INTEGRATING RELEVANCE FEEDBACK IN BOOSTING FOR CONTENT-BASED IMAGE RETRIEVAL

Jie Yu[1], Yijuan Lu[1], Yuning Xu[1], Nicu Sebe[2], Qi Tian[1]

[1]Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

[2]Faculty of Science, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

Many content-based image retrieval applications suffer from small sample set and high dimensionality problems. Relevance feedback is often used to alleviate those problems. In this paper, we propose a novel interactive boosting framework to integrate user feedback into boosting scheme and bridge the gap between high-level semantic concept and low-level image features. Our method achieves more performance improvement from the relevance feedback than AdaBoost does because human judgment is accumulated iteratively to facilitate learning process. It also has obvious advantage over the classic relevance feedback method in that the classifiers are trained to pay more attention to wrongfully predicted samples in user feedback through a reinforcement training process. An interactive boosting scheme called i.Boost is implemented and tested using Adaptive Discriminant Projection (ADP) as base classifiers, which not only combines but also enhances a set of ADP classifiers into a more powerful one. To evaluate its performance, several applications are designed on UCI benchmark data sets, Harvard, UMIST, ATT facial image data sets and COREL color image data sets. The proposed method is compared to normal AdaBoost, classic relevance feedback and the state-of-the-art projection-based classifiers. The experiment results show the superior performance of i.Boost and the interactive boosting framework.

***Index Terms*—**Image classification, Information retrieval, Pattern recognition, Artificial intelligence, Algorithms

1. INTRODUCTION

Content-based image retrieval (CBIR) is a computer vision application that aims at automatically retrieving images of user interest from large image databases based on the visual content. The mapping between high-level semantic concept and low-level image features is obtained by a learning process. The images are often pre-processed to extract statistical features, such as color, texture and shape. An image feature vector represents an image as data point in a high-dimensional space. Although Content-based image retrieval has been successfully applied in many fields, it still faces two major challenges.

Small Sample Set: In CBIR, a set of samples with categorical information are used to train a classifier. Because labeling the training samples requires human interference and could be computational expensive, the size of the training set is often very small. In that case the learning process tends to bias to the training set and overfitting could occur.

High Dimensionality: In many data analysis application, the observed data have very high dimensionality. Specifically the images in CBIR are represented by image feature vector whose dimensionality ranges from tens to hundreds in most cases. Traditional statistical approaches have difficulties in modeling data directly in such a high dimensional space.

Some techniques have been proposed to alleviate the two problems. Relevance Feedback [1] is one of the most widely used techniques to alleviate the small sample set problem. For the high dimensionality problem, it is almost a common practice to conduct dimension reduction to find a compact representation of data in a low dimensional space. Traditional techniques, such as Principal Component Analysis (PCA) [2] and Linear Discriminat Analysis (LDA) [3], have difficulties in finding optimal projection automatically when the data distribution can not be modeled as Gaussian. Boosting could be used to alleviate that problem by combining a set of projection and corresponding classifiers in the projected space [4].

2. BOOSTING AND RELEVANCE FEEDBACK

2.1. Boosting

Boosting algorithms are designed to construct a "strong" classifier from a "weak" learning algorithm, presenting the superior result given by a thresholded linear combination of the weak classifier. A "weak" classifier has probability of misclassification that is slightly below 50%, while a "strong" one achieves much less error rate on test data. This idea was rooted in the framework of PAC learning, where it was theoretically proved. Kearns and Valiant raised the question on how to actually construct such as conversion in [5]. Schapire and Freund took over the idea and worked their way to the invention of AdaBoost [6]. The following couple of years see a great number of empirical work showcasing its ability to improve prediction accuracy. While a broad spectrum of application domains has gone ahead and benefited from boosting, researchers nevertheless have been trying to explain it, resulting in a rich set of satisfactory theory. Yet a complete picture is still out-of-reach.

AdaBoost is often regarded as the generic boosting algorithm, since it is the first practical algorithm that embodies the idea of boosting and has become extremely well-known over the years. Thus a description of the AdaBoost algorithm serves as the introduction to the boosting idea. In order to boost the weak learning algorithm, the data is reweighed (the relative importance of the training examples is changed) before running the weak learning algorithm at each iteration. In other words, AdaBoost maintains a distribution (set of weights) over the training examples and selects a weak classifier from the weak learning algorithm at each iteration. Training examples that were misclassified by the weak classifier at the current iteration then receive higher weights at the following iteration. The end result is a final combined classifier, each component is the weak classifier obtained at each iteration, and each component classifier is weighted according to how this classifier performed during each iteration. AdaBoost performs better than state-of-the-art classification algorithms in many experiments, and it does not seem to overfit. Theories trying to explain this include the margin theory [7] and the additive logistic regression [8]. These explanations have in turns given modifications or improvements over the original AdaBoost.

2.2. Relevance Feedback

Initially developed in documental retrieval [9], Relevance Feedback was transformed and introduced into content-based multimedia retrieval, mainly CBIR. Interestingly, it appears to have attracted more attention in the image field than the text field – a variety of solutions have been proposed within a short period and it remains an active research topic. As we discussed, a challenge in content-based image retrieval is the semantic gap between the high-level semantics in a human mind and the low-level computed features (such as color, texture, and shape). Users seek semantic similarity (e.g., airplane and bird are very similar in terms of low level features such as shape), but the machine can only measure similarity by feature processing.

The early work in Relevance Feedback focused on heuristic techniques, e.g., feature axis weighting in feature space and tree-structured self-organizing map (TS-SOM). The intuition is to emphasize those features that best cluster the positive examples and separate the positive from the negative examples. The assumption of feature independence is rather artificial. Learning in Relevance Feedback has been used in a more systematic way in the framework of optimization, probabilistic models, learning with small samples, pattern classification, active learning, concept learning, and genetic algorithms.

3. INTERACTIVE BOOSTING

3.1. Methodology

Motivated by the strength and success of Boosting and Relevance Feedback, we propose a framework called Interactive Boosting, which can integrate user relevance feedback in the loop of boosting to better bridge the gap between semantic concept and image features.

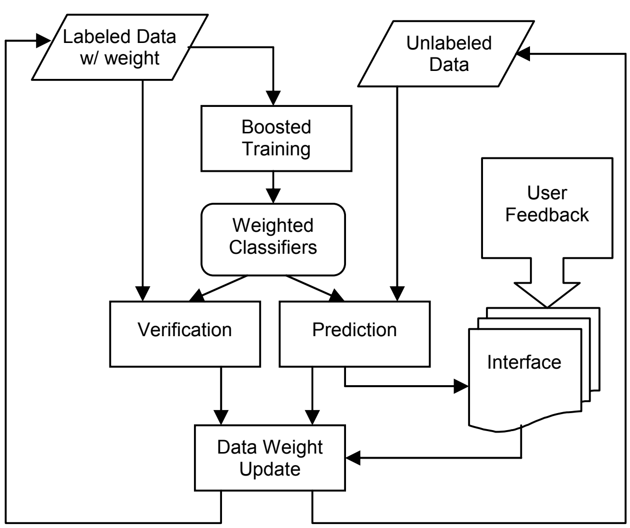


Fig. 1 Interactive Boosting framework

The process can be described in the following steps:

Step 1: Train weak classifiers on the original labeled data set and assign weights to classifiers based on their performance.
Step 2: Predict the labels of unlabelled data and present a subset of unlabeled data with their predicted labels to the user.
Step 3: User gives feedback on the retrieved data.
Step 4: Data obtained from user relevance feedback is added to construct a new labeled data set and removed from unlabeled data set.
Step 5: The labeled data are weighted according to their predicted label correctness.
Step 6: Go back to Step 1.

Figure 1 gives an illustration of the basic idea of the Interactive Learning framework.

2.2. Interactive Boosting for ADP

According to the framework discussed in Section 3.1, we implement a specific technique called i.Boost by using Adaptive Discriminant Analysis (ADP) [4] and K-NN as base classifier. Please note besides ADP classifier (ADP and K-NN classifier in the projected space), any other classification rule can also be plugged into our proposed framework.

The ADP is proposed to provide an accurate model of the complex distribution for positive and negative images by finding an optimal projection in the following way:

$$W_{ADP} = \arg\max_{W} \frac{|W^T[\lambda S_{P->N} + (1-\lambda)S_{N->P}]W|}{|W^T[\eta S_P + (1-\eta)S_N]W|} \quad (1)$$

in which

$$S_{N->P} = \sum_{i \in Negative}(x_i - m_P)(x_i - m_P)^T \quad (2)$$

$$S_{P->N} = \sum_{j \in Positive}(x_j - m_N)(x_j - m_N)^T \quad (3)$$

$$S_P = \sum_{j \in Positive}(x_j - m_P)(x_j - m_P)^T \quad (4)$$

$$S_N = \sum_{j \in Ne}(x_i - m_N)(x_i - m_N)^T \quad (5)$$

The m_P and m_N are the means of positive and negative samples, respectively. The two parameters λ and η controls the bias between positive and negative samples. Proper setting of parameters may fit the real distribution of data better than LDA or PCA [4].

However to find an optimal setting one has to do exhaustive searching in 2D parameter space, which is computationally expensive. Boosting can alleviate that problem by combining and enhancing a set of weak ADP classifiers into a more powerful one. To efficiently incorporate user feedback and enhance the retrieval accuracy, relevance feedback can be integrated in the boosting iterations as the framework suggests in Section 3.1. The brief algorithm below shows how the i.Boost can be implemented.

Algorithm i.Boost with ADP as weak classifiers

Input: Labeled Sample set X and label Y
Unlabeled Sample Set U
K ADP classifiers with different (λ, η)
T: The total number of runs that the classifiers will be trained for.

Initialization: weight $w_{k,t=1}(x) = 1/|X|$

Interactive Boosting

For $t = 1, \ldots, T$

For each classifier $k = 1, \ldots, K$ do

- Train the classifier on labeled samples with weights. Note that $\sum_{x \in X} w_{k,t}(x) = 1$
- Get the probability-rated prediction on labeled and unlabeled sample $\boldsymbol{h}_{k,t}(\boldsymbol{u}), \boldsymbol{h}_{k,t}(\boldsymbol{x}) \in (-1,1)$
- Compute the weights of classifiers based on its classification error rate $\varepsilon_{k,t}$ on labeled samples

$$\alpha_{k,t} = \frac{1}{2} \ln\left(\frac{1-\varepsilon_{k,t}}{\varepsilon_{k,t}}\right)$$

- Present images from the unlabeled data set with their predicted labels to user
- Obtain user feedback on the ground truth labels of images
- Construct new labeled training set by adding data and corresponding labels obtained from user feedback
- Update the weight of training samples

$$w_{k,t+1}(x) = w_{k,t}(x) \exp(-\alpha_{k,t} \cdot h_{k,t}(x) \cdot y)$$

End for each classifier

<u>End for *t*</u>

The final prediction $\boldsymbol{H}(\boldsymbol{u}) = \boldsymbol{sign}(\sum_{k,t} \alpha_{k,t} \boldsymbol{h}_{k,t}(\boldsymbol{u}))$

4. EXPERIMENTS AND ANALYSIS

To evaluate the performance of our proposed method, several experiments are designed and implemented. For simplicity, overall prediction precision and error rate are used as performance measures. Our methods are compared with AdaBoost, ADP with Relevance Feedback and other state-of-the-art projection techniques. The data sets used are UCI benchmark data sets, COREL image data set and three face image data sets, which cover a wide range of data encountered in computer vision applications. During each iteration, the relevance feedbacks on 5 images are fed to the system automatically base on ground trth. The reported results are the average of 50 repeats.

4.1. Comparison to AdaBoost

In the first experiment, we test the performance of our method on UCI benchmark data set and compare it to that of simple AdaBoost with same ADP base classifier (B.ADP [4]) as i.Boost. Due to space limitation only the results on Breast-Caner (B.C.) and Heart data sets are shown here, which mimics medical image retrieval/diagnosis applications. The data dimensions of these two data sets are 13 and 9, respectively. The sizes of the training sets are 170 and 200, and the sizes of the testing data sets are 100 and 77 for these two datasets, respectively. The results are shown in Figure 2.

Several conclusions can be drawn from result in Figure 2: i) When the training set is fixed, the performance improvement of using AdaBoost alone (B.ADP) is less than that of i.Boost on the two data sets. ii) Interactive boosting could improve the performance of ADP iteratively by 24% and 16% on these two data sets respectively. iii) The performance of interactive boosting is consistently better than that of AdaBoost by up to 21% and 10.2%. Similar results are obtained on other data sets from UCI repository, which show the superior performance of i.Boost.

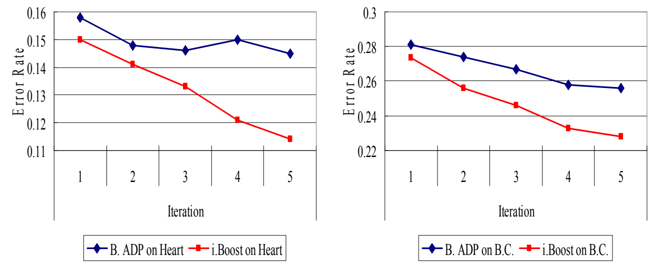


Fig. 2 i.Boost vs AdaBoost on Benchmark data sets

4.2. Comparison to Relevance Feedback

The second experiment is designed to compare the performance of i.Boost and classic Relevance Feedback (RF). The dataset used are COREL image databases. It contains 3000 color images which are roughly categorized into 30 classes. Each class contains 100 images. For simplicity in this experiment we randomly pick up two classes of images for classification. One-thirds of the images are used for training while two-thirds is used for testing. The image features we used to represent the color images are listed in Table 1.

Table 1: Image features extracted for COREL data set

Feature Name	Description	Length
ColorHistNM	Normalized Color Histogram	32
ColorMmtNM	Normalized Color Moments for HSV space	9
WvNM	Normalized Wavelet Moments for texture	10

From the experiment result in Figure 3, we can conclude that: i) i.Boost and Relevance Feedback starts with similar performance in iteration 1; ii) As iteration goes on, simple relevance feedback gain less performance improvement than i.Boost. It could be explained by that the reinforcement training introduced in i.Boost gives it more power in learning from the new data in relevance feedback.

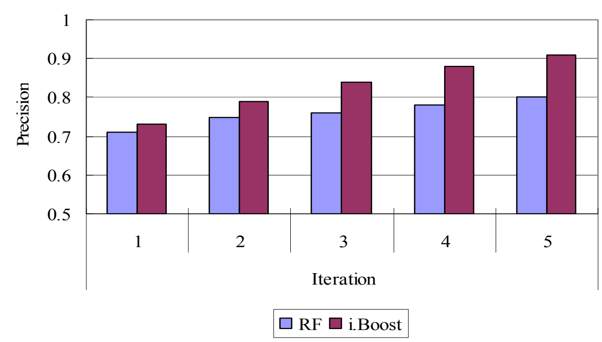
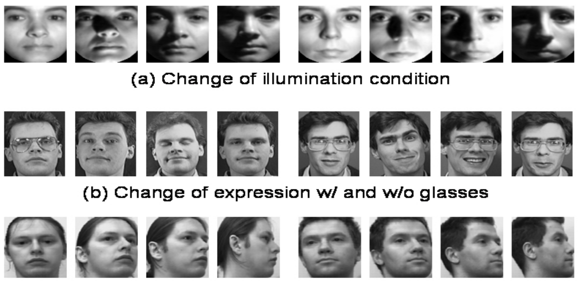


Fig. 3 i.Boost vs simple Relevance Feedback on COREL

4.3 Comparison to state-of-the-art techniques on face recognition

To evaluate how well i.Boost works, we test it on three benchmark face image databases with change of illumination, expression and head pose, respectively. Harvard Face Image database [10] consists of grayscale images of 10 persons. Each person has totally 66 images which are classified into 10 sets based on increasingly changed illumination condition. The ATT Face Image database [11] consists of 400 images for 10 persons. The facial images have resolution of 92×112 with different expressions, with or without glasses under almost same illumination condition. The UMIST Face Database [12] consists of 564 images of 20 people, which covers a range of poses from profile to frontal views. We randomly chose one person's face images as positive and the rest face images of others are considered as negative. In all experiments one third of the images in the database are randomly chosen as training set while the rest are used as testing set. Figure 4 gives some example images from the databases.



(a) Change of illumination condition

(b) Change of expression w/ and w/o glasses

(c) Change of head pose

Fig. 4 Example face images

For comparison purpose, six state-of-the-art projection-based techniques are also tested on the same databases: Eigenface and Fisherface are two of the most widely used techniques in face classification [9], LDA [3], KMDA [13], BDA [14], KBDA [14]. To play fair, the results for these techniques are obtained after 5 iterations of relevance feedback accumulation.

Table 2. i.Boost vs state-of-the-art techniques

Error Rate (%)	Methods	Harvard Dataset			ATT Dataset	UMIST Dataset
		Subset 1	Subset 2	Subset 3		
Methods	Eigenface	1.2	5.4	25.3	28.1	38.3
	Fisherface	0.7	1.4	3.7	19.5	31.2
	LDA	1.5	3.2	7.5	10.3	23.8
	BDA	1.78	3.1	8.9	12.6	20.6
	KDEM	**0.6**	1.1	2.9	8.1	17.5
	KBDA	1.3	1.9	3.4	7.6	16.3
	ADP	0.92	1.67	2.01	9.3	19.6
	i.Boost	0.68	**0.8**	**1.82**	**6.9**	**15.4**

The results are listed Table 2 with smallest error rate in bold. It is clear that i.Boost performs best in 4 out of 5 tests and second to nonlinear classifier KDEM in one test. It is clear that i.Boost provide more robustness to the changes of illumination, expression and pose than other techniques.

5. CONCLUSION AND FUTURE WORK

Content-based image retrieval applications often suffer from small sample set and high dimensionality problem. Relevance feedback and boosting have been widely used to alleviate those problems. In this paper, we propose a novel interactive boosting framework to integrate relevance feedback into boosting scheme for content-based image retrieval. Compared to the traditional boosting scheme, the proposed method obtains more performance improvement from the relevance feedback by putting human in the loop to facilitate learning process. It has obvious advantage over the classic relevance feedback method in that the classifiers are trained to pay more attention to wrongfully predicted samples in user feedback through a reinforcement training process. It is clear that the framework can bridge the gap between high-level semantic concept and low-level image features better.

I.Boost is implemented according to the iterative boosting framework by using ADP as base classifiers. It not only combines but also enhances a set of ADP classifiers into a more powerful one. To evaluate the performance of i.Boost, it is tested on several applications and compared to normal boosting, relevance feedback and 6 state-of-art projection based classifiers. The experiment results show the superior performance of i.Boost and the interactive boosting framework.

We will continue this research work in the following direction: 1) accommodating active learning techniques in the relevance feedback; 2) using different techniques to implement the base classifiers and 3) evaluate the performance difference among different boosting schemes.

Acknowledgement: This work was supported in part by the Army Research Office (ARO) grant under W911NF-05-1-0404, and by the San Antonio Life Science Institute (SALSI).

6. REFERENCE

[1] X. Zhou and T. S. Huang, "Relevance feedback in image retrieval: a comprehensive review," ACM Multimedia Systems Journal, special issue on CBIR, 8(6): pp. 536-544, 2003.

[2] I. T. Jolliffe, *Principal Component Analysis*. 2nd edition, New-York: Springer-Verlag, 2002.

[3] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd edition, John Wiley & Sons, Inc., 2001.

[4] J. Yu and Q. Tian, "Adaptive discriminant projection for content-based image retrieval", *Proc. of Intl. Conf. on Pattern Recognition*, Hong Kong, August 2006.

[5] M. Kearns and L. Valiant, "Cryptographic limitations on learning Boolean formulae and finite automata," J*ournal of the ACM*, 1994.

[6] Y. Freund and R. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence,* 14(5):771-780, Sep., 1999.

[7] L. Reyzin and R. Schapier, "How Boosting the Margin Can Also Boost Classifier Complexity," *Intl. Conf. on Machine Learning*, 2006.

[8] J. Friedman *et al.*, "Additive logistic regression: a statistical view of boosting," *the Annals of Statistics*, April, 2000.

[9] G. Salton and M. J. McGill, *Introduction to modern information retrieval*, New York: McGraw-Hill Book Company, 1992.

[10] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. PAMI*, Vol. 19, No. 7, July 1997.

[11] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. PAMI,* Vol. 20, 1998.

[12] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," *IEEE Workshop on Applications of Computer Vision,* Sarasota FL, December 1994

[13] Q. Tian, Y. Wu, J. Yu, and T.S. Huang, "Self-supervised learning based on discriminative nonlinear features for image classification," *Pattern Recognition, Special Issue on Image Understanding for Digital Photographs*, Vol. 38, 2005.

[14] X. Zhou and T.S. Huang, "Small sample learning during multimedia retrieval using biasMap," *Proc. of IEEE Conf. CVPR*, Hawaii, December 2001.