BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
Vol. 54, No. 1, 2006

Support vector clustering algorithm for identification of glaucoma in ophthalmology

K. STĄPOR*

Institute of Computer Science, Silesian Technical University, 16 Akademicka St., 44-100 Gliwice, Poland

Abstract. This paper presents the improved version of the classification system for supporting glaucoma diagnosis in ophthalmology. In this paper we propose the new segmentation step based on the support vector clustering algorithm which enables better classification performance.

Key words: clustering, image segmentation, kernel-based learning, glaucoma.

1. Introduction

This paper presents the improved version of the classification system for supporting glaucoma diagnosis in ophthalmology, proposed in [1]. Glaucoma is a group of ocular diseases characterized by the proceeding optic nerve neuropathy which leads to the rising diminution in vision field, ending with blindness. The optic disk structure (i.e. the exit of the optic nerve from the eye known as "blind spot" is comprised of a yellowish cup surrounded by a neuroretinal pink rim [2] (e.g. see Fig. 1a)). Glaucomatous changes in the retina appearance embrace various changes in the cup, as the result of nerve fibers damages. The method proposed in [1] enables automatic classification of digital fundus eye images (FEI) taken from classical fundus-camera into normal and glaucomatous ones.

In this paper we propose the new segmentation method based on the support vector clustering (SVC) algorithm which improves the accuracy of the method for supporting glaucoma diagnosing, proposed in [1]. The modified method is composed of the following three main stages:

i. Segmentation of the cup region using support vector clustering.
ii. Selection of the cup features using genetic algorithms.
iii. Classification of FEI using the support vector machine (SVM) classifier.

2. Support vector clustering algorithm

The support vector clustering (SVC) algorithm [3] is a recently emerged kernel-based unsupervised learning method [4] inspired by the support vector machines [5] and consists of two main steps: 1) training step for constructing a trained kernel radius function and 2) a cluster labelling step for assigning to each data point a cluster index determined by its trained kernel radius function. Training step Let $\{x_i\} \subset R^d$ be a data set of N points (R^d being a data space). Using a nonlinear transformation $\Phi : R^d \to Q$ to some high-dimensional feature space Q, we look for the smallest enclosing sphere of radius r described by the constraints:

$$\forall_j ||\Phi(x_j) - a||^2 \leqslant R^2 + \xi_j \tag{1}$$

where $||.||$ is the Euclidean norm, a is the centre and $\xi_j \geqslant 0$ are some slack variables allowing for soft boundaries. By introducing the Lagrangian with a regularization constant C in its penalty term:

$$\begin{aligned} L = r^2 - \sum_j \left(r^2 - \xi_j - ||\Phi(x_j) - a||^2\right) \beta_j \\ - \sum_j \xi_j \mu_j + C \sum_j \xi_j \end{aligned} \tag{2}$$

where $\beta_j, \mu_j \geqslant 0$ are Lagrange multipliers the solution of the primal problem (1) can be obtained by solving its dual problem:

$$\max W = \sum_j \Phi(x_j)^2 \beta_j - \sum_{i,j} \beta_i \beta_j \Phi(x_i) \Phi(x_j) \tag{3}$$

subject to: $0 \leqslant \beta_j \leqslant C$, $\sum_j \beta_j = 1$, $j = 1, ..., N$.

Only those points with $0 < \beta_j < C$ lie on the boundary of the sphere and are called support vectors (SV). Points with $\beta_j = C$ lie outside the boundaries and are called bounded support vectors (BSV). All other points lie inside the boundary. Following the SV method [5] we can represent the dot products by an appropriate Mercer kernel:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \tag{4}$$

The Lagrangian can now be written as:

$$\max W = \sum_j K(x_j, x_j) \beta_j - \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \tag{5}$$

Setting to zero the derivative of L with respect to a leads to: $a = \sum_j \beta_j \Phi(x_j)$. The trained kernel radius function, defined by the squared distance of the image of x from the sphere

*e-mail: delta@ivp.iinf.polsl.gliwice.pl

centre, is then given by:

$$r^2(x) = ||\Phi(x) - a||^2 = K(x,x) - 2\sum_j \beta_j K(x_j, x) + \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \quad (6)$$

The radius of the sphere is $\hat{r} = r(x_i)$ where x_i is a support vector.

Labelling step. To differentiate between points that belong to different clusters a geometric approach involving r(x) is used. It is based on the following observation. Given a pair of data points that belong to different clusters, any path that connects them must exit from the sphere in feature space. Therefore, such a path contains a segment of points y such that r(y)>r. This leads to the definition of the adjacency matrix Aij between pairs of points xi and xj whose images lie in or on the sphere in feature space:

$$A_{ij} = \begin{cases} 1 \text{ if for all } y \text{ on the line segment} \\ \quad \text{connecting } x_i \text{ and } x_j r(y) < r \\ 0 \text{ otherwise} \end{cases} \quad (7)$$

Clusters are defined as the connected components of the graph induced by A. Checking the line segment is implemented by sampling a number of points (20 in our experiments).

3. Feature selection using genetic algorithms

In our approach, 30 geometric features were computed on the extracted cup region ([1]). These are: Hu moment invariants, compound invariant moments, circular coefficients, different shape coefficients including Danielsson, Haralick and Feret ones. Genetic algorithms were used to select the most significant features characterizing the shape of a cup region. A given feature subset was represented as a binary string with a zero or one in position i, denoting the absence or presence of feature i in the set. The initial population was randomly generated. We used the following fitness function:

$$\text{Fitness} = 10^4 \text{ accuracy} + 0.4 \text{ zeros} \quad (8)$$

where "accuracy" is the accuracy rate that the given subset of features achieves (i.e. the performance of a classifier on a given subset of features), "zeros" is the number of zeros in the chromosome. Overall, higher accuracy implies higher fitness. Fewer features used imply a greater number of zeros, and as a result, the fitness increases. As a classifier we used SVM with Gaussian kernel ([5]). The accuracy of the SVM classifier on a given subset of features required for the calculation of the fitness function is measured as a generalization error G_e, calculated using the k-fold cross-validation method (k = 10) [6]:

$$G_e = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (9)$$

where TP – true-positive, FN – false-negative, TN – true-negative, FP – false-positive. The parameters we used in all the experiments are as follows:

1) the length of each chromosome: 30,
2) the population size: 120,
3) the maximum number of generations: 500,
4) the cross-over rate: 0.6,
5) the mutation rate: 0.005.

The best chromosome (i.e. the best feature subset) is the one which is the most frequent among the chromosomes in the last generation.

4. SVM classifier

Having a training set $S = \{(x_i, y_i), 1 \leqslant i \leqslant N\}$ composed of the examples $x_i \in R^n$, each belonging to a class labelled by $y_i \in \{1, -1\}$, the goal of the SVM classifier [5] is to find the optimal separating hyperplane (OSH) – i.e the one which maximizes the separation margin which is a distance between the hyperplane and the closest data point. In the case when the data points are not linearly separable, a non-linear transformation $\Phi(x)$ is used to map the data vector x into a higher dimensional space using a kernel function. In our experiment, a nonlinear SVM with a Gaussian radial basis kernel:

$$K(x,z) = \exp\left(-\gamma \cdot |x - z|^2\right) \quad (10)$$

where γ is a constant, was used. The problem of finding the OSH in general is equivalent to the maximization of the function:

$$W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (11)$$

subject to the constraints:

$$\sum_{i=1}^{N} y_i \alpha_i = 0, \;\; 0 \leqslant \alpha_i \leqslant C \quad (12)$$

where α_i are the N nonnegative Lagrange multipliers, C is a regularization parameter. Finally, the decision function for classifying a new data point x can be written as follows:

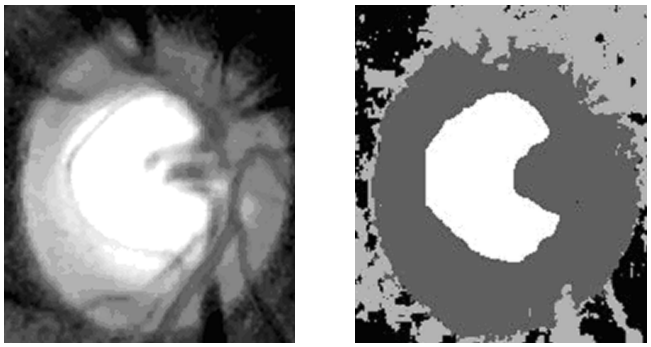
$$f(x) = \text{sgn}\left(\sum_{i=1}^{N_s} y_i \alpha_i K(x_i, x) + b\right) \quad (13)$$

where N_s is the number of support vectors, α_i, b are constants, all determined through the numerical optimization during learning.

5. Experiments

5.1. Segmentation of the cup region. The data set used for this research consists of 100 digital fundus eye images of patients with glaucoma and 100 images of normal patients. These images are part of the data set acquired from the Department of Ophthalmology, Friedrich-Alexander-University Erlangen-Nuremberg, Prof. Dr George Michelson. To produce a "gold standard" segmentation, an ophthalmologist marked manually the boundary of the cup in each of the images. To decrease the computational time, the cup segmentation was performed in a window, automatically computed based on the cup localization procedure described in [1]. Moreover, we performed the subsampling procedure of the computed window, i.e. we chosen every 10-th pixel. The 3-dimensional feature space (L, a, b) was used for clustering, i.e. each image pixel was described by three components of Lab colour model. All features

were normalized using z-score normalization [7]. The remaining pixels in the window were assigned to the groups revealed during clustering based on the distance from representatives of the groups. The cup in the segmented image was chosen as the region having the smallest value of a. Fig. 1b) presents the segmented image from the FEI shown in Fig. 1a) with the segmented white cup region in the central part.



(a) (b)

Fig. 1. The automatically selected window from input FEI with the cup in the central part a), the segmentation result b)

5.2. Model selection and testing. The set of 200 segmented cup regions was divided into two disjoint subsets: 1) the training set: 150 images, 2) the testing set: 50 images. In each of those sets there were equal numbers of glaucomatous and normal cups. The training set was used for model selection: the suboptimal feature vector calculation based on genetic algorithms, setting SVM classifier parameters (performed by 10-fold cross-validation method) and final SVM learning. The feature selection described in subsection 3 was performed for different combinations of the classifier parameters C, a regularization parameter and γ, a Gaussian kernel one. For each such combination we noted down the best subset of features with the corresponding value of the generalization error G_e. As the final subset of features we took the one with the smallest value of G_e:

$$\upsilon_0 = (\phi_2, I_3, R_F) \tag{14}$$

where:

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \tag{15}$$

is Hu invariant moment, in which $\eta_{20}, \eta_{02}, \eta_{11}$ are normalized central moments all computed on an image function $f(x, y)$,

$$\begin{aligned} I_3 = \mu_{20}\left(\mu_{21}\mu_{03} - \mu_{12}^2\right) - \mu_{11}\left(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}\right) \\ + \mu_{02}\left(\mu_{30}\mu_{12} - \mu_{21}^2\right) \end{aligned} \tag{16}$$

is compound, invariant moment,

$$R_F = \frac{L_h}{L_v} \tag{17}$$

is Feret coefficient, where:

L_h – the maximal diameter in the horizontal direction,

L_v – the maximal diameter in the vertical direction.

The selected feature vector υ_0 corresponds to the combination of the classifier parameters: $C = 100$, $\gamma = 2.5$. Finally, the classifier was trained on the set composed of feature vectors υ_0 computed on the training set.

Classifier performance was tested on the feature vectors υ_0 calculated on the testing set. The following results were obtained: the mean sensitivity which is the percent of the correctly classified glaucomatous cases:

$$\text{sensitivity} = \frac{TP}{TP + FP} = 94,5\% \tag{18}$$

and the mean specificity which is the percent of the correctly classified normal cases:

$$\text{specificity} = \frac{TN}{TN + FN} = 97,5\%. \tag{19}$$

6. Conclusions

In this paper we described a novel cup segmentation method FEI which is based on support vector clustering algorithm for the purpose of supporting glaucoma diagnosing in ophthalmology. The proposed method can find clusters non-linearly separable as well as clusters of varying shapes and sizes. The proposed method enables automatic classification of digital FEI into normal and glaucomatous ones. The obtained classification results are encouraging. It is expected that the new method, after clinical tests, would support glaucoma diagnosis based on digital FEI obtained from fundus camera.

Acknowledgements. The authors would like to thank prof. George Michelson from Department of Ophthalmology, Friedrich-Alexander-University Erlangen-Nuremberg for the possibility to access a database with fundus eye images, as well as for medical consultation.

REFERENCES

[1] K. Stapor and A. Switonski, "Automatic analysis of fundus eye images using mathematical morphology and neural networks for supporting glaucoma diagnosis", *Machine Graphics & Vision* 13 (1/2), 65–79 (2004).

[2] J. Kanski, *et al.*, *Glaucoma: a Color Manual of Diagnosis and Treatment*, Butterworth-Heinemann Medical, 1996.

[3] A. Ben-Hur et al, "Support vector clustering", *J. Machine Learning Research* 2, 125–137 (2001).

[4] C.R. Mueller, et al., "An introduction to kernel-based learning algorithms", *IEEE Trans. Neural Networks* 12 (2), 2001.

[5] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 1995.

[6] K. Stąpor, *Automatic Classification of Objects*, Exit, Warszawa, 2005, (in Polish).

[7] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, N.J., 1988.