathrm{K}\ ,n \end{aligned} \tag{3}$$

where α_i is the Lagrange multiplier. The decision function of the SVM can be represented as:

$$f(\mathbf{x}) = \mathbf{w}\ \ \mathbf{z} + b = \sum_{i=1}^{n} \alpha_i y_i \varphi(\mathbf{x}_i)\ \ \varphi(\mathbf{x}) + b = \sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{4}$$

where $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function in the input space that computes the inner product of two data points in the feature space. Using this kernel trick, we can construct the optimal hyperplane in the feature space without having to know about the mapping φ in explicit form. There are three common types of kernels used in SVM including polynomial kernel, radial basis function kernel and sigmoid kernel.

In FSVM, each training sample is associated with a fuzzy membership value $\{\ _i\}_{i=1}^{n} \in [0,1]$. The membership value $_i$ reflects the fidelity of the data, or 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 [6]:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\ _i\xi_i \\ &\text{subject to } y_i(\mathbf{w}\ \ \mathbf{z}_i + b)\ \ 1-\xi_i, \quad \xi_i\ \ 0, \quad i = 1,\mathrm{K}\ ,n \end{aligned} \tag{5}$$

It is noted that the error term ξ_i is scaled by the membership value $_i$. 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.

Similar to the conventional SVM, the optimization problem of FSVM can be transformed into its dual problem as follows:

$$\begin{aligned} &\text{maximize } \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &\text{subject to } \sum_{i=1}^{n} y_i\alpha_i = 0, \quad 0 \le \alpha_i \le\ _i C, \quad i = 1,\mathrm{K}\ ,n \end{aligned} \tag{6}$$

Solving equation (6) will lead to a decision function similar to (4), but with different support vectors and corresponding weights α_i.

B. Unlabeled Image Selection and Pseudo-Label Estimation

Appropriate selection of the unlabeled images to aid retrieval is vital as the unlabeled images may not help or even degrade the performance if chosen improperly. In this work, we present a method to select the unlabeled images for pseudo-labeling by studying the characteristics of the labeled images. The selection criterion is to determine certain informative samples among the unlabeled ones which are 'similar' to the labeled images in terms of the visual features for pseudo-labeling and fuzzy membership estimation. The enlarged hybrid data set consisting of both pseudo-labeled and explicitly labeled samples is then utilized to train the FSVM.

It is observed that the labeled images usually exhibit local characteristics of image similarity. To exploit this property, it is desirable to adopt a multi-cluster local modeling strategy. Taking into account the local multi-cluster nature of image similarity, we employ a two-stage clustering process to determine the local clusters. The labeled samples are clustered according to their types: relevant or irrelevant. K-means clustering is one of the most widely used clustering algorithms. It groups the samples into K clusters by using an iterative algorithm that minimizes the sum of distances from each sample to its respective cluster centroid for all the clusters. Notwithstanding its attractive features, K-means clustering requires a specified number of clusters in advance and is sensitive to the initial estimates of the clusters. To rectify this difficulty, we adopt a two-stage clustering strategy in this work. First, subtractive clustering is employed as a preprocessing step to estimate the number and structure of clusters as it is fast, efficient and does not require the number of clusters to be specified *a priori* [8]. These estimates are then employed by K-means to perform clustering based on iterative optimization in the second stage.

Two sets of separate clusters are obtained, relevant and irrelevant sets after clustering. Unlabeled image selection and pseudo-label assignment is then based on a similarity measure analogous to the k-nearest neighbor (K-NN) technique. That is, samples close in distance will potentially have similar class labels. For each cluster formed by the labeled images using the two-stage clustering scheme, K nearest unlabeled neighbors are chosen based on their Euclidean distances to the center of the respective labeled cluster. The label (relevant or irrelevant) of each labeled cluster is then propagated to the unlabeled neighbors. This is referred to as pseudo-labeling process. As the computational cost will increase with respect to the number of pseudo-labeled images, therefore, only the most 'similar' neighbor for each cluster is selected in this work.

C. Estimation of Soft Relevance Membership Function for Pseudo-Labeled Images

In consideration of the potential fuzziness of the pseudo-labeled images, our objective here is to determine a soft relevance membership function $g(\mathbf{x}_P): \Re^m \to [0,1]$ that assesses each pseudo-labeled image $\mathbf{x}_P$ and assigns it a proper relevance value between [0, 1]. Since clustering has been performed on each positive (relevant) and negative (irrelevant) class separately to get multiple clusters per class, the obtained clusters in each class can be employed to generate the membership value. Intuitively, the closer a pseudo-labeled image is to the nearest cluster of the same class label, the higher is its degree of relevance. In contrast, the closer a pseudo-labeled image is to the nearest cluster of the opposite class label, the lower is its degree of relevance.

Based on this argument, an exponentially-based fuzzy function is selected:

$$w_1(\mathbf{x}_P)=\begin{cases}\exp\left(-a_1\dfrac{\min_i(\mathbf{x}_P-\mathbf{v}_{Si})^{\mathrm{T}}(\mathbf{x}_P-\mathbf{v}_{Si})}{\min_j(\mathbf{x}_P-\mathbf{v}_{Oj})^{\mathrm{T}}(\mathbf{x}_P-\mathbf{v}_{Oj})}\right), & \text{if } \dfrac{\min_i(\mathbf{x}_P-\mathbf{v}_{Si})^{\mathrm{T}}(\mathbf{x}_P-\mathbf{v}_{Si})}{\min_j(\mathbf{x}_P-\mathbf{v}_{Oj})^{\mathrm{T}}(\mathbf{x}_P-\mathbf{v}_{Oj})}<1\\ 0 & \text{otherwise}\end{cases} \tag{7}$$

where $\mathbf{v}_{Si}$ denotes the center of the *i*th cluster with the same class label as the pseudo-labeled image $\mathbf{x}_P$, while $\mathbf{v}_{Oj}$ denotes the center of the *j*th cluster with the opposite class label to the pseudo-labeled image $\mathbf{x}_P$. $\min_i(\mathbf{x}_P-\mathbf{v}_{Si})^{\mathrm{T}}(\mathbf{x}_P-\mathbf{v}_{Si})$ and $\min_j(\mathbf{x}_P-\mathbf{v}_{Oj})^{\mathrm{T}}(\mathbf{x}_P-\mathbf{v}_{Oj})$ represent the distance between $\mathbf{x}_P$ and the nearest cluster centers with the same and opposite class labels, respectively. $a_1>0$ is a scaling factor. This membership function is divided into two scenarios. If the distance ratio is smaller than 1, suggesting that the pseudo-labeled image is closer to the nearest cluster with the same class label, then we will estimate its soft relevance. Otherwise, if the pseudo-labeled image is closer to the nearest cluster with the opposite class label, a zero value is assigned.

Further, the agreement between the predicted label obtained in subsection 3-B and the predicted label obtained from the trained FSVM can also be utilized to assess the degree of relevance of the pseudo-labeled samples. The second factor of the fuzzy function is chosen as a sigmoid function as follows:

$$w_2(\mathbf{x}_P)=\begin{cases}\dfrac{1}{1+\exp(-a_2 y)} & \text{pseudo-label is positive}\\ \dfrac{1}{1+\exp(a_2 y)} & \text{otherwise}\end{cases} \tag{8}$$

where $a_2>0$ is a scaling factor. y is the directed distance of the pseudo-labeled image $\mathbf{x}_P$ to the FSVM boundary (the decision function output of FSVM for the pseudo-labeled image $\mathbf{x}_P$). We will explain the rationale of the fuzzy expression in (8) by first considering that the pseudo-label of the selected image has been determined as positive in subsection 3-B. In this case, the upper equation in (8) will be used. If y has a large positive value, this will suggest that it is most likely to be a relevant image. Since there is a strong agreement between the predicted pseudo-label from subsection 3-B and the predicted class label using the trained FSVM, its fuzzy membership value should be set to a large value close to unity. If y has a large negative value, this will suggest that it is most likely to be an irrelevant image. Since there is a strong disagreement between the predicted pseudo-label from subsection 3-B and the predicted class label using the trained FSVM, its fuzzy membership value should be set to a small value close to zero. The same arguments apply when the pseudo-label of the selected image has been determined to be negative in subsection 3-B.

Finally, these two measures affecting the fuzzy membership are combined together to produce the final soft relevance estimate, namely:

$$g(\mathbf{x}_P)=w_1(\mathbf{x}_P)w_2(\mathbf{x}_P) \tag{9}$$

The estimated soft relevance of the pseudo-labeled images is then used in FSVM training.

4. Experimental Results

The performance of the PLFSVM is evaluated on an image database consisting of 10,000 natural images with 100 different categories obtained from the Corel Gallery product. Color histogram, color moments and color auto-correlogram are used to represent the color feature, while Gabor wavelet and wavelet moments are used to represent the texture feature.

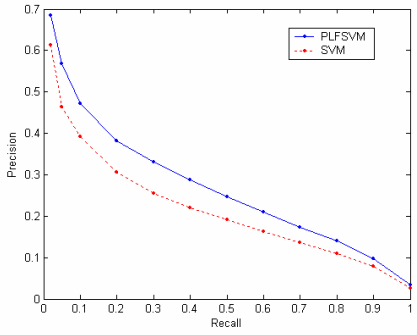
In our experiment, we use objective measure to evaluate the performance of the proposed PLFSVM method, and compare it with active learning using SVM [7]. The objective measure is based on the Corel's predefined ground truth. That is, the retrieved images are judged to be relevant if they come from the same category as the query. 100 queries with one from each category are selected for evaluation. Retrieval performance is evaluated by ranking the database images according to their directed distances to the SVM boundary after each active learning iteration. Five iterations of feedbacks are recorded. Precision-versus-recall curve is adopted in our experiment. The precision and recall rates are averaged over all the queries. The average precision-versus-recall (APR) graphs after the first iteration of active learning are shown in Fig. 1. We have shown the results for two different numbers of initially labeled images. From the figures, we observe that the PLFSVM method outperforms the standard SVM method in both cases. The PLFSVM method achieves higher recall rate at the same precision level. It also offers higher precision rate for the same recall level. This indicates the superiority of the proposed PLFSVM method. In our experiments, it is observed that PLFSVM consistently achieves better performance than SVM for different values of initial labeled images.

In addition, we have adopted another measure called retrieval accuracy to evaluate the retrieval system:

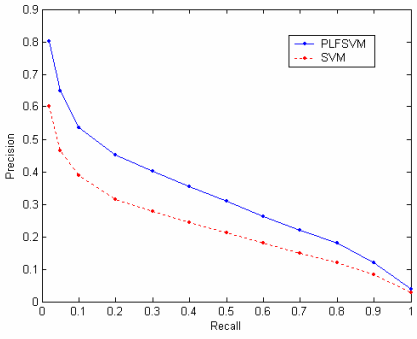
$$\text{Retrieval accuracy}=\frac{\text{relevant images retrieved in top } T \text{ returns}}{T} \tag{10}$$

where T is the number of top returned images with $T = 10$ in the experiment. The performance comparison of the PLFSVM method and the SVM method is given in Fig. 2 for the case of 10 initial labeled images. The retrieval accuracy is averaged over the 100 queries. We observe that PLFSVM

method achieves higher retrieval accuracy than the SVM method. Further, the retrieval accuracy of the PLFSVM method increases quickly in the initial stage. This is a desirable feature since the user can obtain satisfactory results quickly. It is worth emphasizing that the initial retrieval performance is very important since users often expect quick results and are unwilling to provide much feedback. Hence, reducing the amount of user feedback while providing good retrieval results is of great interests for many CBIR systems. It is observed that our method offers an improvement of 16% over the SVM method after the first iteration of active learning. The superiority of our method over the SVM method mainly lies in the incorporation of pseudo-labeled images for effective learning.



(a)



(b)

Fig. 1. The average precision-versus-recall graphs (after the first iteration of active learning). (a) APR for 5 initial labeled images, (b) APR for 10 initial labeled images.

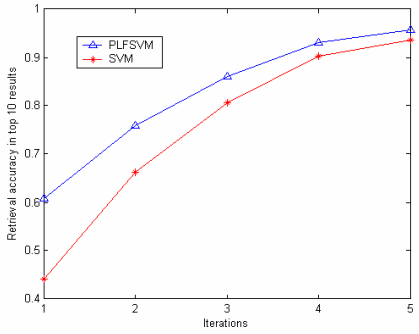


Fig. 2. Retrieval accuracy of the PLFSVM and SVM methods in top 10 results.

5. CONCLUSION

This paper addresses the small sample problem in interactive CBIR systems by incorporating pseudo-labeled images into FSVM along with labeled images for effective retrieval. By exploiting the characteristics of the labeled images, pseudo-labeled images are selected through an unsupervised clustering algorithm. Further, the relevance of the pseudo-labeled images is estimated using the fuzzy membership function. FSVM-based active learning is then performed based on the hybrid of pseudo-labeled and explicitly labeled images. Experimental results confirm the effectiveness of our proposed method.

REFERENCES

[1] Y. Wu, Q. Tian, and T. S. Huang, "Discriminant-EM algorithm with application to image retrieval", *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 222–227, South Carolina, 2000.

[2] T. Joachims, "Transductive inference for text classification using support vector machines," *Proc. Int. Conf. Machine Learning,* pp. 200–209, 1999.

[3] L. Wang and K. L. Chan, "Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition,* pp. 629–634, 2003.

[4] L. Wang and K. L. Chan, "Incorporating prior knowledge into SVM for image retrieval," *Proc. IEEE Int. Conf. Pattern Recognition,* pp. 981–984, 2004.

[5] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[6] C. F. Lin and S. D. Wang, "Fuzzy support vector machines," *IEEE Trans. Neural Networks*, vol. 13, no. 2, pp. 464–471, Mar. 2002.

[7] S. Tong and E. Chang, "Support vector machine active leaning for image retrieval," *Proc. ACM Int. Conf. Multimedia*, pp. 107–118, Ottawa Canada, 2001.

[8] S. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent & Fuzzy Systems*, vol. 2, no. 3, pp. 267–278, Sept. 1994.