Katarzyna STAPOR[*], Adrian BRUCKNER[**]

KERNEL K-MEANS CLUSTERING ALGORITHM
FOR IDENTIFICATION OF GLAUCOMA IN OPHTHALMOLOGY

This paper presents the improved version of the classification system for supporting glaucoma diagnosis in ophthalmology, proposed in [4]. In this paper we propose the new segmentation step based on the kernel K-Means clustering algorithm which enable for better classification performance.

1. INTRODUCTION

This paper presents the improved version of the classification system for supporting glaucoma diagnosis in ophthalmology, proposed in [4]. 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 [4] 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 cup segmentation method based on the kernel K-Means clustering algorithm which improves the accuracy of the method for supporting glaucoma diagnosing, proposed in [4]. The modified method is composed of the following three main stages:

1. Segmentation of the cup region using kernel K-Means.
2. Selection of the cup features using genetic algorithms.
3. Classification of FEI using the support vector machine (SVM) classifier.

2. KERNEL BASED K-MEANS CLUSTERING ALGORITHM

Traditional K-Means clustering algorithm [1] aims to partition the data set composed of N samples $x_1,...,x_N$ into K clusters: $G_1,...,G_K$, and then returns the centre of each cluster: $c_1,...,c_k$ as the representatives of the data set. The assumption behind this algorithm is the belief that the data space consists of isolated elliptical regions. To tackle the problem when that assumption is not held, one idea is to apply the transformation $\Phi: R^d \rightarrow Q$ that maps each data x_i from the input space R^d

[*] Institute of Computer Science, Silesian University of Technology, Gliwice, Poland. delta@ivp.iinf.polsl.gliwice.pl
[**] Institute of Mathematics, Silesian University, Katowice, Poland

to a new space Q that satisfies the shape requirement. The transformation is done implicitly by means of a kernel function k, being the dot product in the new space Q (which should be a Hilbert space):

$$k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \qquad (1)$$

Let $u_i = \Phi(x_i)$ denotes x_i's transformation into new space. Using kernel function the Euclidean distance between u_i and t_k, the cluster center in the transformed space can be written as:

$$D^2(u_i, t_k) = k(x_i, x_i) + h_1(x_i, G_k) + h_2(G_k), \qquad (2)$$

where:

$$h_1(x_i, G_k) = -\frac{2}{|G_k|} \sum_{j=1}^{N} \omega(u_j, G_k) + k(x_i, x_j) \qquad (3)$$

$$h_2(G_k) = \sum_{j=1}^{N} \sum_{l=1}^{N} \omega(u_j, G_k) \omega(u_l, G_k) k(x_j, x_l) \qquad (4)$$

$$\omega(x_i, G_k) = \begin{cases} 1 & \text{if } \forall j \neq k \ D(x_i, c_k) < D(x_i, c_j) \ j = 1,...K \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

is the indicator function, $D(x_i, c_k)$ is the Euclidean distance, K, N are the number of clusters and data points, respectively. The kernel-based K-Means algorithm can be stated as follows:

1. Assign $\omega(x_i, G_k)$ $(1 \leq i \leq N, \ 1 \leq k \leq K)$ with initial value, forming K initial clusters
2. For each cluster G_k compute $|G_k|$ and $h_2(G_k)$
3. For each training sample x_i and cluster G_k, compute $h_1(x_i, G_k)$ and then assign x_i to the closest cluster:

$$\omega(x_i, G_k) = \begin{cases} 1 & \text{if } \forall j \neq k \ h_1(x_i, G_k) + h_2(G_k) < h_1(x_i, G_j) + h_2(G_j) \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

4. Repeat steps 2-3 until convergence
5. For each cluster G_k select the sample that is closest to the center as the representative of G_k:

$$c_k = \underset{x_i : \omega(x_i, G_k) = 1}{\arg \min} D(\Phi(x_i), t_k) \qquad (7)$$

Mercer's theorem [5] guarantees that as long as the kernel function exhibits certain mathematical properties (it is positive definite), then the algorithm implicitly operates in a higher dimensional space.

3. FEATURE SELECTION USING GENETIC ALGORITHMS

In our approach, 30 geometric features were computed on the extracted cup region ([4]). 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:

$$Fitness = 10^4 \, accuracy + 0.4 \, zeros \qquad (8)$$

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. As a classifier we used SVM with Gaussian kernel ([6]). 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):

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

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 $S = (x_i, y_i, 1 \leq i \leq N)$ composed of the examples $x_i \in R^n$, each belonging to a class labeled by $y_i \in \{1, -1\}$, the goal of the SVM classifier [6] 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:

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

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

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

subject to the constraints:

$$\sum_{i=1}^{N} y_i \alpha_i = 0, \ 0 \le \alpha_i \le C \qquad (12)$$

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:

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

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 [4]. 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 color model. All features were normalized using z-score normalization [1]. The number of clusters was chosen as 4, i.e. corresponding to the 4 anatomical parts of the FEI: the retina, blood vessels, neuroretinal rim and the cup. 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 contour of the cup region imposed on it.

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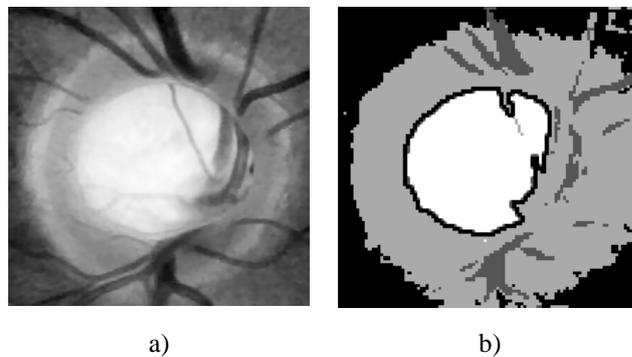
Fig. 1. a) The automatically selected window from input FEI with the cup in the central part
b) the segmentation result with the contour of the cup region imposed

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:

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

where:

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

is Hu invariant moment, in which $\eta_{20}, \eta_{02}, \eta_{11}$ are normalized central moments,

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

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:

$$sensitivity = \frac{TP}{TP + FP} = 93\% \tag{18}$$

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

$$specificity = \frac{TN}{TN + FN} = 97\% \tag{19}$$

6. CONCLUSIONS

In this paper we described a novel kernel K-Means clustering algorithm that can find clusters non-linearly separable as well as clusters of varying shapes and sizes. We also demonstrated application of the proposed clustering algorithm in the segmentation of the cup region on FEI taken from classical fundus camera for the purpose of supporting glaucoma diagnosing in ophthalmology. 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.

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.

BIBLIOGRAPHY

[1] JAIN A.K., DUBES R.C., Algorithms for clustering data, Prentice-Hall, Englewood Cliffs N.J, 1988.
[2] KANSKI J., McALLISTER J.A., SALMON J.F., TARRANT T.R., Glaucoma: a color manual of diagnosis and treatment, Butterworth–Heinemann Medical, 1996.
[3] MUELLER C.R., et al., An introduction to kernel-based learning algorithms, IEEE Trans. Neural Networks, 12(2), 2001.
[4] STAPOR K., SWITONSKI A., 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.
[5] STAPOR K, Automatic classification of objects. EXIT, Warszawa 2005. (in Polish)
[6] VAPNIK V., The nature of statistical learning theory, Springer Verlag, New York, 1995.