RELEVANCE FEEDBACK USING RANDOM SUBSPACE METHOD*

Wei Jiang
Department of Automation
Tsinghua University
Beijing 100084, China

Mingjing Li, Hongjiang Zhang
Microsoft Research Asia
49 Zhichun Road
Beijing 100080, China

Jie Zhou
Department of Automation
Tsinghua University
Beijing 100084, China

ABSTRACT

The relevance feedback process in content-based image retrieval is generally treated as a classification problem, where the small sample size learning difficulty and the fast response requirement make it difficult for most classifiers to achieve a satisfying performance. In this paper, we incorporate the stochastic classifier ensemble method as a solution to alleviate this problem. In particular, the random subspace method is adopted in relevance feedback process to both improve the retrieval accuracy and decrease the processing time. Experimental results on 5,000 images demonstrate the effectiveness of the proposed method.

1. INTRODUCTION

Content-based image retrieval (*CBIR*) is becoming more important with the increasing amount of images that can be accessed. In CBIR context, an image is represented by a low-level visual feature vector, which can be viewed as a point in high-dimensional feature space. However the gap between high-level semantics and low-level features limits the effectiveness of CBIR systems. To bridge this gap, relevance feedback is introduced into CBIR systems [13, 14].

Generally the relevant feedback process can be treated as a learning and classification problem [1, 6, 9], and is usually viewed as a binary classification case. During the feedback process, a classifier is constructed using the feedback images labeled by users, and classifies all the images in the database into two class–the required images ("*relevant*" class), and the rest images ("*irrelevant*" class).

There are two special difficulties for CBIR classifiers: one is the small training set size, the other is the fast response requirement. The training samples are the labeled images from user's feedback, and are usually too few compared with the feature dimension and the whole database size. Most classifiers are weak and unstable in such cases [10]. More over, the CBIR system must give out the online retrieval result with a tolerable time cost. Thus, too complicated classifier will not be practical.

The stochastic discrimination (*SD*) theory [2] gives a solution to alleviate the above difficulties, which combines several stochastically created component classifiers with weak discriminant power to generate a strong classifier with nearly monotonic accuracy increase [8]. Random Subspace Method (*RSM*) is one of the SD methods based on a stochastic feature space subsampling process [7]. RSM is very suitable for small sample learning problems, and for practical application that needs fast learning [4]. Thus, RSM is very suitable for enhancing CBIR classifiers.

In this paper we incorporate RSM into the relevance feedback process to improve the retrieval performance. To make the underlying SD theory play, the component classifier should satisfy the *projectability*, *enrichment* and *uniformity* conditions. In our system, support vector machine (*SVM*) is adopted as the component classifier to satisfy the projectability and enrichment requirements, because of its large margin characteristic with good generalization ability, and its advantage for small sample learning. And based on the analysis of characteristic of CBIR training set, a sample re-weighting mechanism is proposed to promote the uniformity condition. Experimental results on 5,000 images show that the proposed method can achieve significant performance improvement with a decreased processing time.

The rest of this paper is organized as follows. Since the SVM$_{Active}$ feedback mechanism [6] is adopted in our system, Section 2 formulates our problem and introduces the SVM$_{Active}$ classifier. In Section 3 we introduce the SD theory briefly and describe our method in detail. Experimental results are presented in Section 4. Finally, we give our conclusion in Section 5.

2. RELATED WORK

Assume that in CBIR system, the M images in the database are denoted by $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M]$. Each image is a d-dimensional feature vector $\mathbf{x}_i = [x_{i1}, \ldots, x_{id}]^T$, which is a point in d-dimensional feature space, $\mathbf{x}_i \in \mathcal{F}^d$. The relevance feedback process of SVM$_{Active}$ is as follows. In feedback round t, we have a training set $\mathcal{S}^t$. An SVM classifier is constructed over $\mathcal{S}^t$, which then classifies images $\mathbf{X}$ in

*This work was performed at Microsoft Research Asia the database, and outputs a set of distance $\mathcal{D}^t = \{D^t(i), i = 1, \dots, M\}$. $D^t(i)$ denotes the distance of image $\mathbf{x}_i$ to the decision boundary ($D^t(i) > 0$ represents "relevant", and $D^t(i) < 0$ represents "irrelevant"). This classifier construction and classification process is denoted by $\mathcal{D}^t = \mathrm{C}(\mathcal{S}^t, \mathbf{X})$ in rest of this paper. The retrieval result is $\mathcal{R}^t$, called *return set*, which is formed by images with largest $D^t(i) > 0$. If user doesn't satisfy with $\mathcal{R}^t$, images with smallest $|D^t(i)|$ forms a *label set*, $\mathcal{L}^{t+1}$, for user to label. The label result is $\mathbf{Y} = \{y_i, i = 1, \dots, M\}$, where $y_i = 1$ if $\mathbf{x}_i$ is "relevant", $y_i = -1$ otherwise. Then we update $\mathcal{S}^{t+1} = \mathcal{S}^t \bigcup \mathcal{L}^{t+1}$ and go to next feedback round. User provides one query image $\mathbf{x}_q$ to start the query, and $\mathcal{L}^1$ is randomly selected from database. The initial training set is $\mathcal{S}^1 = \mathcal{L}^1 \bigcup \{\mathbf{x}_q\}$.

3. RELEVANCE FEEDBACK USING RSM

3.1. Stochastic Discrimination

The stochastic discrimination theory constructs an ensemble classifier by many stochastically created weak component classifiers, and theoretically predicts that if the component classifiers satisfy the three conditions, namely, projectability, enrichment and uniformity, the accuracy of the ensemble classifier will increase monotonically as the number of component classifiers increases. Enrichment requires that the component classifiers have the error rate of no more than 0.5, and projectability refers that the classifier has generalization ability to unseen samples (test set). Uniformity means that for any two unseen positive (negative) samples, the number of component classifiers which classify each of them into the positive (negative) class is identical. Actually it is quite easy for most classifiers to meet the enrichment and projectability requirement, but strict uniformity condition is difficult to obtain. In [5] theoretical analysis and experiments show that satisfactory results can be attained with approximate uniformity setting. Boosting algorithm [12] is used as a uniformity forcing method in [3], which iteratively modifies training set's distribution to emphasize"hard" samples (the samples misclassified by earlier classifiers). Although the hypothesis fusion strategy of boosting is not relevant under SD's paradigm, its sample re-weighting mechanism provides an approach toward promoting uniformity.

In our system, RSM is incorporated to enhance the retrieval performance. We use SVM as the component classifier, which fits the projectability and enrichment conditions. The sample re-weighting mechanism of boosting is adopted as the uniformity promoting method. Now we describe our enhancing scheme in detail.

3.2. The Combination Scheme

RSM relies on a parallel stochastic process which creates stochastic training sets by sampling on feature space $\mathcal{F}^d$. Assume that in each feedback round, totally n_{en} component classifiers will be combined. The ensemble mechanism of RSM in the relevance feedback process can be described as follows. For the k-th component classifier, the system randomly selects d_r feature axes from $\mathcal{F}^d$, and forms a d_r-dimensional feature subspace $\mathcal{F}^{d_r}$. Then it projects $\mathcal{S}^t$ into $\mathcal{F}^{d_r}$ to be $\mathcal{S}^t_k$, denoted by $\mathcal{S}^t \mapsto \mathcal{S}^t_k$. Also it projects $\mathbf{X} \mapsto \mathbf{X}_k$. Then an SVM classifier $\mathcal{D}^t_k = \mathrm{C}(\mathcal{S}^t_k, \mathbf{X}_k)$ is constructed. The decision ensemble method is to combine the output of each classifier, that is, to combine $\mathcal{D}^t_k, k = 1 \dots, n_{en}$ into a set of ensemble distance $\mathcal{D}^t$. The element of $\mathcal{D}^t$ is:

$$D^t(i) = \frac{1}{n_{en}} \sum_{k=1}^{n_{en}} D^t_k(i) \tag{1}$$

According to this ensemble $\mathcal{D}^t$, $\mathcal{R}^t$ and $\mathcal{L}^{t+1}$ are selected following the criterion of SVM$_{Active}$: $\mathcal{R}^t$ is formed by images with largest $\mathcal{D}^t > 0$, $\mathcal{L}^{t+1}$ is formed by images with smallest $|\mathcal{D}^t|$.

RSM is suitable for enhancing small sample size learning cases. This can be intuitively explained as the random subspace method actually constructs each classifier in a lower dimensional feature space with the training sample size unchanged, then the ratio of training sample size versus feature dimensionality increases compared with original training set, and better performance may be achieved. More over, RSM can alleviate the curse of dimensionality, and can take advantage of high dimensionality. Also, it is a parallel algorithm which can be processed by parallel computation method, thus is suitable for cases with fast learning requirement. All these characteristics make RSM fit for enhancing CBIR classifiers.

3.3. Uniformity Promoting

The central idea of sample re-weighting in boosting algorithm gives a way for uniformity promoting. Here we realize the sample re-weighting target by sampling method. If a training sample is important, we generate more samples around it. This can be viewed as another way to re-weight the training samples.

Suppose in feedback round t ($t > 1$), the training set for the previous $t-1$ round is $\mathcal{S}^{t-1}$, and the classification result of the ensemble classifier in $t-1$ round is $\mathcal{D}^{t-1}$ (generated by Eqn(1)). $\mathcal{L}^t$ is the label set selected according to $\mathcal{D}^{t-1}$. Define the *important set* for feedback round t as:

$$\begin{aligned} \mathcal{T}^t \;&=\; \left\{\mathbf{x}_i : \mathbf{x}_i \in \mathcal{S}^{t-1}, y_i D^{t-1}(i) < 0\right\} \\ &\;\; \bigcup \left\{\mathbf{x}_i : \mathbf{x}_i \in \mathcal{L}^t, y_i = 1\right\} \end{aligned} \tag{2}$$

Eqn(2) indicates that, besides the training samples misclassified by the previous classifiers, the newly labeled "relevant" samples are also contained in $\mathcal{T}^t$. This is because we have a training set whose size is increasing during the feedback rounds, and newly labeled training samples should also be weighted. Since in the CBIR context, the "relevant" images usually share some semantic cues which reflect the query concept, while the "irrelevant" ones come from different semantic categories and have little correlation. The "relevant" images are important to grasp user's query concept, and need to be emphasized.

Then the updated training set is given by:

$$\mathcal{S}^t = \mathcal{S}^{t-1} \bigcup \mathcal{T}^t \bigcup \mathcal{L}^t \tag{3}$$

Pseudo-code for the proposed classifier ensemble method is given in Fig.1.

Initialize: Get the user's query image $\mathbf{x}_q$, randomly select $\mathcal{L}^1$ from the database, and set $\mathcal{S}^1 = \mathcal{L}^1 \bigcup \{\mathbf{x}_q\}$
Recursion: for each feedback round t

1. if $t=1$, construct $\mathcal{D}^t = \mathrm{C}\left(\mathcal{S}^t, \mathbf{X}\right)$, go to step 5
2. Uniformity promoting
   - Calculate $\mathcal{T}^t$ by Eqn(2)
   - $\mathcal{S}^t = \mathcal{S}^{t-1} \bigcup \mathcal{T}^t \bigcup \mathcal{L}^t$
3. For $k=1,\ldots,n_{en}$
   - Randomly select $\mathcal{F}_k^{d_r}$, $\mathcal{S}^t \mapsto \mathcal{S}_k^t$, $\mathbf{X} \mapsto \mathbf{X}_k$
   - Get $\mathcal{D}_k^t = \mathrm{C}\left(\mathcal{S}_k^t, \mathbf{X}_k\right)$
4. Get $\mathcal{D}^t$ by Eqn(1)
5. Get $\mathcal{R}^t$ and $\mathcal{L}^{t+1}$. If user satisfy with $\mathcal{R}^t$, stop; otherwise, label $\mathcal{L}^{t+1}$, go to next round

Fig. 1. Pseudo-code for enhancing scheme.

4. EXPERIMENTS

The proposed method is evaluated on 5,000 real world images from 50 semantic categories, with 100 images for each category. All the images are collected from Corel CDs. The influence of different d_r and n_{en} are also investigated. The low-level features used in the experiment are the color coherence in HSV color space, the first three color moments in LUV color space, the directionality texture, and the coarseness vector texture, which comprise a 155-dimensional feature space in total. Details about these features can be found in [11].

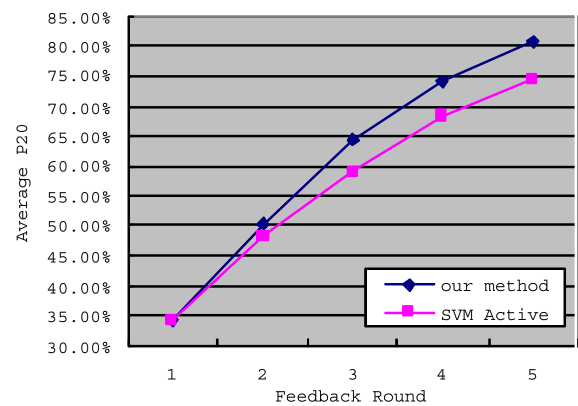
Assume that the user is looking for one semantic category in each query session, and will perform 5 rounds of feedback. In each round $|\mathcal{L}^t| = 10$. The initialization of our algorithm is similar to that of SVM$_{Active}$ (Fig.1). The performance measurement used is the average top-k precision:

$$P_{|\mathcal{R}^t|=k} = \frac{\text{the number of "relevant" images in } \mathcal{R}^t}{|\mathcal{R}^t|} \tag{4}$$

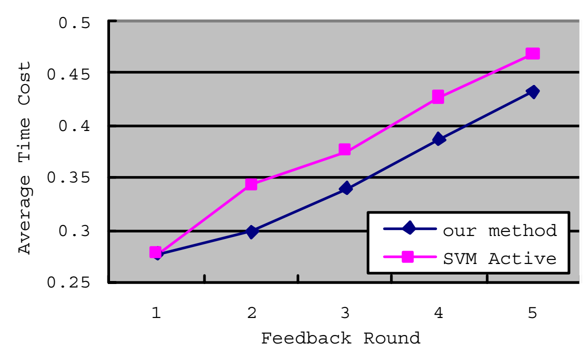
Each result listed in the experiments is the average result of 500 independent search processes. The kernel function for SVM is RBF kernel: $K\left(\mathbf{x}, \mathbf{x}_i\right) = \exp\left\{-\,||\mathbf{x} - \mathbf{x}_i|\,|^2/f\right\}$, where f is the dimensionality of $\mathbf{x}$.

4.1. Comparison with SVM$_{Active}$

To show the performance improvement achieved by our enhancing method , the proposed method is compared with the SVM$_{Active}$ algorithm in this experiment. We fix $d_r = 60$, $n_{en} = 2$ here. Fig.2 (a) gives the average P_{20} of these two algorithms after 5 rounds of feedback, and Fig.2 (b) lists their corresponding time cost. The figures indicate that the precision curve of our method is above the corresponding one of SVM$_{Active}$, which means that our method can consistently improve the retrieval performance. The precision improvements achieved after the feedback round 3, 4 and 5 are 9.45%, 8.84% and 8.72% respectively. Note that the results listed are for $n_{en} = 2$. That is, we only construct two SVM classifiers during each feedback round, and each classifier is constructed over a feature subspace with a dimensionality less than half of the original space. Thus our method costs less time than original SVM$_{Active}$ in this parameter setting. This experiment shows that the proposed enhancing algorithm can achieve better retrieval results within a shorter processing time.



(a)P_{20}



(b)Time Cost

Fig. 2. (a) The average precision of our method and SVM$_{Active}$ after the 5 feedback rounds. (b) The corresponding processing time. In this experiment, $d_r = 60$, $n_{en} = 2$ for our method.

4.2. The Influence of d_r and n_{en}

To test the influence of d_r and n_{en} on the retrieval performance, we first fix $n_{en} = 2$ and let d_r change from 50 to 90, then fix $d_r = 60$ and let n_{en} change from 2 to 6. Table 1 and 2 show the average P_{20} of these two cases respectively, and the corresponding time cost. The tables indicate that, when n_{en} is fixed, the retrieval result gets better as d_r increases when $d_r \leq 70$, and attains a maximum when $d_r = 70$. The time cost also increases as d_r increases. Also, when d_r is fixed, the precision increase when more component classifiers are combined, so is time cost. Thus, there is a trade off between the processing time and precision achieved. If we want more accurate result, more component classifiers should be combined.

Table 1.
The influence of d_r with $n_{en} = 2$

d_r	Round 2		Round 3		Round 5	
	P_{20}	Time	P_{20}	Time	P_{20}	Time
50	49.92	0.261	62.72	0.297	79.96	0.341
60	50.42	0.298	64.10	0.339	80.70	0.386
70	51.02	0.337	64.62	0.363	81.06	0.437
80	50.88	0.377	63.42	0.405	80.66	0.500
90	49.52	0.432	61.26	0.482	79.56	0.565

Table 2.
The influence of n_{en} with $d_r = 60$

n_{en}	Round 2		Round 3		Round 5	
	P_{20}	Time	P_{20}	Time	P_{20}	Time
2	50.42	0.298	64.10	0.339	80.70	0.386
3	51.20	0.430	64.92	0.465	81.30	0.563
4	51.92	0.574	66.80	0.656	82.60	0.732
5	52.10	0.720	67.42	0.801	83.12	0.992
6	52.72	0.900	67.86	0.996	83.66	1.241

5. CONCLUSION

In this paper, the random subspace method is incorporated into the relevance feedback rounds to improve the retrieval result and decrease processing time cost. The SVM classifier is selected to be the component classifier to fit the projectability and enrichment requirement for SD theory, and a sample re-weighting method is proposed to promote uniformity for our algorithm. The mechanism of SVM$_{Active}$ is also adopted in our system. Experiments on 5,000 images show that the proposed method can achieve accuracy improvement with less processing time.

6. REFERENCES

[1] G.D. Guo. et al, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. Neural Networls*, 13(4), pp.811-820, 2002.

[2] E.M. Kleinberg, "Stochastic Discrimination," *Annals of Mathematics and Artificial Intelligence*, vol.1, pp.207-239, 1990.

[3] E.M. Kleinberg, "On the algorithm implementation of stochastic discrimination," *IEEE Trans. Pattern Recognition and Machine Intelligence*, 22(5), pp.473-490, 2000.

[4] M. Skurichina, and R.P.W. Duin, "Bagging and the random subspace method for Redundant Feature Spaces," *Multiple Classifier System*, LNCS 2096, pp.1-10, 2001.

[5] R. Berlind, "An alternative method of stochastic discrimination with application to pattern recognition," PhD thesis, SUNY/Buffalo, New York, 1994.

[6] S. Tong, and E. Chang, "Support vector machine active learning for image retrieval," *ACM Multimedia*, Ottawa, Canada, 2001.

[7] TK. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(8), pp.832-844, 1998.

[8] TK. Ho, "Multiple classifier combination: lessons and next steps," *Hybrid methods in pattern recognition*, pp.171-198, World Scientific, 2002.

[9] T.S. Huang. et al, "Learning in content-based image retrieval," *IEEE Proc. of the* 2^{nd} *international conference on Development and Learning*, pp.155-164, Cambridge, Massachusetts, 2002.

[10] X.S. Zhou, and T.S. Huang, "Relevance Feedback in Image retrieval: a comprehensive review," *Multimedia Systems*, 8(6), pp.536-544, 2003.

[11] X.Y. Jin. et al, "Feature evaluation and optimizing feature combination for content-based image retrieval," technique report, http://www.cs.virginia.edu.cn/xj3a/publication/feature_selection.pdf

[12] Y. Freund, and R. Schapire, "Experiments with a new boosting algorithm," *Proc. of the* 13^{th} *International Conference on Machine Learning*, pp.148-156, Bari, Italy, 1996.

[13] Y. Ishikawa, R. Subramanya and C. Faloutsos, "MindReader: querying database through multiple examples," *Proc. of* 24^{th} *VLDB Conference*, pp.218-227, New York, USA, 1998.

[14] Y. Rui. et al, "Relevance feedback: A powerful tool in interactive content-based image retrieval," *IEEE Trans. Circuit and System for Video Technology, Special Issue on Segmentation, Description and Retrieval of Video Content*, 8(5), pp.644-655, 1998.