Asymmetric Real Adaboost

Zhanjun Wang, Chi Fang, Xiaoqing Ding
State Key Laboratory of Intelligent Technology and System,
Tsinghua National Laboratory for Information Science and Technology
Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, P.R.China
$\{$*wangzhanjun, fangchi, dxq*$\}$*@ocrserv.ee.tsinghua.edu.cn*

Abstract

A cost-sensitive extension of Real Adaboost denoted as asymmetric Real Adaboost(RAB) is proposed. The two main differences between Asymmetric RAB and the naïve RAB are (1) a Chernoff measurement is used to evaluate the best weak classifier during training, rather than a Bhattacharyya measurement used in naïve RAB, and (2) the weights are updated separately for positives and negatives at each boosting step. The upper bound on training error is also provided. Experiment results are shown to demonstrate its cost-sensitivity when selecting weak classifiers, and also show that it outperforms previously proposed cost-sensitive extensions of Discrete Adaboost(DAB) and several extensions of Real Adaboost. Besides, it also consumes much less time than previously proposed DAB extensions.

1. Introduction

Rare events detection has become a universal concern in computer vision and pattern recognition, especially in detection problems, such as face detection [9], pedestrian detection. For all those problems, the number of uninterested areas is often millions of times more than that of the target area. Cost-sensitive classification methods, which pay unequal attention to different kinds of samples, are more suitable to those problems.

Adaboost [1] is one of the most successful algorithms in computer vision, especially in detection problems. Various cost-sensitive extensions for Discrete Adaboost [1] have been proposed, including asymmetric-Adaboost [10] and Cost-Sensitive Adaboost(CS-DAB) [6] etc. All of these algorithms are heuristic methods, trying to achieve cost-sensitivity by direct manipulation on the weights and confidence parameters of Adaboost. Masnadi-Shirazi et.al [7] recently proposed a new cost-sensitive method for Discrete Adaboost called Asymmetric Boosting and achieved better results than those heuristic methods. However, there is no analytical solution for the parameter α used in Asymmetric Boosting for weight-updating, and a time-consuming and imprecise bisection search is used to find the optimal α. Compared with so many cost-sensitive extensions of Discrete Adaboost, few has been done to Real Adaboost[8]. In this paper, we propose an asymmetric extension for Real Adaboost. A theoretical analysis is provided, showing that a Chernoff distance is used to select the best weak learner during training, compared with a Bhattacharyya distance used by naïve Real Adaboost.

The rest of the paper is organized as follows: Section 2 introduces the Asymmetric Real AdaBoost algorithm and provides the new upper bound on training error. Section 3 compares our method with several previously proposed cost sensitive DAB methods and several extensions of RAB methods. Section 4 gives conclusions.

2. Asymmetric Real Adaboost

AdaBoost, mainly developed by Freund and Schapire [1], is a general method for improving the classification performance of any given learning algorithm. Schapire et.al [8] extended discrete weak classifiers to real-valued confidence-rated classifiers. Friedman [2] showed the Real AdaBoost algorithm fits an additive logistic regression model by stagewise and approximate optimization of the loss function

$$J(F) = E\left[\exp(-yF(x))\right]$$

where $F(x)$ is the strong classifier of Adaboost, and y is the label.

For an asymmetric loss, we pay different attentions to different training sets, with cost factors of C_1 for misses and C_2 for false positives. And the new asym-

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

metric loss function is

$$J(F) = E[I(y=1)\exp(-yC_1F(x)) + I(y=-1)\exp(-yC_2F(x))] \quad (1)$$

We propose an Asymmetric Real Adaboost (Asymmetric RAB) to minimize this asymmetric loss with normalized costs of $c_1(c_1 = \frac{C_1}{C_1+C_2})$ and $c_2(c_2 = 1 - c_1)$, as shown in Algorithm 1.

Algorithm 1 Asymmetric Real Adaboost

- Given dataset $S = \{(x_1, y_1), ..., (x_m, y_m)\}$, where $(x_i, y_i) \in \chi \times \{-1, +1\}$, the weak classifier pool $\mathbf{H}$ and the number of weak classifiers to be selected T.
- Assign different costs $c_1, c_2 (c_1, c_2 \in (0,1), c_1 + c_2 = 1)$ to $S_+ = \{(x_i, y_i) \mid y_i = 1\}, S_- = \{(x_i, y_i) \mid y_i = -1\}$
- Initialize the sample distribution $D_1(i) = \frac{1}{m}$.
- For $t = 1, ..., T$
  1. For each weak classifier h in $\mathbf{H}$ do:
     (a) Partition χ into several disjoint blocks $X_1, ..., X_n$
     (b) Under the distribution D_t calculate W_b^j as Eq.3.
     Set the output of h on each X_j as Eq.8
     (c) Calculate the Chernoff measurement as Eq.7
  2. Select the h_t minimizing Z, i.e.
  $$Z_t = \min_{h \in \mathbf{H}} Z, h_t = \arg\min_{h \in \mathbf{H}} Z$$
  3. Update the sample weight distribution
  $$D_{t+1}(i) = \begin{cases} D_t(i)\exp\left[-c_1 h_t(\mathbf{x}_i)\right], i \in S_+ \\ D_t(i)\exp\left[c_2 h_t(\mathbf{x}_i)\right], i \in S_- \end{cases}$$
- The final strong classifier F is
$$F(\mathbf{x}) = sign\left[\sum_{i=1}^{T} h_t(\mathbf{x}) - b\right]$$
where b is a threshold whose default value is zero. The confidence of F is defined as
$$Conf_F(\mathbf{x}) = \left|\sum_{T} h_t(\mathbf{x}) - b\right|$$

2.1. Criterion for finding asymmetric weak hypotheses

According to Schapire et.al, Adaboost greedily minimizes the upper bound on training error by minimizing Z_t on each round [8]. Thus, given the asymmetric loss function in Eq.1, the weak learner for Asymmetric RAB should attempt to find a weak hypothesis h_t which minimizes (omitting t subscripts)

$$Z = \sum_i [I(y_i = 1)D(i)\exp(-c_1 h(x_i)) + I(y_i = -1)D(i)\exp(c_2 h(x_i))] \quad (2)$$

Using the domain-partitioning weak hypotheses given by Schapire et.al [8], let

$$W_b^j = \sum_{i; x_i \in X_j \wedge y_i = b} D(i) = \mathbf{Pr}_{i \sim D}[x_i \in X_j \wedge y_i = b] \quad (3)$$

which is the weighted fraction of examples which fall in block j with label b. Then Eq.2 can be rewritten as

$$Z = \sum_j \left[W_+^j e^{-c_1 h^j(x)} + W_-^j e^{c_2 h^j(x)}\right] \quad (4)$$

Using standard calculus, Eq.4 is minimized when

$$h^j = \frac{1}{c_1 + c_2} \ln\left(\frac{c_1 W_+^j}{c_2 W_-^j}\right) \quad (5)$$

Substituting it into Eq.4, then

$$Z_{\min} = \left[\left(\frac{c_1}{c_2}\right)^{-c_1} + \left(\frac{c_1}{c_2}\right)^{c_2}\right] \sum_j \left(W_+^j\right)^{c_2} \left(W_-^j\right)^{c_1} \quad (6)$$

This equation can be rewritten by eliminating the constant factor as

$$Z'_{\min} = \sum_j \left(W_+^j\right)^{1-c_1} \left(W_-^j\right)^{c_1} \quad (7)$$

which is a Chernoff distance.

To limit the magnitudes of the predictions in Eq.5, a “smoothed” value of h^j is

$$h^j = \frac{1}{c_1 + c_2} \ln\left(\frac{c_1 W_+^j + \varepsilon}{c_2 W_-^j + \varepsilon}\right) \quad (8)$$

Substituting Eq.7 into Eq.4, then

$$Z_{\min} < \left[(c_1/c_2)^{-c_1} + (c_1/c_2)^{c_2}\right] \sum_j W_+^{j\,c_2} W_-^{j\,c_1} + 2Nc^{-c}\varepsilon^c, \quad c = \min(c_1, c_2) \quad (9)$$

where N is the number of blocks in the partition, and Z will not be greatly degraded by smoothing if we choose $\varepsilon \ll (2N)^{-c}c$.

Asymmetric RAB will be the naïve RAB with $c_1 = 0.5$, $c_2 = 0.5$, and $Z_{\min} = \sum_j \sqrt{W_+^j W_-^j}$ which is a Bhattacharyya distance. If the underlying distributions are Gaussian, the Chernoff error bound is never looser than the Bhattacharyya bound. Thus we could get a lower upper error bound by selecting optimal asymmetric weak classifiers with appropriate cost factors. Besides, Kullback-Leibler divergence [5] and Jensen-Shannon divergence [4] have also been proposed for selecting best weak classifiers for Real Adaboost on each training round, as shown in Table 1. Further experiments are provided to compare these measurements.

3. Experiments

We adopted Viola & Jones's protocol to compare Asymmetric RAB with several previously proposed cost-sensitive Discrete Adaboost, and several extensions of Real Adaboost algorithms. Haar features [9] and LUT method [3] are used as weak classifiers.

3.1. Chernoff measurement to select asymmetric weak classifiers

Using a training set of 10000 face images and 10000 nonface images and a test set of 10000 face images and 10000 nonface images, we trained Asymmetric RAB($C_1 = 2$, best performance on training set) and naïve RAB.

The first feature selected by each method is shown in Fig.1. Feature (a) was chosen by the naïve RAB and feature (b) was chosen by Asymmetric RAB. The middle plot shows the weight distributions on positive and negative samples of the two features. The lower plot shows the values of Chernoff measurement with different costs. When $c_1 = 0.5$ it is a Bhattacharyya measurement used in naïve RAB. We can see feature (a) achieves a lower value with a Bhattacharyya measurement($c_1 = 0.5$). However, when a larger cost is given on positive set($c_1 > 0.55$), feature (b) performs better than feature (a). This means if the cost of misses is much larger than that of false alarms which is common in detection problems, a Bhattacharyya measurement cannot select the optimal feature. The experiment result also validates this: after selecting the first feature, Asymmetric RAB achieves a FAR of 98.62% while naïve RAB still 100% when FRR is 0. This intrinsic asymmetric attention in feature selection results in Asymmetric RAB's cost-sensitive properties on different training sets.

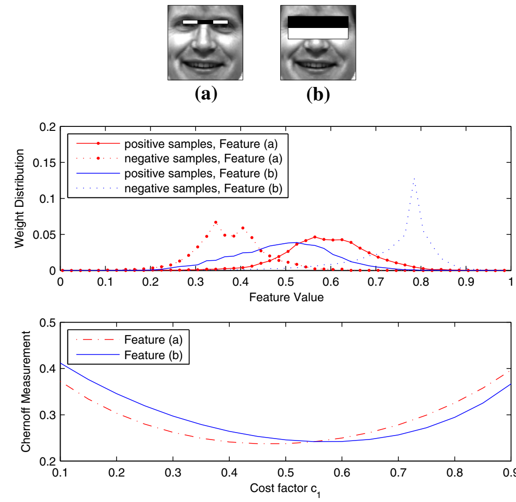

Figure 1. Different features selected by naïve RAB and Asymmetric RAB

Table 2. FAR of different boosting methods on training and testing sets

Features	10		70	
	Train	Test	Train	Test
CS-DAB	0.5244	0.8690	0.0707	0.1898
ADAB	0.4981	0.4852	0.0576	0.3193
ARAB	**0.3191**	**0.3549**	0.0043	0.0833
RAB	0.6099	0.6544	0.0243	**0.0587**
KL-RAB	0.4067	0.8482	**0.0004**	0.0895
JS-RAB	0.4577	0.4955	**0.0004**	0.2198

3.2. Comparison with different boosting algorithms

Using the same sets in last experiment, we trained 6 strong classifiers using CS-DAB($C = 3$), Asymmetric Boosting(ADAB, $C_1 = 3$), naïve RAB, Asymmetric RAB(ARAB, $c_1 = 0.75$), KL Boost(KL-RAB) [5] and JS Boost(JS-RAB) [4] for 70 weak classifiers. The number of bins for LUT method used in Real Adaboost based methods is 16, and the best cost factors are selected for each cost-sensitive methods when FRR is 0.

Fig.2 shows the error curves on training and testing sets, and Table 2 gives the false alarm rate with 10 and 70 weak classifiers of each method respectively. Compared with cost-sensitive boosting methods, ARAB achieves the best results both on training and testing sets. ADAB converges fast during the beginning stage of training, but in the end of training, it

Table 1. Different measurements for choosing weak classifiers

Measurements	Description
Bhattacharyya	$BH(x) = -\log \sum_j \left(W_+^j\right)^{0.5} \left(W_-^j\right)^{0.5}$
Chernoff	$CH(x) = -\log \sum_j \left(W_+^j\right)^{1-c_1} \left(W_-^j\right)^{c_1}$ $(0 < c_1 < 1)$
Kullback-Leibler	$KL(x) = \sum_j \left(W_+^j - W_-^j\right) \log \frac{W_+^j}{W_-^j}$
Jensen-Shannon	$JS(x) = \sum_j W_+^j \log W_+^j + \sum_j W_-^j \log W_-^j - 2\sum_j \left(\frac{W_+^j + W_-^j}{2}\right) \log \left(\frac{W_+^j + W_-^j}{2}\right)$

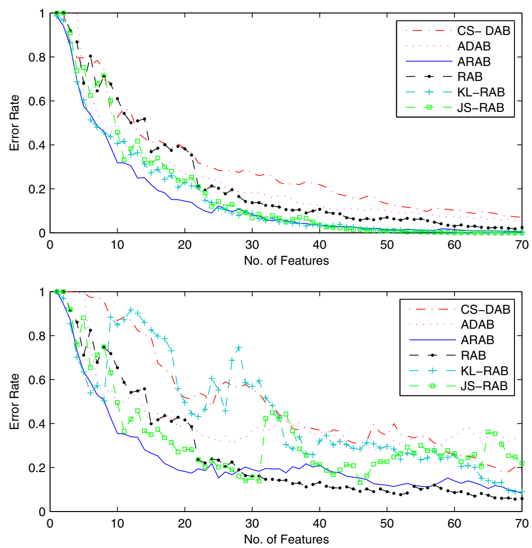

Figure 2. Error curves of different boosting algorithms

converges slower on testing set. KL-RAB and JS-RAB have comparable performances with ARAB on training set, however the generalization abilities are not as good as them on training set. ARAB converges faster than naïve RAB both on training and testing sets, and they achieve comparable results during the end of training. Besides, our method, ARAB, is much faster than CS-DAB and ADAB. Our method needs less than 7s to train a weak classifier on a PIV 2.6G PC, while CS-DAB and ADAB needs 664.7s and 1019.8s respectively.

4. Conclusions

An asymmetric extension for Real Adaboost is proposed in this paper. The new cost-sensitive algorithm minimizes the asymmetric exponential loss function, which leads to unequal attention to different samples intrinsically. It is shown that this method uses a Chernoff distance to select weak hypotheses, and experiment results validated this asymmetric feature selection criteria. Experiments also showed Asymmetric RAB(ARAB) outperforms previously proposed cost-sensitive extensions of Discrete Adaboost and several kinds of Real Adaboost. Due to this, we believe that asymmetric approach will promote Adaboost's performance in rare events detection and further applications could be conducted in detection problems.

References

[1] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *European Conference on Computational Learning Theory*, pages 23–37, 1995.

[2] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 28(2):337–407, 2000.

[3] C. Huang, H. Ai, B. Wu, and S. Lao. Boosting nested cascade detector for multi-view face detection. *ICPR*, 2:415–418, 2004.

[4] X. Huang, S. Z. Li, and Y. Wang. Jensen-shannon boosting learning for object recognition. *CVPR*, 2:144–149, 2005.

[5] C. Liu and H.-Y. Shum. Kullback-leibler boosting. *CVPR*, 1:587–594, 2003.

[6] Y. Ma and X. Ding. Robust real-time face detection based on cost-sensitive adaboost method. *ICME*, 1:465–468, 2003.

[7] H. Masnadi-Shirazi and N. Vasconcelos. Asymmetric boosting. *ICML*, pages 609–619, 2007.

[8] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, 1999.

[9] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *CVPR*, 1:511 –518, 2001.

[10] P. Viola and M. Jones. Fast and robust classification using asymmetric adaboost and a detector cascade. *NIPS*, 14, 2002.