Online Learning Objectionable Image Filter Based on SVM

Yang Liu, Wei Zeng, and Hongxun Yao

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
{yliu,wzeng}@jdl.ac.cn; yhx@vilab.hit.edu.cn

Abstract. In this paper we propose an on-line learning system for objectionable image filtering. Firstly, the system applies a robust skin detector to generate skin mask image, then features of color, skin texture and shape are extracted. Secondly these features are inputted into an on-line incremental learning module, which derives from support vector machine. The most difference between this method and other online SVM is that the new algorithm preserves not only support vectors but also the cases with longest distance from the decision surface, because the more representative patterns are the farthest examples away from the hyperplane. Our system is tested on about 70000 images download from the Internet. Experimental results demonstrate the good performance when compared with other on-line learning method.

1 Introduction

With the rapid development of internet, more and more digital information can be acquired from the world-wide web. Although the new tool bring us convenience for digital communication, it also makes it more easy for young persons to browse contents on adult web site. This side effects of internet has brought increasing severe social problems, such as sexy crime. It is urgent for us to set up effective safety net between internet and personal computer.

Some schemes have been implemented to solve the problem. All these methods can be divided into two types. One is to collect IP addresses within which adult contents exist. Once the web server finds these web sites are being accessed, it will block this behavior. The other is to stop the links through text semantic analysis. Some search engines clean the sexy web sites by this method. Although these two methods are effective partially, they face a lot of problems, since much more new sexy web sites are set up every day and text semantic analysis is difficult, even more, some famous search engine does not take measures, such as Google.

The two mentioned approaches both use net tools to avoid adult web site being accessed. In order to more effectively screen these objectionable images, a new method is being developed in which image content analysis technology is used. Forsyth's research group has done the pioneer work[1]. They designed and implemented a system that can tell whether an image contain warm description

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3331, pp. 304–311, 2004.
© Springer-Verlag Berlin Heidelberg 2004

content. In their scheme, skin detection and shape processing were combined to detect human body. Wang et al. presented a system of screening objection images for practical applications [2]. Wang's method employed a combination of an icon filter, a graph-photo detector, a color histogram filter, a texture filter and a wavelet-based shape matching algorithm. The image that passed histogram analysis, texture analysis and shape matching was classified as objectionable image. In [3] Jones and Rehg used statistical method for skin detection and artificial neural network to classify.

As our system is designed for gateway to identify objectionable image, system with on-line learning capacity are necessary. Meanwhile, through our analysis, objectionable image filtering system is unlike other two classes pattern recognition problem, because the class correspond to benign images is an open field. It is difficult to select the negative cases, if objectionable images are regarded as positive class. What type and how many images are selected? All the problems urge us to develop an on-line learning system for filtering objectionable images.

In this paper, we propose a new method for incremental learning support vector machine. In our algorithm, not only the support vectors but also the cases farthest from decision surface are reserved for the next learning step. Our method is different from the approaches in [4], [5], in which only the support vectors are preserved. Experimental results show that the proposed method has better performance.

2 Incremental Learning Support Vector Machine

In this section support vector machine will be introduced briefly, then we will give our new incremental learning support vector machine.

2.1 Support Vector Machine

Support vector machine is derived from Vapnik's work on statistic learning theory[6],[7], in which he regard the procedure of searching a decision rule as a functional problem. The goal is to find a function that can minimize the expected risk, described as (1).

$$R(u) = \int\int L\left(y, f\left(X\right)\right)\ dp(y|X)\ dp(X). \tag{1}$$

where $p(X)$, $p(y|X)$ are the distribution of the examples and their classification respectively. Unfortunately, these two functions can not be known in general and SVM has to estimate them from example set $(X_i, y_i), i \in N$. Instead of computing formula (1), SVM converts this problem into finding a hyperplane that maximize the minimal distance between example and the plane. The decision function can be written as $f(X) = sign(\sum_i^N \lambda_i K(X_i, X) + b)$, where all λ_i are calculated through the optimization process. The examples with non-zero λ_i are called support vectors and K is a kernel function that can convert non-linear problem into linear problem.

2.2 Incremental Support Vector Machine

Support vectors have an important property, that if SVM is trained only on support vectors, the same result will be acquired as the one trained on the whole data set. However it does not mean that the support vectors can represent the whole data set in all directions, since support vectors are only description of the decision boundary between the two classes examples, but not of the examples distribution. Well then, how can we select a small part of examples which can represents the complete data set more effectively? What are the more representative ones except for the support vectors? In our experiments, we found that some support vectors are ambiguous ones that can be classified into either class. Are such examples representative? In people's intuitive opinion, the more "classic" ones are the more representative. What are the "classic" ones in support vector machine in our objectionable image filtering system? It is observed that most of the examples with largest absolute value of decision function are easier to be "classified" since they are clear semantically. It means that these examples are more representative than support vectors for complete data set. This property motivates us to reserve these farthest examples for the next learning step in incremental SVM.

The second reason that we do like this is of concept inconsistent. In online learning process learning machines have to face the problem of drifting concept, because the distribution of the incremental data given to the machine in continuous learning steps may be different. Only the support vectors are not sufficient in this situation. We may ask, what if the farthest examples are put into incremental data set? Thus, support vectors depict one margin of the distribution of previous training data, while these farthest ones depict the other margin.

The upper mentioned reason drive us to update the SVM's optimization object as (2).

$$
\begin{aligned}
&Min : \Phi(W,\xi) = \frac{1}{2}(W \bullet W) + C(\sum_{i=1}^{N}\xi_i + \sum_{i\in S}\xi_i + \sum_{i\in F}\xi_i) \\
&subject \quad to : y_i(W^T X_i + b) \le 1 - \xi_i, i = 1,\cdot\cdot\cdot, N \\
&\qquad\qquad\quad \xi_i \ge 0, i = 1,\cdot\cdot\cdot, N \qquad\qquad\qquad .
\end{aligned}
\tag{2}
$$

Where S and F are support vectors set and farthest cases set respectively. In comparison with SV incremental support vector machine, our new incremental support vector machine preserves not only the support vectors but also the vectors with the longest distance far from the decision surface for next learning step. Although the optimization function is different only in the last term $\sum_{i\in F}\xi_i$ from Stefan's method, the idea has proceeded. The farthest examples incorporate new meanings.

As described in [5], we also make the preserved examples have more punishment when an error occurs on them than an error on new incremental training data. Thus, the formula (2) is rewritten as (3).

$$\Phi(W,\xi) = \frac{1}{2}(W \bullet W) + C(\sum_{i=1}^{N}\xi_i + L\sum_{i\in S}\xi_i + L\sum_{i\in F}\xi_i), \tag{3}$$

where L is defined as:

$$L = \frac{\#All\_Trainning\_Examples}{\#SV + \#Farthest\_Examples}$$

3 Skin Detection and Feature Extraction

3.1 Skin Detection

Generally, objectionable images have much area of skin, so we decide to extract features from skin mask image, in which skin pixels are labeled as 1, and non-skin pixels as 0. Then, skin detection process can be treated as a two class classifying problem.

As described in [11], a skin color detector with the same performance can always be found no matter what color space is chosen, if there exists an invertible transformation between the color spaces. Since, statistical histogram of skin can describe the skin distribution accurately, the non-parametric method is adopted. Experiments show that it can acquire better performance than that of Gauss Mixture Model (GMM). Due to the space limitation, the result is not given out. According to the theory of Bayesian classification, the classifier for skin detection can be described as formula(4).

$$P(skin|rgb) = \frac{[P(rgb|skin)P(skin)]}{[P(rgb|skin)P(skin) + P(rgb|!skin)P(!skin)]}. \tag{4}$$

Where $P(skin|rgb)$ represents the probability of the pixel with the value rgb being skin. When the probability is larger than a threshold, then the pixel is regarded as pixel. Varying the threshold, we can acquire the ROC (receiver operating characteristic) curve. In order to save space, the ROC curve is not depicted. At last, the threshold corresponding to the rate of correct detection is 0.841 and the error rate 0.143 is adopted. All these results are tested on one hundred million skin pixels.

Because the reflection of the skin, some skin pixels are desaturated. Those skin pixels will be classified as non-skin pixel. Some pixels in shadow region will also be miss-classified because the color is very close black. These two types errors will produce many skin holes in the detected skin regions. Therefore, we apply a region analysis to eliminate the small holes. If the small region is surround by skin color and its color is closed to white or black. The region is marked as skin region.

At the same time, some pixels that belong to the background are detected as skin region. It produces some small skin regions in the background. To remove these small skin regions, we count the size of the connected small skin region. If connecting small skin area is less than 0.05 of the whole detected skin pixel number, the region is classified as background. The median filter is used to filter out the noise at the last step. Figure 1 gives some detection results.

Fig. 1. The results of skin detection

3.2 Feature Extraction

Following the output of skin detection process, feature extraction module is performed. To represent the objectionable images effectively, three kind of features are selected, including color feature, texture feature and shape feature. Details are described as the following.

Color feature: Such as mean of skin color probability, variance of skin color probability; The mean and variance of skin probability reflect the distribution of skin in the skin probability space. The two measures are calculated by (5)and(6).

$$P_{mean} = \frac{1}{N} \Sigma_{j \in skin} P_j(skin|rgb), \tag{5}$$

$$P_{var} = \frac{1}{N} \Sigma_{j \in skin} P_j(skin|rgb) - P_{mean}, \tag{6}$$

where the N is the number of detected skin pixels.

Texture feature: Include texture contrast, texture coarseness. The texture contrast is defined as (7)

$$T_{con} = \frac{\sigma}{(\alpha_4)^4}, \tag{7}$$

where $\alpha_4 = \mu_4/\sigma^4$ is kurtosis. μ_4 is the forth order moment and σ^2 is variance; And the texture coarseness is formulated as (8).

$$T_{crs} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} S_{best}(i,j), S_{best}(i,j) = 2^k. \tag{8}$$

where m and n is the size of the window used to detect skin region. More details can be found in [12].

Shape feature: Include skin region area, region edge intensity, zernike moment.The area feature describes the proportion of skin pixel to the image. The compactness feature is calculated by (9).

$$C = \frac{(region\_border\_length)^2}{area}. \tag{9}$$

The zernike moment descriptor is represented by a set of ART (Angular Radial Transform) coefficients, and the definition can be formularized as (10).

$$F_{nm} = \langle V_{nm}(\rho,\theta), f(\rho,\theta) \rangle = \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho,\theta) f(\rho,\theta) \rho d\rho\theta d\theta, \tag{10}$$

where F_{nm} is an ART coefficient of order n and m, $f(\rho, \theta)$ is an image function in polar coordinates, and $V_{nm}(\rho, \theta)$ is the ART basis function that are separable along the angular and radial directions. [13] describes more details.

The three type features involve low level and high level features. The former two are low level, and the latter is high level.

The third module is the classifier. There are two available incremental learning algorithms for us: incremental decision tree [10] and incremental support vector machine. For the sake of implementing the incremental learning capacity, decision tree has to maintain an information list for every node that stores prior classification info. It is very difficult to deal with the list, because it will cost so much memory when the data set is large. While incremental learning support vector machine has not this shortcoming. In the filtering system our incremental support vector machine is adopted.

4 Experiments

4.1 Experiment Design

In our experiment, we downloaded 70,406 nature images from the Internet and classified them into 11,349 objectionable images and 59,057 benign images by hand. For convenience, the objectionable images set is called positive case set, while the other one is called negative case set. The three kinds of features comprise a vector containing 49 components. Before training, every dimension attribute is scaled to expectancy 0 and variance 1.

To check the performance of our incremental learning support vector machine, all collected images are divided into six groups uniformly. Every group contains 1891 objectionable images and 9842 benign images. The former 5 groups are used for incremental training data set and the last group used for the test set. In order to eliminate contingency factor, all the following results are obtained on six test processes. Every group is used as test set rotatively and other groups are used as training data for five incremental learning steps. Before training the SVM in the new learning step, we incorporate all support vectors and five percent of farthest examples into the incremental training data set. Radial basis function is used as kernel with gama=0.01.

4.2 Experimental Results and Analysis

In this section the accuracy, precision and recall are defined as the followings respectively.

$$Accuracy = \frac{Num(o) + Num(n)}{N},$$

where $Num(0)$ is the number of truly classified objectionable images and $Num(n)$ is the number of truly classified non-objectionable images. The other two terms are defined as followings,

$$recall = \frac{a}{N_1}, \quad precision = \frac{a}{N_2}.$$

Table 1. Accuracy of the three earning algorithm (Method 1 is batch SVM, method 2 is our method and method 3 is the method in [5])

Test number	Method 1	Method 2	Method 3
1	91.52 %	91.52%	91.52%
2	91.73 %	91.01%	89.27%
3	91.40 %	91.40%	90.05%
4	91.65 %	91.00%	88.85%
5	91.50 %	90.85%	88.50%

Table 2. Precision of the three methods (Method 1 is batch SVM, method 2 is our method and method 3 is the method in [5])

Test number	Method 1	Method 2	Method 3
1	75.66 %	75.66%	75.66%
2	74.10 %	70.28%	63.35%
3	72.62 %	73.14%	67.72%
4	74.17 %	70.66%	63.09%
5	71.88 %	70.60%	61.82%

In these two formulas a is the number of truly classified objectionable images, N_1 is the number of all the objectionable images, and N_2 is the number of images that are classified as objectionable images.

Table 1 illustrates the accuracy of the two different incremental support vector machines and the batch SVM on objectionable images filtering system. From the test result, we find that the accuracy of incremental learning only with support vectors drops as the learning process continues. Our new method's accuracy has drops and rises, but it is always better than SV-incremental SVM.

Table 2 and 3 describe the precision and recall of the three methods. The same as the accurate our new methods has better performance than only support vector incremental learning algorithm.

Table 3. Recall of the three methods (Method 1 is batch SVM, method 2 is our method and method 3 is the method in [5])

Test number	Method 1	Method 2	Method 3
1	69.85 %	69.85%	69.85%
2	74.87 %	74.00%	73.21%
3	74.87 %	73.72%	73.09%
4	76.33 %	75.45%	74.20%
5	77.59 %	76.53%	74.90%

All the results tell us our incremental learning SVM has similar performance to that of batch SVM, But it is faster than batch SVM in learning process since it need little examples to be trained

5 Conclusion and Future Work

Our system makes a profitable attempt at to filter objectionable images through image contents analysis, although we are facing a lot of problems, for instance, how to make skin detection more robust in variant light condition and to extract more helpful visual feature? Our new incremental support vector machine has better performance than the traditional incremental learning method.

In a word, the proposed filtering system is effective and efficient. The most difference from other system is that it realizes on-line learning.

References

1. Margaret Fleck, David A. Forsyth, Chris Bregler: Find Naked People. Pro. 4th European Conf. on Computer Vision, UK, vol. 2, pp. 593-602, 1996.
2. James Z. Wang, Jia Li, Gio Wiederhold, Oscar Firschein: System for screening objectionable images. Computer Communications, vol. 21, no. 15, pp. 1355-1360, Elsevier, 1998.
3. Michael J. Jones, James M. Rehg: Statistical Color Models with Applications to Skin Detection. Technical report of the Cambridge Research Laboratory, No. 98-11, Dec, 1998.
4. N.A. Syed, H. Liu and K.K. Sung: Incremental Learning with Support Vector Machines. inProc. Int. Joint Conf. on Artificial Intelligence (IJCAI-99), 1999.
5. Stefan Ruping: Incremental Learning with Support Vector Machines. ICDM 2001.
6. V. Vapnik: The nature of statistical learning theory. Springer, New York, 1995.
7. V. Vapnik: Statistical learning theory. John Wiley and Sons, inc., 1998.
8. Joachims, T.: Making large-scale SVM learning practical. In B. Schölkopf, C. Burges & A. Smola, Advances in kernel methods-Support vector learning, Cambridge, MA, MIT Press.
9. Qixiang Ye, Wen Gao, Wei Zeng, Tao Zhang, Weiqiang Wang, Yang Liu: Objectional image detection on compression domain. International Conference on Intelligence Data Engineering and Automatic Learning, 2003, HongKong.
10. P. E. Utgoff: Incremental induction of decision tree. Machine Learning, 4:161–186, 1989.
11. Alberto Albiolt, Luis Torres, Edward J. Delp. Optimum: Color Spaces For Skin Detection. in Proceedins of ICIP2001.
12. Hideyuki Tamura, Shunji Mori, and Yamawaki: Textural Features Corresponding to Visual Perception. IEEE Trans. on Systems, Man, and Sybernetics, vol. SMC-8, no. 6, June 1978.
13. Rafael C. Gonzalez, Richard E. Woods: Digital Image Processing. Prentice Hall, 2002.