COMBINING MULTIPLE 2ν-SVM CLASSIFIERS FOR TISSUE SEGMENTATION

Yusuf Artan

Department of Electrical Engineering
Lehigh University
yua206@lehigh.edu

Xiaolei Huang

Department of Computer Science and Engineering
Lehigh University
xih206@lehigh.edu

ABSTRACT

In image classification problems, especially those involving tumor or precancerous lesion, we are usually faced with the situation in which the cost of mistakenly classifying samples in one class is much higher than that of the opposite mistake in the other class. Therefore it is essential to include cost information about classes in our classification methods. This paper applies a cost-sensitive 2ν-SVM classification scheme to cervical cancer images to separate diseased regions from healthy tissue. Using this method, we are able to specify a higher weight to the class that is deemed more important. To the best of our knowledge, cost-sensitive SVM based medical image classification has not been done before. We specifically target segmenting disease regions in digitized uterine cervix images in a NCI/NLM archive of 60,000 images. Our second contribution is the introduction of a multiple classifier scheme instead of the traditional single classifier model. Using the multiple classifier scheme improves significantly classification accuracy as demonstrated by our experiments.

Index Terms— Image classification, classification cost, cost-sensitive classifiers, tissue segmentation, multiple classifier system, segmentation evaluation, support vector machines

1. INTRODUCTION

To make images searchable by content in large medical archives, it is very important to reliably segment and label different tissue regions, especially biomarker regions, in images. We consider the automated segmentation problem in a very large archive of 60,000 digitized uterine cervix images, created by the National Library of Medicine (NLM) and the National Cancer Institute (NCI). These images are optical *cervigram* images acquired by Cervicography using specially-designed cameras for visual screening of the cervix, and they were collected from the NCI Guanacaste project for the study of visual features correlated to the development of precancerous lesions. The most important observation in a cervigram image is the Acetowhite (AW) region, which is caused by whitening of potentially malignant regions of the cervix epitheliuem (CE), following application of acetic acid to the cervix surface.

The goal of our research is to develop robust automatic segmentation methods for finding the Acetowhite and other important tissue region boundaries in a large portion, if not all, of the cervigrams in the NCI/NLM archive. This problem is very challenging because images in the archive has large variations in their appearance due to irregular illumination, shading and specularity, as well as differences in subject, camera positioning and other acquisition conditions.

In this paper, after reviewing the earlier works done in cervigram image analysis [1][2] and evaluating the potential of several classical color segmentation methods including clustering, model-based segmentation [3], Markov Random Fields [4] and Graph Cut [5], we turn to a discriminative color segmentation approach that exploits machine learning techniques such as Support Vector Machines (SVM) [6] to learn, from a ground-truth database, critical visual signs and construct efficient classifiers to distinguish different tissue types.

Tissue classification in cervical images involves the identification of precancerous lesions and biomarker regions such as Acetowhite. Since the impact of a false alarm is much less than a miss, it might be intuitive to build classifiers that are lenient on false alarms. However this does not mean that the false alarm rate could be arbitrarily high. On this complex pattern recognition problem, we can take an approach that is similar to Neyman-Pearson (NP) [11] based classification, with the objective of segmenting the diseased areas as accurately as possible while constraining false alarms under a threshold value. In this paper, we employ cost-sensitive 2ν-SVM classifiers that give a higher weight to the more important diseased (or biomarker) tissue class.

Traditional pattern classification systems use a feature descriptor and a particular classification procedure to determine the true class of a pattern. However, one often needs to find a large number of features to build accurate classifiers for given arbitrary problems. Classifiers that use a small set of features may result in unstable classification performance, depending on the discrimination power of the feature set and the complexity of the classifier. In cervical image classification, classifiers generally perform well only on a limited type of data due to large variations in image appearance. One question that arises is whether we can combine the decisions of several classifiers and obtain a new classifier with performance better than a single classifier working in isolation.

Multiple classifier systems (MCS) are developed for situations in which individual classifiers perform well only for limited types of data. By using an estimation algorithm, STAPLE [7], on the outputs of multiple classifiers, it is possible to get better classification performance than using a traditional single classifier.

The STAPLE algorithm uses iterative likelihood information about individual pixels to estimate the hidden true segmentation of diseased areas (AW) of a given test image. In our experiments we used a large number of classifiers for MCS such that the set of classifiers are sufficiently rich and all types of inputs are included in their specialty.

2. METHODOLOGY

For every image in a set of 939 cervigrams in the NLM/NCI archive, the ground truth boundaries of diseased AW and healthy CE tissue regions are drawn by medical experts. In this paper, we consider boundaries drawn by only one expert per image. After our initial experiments with these images, we discovered that the RGB color space does not perform very well separating AW and CE regions.

Table 1. Classification performances using different kernels

Classification Accuracy				
Image #	State	Linear	Poly	Rbf
Image 1	TP	% 70.2	% 66	% 67.6
	FP	% 7.65	% 6.3	% 4.75
Image 2	TP	% 61.2	% 58	% 59.7
	FP	% 5	% 4.1	% 4.2
Image 3	TP	% 47.2	% 37.9	% 42.3
	FP	% 3.2	% 1.74	% 1.45

Next, we converted from RGB to HSV color space and obtained nice separation between AW and CE regions. Thus the feature space is built by using the HSV color space information as well as gray-level intensity information following a normalization process. Results obtained using these set of features are very promising.

2.1. Image Preprocessing

Illumination variation causes specular reflections in the cervigram images. These specular reflection (or highlight) areas need to be removed in order to achieve high classification accuracies. Although we were able to remove most of the reflections in cervix regions of the images by using the reflection removal algorithm mentioned in [8], we also applied k-means clustering to further eliminate the reflections that are missed by the algorithm.

2.2. Training Set

The training set is built by taking samples from the Acetowhite (AW) region and the cervix (CE) region of images with expert boundary markings. For each image, 100 points are collected from the Acetowhite region as diseased (or positive) samples and another 100 points are collected from the CE region as healthy (or negative) samples. This is a time consuming process that is done manually. For the following experiments, a total of 35 images are used, of which, 18 images are used for training and 17 images are used for testing.

Using the training data, we built cost-sensitive support vector machine classifiers to separate the AW region from CE region. We have tested several kernels (i.e. linear, polynomial, Rbf kernels) to observe the variation in classification performances. Linear and Rbf kernels achieve good separation between different classes. Figure 1 shows the visual comparison between 3 different kernels on 3 sample test images shown in Row 1 of Figure 1, and Table 1 provides the numerical comparison values for these 3 sample images. It can be observed that the Rbf kernel has comparably good performance, with high True Positive (TP) fraction and low False Positive (FP) fraction.

Therefore, in the following sections, we will present our results using Rbf kernels. We have collected several thousand training points from our training images. In order to reduce the amount of computation during testing, we extracted a section of the images that contains both the cervix region (CE) and diseased region (AW).

2.3. Classification method

Support Vector Machine (SVM) is a classification method that is frequently used in image processing and pattern recognition applications. If we let $(\mathbf{x_i}, y_i)$, $i = (1, \cdots, n)$ denote the training data,where $\mathbf{x_i} \in R^d$ is a d-dimensional vector and $y_i \in \{-1, 1\}$. The objective of SVM is to build a decision hyperplane,

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{1}$$

that separates different classes with the maximum *margin*-distance between decision hyperplane and the closest point to this decision

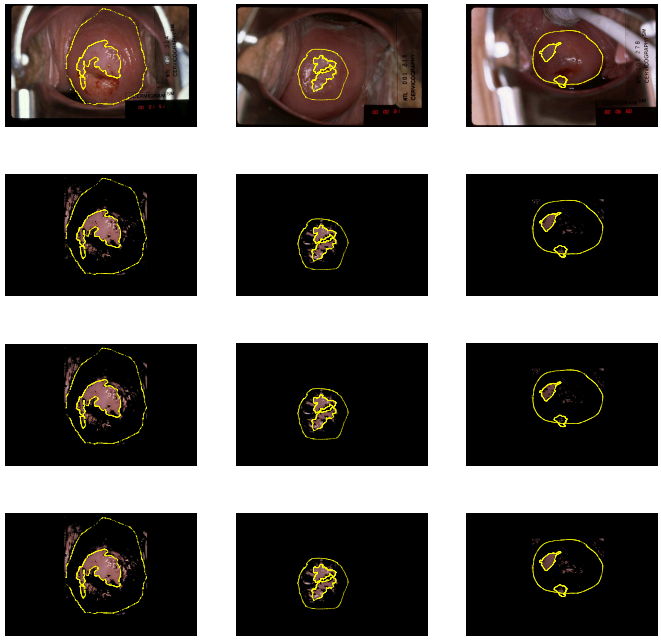

Fig. 1. Original test images (Row 1), Performances of Rbf (Row 2) kernel, polynomial (d=2) kernel (Row 3) and linear kernel (Row 4)

hyperplane [6]. Decision hyperplane is obtained after solving a convex quadratic programming problem shown in (2). The originally proposed SVM formulation (2) is usually referred as the C-SVM

$$\begin{aligned} \min \quad & \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\epsilon_i \\ \text{st.} \quad & y_i(k(\mathbf{w}, \mathbf{x_i}) + b) \geq 1 - \epsilon_i \quad \text{for} \quad i = 1, 2, \cdots, n \\ & \epsilon_i \geq 0 \quad \text{for} \quad i = 1, 2, \cdots, n \end{aligned}$$

where $C \geq 0$ is an error penalty parameter that controls overfitting.

In certain classification applications, there is a disadvantage to the above C-SVM formulation (Eqn. 2), which penalizes the errors of both classes equally since the C value is the same for all classes. However, there are cases where one class is more important than the other, therefore it is necessary to assign a higher cost to misclassification for one class than the other. For instance, classifying an AW sample as CE is much more costly than labeling a CE sample as AW. A weighted SVM method, called 2C-SVM, is proposed [9] for classification applications where a classification error of one type is more expensive than the other. Recently, a variation of the 2C-SVM method, 2ν-SVM, is presented [10] and it gives different weights to different classes. It is proven [10] that the P_{2C} and P_{2v} are closely related and equivalent.

2ν-SVM searches for the best ν parameter on a $[0, 1]^2$ grid, and it has the following primal formulation:

$$\begin{aligned} \min \quad & \frac{1}{2}||\mathbf{w}||^2 - \nu\rho + \frac{\gamma}{n}\sum_{i\in I_+}\epsilon_i + \frac{1-\gamma}{n}\sum_{i\in I_-}\epsilon_i \\ \text{st.} \quad & y_i(k(\mathbf{w}, \mathbf{x_i}) + b) \geq \rho - \epsilon_i \quad \text{for} \quad i = 1, 2, \cdots, n \\ & \epsilon_i \geq 0 \quad \text{for} \quad i = 1, 2, \cdots, n \\ & \rho \geq 0 \end{aligned}$$

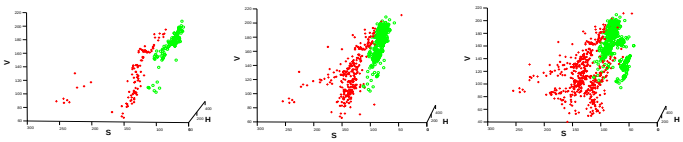

Fig. 2. AW vs. Cervix color sample distributions in HSV space. a)Samples from one image b)Samples from three images c) Samples from six images.

Table 2. Results using different classification schemes

Classification Accuracy			
Image #	State	Standard	Multi-classifier system
Image 1	TP	% 65.3	% 57
	FP	% 30.6	% 9.5
Image 2	TP	% 66	% 67.6
	FP	% 7.2	% 4.75
Image 3	TP	% 32.4	% 50.6
	FP	% 5.6	% 6.3
Image 4	TP	% 66.6	% 67.5
	FP	% 28.4	% 19.9

A cost-sensitive extension of 2ν-SVM [10] is proposed which, instead of parameters ν and γ, is formulated using ν_+ and ν_-, where

$$\nu_+ = \frac{\nu\eta}{2\gamma\eta_+},\ \nu_- = \frac{\nu\eta}{2(1-\gamma)\eta_-} \tag{2}$$

ν_+ and ν_- show the upper bound on the fraction of margin errors and the lower bound on the fraction of support vectors from the class $+1$ and the class -1, respectively. In summary, 2ν-SVM conducts a grid search over the SVM parameters, and we choose the parameters that minimize FP fraction and maximize TP fraction.

2.4. Multiple Classifier System

The colors of AW and CE regions in the cervical images vary greatly from one image to another. When all of the training data are used at once for training a single classifier, the classifier makes a tremendous amount of misclassification. We can get an intuition about the high misclassification rate by looking at the color sample distribution of images. Figure 2 shows that we lose the nice separation between AW and CE regions if we include more images to train the classifier. Therefore, we opt not to use the all-data-at-once training procedure after observing its detrimental effects on classification performance.

We train a set of classifiers by dividing the training set into smaller training sets, and then we can use the output of each classifier to estimate the true segmentation of a given test image. For instance, if we have N training images, we will build N different classifiers, one for each training image. Given a test image, the outputs of these N parallel and simultaneous classifier segmentations are used in the STAPLE Algorithm [7] to estimate the hidden true segmentation of the test image.

Our results indicate that the performance of this new classification scheme is better than the standard single-classifier method. Figure 3 shows some comparison between the standard and the Multiple Classifier System classification and Table 2 presents the numerical values corresponding to Figure 3.

2.4.1. STAPLE Algorithm

This section describes a version of an EM algorithm for estimating the hidden true segmentation from a collection of segmentations by iterative estimation. Each segmentation is obtained as the output of the 2ν-SVM classifier, and using these set of known segmentations, we want to estimate the performance parameters and recover the hidden true segmentation as accurately as possible.

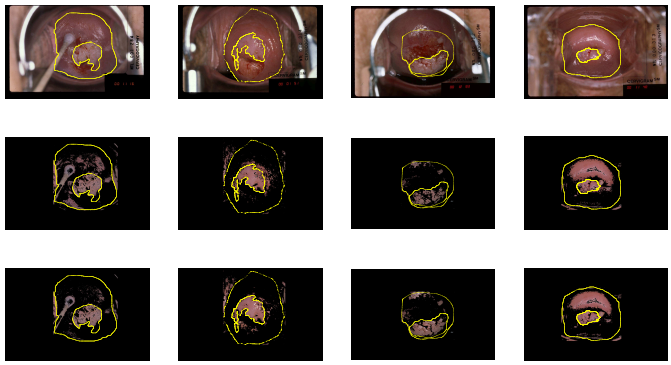

Fig. 3. Original test images (Row 1), Performances of the Standard (using all training data at once) classifier (Row 2) and the Multiple Classifier System (Row 3)

Suppose we have an image of N voxels and the task is to segment a structure in that image by indicating the presence or absence of the structure at each voxel. And suppose we have R segmented images obtained as the outputs of 2ν-SVM classifiers. Now, we want to recover the hidden true segmentation of a test image from these collection of classifier segmentations.

Let $\mathbf{p} = (p_1, p_2, \cdots, p_R)$ denote the sensitivity parameter characterizing the performance (or contribution) of each one of the R segmentations, and $\mathbf{q} = (q_1, q_2, \cdots, q_R)$ denote the specificity parameter.

Let $\mathbf{D}$ be the $N \times R$ matrix describing the binary decisions made by each segmentation at each voxel of the image and let $\mathbf{T}$ denote the hidden binary true segmentation that we need to estimate. The objective of the STAPLE algorithm is to estimate the performance level parameters $(\mathbf{p},\mathbf{q})$ of observed (classifier) segmentations which maximize the log likelihood function,

$$(\hat{\mathbf{p}}, \hat{\mathbf{q}}) = \arg\ \max_{\mathrm{p,q}}\ \ln f(D, T|p, q) \tag{3}$$

$\underline{\mathbf{E-step}}$: we derive an estimator for the unobserved true segmentation. We first derive an expression for the conditional probability density function at each voxel $\mathbf{i}$ assuming voxel-wise independence and the previous estimate of the performance parameters $(\mathbf{p}, \mathbf{q})$.

$$f(T_i|D_i, p^{(k-1)}, q^{(k-1)}) = \frac{\prod_j f(D_{ij}|T_i, p_j^{(k-1)}, q_j^{(k-1)}) f(T_i)}{\sum_{T_i'} \prod_j f(D_{ij}|T_i', p_j^{(k-1)}) f(T_i'))} \tag{4}$$

where $f(T_i)$ is the prior probability of T_i. Now if we consider the above equation for $T_i = 1$ and $T_i = 0$.

$$a_i^k = f(T_i = 1) \prod_j f(D_{ij}|T_i = 1, p_j^{(k)}, q_j^{(k)}) = f(T_i = 1) \prod_{j:D_{ij}=1} p_j^k \prod_{j:D_{ij}=0} (1 - p_j^k) \tag{5}$$

$$b_i^k = f(T_i = 0) \prod_j f(D_{ij}|T_i = 0, p_j^{(k)}, q_j^{(k)}) = f(T_i = 0) \prod_{j:D_{ij}=0} q_j^k \prod_{j:D_{ij}=1} (1 - q_j^k) \tag{6}$$

Table 3. Results using different classification schemes

Classification Accuracy			
Image #	State	Standard	Multi-classifier system
10 Images	TP	% 47.9 $\pm$ 13.9	% 58.95$\pm$ 12.1
	FP	% 8.54 $\pm$ 6.28	% 7.43$\pm$ 7.19

where $j : D_{ij} = 1$ denotes the set of indices for which the decision of SVM classifier j at voxel i has value 1. Using a_i^k and b_i^k we write the conditional probability of the true segmentation at each voxel,

$$W_i^{k-1} = f(T_i = 1|D_i, \mathbf{p}^{(k-1)}, \mathbf{q}^{(k-1)}) = \frac{a_i^{k-1}}{a_i^{(k-1)} + b_i^{(k-1)}} \quad (7)$$

This weight $W_i^{(k-1)}$ indicates the probability of the true segmentation at voxel i being equal to 1.

$\underline{\mathbf{M-Step}}$: Once we obtained the estimated weight variables $W_i^{(k-1)}$ then we can find the values of the classifier performance-level parameters($\mathbf{p}^k$, $\mathbf{q}^k$) that maximize the conditional expectation of the complete data likelihood function:

$$p_j^{(k)} = \frac{\sum_{i:D_{ij}=1} W_i^{(k-1)}}{\sum_i W_i^{(k-1)}} \quad (8)$$

$$q_j^{(k)} = \frac{\sum_{i:D_{ij}=0} (1 - W_i^{(k-1)})}{\sum_i (1 - W_i^{(k-1)})} \quad (9)$$

We keep iterating until a convergence criterion is satisfied to stop the EM iterations [7].

3. EXPERIMENTS

In certain classification applications, the cost of mislabeling a sample that truly belongs to one class (e.g. AW) is much higher than the cost of mislabeling a sample in the other class (CE). Therefore it might be intuitive to build classifiers that have a bias toward the less costly class. However, we need to avoid the false alarm rate from being arbitrarily high. So the overall objective is to find the diseased area as accurately as possible while minimizing the overall cost. So far we have presented the performances of different kernels, and advantage of using a multiple classifier system. Table 3 shows the benefit of using cost-sensitive SVM and the multiple classifier system over the simple single C-SVM classifier. In Table 3, we used 10 test images and the table shows the mean classification accuracy and standard deviation for the 10 test image set.

In all our experiments we use a radial basis function kernel and search the σ parameter from 10^{-2} to 10^3. For the 2ν-SVM method we considered a 10×10 regular grid of (ν_+,ν_-)$\in [0, 1]^2$. During the training stage, we manually determine the best performing parameters and then we use these parameters during the testing stage. For example, given a new test image, we apply to it the cost-sensitive 2ν-SVM classifiers built using pre-determined σ and (ν_+,ν_-) values. Each 2ν-SVM classifier gives us a segmentation, and these collection of segmentations are used in the STAPLE algorithm to estimate the hidden true segmentation.

4. CONCLUSIONS

SVM is a classification method that is widely used in image classification. However, there are situations where we have to assign a higher weight to one class than another class. In this paper, a cost-sensitive SVM is used to classify tissue regions (i.e AW, CE) in a cervix image. Furthermore we proposed the use of a cost-sensitive SVM based multiple classifier system to improve classification performance. The methods have shown promising results on a small set of images, with each image giving us hundreds of AW and CE pixel samples. Classifiers can also cooperate in ways other than a parallel combination. Given that we are more concerned with segmentation accuracy than efficiency, we are currently working on dynamic classifier selection, in which the system will dynamically select the most suitable classifiers for a given test image.

5. ACKNOWLEDGMENTS

The authors would like to thank the Communications Engineering Branch, National Library of Medicine - NIH, and the Hormonal and Reproductive Epidemiology Branch, National Cancer Institute - NIH, for providing the data and support of this work. The authors are thankful to Yaoyao Zhu, Zhiyun Xue, L. Rodney Long, Sameer Antani, and Dan Lopresti, for stimulating discussions on the STAPLE and Support Vector Machines algorithms.

6. REFERENCES

[1] Gordon, S., Zimmerman, G., Long, L.R., Antani, S., Jeronimo, J., Greenspan, H., "Content analysis of uterine cervix images: Initial steps towards content based indexing and retrieval of cervigrams," In: Proc. of SPIE. Volume 6144. (2006)

[2] Tulpule, B., Hernes, D., Srinivasan, Y., Yang, S., Mitra, S., Sriraja, Y., Nutter, B., Phillips, B., Long, L.R., Ferris, D., "A probabilistic approach to segmentation and classification of neoplasia in uterine cervix images using color and geometric features," In: Proc. of SPIE Medical Imaging. Volume 5747. (2005) 995–1003

[3] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., "Active shape models - their training and application," Computer Vision and Image Understanding **61**(1) (1995) 38–59

[4] Manjunath, B., Chellapa, R., "Unsupervised texture segmentation using markov random field Models," IEEE Trans. on Pattern Analysis and Machine Intelligence **13** (1991) 478–482

[5] Boykov, Y., Jolly, M., "Interactive organ segmentation using graph cuts," In: Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention. Volume LNCS-1935. (2000) 276–286

[6] V.Vapnik, *Statistical Learning Theory*, Wiley, New York, 1st Edition, 1998.

[7] Simon K. Warfield, K. Zou, W.M. Wells, "Simultaneous Truth and Performance Level Estimation(STAPLE): An Algorithm for the Validation of Image Segmentation," IEEE Trans. on Medical Imaging, 23(7): 903-921, 2004.

[8] G. Zimmerman-Moreno, H. Greenspan, "Automatic Detection of Specular Reflections in Uterine Cervix Images," SPIE Medical Imaging Symposium, 2006.

[9] E. Osuna, R. Freund, F. Girosi, "Support Vector Machines:Training and Applications," Tech. Rep. A.I Memo No. 1602, MIT Artificial Intelligence Laboratory, March 1997

[10] H.G. Chew, R.E. Bougner, C.C. Lim, "Dual-ν support vector machine with error rate and training size biasing," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp.1269-1272, 2001.

[11] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 2nd Edition, 1994.