Confidence Rated Boosting Algorithm for Generic Object Detection

Nayyar A.Zaidi, David Suter
Department of Electrical and Computer Systems Engineering
Monash University, Clayton 3800 VIC, Australia
{*nayyar.zaidi,d.suter*}*@eng.monash.edu.au*

Abstract

In this paper we propose a confidence rated boosting algorithm based on Ada-boost for generic object detection. Confidence rated Ada-boost algorithm has not been applied to generic object detection problem; in that sense our work is novel. We represent images as bag of words, where the words are SIFT descriptors extracted over some interest points. We compare our boosting algorithm to another version of boosting algorithm called Gentle-boost. Our approach generalizes well and performs equal or better than Gentle-boost. We show our results on four categories from the Caltech data sets, in terms of ROC curves.

1. Introduction

Object recognition problem is as old as computer vision itself. A large number of techniques have been developed to efficiently solve this problem. With the emergence of machine learning dream of a major break through which was expected still has to be realized, but a lot of improvements has been seen in case of object recognition [3, 4, 13, 15]. There has been significant debate over the potential of generative versus discriminative approaches to classification in machine learning community. The ideal choice of classifier depends a lot on the training data and also on the problem in hand. Though in the text mining community, discriminative classifiers have shown better results; it is quite difficult to pin point one against the other in object recognition research.

Generative classifiers model the distribution of data. Their aim is to come up with the model that best describes the population; once the model is created samples can be generated from the model [2, 7]. On the other hand discriminative classifiers finds the features in the distribution that discriminates it from the rest of classes [6]. In this paper we propose a discriminative classifier based on the confidence rated boosting algorithm [11] which has been applied successfully for facial expressions recognition and related tasks [19, 18], but has not been explored for object detection tasks. We also train a classifier using Gentle-boost based on [15] and compare its performance with our classifier.

In designing a classifier the choice of feature set is always critical. Features vary from modelling shape to modelling appearance and from being specific to being generic. For examples boundary fragments [12], k-Adjacent segments [5], SIFT, color, Tetons, Haar wavelets etc. Relying on one set of feature is not a good idea, there has been some study that discuss merits/demerits of features [8] and some that combine these feature sets to improve recognition performance [9]. This is not our goal. The reasons for using SIFT descriptors as described in section 3 has been primarily didactic, also SIFT features are proved to perform best [8].

The bag of words model, which originated from the text community, has proved very useful in images research [13]. The idea is to represent a document as the frequency of words in that document. An image in a bag of words model is the histogram of features present in the image. One problem with the bag of words model is that it ignores the interconnection of words that were present in the document or image; but still they can be very useful as they are an excellent representation of feature space. Similarly, learning a codebook of appearance parts or boundary fragments has been shown to be very useful for object categorization and detection tasks [3, 1]. The methods for creating the codebooks differ in the way as to which features are used in the codebook, selection of the features in the codebook, size of the codebook and also how different features inside the codebook are connected or related. In this paper we learn a codebook of visual words and represent each image as the bag of words representation and then train a discriminative classifier using confidence rated Ada-boosting. We test our algorithm with different codebook

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

sizes.

Boosting algorithms have been prevalent in object detection research in different forms and in different applications [15, 10, 16]. Boosting is a general learning algorithm. The idea is to combine a number of weak classifiers in a way to produce single strong classifier, which has performance much better than single monolithic classifiers. During training, the samples are reweighted according to the training error so weak classifiers, trained later, concentrate on harder training samples having higher weights. A number of variants of boosting exist in the literature. Sometimes they differ on the way weights are modified and sometimes on the choice of weak classifiers.

The discrete version of Ada-boost defines strong binary classifier F as $F(v) = sign(\sum_{t=1}^{T} h_t(v))$, where v is the input feature vector, T is the number of boosting rounds and h(v) are the weak learners. The weak learner in each round proposes feature dimension along with threshold to classify (sign of h(v) indicating the class whereas its value showing the confidence on the prediction). In domain partitioning confidence rated boosting algorithm, each weak classifier partitions the input space into finite bins and gives the prediction related to each partition along with the confidence values of each prediction. Rather than proposing single threshold and giving prediction signs, domain partitioning confidence rated prediction gives predictions for the full range of input feature space. We compare this version of boosting with another popular variant of boosting called Gentle-boost.

Rest of paper is organized as follows: In section 2 we present the overall approach in detail, experimental results then follows in section 3, in section 4 we conclude with future works.

2. Method

Instead of directly processing input images, we map the input image into the feature space. The classifier is trained on the feature space whose dimensionality is vastly reduced as compared to the input space. Assuming the sample input images (x) are given along with their labels (y) as $\{x_i, y_i\}_{i=1}^{N}$, where $x_i \in R^d$ and $y_i \in \{-1, +1\}$. As explained above we project this input space into feature space using projection functions $\phi_j \in \Phi : R^d \rightarrow R, j = 1, 2, ...M$. The goal of a discriminative classifier is to select a subset of discriminative features $\phi_j(x)$, which distinguish one class from the others.

Our goal is to learn the most discriminative features ϕ_j that produce the lowest error rates. We use a modified version of Ada-boost using confidence rated predictions from [11]. Details are given in algorithm 1.

Algorithm 1: Confidence Rated Boosting Algorithm

Given data set $(x_1, y_1), ..., (x_m, y_m)$, where $x \in \chi$ and $y \in -1, +1$, and m is the number of training samples. Initialize the distribution, such that for all training samples $D_t(i) = \frac{1}{m}$

for $t = 1, 2, ...T$
- Normalize the weights.
- Train a weak classifier $f_j(x|d_t)$ on feature $\phi_j(x)$
- Evaluate cost of the feature using eq.8
- Find the best feature ϕ_t using eq.9
- Update weights as: $d_{t+1} = d_t exp[-y_i f_t(x|d_t)]$

Output: Final Strong Classifier

$$F(x) = sign(\sum_{t=1}^{T} f_t(x|d)) \tag{1}$$

with feature set $\{\phi_t : t = 1, 2, ..T\}$

Representing an image as a bag of words model projects input space into a feature space. The input to the training algorithm will be the histogram of images (Algorithm 1) along with appropriate labels. For each feature ϕ_j the weighted distribution of positive and negative samples is defined as

$$h_j^+(x|d) = [p(\phi_j(x)|y = +1) * d(x|y = +1)]/D_j^+ \tag{2}$$

$$h_j^-(x|d) = [p(\phi_j(x)|y = -1) * d(x|y = -1)]/D_j^- \tag{3}$$

Here D_j^+ is the normalization factor to make h_j^+ a distribution. A weak classifier of feature ϕ_j is defined as

$$f_j(x|d) = \frac{1}{2} log \frac{h_j^+(x|d) + \epsilon}{h_j^-(x|d) + \epsilon} \tag{4}$$

For each feature ϕ_j we partition the region $[min(\phi_j(x)), max(\phi_j(x))]$ into several disjoint k bins $X_j^1, X_j^2, ...X_j^k$ where

$$h_j^+(k) = \sum_{\phi_j(x_i) \in X_j^k \wedge y_i = +1} (d_i / D_j^+) \tag{5}$$

$$h_j^-(k) = \sum_{\phi_j(x_i) \in X_j^k \wedge y_i = -1} (d_i / D_j^-) \tag{6}$$

where D_j^{+-} are the normalization factors and $k = 1, 2, ...K$. Equation 4 can be written as

$$f_j(k) = \frac{1}{2} log \frac{h_j^+ + \epsilon}{h_j^- + \epsilon} \tag{7}$$

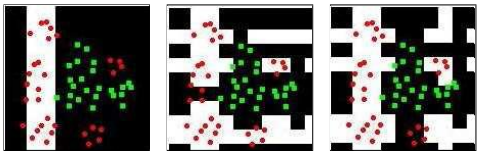

Figure 1. Shows the application of classifier after first, third and fifth boosting iteration on two dimensional data. White area is classified as belonging to class A (red blocks) and black belonging to class B (green blocks)

The cost of each feature ϕ_j is calculated as

$$C_j(k) = 2\sum_k \sqrt{h_j^+(k)h_j^-(k)} \tag{8}$$

The best feature is selected by

$$t = argmin_j C_{t,j}(x) \tag{9}$$

where t is the boosting iteration.

We show a demonstration of algorithm1 in figure 1, where red and green blocks are randomly generated two-dimensional data. We show the first few iterations of the boosting algorithm. As can be seen that the few weak classifier can classify data quite efficiently.

3. Experimental Details

Given some unlabeled images our goal is to classify images as if containing certain category or not. We have used Caltech data set for training and testing our learning algorithm as it is moderately challenging and has significant scale variations. Statistics of the training and testing images are shown in following table 1.

From the training images we extract visual words (explained below) and form a visual codebook. We used SIFT descriptors as described in section 1 extracted around Harris-Affine interest points as visual words. The vector quantization of images is carried out by k-means clustering. The code for the interest point detector, and for feature extraction were downloaded from VGG Oxfords website. Once the visual words are computed for an image, each image is represented by a histogram of visual words contained within the image (the bag of words model).

Table 1. Training/Testing Statistics

	Training	Testing
Motorbikes	248	578
Airplanes	322	752
Faces	135	315
Cars Back	158	368

For each category in the Caltech data set, images are split into training and testing images (table 1). Testing images are not used in any form during the training procedure. Around 200k patches are extracted from the training images of four categories and clustered into 250, 500 and 1000 size codebook. Each image is represented as a histogram of codebook words present in the image, thus representing each image as 250, 500 and 1000 size vector. The categories are trained in a one against the all fashion. We have not used any background images. We report classification results in form of ROC curves.

It is not the goal of this study to show the effect of codebook size on the classification but as discussed in section 4 this can be a very interesting area of research (some work has been done in [17]). For training using Gentle-boost we used the code from [14].

Airplanes: classification results (ROC) for airplanes are shown in the first column of figure 2. Both Ada-boost and Gentle-boost has similar performance with codebook sizes 250, 500 and 1000.

Cars: classification results (ROC) for cars are shown in second column of figure 2. Like airplanes category Ada-boost and Gentle-boost give similar performance with codebook 250, 500 and 1000.

Faces: classification results (ROC) for faces are given in the third column of figure 2. Gentle and Ada-boost performs similarly in case of codebook of size 250 but Ada-boost outperforms Gentle-boost if size of the codebook is increased to 500 and 1000.

Motorbikes: classification results for motorbikes are shown in fourth column of figure 2. As with the faces category Ada-boost and Gentle-boost performs similarly in case of codebook of size 250 but Ada-boost outperforms Gentle-boost if size of the codebook is increased to 500 and 1000.

4. Conclusion and Future Works

In this paper we have proposed domain partitioning Ada-boost learning algorithm for generic object classification and has compared its performance with Gentle-boost. Our algorithm takes more time to train (almost double as compared to gentle boost), testing time is almost the same. It has similar or better performance to Gentle-boost for codebook size 250 but performs much better than Gentle-boost when the size of the codebook is increased from 250 to 500 and 1000, this can be extremely interesting area of research. Our proposed al-

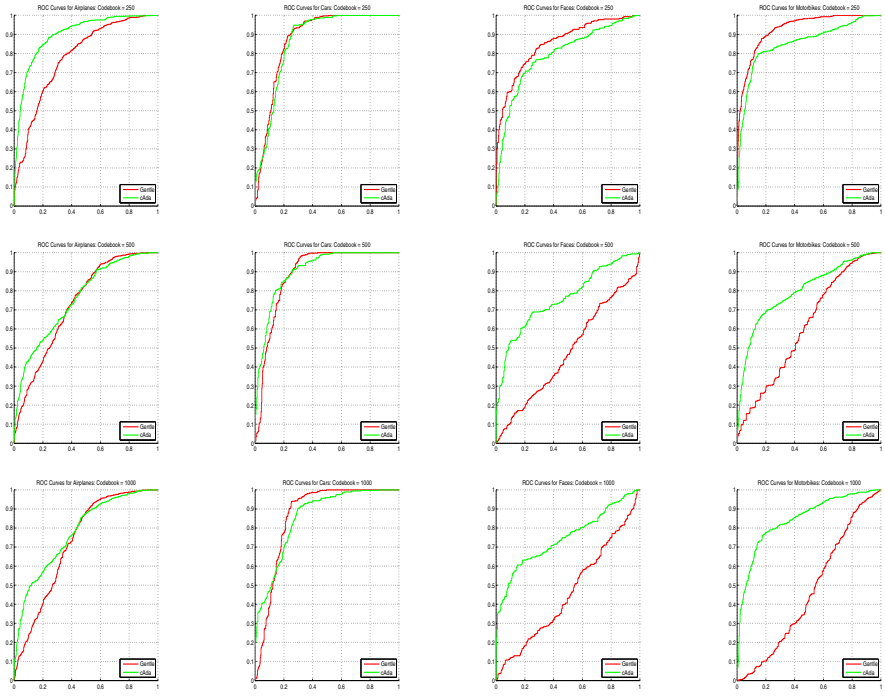

Figure 2. ROC Curves for Airplanes, Cars, Faces and Motorbikes trained with Ada and Gentle boosting.

gorithm has the advantage that it can be easily modified for producing discriminative codebooks by selecting the most discriminative codeword at each stage [17]. We are currently working on adapting our algorithm to produce such codebooks. We are also modifying our algorithm for learning jointly [15] so that features are shared across different categories.

References

[1] S. Agarwal and D. Roth. Learning a sparse representation for object detection. In *ECCV*, 2002.

[2] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning*, 2003.

[3] C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka. Visual categorization with bags of keypoints. In *ECCV*, 2004.

[4] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsuprevised scale invariant learning. In *CVPR*, 2003.

[5] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *IEEE PAMI*, 2007.

[6] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 2000.

[7] T. Hoffmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 2001.

[8] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE PAMI*, 2005.

[9] A. Opelt and A. Pinz. Fusing shape and appearance information for object category detection. In *BMVC*, 2006.

[10] A. Opelt, A. Pinz, and A. Zisserman. A boundary fragment model for object detection. In *ECCV*, 2006.

[11] R. Shapire and Y. Singer. Improved boosting algorithms using confidence rated predictions. In *Conf. on Computational Learning Theory*, 1998.

[12] J. Shotton, A. Blake, and R. Cipolla. Contour based learning for object detection. In *ICCV*, 2005.

[13] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. Discovering objects and their location in images. In *ICCV*, 2005.

[14] A. Torralba. A simple object detector with boosting. In *ICCV shortcourses*, 2005.

[15] A. Torralba, K. Murphy, and W. Freeman. Sharing visual features for multiclass and multiview object detecion. *IEEE PAMI*, 2007.

[16] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple featrues. In *CVPR*, 2001.

[17] L. Wang. Towards a discriminative codebook: Codeword selection across multi resolution. In *CVPR*, 2007.

[18] Y. Wang, H. AI, B. Wu, and C. Huang. Real time facial expression recognition with adaboost. In *ICPR*, 2004.

[19] R. Xiao, W. Li, Y. Tian, and X. Tang. Joint boosting feature selection for robust face recognition. In *CVPR*, 2006.