support vector machines,
genetic algorithms,
human face recognition

Michał KAWULOK*

GENETIC ALGORITHMS FOR CLASSIFIERS' TRAINING SETS OPTIMISATION APPLIED TO HUMAN FACE RECOGNITION

Human face recognition is a multi-stage process within which many classification problems must be solved. This is performed by learning machines which elaborate classification rules based on a given training set. Therefore, one of the most important issues is selection of a training set which would properly represent the data that will be further classified. This paper presents an approach which utilizes genetic algorithms for selecting classifiers' training sets. This approach was implemented for the Support Vector Machines which is applied in two areas of automatic human face recognition: face verification and feature vectors comparison. Effectiveness of the presented concept was confirmed with appropriate experiments which results are described in this paper.

1. INTRODUCTION

Face recognition [7, 13, 14] is among the most popular biometric techniques which are being developed nowadays and it is worth noticing that this is the method which is the most frequently used naturally by humans. Automatic face recognition is characterized by a low level of required interaction with a person who is being recognized, but offers relatively low effectiveness comparing to other biometric methods [4, 9]. A face recognition system discriminates between different faces and fulfils four main identification tasks [4]: classification, known-unknown problem (checking whether an image belongs to one of the defined classes or to none of them), verification and full identification. Such systems can be applied in many areas, ranging from entertainment to access control and surveillance tracking.

Face recognition algorithms as well as other tasks concerned with image recognition can be significantly improved by application of classifiers, such as the Support Vector Machines (SVM) [2]. A concept of applying the SVM in automatic face recognition was developed and described by the author in [6]. Results of the experiments showed that available improvement depends strongly on how well-trained the classifier is and how the training data set is selected.

In the analysed case the training data are generated in automatic or semi-automatic way and their representativeness can be assessed only by measuring the effectiveness of the trained and working classifier. Hence, it is virtually impossible to select the training data properly independently from the tests with the classifier. This paper presents an approach in

* Silesian University of Technology, Institute of Computer Science, Akademicka 16, 44-100 Gliwice, Poland

which the training sample is optimised using genetic algorithms which make it possible to manage populations of potential solutions in order to generate the best one.

2. SUPPORT VECTOR MACHINES

The Support Vector Machines (SVM) [1, 2] is a learning machine which solves two-group classification problems and may be enhanced to multi-class cases as well. The SVM-based processing consists of two phases: learning and classification.

The learning aims at finding an optimal hyperplane which separates a classified, linearly separable training data set: $(y_1,\mathbf{x_1}), ..., (y_n,\mathbf{x_n})$, $y_i \in \{-1, 1\}$, where $\mathbf{x_i}$ are vectors in N-dimensional input space and y_i are class labels. The optimal hyperplane is found by maximizing the margin ρ between classes:

$$\rho(\mathbf{w}_O,b) = \min_{\{\mathbf{x}:y=1\}} \frac{\mathbf{x}\cdot\mathbf{w}_O}{|\mathbf{w}_O|} - \max_{\{\mathbf{x}:y=-1\}} \frac{\mathbf{x}\cdot\mathbf{w}_O}{|\mathbf{w}_O|} \;\; and \;\; \mathbf{w}_O = \sum_{i=1}^{n} y_i \alpha_i^0 \mathbf{x_i}\,, \tag{1}$$

where $\mathbf{w}$ is the hyperplane normal vector. A normal vector of the optimal hyperplane ($\mathbf{w}_O$) is expressed as a linear combination of vectors from the training set, where α_i^0 are non-negative Lagrange multipliers which are obtained during the optimisation process. Each vector from the training set is associated with one α and usually a relatively small number of α has non-zero values. Vectors with non-zero α are termed support vectors and are used further for classification.

When the training process is finished, the SVM allows for classification of any N-dimensional vector $\mathbf{x}$ based on the calculated decision surface $f(\mathbf{x})$:

$$f(\mathbf{x}) = \sum_{i=1}^{n} y_i \alpha_i \,\mathbf{x}\cdot\mathbf{x_i}\,. \tag{2}$$

The main disadvantage of this solution is an assumption that the input data set must be linearly separable. However, if this requirement is not satisfied, linear separability can be achieved by increasing dimensionality of the input space. There is no need for mapping the input vectors to higher dimensions straightforwardly, instead of this their dot-products are substituted with kernel functions $K(\mathbf{u},\mathbf{v})$ and the decision surface is calculated in this way:

$$f(\mathbf{x}) = \sum_{i=1}^{n} y_i \alpha_i K(\mathbf{x},\mathbf{x_i})\,. \tag{3}$$

The most effective kernel functions which make it possible to achieve linear separability in the majority of classification problems are: linear ($K(\mathbf{u},\mathbf{v}) = \mathbf{u}\cdot\mathbf{v}$), p-th degree polynomial ($K(\mathbf{u},\mathbf{v}) = (\mathbf{u}\cdot\mathbf{v}+1)^p$) and radial basis functions ($K(\mathbf{u},\mathbf{v}) = \exp[-\|\mathbf{u}-\mathbf{v}\|^2/\sigma^2]$, where σ is a width parameter).

A crucial problem concerned with the SVM is a choice of a representative training set. Depending on it, the classification rules established by the SVM during the learning phase are general or not, which determines the effectiveness of solving classification problems. There exists a possibility that for a training data set of high cardinality, its subsets are more representative than the whole set, especially if the data were acquired in an automatic, uncontrolled way. A representative training set can be chosen randomly by drawing various pairs of sets which consist of samples belonging to opposite classes. The SVM can be trained with every drawn set and its classification effectiveness can be measured based on classification of a testing set which is independent from the training data. However, the developed approach in which genetic algorithms are used for the training set optimisation is more effective which was confirmed by the conducted experiments.

3. GENETIC ALGORITHMS

Genetic algorithms [3, 8] implement a heuristic approach which supports the search for solutions close to the optimal based on evolution strategy. At first a population of individuals characterized by their genotypes is generated and transformed using three genetic operators: selection, mutation and crossover. A genotype contains a set of parameters to be optimised and defines unambiguously a single solution.

The first population is initialised with a set of individuals whose genotypes are generated randomly. After that every individual is assessed and its fitness is measured which is required to perform the selection. The subsequent generation is created based on selected individuals with highest fitness transformed with mutation and crossover operators.

Mutation means a random modification to the genotype of an individual and it is aimed at sustaining the genotype variety. Otherwise there would be a risk that the individuals within every generation become too similar to each other which would result in finding a local minimum.

Crossover operates on two individuals which genotypes are randomly merged into one. In this way subsequent generations are created and the fitness of the best individual should be increased in every generation compared with the previous.

Genetic algorithms are particularly useful in the areas where standard optimisation techniques cannot be applied, but the effectiveness of a single solution can be easily measured automatically.

3.1. TRAINING SET OPTIMISATION WITH GENETIC ALGORITHMS

The research described in this paper was based on a concept that the genetic algorithms approach can be applied to optimise the training sets of the Support Vector Machines in cases when cardinality of the training data sets is large, but quality of the data is low. Every classified training set created as a subset of the main set forms a single individual, whose genotype defines the content of the training data. Number of vectors in each class was constant for every individual in the analysed case.

The fitness of each individual is based on classification effectiveness, which is measured for a test set. In the cases when the SVM is applied to solve a specific task which effectiveness is unambiguously measurable, this effectiveness may be treated as the fitness. The individuals with the highest fitness are selected to create the subsequent generation.

Mutation is performed by random changes in the training sets of the individual. In order to crossover two individuals, the sets defined by their genotypes are at first summed within each class and then randomly chosen elements are removed from each class, so that the number of elements in each class do not increase.

The described process is iteratively repeated as long as the maximal effectiveness is increasing in subsequent generations or until the effectiveness reaches a desired level.

4. APPLICATION OF SUPPORT VECTOR MACHINES IN FACE RECOGNITION

Automatic face recognition [7] is a multi-stage process which aims at solving the identification tasks. This requires an ability to measure similarity between any two given face images. The face recognition process consists of the following stages: detection, normalization, feature extraction and feature vectors comparison.

The process operates on digital images which may contain human faces of various sizes. The first stage is face detection which aim is to determine whether there are any faces in the input image and to find their exact locations afterwards. Face location in the case of frontal images is unambiguously defined by the central points of eyes and therefore for every image the detection stage outputs a list of eyes coordinates.

The images with detected faces are scaled, rotated and clipped to generate normalized images of fixed size in which the eyes are always located in the same positions. Additionally, the normalization phase includes image filtering and histogram-based operations aimed at lighting compensation. Ideally, any two different face images of the same person acquired in various conditions should be transformed into identical normalized images. Noteworthy, the extrapersonal features which make it possible to distinguish between faces should not be eliminated during the normalization.

Feature extraction operates on a normalized image and should generate a feature vector which describes the face. Moreover, there must be a similarity measure defined in order to calculate the similarity between any two feature vectors.

The described process makes it possible to compare any two given face images in order to decide whether they present the same person.

4.1. THE SVM FOR FACE VERIFICATION

In the analysed case face detection was implemented to work in two steps [6, 7]: preliminary selection and verification. Purpose of the selection is to find face candidates, which are the areas in the image where faces may be located. The set of the candidates may include many false positives, but the crucial issue is that all the faces in the given image should be covered by the set. The candidates are selected based on ellipse detection performed with the Hough transform.

After the candidates are detected, they are normalized and verified to reject all the false cases and accept face images. Ellipse detection is an effective technique for the preliminary selection, but the detection effectiveness depends strongly on the verifier as well. The SVM was trained with two sets of the images: normalized face images (upper row in Fig. 1) and normalized images of false candidates which do not contain faces (bottom row in Fig. 1).

Fig.1. Examples of normalized face candidates

Face detection effectiveness is often defined based on the percentage of correctly detected faces compared to the number of false positives. However, this approach does not take into account the detection precision which is an important factor that influences the recognition effectiveness. When the face detection is applied as a part of the recognition process, rather than a task itself, the detection error is propagated further and therefore it is crucial to take the precision into account as well. In the presented approach this was done by calculating a relative detection error (δ_d) for a single face in the following way [7]:

$$\delta_d = \frac{\Delta_l + \Delta_r}{2D}, \tag{4}$$

where Δ_l and Δ_r are the distances in pixels between real and detected positions for left and right eye respectively and D is the distance between real positions of eyes. In the case when a face is not detected at all, the relative error is set to 1. Finally, the effectiveness for a set is calculated as a reversed sum of average partial errors:

$$E = \frac{N}{\sum \delta_d}, \tag{5}$$

where N is the real number of faces in the processed images.

4.2. THE SVM FOR FEATURE VECTOR COMPARISON

In order to estimate similarity between two face images it is necessary to generate and compare the feature vectors which describe these faces. Depending on the feature extraction algorithm there are various methods of measuring the similarity. In the case of the Eigenfaces method [12], which was implemented and analysed, Euclidean or Mahalanobis distance between two vectors is calculated and used to assess the similarity.

The Support Vector Machines is used in this area [6, 7] to decide whether two feature vectors were derived from the images of different faces or from the same person. At first, two feature vectors (ν_1 and ν_2) are subtracted from each other to obtain a difference vector (ν_d) which is classified by the SVM.

For this application, the SVM is trained based on two sets of difference vectors: intrapersonal derived from the same person and extrapersonal derived from different people. The experiments proved that application of a well-trained classifier gives better results than a standard distance-based method.

The effectiveness of feature vectors comparison is measured based on the classification task. There is a set of classified face images in which every class gathers all images of a single person whose images are in the set. This set is divided into two exclusive subsets [5]: a gallery (*G*) and a query set (*Q*). The gallery contains exactly one image per person and the query set gathers the rest of the images. During the classification every image from the query set is compared with all the images from the gallery and the most similar image is chosen from the gallery. If both images belong to the same class, the classification is correct. If the image is *n*-th on the list of the most similar images from the gallery, it is said to be classified with *n*-th rank. Percentages of the images classified with subsequent ranks are often presented in a form of Cumulative Match Characteristic (CMC) [5].

5. EXPERIMENTS

The experiments were conducted for face verification and for feature vectors comparison. In both cases the effectiveness was tested based on 1000 images from the Feret [11] face image database of 395 various individuals.

5.1. FACE VERIFICATION

The SVM for face verification was trained based on a classified set of 2000 normalized face images and 2000 normalized non-face images selected as face candidates (false positives) during the ellipse detection phase. The radial basis functions with σ=0.7 were used as the SVM kernel.

At first, no selection was done, so the SVM was trained with all available images and face detection effectiveness was tested for this case. Afterwards, training samples containing 200 vectors (100 of each class) were being drawn and finally a genetic algorithm was applied to optimise the training sample.

In the genetic approach there were 10 individuals in every generation and the optimisation was performed for 21 subsequent generations. In Fig. 2 fitness of the best individual in subsequent generations is presented. Additionally, a result for the case when the SVM was trained without any selection, as well as the best result obtained by drawing the training sample randomly are presented in the graph. Noteworthy, the number of single classification tests was equal in both cases, so the results can be compared with each other. An average detection error was reduced from 5.4 % in the case of no selection to 4.1 % for the genetic algorithm which means that the error was reduced by 24 %.

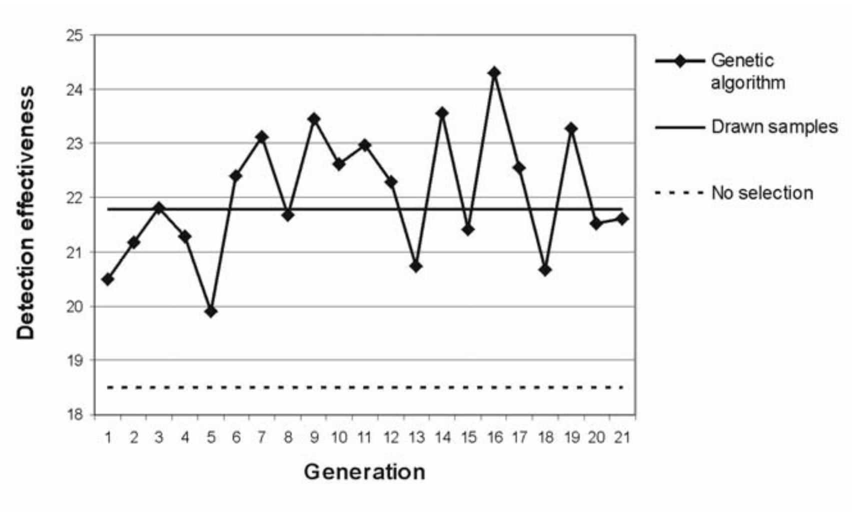

Fig.2. Detection effectiveness in subsequent generations of training sample optimisation

5.2. FEATURE VECTORS COMPARISON

Feature vectors comparison was tested based on a set of 761 face images from the Notre-Dame database [10] which was used to generate the data for the SVM training. The feature vectors were generated using the Eigenfaces method [12] and all possible vector pairs were created: 286066 extrapersonal and 3115 intrapersonal pairs including one pair of identical vectors. The pairs were transformed into difference vectors which can be processed by the SVM. Two kernel functions were used in this case: third degree polynomial and linear.

It is remarkable that in this case it was virtually impossible to train the SVM with all available vectors due to large and unequal cardinality of the training sets. The SVM was therefore trained with 100 intrapersonal and 100 extrapersonal difference vectors selected from the set of all available vectors. Similarly as in the case of face verification, the training samples were being selected in two ways: randomly and using the genetic algorithm and the number of single classification tests was equal in both cases.

The best classification effectiveness achieved in subsequent generations is presented in Fig. 3 for polynomial and linear kernel. Application of the genetic approach allowed for effective training set optimisation and the results achieved were much better comparing with the draw-based approach. The best cases for the polynomial and linear kernels are compared in Fig. 4 in the form of Cumulative Match Characteristics. The best classification results for each case, as well as in the case of Euclidean-distance-based comparison are presented in Tab. 1. The genetic algorithms made it possible to find more representative training set which reduced the classification error significantly.

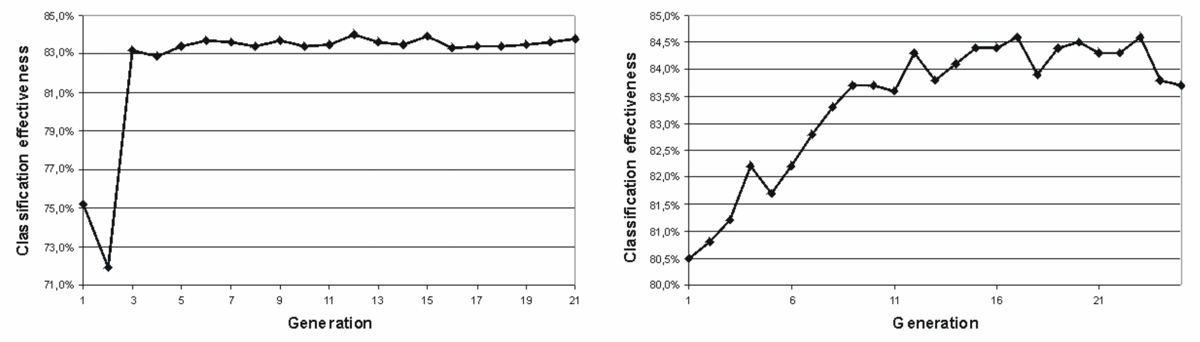

Fig.3. Classification effectiveness in subsequent generations for the SVM with polynomial (left) and linear (right) kernel

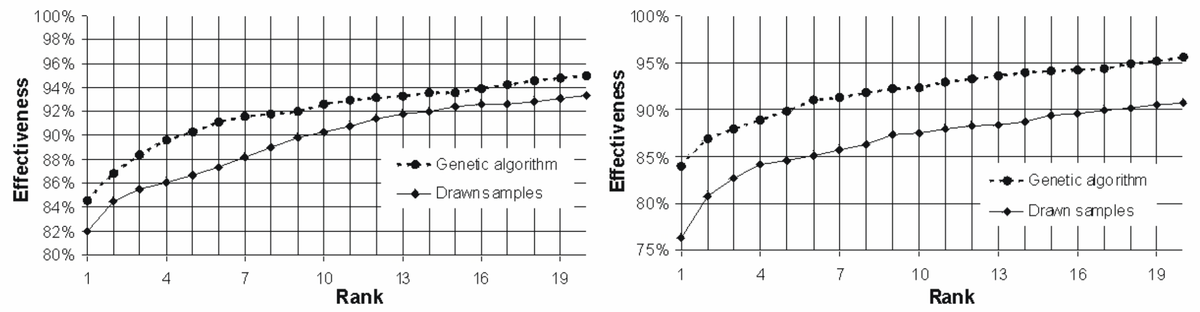

Fig.4. Cumulative Match Characteristic [5] for the SVM with polynomial (left) and linear (right) kernel

Table .1 Classification error for various feature vector comparison methods

	SVM-based comparison		Euclidean-distance-based comparison
Kernel type	Drawn samples	Genetic algorithm	
Polynomial	23.7 %	16.0 %	17.1 %
Linear	18.0 %	**15.8 %**	

6. CONCLUSIONS AND FUTURE WORK

This paper presents an evolutionary approach for optimising the SVM training set which was implemented and tested in the area of human face recognition. The conducted experiments which results are presented in this paper proved that the proposed method is effective and can be applied in the cases where amount of training data is high, but their quality is variable.

There are two ways of developing the presented concept in the future. In the conducted experiments the training parameters including cardinality of training sets and kernel characteristics were chosen before starting the genetic optimisation. It may be possible to obtain better results if these parameters are included in the genotype as well and optimised together with the content of the training samples.

The second direction of the future works is concerned with applying the presented approach to other areas where the classifiers are used, including many cases of medical data processing.

BIBLIOGRAPHY

[1] ABE S., Analysis of Support Vector Machines. In Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing, pages 89-98, 2002.
[2] CORTES C., VAPNIK V., Support vector networks, Machine Learning, 20:1-25, 1995.
[3] GOLDBERG D.E., Genetic Algorithms in Search, Optimisation, and Machine Learning, 1989, Addison-Wesley Publishing Co.
[4] GONG S., MCKENNA S. J., PSARROU A., Dynamic Vision: From Images to Face Recognition, Imperial College Press 1999.
[5] GROTHER P., MICHEALS R., PHILLIPS P.J., Face recognition vendor test 2002 performance metrics. In Proceedings of the Fourth International Conference on Audio-Visual Based Person Authentication, June 2003.
[6] KAWULOK M., Application of Support Vector Machines in Automatic Human Face Recognition, Medical Informatics & Technologies, 9:143-150, October 2005.
[7] KAWULOK M., Wybrane metody poprawy skuteczności automatycznego rozpoznawania obrazów twarzy (Selected methods of improving automatic face recognition effectiveness), PhD Thesis, Silesian University of Technology, Gliwice, 2006.
[8] MICHALEWICZ Z., Genetic Algorithms + Data Structures = Evolution Programs, Springer, Berlin, 1996.
[9] Phillips P.J., Grother P., Micheals R.J., Blackburn D.M., Tabassi E., Bone J.M., FACE RECOGNITION VENDOR TEST 2002: EVALUATION REPORT. NISTIR 6965, 2003.
[10] PHILLIPS P.J., FLYNN P.J., SCRUGGS T., BOWYER K.W., CHANG J., HOFFMANN K., MARQUES J., MIN J., WOREK W., Overview of the Face Recognition Grand Challenge. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 947-954, June 2005.
[11] PHILLIPS P.J., WECHSLER H., HUANG J., RAUSS P., The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing J, Vol. 16, No. 5, pages 295-306, 1998.
[12] TURK M., PENTLAND A., Face Recognition Using Eigenfaces. In Proceedings of Computer Vision and Pattern Recognition 1991, p.586-591.
[13] WECHSLER H., PHILLIPS P.J., BRUCE V., SOULIE F.F., HUANG T.S., Face Recognition: From Theory to Applications. Springer-Verlag, Berlin, 1998.
[14] ZHAO W., CHELLAPPA R., PHILLIPS P. J., ROSENFELD A., Face Recognition: A Literature Survey, Technical Report CARTR-948, Center for Automation Research, University of Maryland, College Park, MD, 2000.