Development of the SVM Classifier by means of the Hybrid Versions of the Particle Swarm Optimization Algorithm Based on the Grid Search

Liliya Demidova, Irina Klyueva*

Ryazan State Radio Engineering University Gagarin Str., 59/1, Ryazan, Russian Federation

Corresponding author's e-mail: i.aleschenko@yandex.ru

Received 3 April 2017, www.cmnt.lv

Abstract

In this article the approaches to the problem solving of searching of the parameters of the SVM classifier based on the hybridization of the particle swarm optimization algorithm (PSO algorithm) and the grid search algorithms with the aim of providing of high quality classification decisions have been considered. The paper presents two hybrid versions of the basic PSO algorithm, involving the use of the classical Grid Search (GS) algorithm and Design of Experiment (DOE) algorithm correspondingly. It is proposed to use the canonical PSO algorithm as the basic algorithm. The results of experimental studies confirm the application efficiency of the hybrid versions of the basic PSO algorithm with the aim of reducing of the time expenditures for searching the optimum parameters of the SVM classifier while maintaining of high quality of its classification decisions.

Keywords:

classification,
particle swarm optimization algorithm,
grid search algorithm,
SVM classifier,
radial basis kernel function

1 Introduction

Data classification is one of the most common problems of machine learning [1–7]. The solution of this problem requires creation of a classifier that assigns each input dataset the value of the label of one of the classes. Classification of new data is produced after passing through the stage of "learning", in which the input of the learning algorithm serves the data with already assigned labels of classes.

Currently, the SVM algorithm (Support Vector Machine, SVM) [1–7] is successfully applied for solving a wide spectrum classification problems in various applications. The SVM algorithm is a machine learning algorithm by precedents. The SVM algorithm implements the construction of binary SVM classifier.

The SVM algorithm implements creation of the separating hyperplane that divides objects with different class. Herewith, two parallel hyperplanes defining the boundaries of classes and locating at the greatest possible distance from each other are constructed on both sides of the separating hyperplane. It is assumed that the greater the distance between these parallel hyperplanes the smaller the average error of the SVM classifier. The vectors of features of the classified data nearest to the parallel hyperplanes are called the support vectors.

In most cases, the linear separability of objects of real datasets into classes is impossible. In this regard, the main feature of the SVM classifier in case of nonlinear separability of objects is the use of special function, called the kernel function. The kernel function is used to transfer the experimental dataset from the original space of features to the space of higher dimension in which the separating hyperplane is build.

In the process of learning of the SVM algorithm the one of the priority problems is to configure the parameters of the SVM classifier, the most important of which are the kernel function type, the values of the kernel parameters and the value of the regularization parameter.

The one of the following functions [6] is usually used as a kernel function that allows to separate the objects of different classes: linear function, polynomial function, radial basis function, sigmoid function.

The regularization parameter C allows finding a compromise between the maximizing of the gap separating the classes and the minimizing of the total error. In other words, the regularization parameter controls the ratio between the smooth boundary and the corrects data classification.

In case of radial basis kernel function (RBF) [6] it is necessary to determine the value of the coefficient σ of this function.

The simplest approach to settings of the SVM classifier parameters is based on a simple enumeration of the different combinations of the parameter values. For the purpose of setting parameters of the SVM classifier the grid search algorithms (in particular, the Grid Search algorithm) are most often applied [2]. Herewith, the cross-validation on the training dataset is used for each parameters combination corresponding to the specific grid node. As the result, the best combination of the parameters values is selected. This combination defines the certain grid node which is characterized by the best value of the cross-validation indicator.

Finding the optimal set of values for the parameters of the SVM classifier allows avoiding the problems of overfitting or the problems of underfitting of the SVM classifier. If the errors on the training and testing datasets are close to each other and small in value, such the SVM classifier is recognized the sought for the solution of classification problems.

Since a complex, multi-extreme and multi-parameter objective function is used for the construction of SVM classifiers, it is advisable to use the search for its optimum from the whole space of possible solutions.

Currently, the optimization algorithms inspired by the natural biological systems are used widely. Such algorithms are the bioinspired algorithms for the stochastic optimization: genetic algorithm, particle swarm optimization algorithm, ant colony optimization algorithm, bee algorithm. These algorithms operate with sets of simple entities in the search space, simulating the intellectual behavior of a population in which each individual represents some alternative approximate solution.

In recent years, the particle swarm optimization algorithm (PSO algorithm) [4–9] is used in the solution of various applied optimization problems, based on the idea of possibility of the optimization problems solving by modeling of behavior of the animals groups.

The PSO algorithm is characterized by simplicity of implementation and, consequently, low algorithmic complexity. It is sufficient to determine only the value of the optimized function for the implementation of the PSO algorithm. In this regard, the PSO algorithm can be recommended for the search of the optimum parameters values of the SVM classifier.

Currently, there are various ways to improve the efficiency of the basic PSO algorithm, which can be divided into metoptimazine and combinational.

In this paper we propose to implement a combinational method of improving for the basic PSO algorithm by developing the hybrid versions with the use of the grid search algorithms. It is plan to use two grid search algorithms: the classic Grid Search algorithm (GS algorithm) and Design of Experiment algorithm (DOE algorithm) [2, 6, 7].

The aim of this paper is the development of the hybrid versions of the basic PSO algorithm based on the grid search algorithms and the comparison of their search characteristics. It is planned to test the developed hybrid versions of the PSO algorithm on real datasets in the framework of the problem solving of search of the optimum parameters values of the SVM classifier. The main indicators to measure the effectiveness of the developed algorithms are the search time of the optimum parameters values of the SVM classifier, the quality of data classification (overall accuracy, sensitivity, specificity, the number of support vectors). Herewith, the problem of binary classification has been considered.

2 Principles of the SVM algorithm implementation

As a result of SVM classifier learning, the separating hyperplane is defined (Figure 1) [6]. It can be represented by equation $<w,z>+b=0$, where w is the vector-perpendicular to the separating hyperplane; b is the parameter which corresponds to the shortest distance from the origin of coordinates to the hyperplane; $<w,z>$ is the scalar product of vectors w and z. The condition $-1<<w,z>+b<1$ specifies the strip that separates the classes. The wider the strip is, the more confidently we can classify objects.

The objects closest to the separating hyperplane and located on the bounders of the separating strip are called support vectors. They carry basic information about the separation of the classes.

One of the main problem in case of nonlinear separability of objects is to define the rectifiable type of the kernel function and select the optimal values for some set of parameters in order to build the effective SVM classifier.

The classification of the specific object can be performed using the following rule [6]:

$$F(z)=sign\ \left(\sum_{i=1}^{S}\lambda_i y_i \kappa(z_i,z)\ +\ b\right), \tag{1}$$

where λ_i is a dual variable of the Lagrange function; z_i is the object of the training dataset; $y_i \in Y=\{-1;+1\}$ is the number, which characterizes the class of the object z_i from the training dataset; $\kappa(z_i,z)$ is the kernel function; C is the regularization parameter ($C>0$); S is the quantity of objects in the training dataset; $i=\overline{1,S}$.

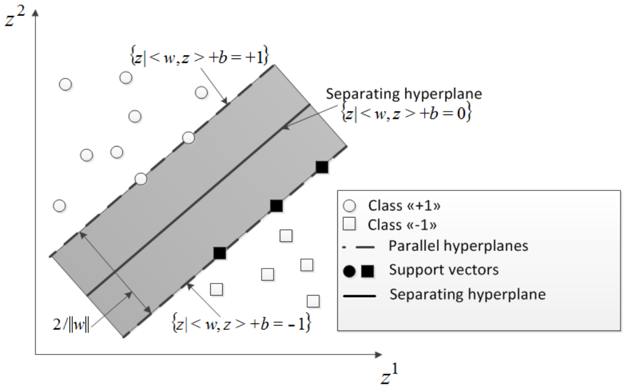

FIGURE 1 The separating hyperplane in the space D-2

The most complete mathematical description of the SVM algorithm is given in [5, 6].

The main problem of the SVM classifier learning is the absence of the recommendations for choice of the value of the regularization parameter C, the kernel function type $\kappa(z_i,z)$, and the kernel function parameters values, which provide the high data classification accuracy. This problem can be solved by means of the application of different optimization algorithms, in particular, with the use of the PSO algorithm.

The radial basis kernel (RBF) [6] is often applied for the SVM classifier development in case of nonlinear separation of objects into classes:

$$\kappa(z_i,z)=\exp(-\|z_i-z\|^2/(2\sigma^2)), \tag{2}$$

where $\sigma>0$.

Herewith, it is necessary to determine the value of the parameter σ of the radial basis kernel function along with the value of the regularization parameter C.

3 The principles of implementation of the PSO algorithm and its hybrid versions

The search space in the PSO algorithm is filled with a population of particles each of which has some location and velocity in the space of the problem parameters at the concrete moment of time. In addition, each particle can remember its best location in the swarm and communicate with other particles about the globally "best" location among all particles.

The value of the objective function is calculated for each particle. The particle location and the velocity are changed according to the certain rules [6, 7] after calculation of the new value of the objective function.

The basic principles of calculating of the new location and the new velocity of particles are given in [6, 7].

Currently, the different versions of the PSO algorithm are known. The canonical version received the traditional application and it's one of the most common versions of the PSO algorithm. In this version of the PSO algorithm it is proposed to perform the normalization of the speedup' coefficients in such way that the convergence of the algorithm not so much depends on the choice of their values [6, 7].

In recent years, the approaches implementing the hybridization of the PSO algorithm with other optimization algorithms in order to increase the efficiency of the classical PSO algorithm are widely used [6–8].

In this paper we present two hybrid versions of the PSO algorithm, involving the use of the classical "Grid Search" (GS) algorithm and the "Design of Experiment" (DOE) algorithm [6–8].

The proposed hybrid versions of the PSO algorithm were developed, primarily, to solve the problem of search of the optimum parameters values of the SVM classifier based on the radial basis kernel function. These algorithms operating with the set of particles in the search space of D-2 can be applied for solving other optimization problems of the appropriate dimension. Also, these algorithms can also be adapted to the case of the search space with more higher dimensional.

At the creating of the hybrid version of the PSO algorithm it is proposed to execute the clarification of the position (coordinates) of the globally best particle in the swarm at the each iteration of the PSO algorithm with the use of the grid search algorithm and update the current swarm particle population. A herewith the "worst" particle should be removed from the swarm (the particle with the "worst" value of objective function), and the "best" particle founding by the grid search algorithm must be added instead.

In the hybrid PSO-GS algorithm, acceleration of search of the globally optimal solution in swarm is achieved by:

- additional grid search in the area of the potential globally "best" location in swarm;
- updating of the particle swarm population and removal of the "worst" particles.

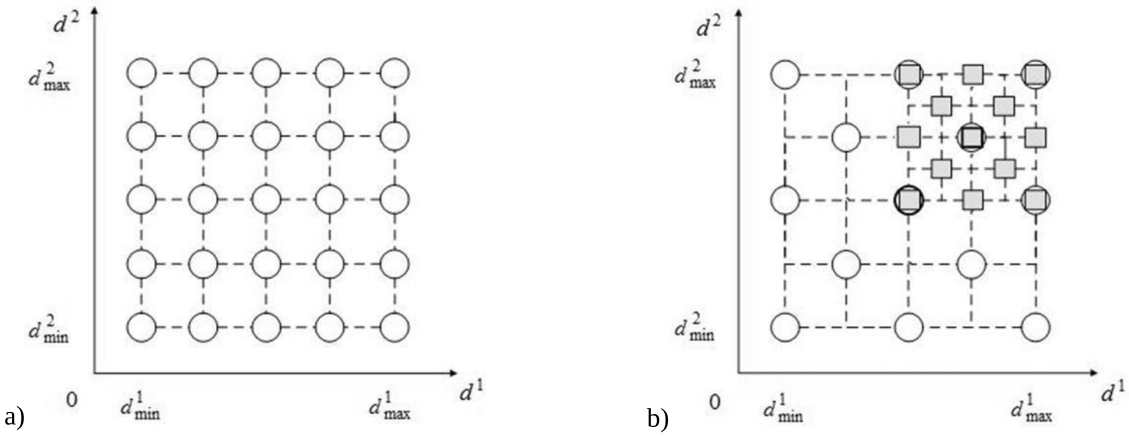

FIGURE 2 The grid formation: a) in the GS algorithm; b) in the DOE algorithm

In the GS algorithm the grid is created on the variation's ranges of the optimization parameters with a certain step for each parameter (Figure 2, a, a special case of the search space D-2) and the efficiency of all combinations of values of the optimization parameters on the grid is evaluated. A herewith all nodes of the grid are explored. The advantage of the search over all grid nodes is the thoroughness of the finding of the globally optimal solution.

The DOE algorithm is an alternative grid search algorithm. The advantages of the DOE algorithm (Figure 2, b) are the following.

- The search boundaries are iteratively improved until the conditions for stopping the search are not satisfied. After each iteration of the DOE algorithm, the search space is narrowed and refined so that "the best" founding node corresponding to the best value of the objective function will be the center of the search space.
- If the search process goes beyond the initially specified (acceptable) search ranges, the new boundaries of the search ranges will be defined in such way that the new search space in the DOE algorithm will be kept within the permissible boundaries of the search ranges.

The hybrid version of the PSO algorithm can be presented by the following sequence of steps.

Step 1. To determine the initial characteristics of particles in the swarm (coordinates and velocities). To determine the customizable parameters of the PSO algorithm (the number of particles in a swamp, the maximum iterations number of the PSO algorithm, the boundaries of the search ranges).

Step 2. To realize one step of the PSO algorithm. To correct the velocity $v_i \in R^n$ and the current coordinates $x_i \in R^n$ for each i-th particle ($i = \overline{1, m}$), to determine the coordinates of the globally "best" particle in the swarm.

Then the objective function value is calculated at each new point of the search space and check of each point is carried out in order to determine if its location is the best in swarm. For the problem to search a minimum of the function in the form $f(x) \to \min\limits_{x \in R^n}$ the best particle location will be considered as a point in the search space where the minimum value of this function is achieved at all of the algorithm iteration starting from the first iteration to the current.

Step 3. To determine the boundaries of the search ranges for the grid search algorithm (the classical GS algorithm or the DOE algorithm). A herewith to determine the grid size $[d^j_{\min}, d^j_{\max}]$ ($j = \overline{1,n}$) taking into account the maximum straggling $[r^j_{\min}, r^j_{\max}]$ of particles in the swarm at the current generation of the PSO algorithm. The values of coordinates of the globally "best" particle in swarm can be used as the values of coordinates χ^j ($j = \overline{1,n}$) of the "main" (central) grid node.

The minimum distance from the "best" particle in the swarm (the centroid of the grid) to the boundaries of the straggling range can be calculated as:

$$l^j = \min\{\chi^j - r^j_{\min}, r^j_{\max} - \chi^j\} \tag{3}$$

and the boundaries of the search ranges for the grid are defined as:

$$d^j_{\min} = \chi^j - l^j, \tag{4}$$

$$d^j_{\max} = \chi^j + l^j. \tag{5}$$

Step 4. To specify the coordinates of the globally "best" particle in the swarm using the grid search algorithm (the classical GS algorithm or the DOE algorithm). To check if the really clarification of coordinates of the globally "best" particle of the swarm is achieved. If the clarification is achieved (a new solution is obtained), then to transfer to step 5 or otherwise to transfer to step 6.

Step 5. To take as the new globally optimal solution at the current iteration of the PSO algorithm the solution obtained by implementing of the grid search algorithm at the step 4.To update the swarm particle population: to delete the "worst" particle and to add the "best" particle founding at the step 4.

Step 6. In the case of achievement of the algorithm breakpoint determined according to the maximum number of iterations or the finding of the global optimum with the given accuracy, to transfer to step 7 or otherwise to transfer to step 2.

Step 7. To accept the values of coordinates of the "best" particle in the swarm as the found value of the globally optimal solution and to complete the algorithm execution.

The features of implementation of the grid search algorithms which are used at the step 4 of the proposed hybrid version of the PSO algorithm are considered further.

The search ranges of the GS algorithm $[d^j_{\min}, d^j_{\max}]$ ($j = \overline{1,n}$) found on the base of formulas (4) and (5) at the step 3 of the hybrid version of the PSO algorithm are divided into the specified number of intervals, and the grid nodes are determined.

Then the value of the optimized (target) function in each grid node is calculated. As a result of implementation of the GS algorithm the "best" node with the "best" value of the objective function will be determined. The coordinates of this node can be used as the coordinates of the new globally best particle in the swarm.

The DOE algorithm is used to solve the optimization problems in the search space D-2 typically, but it can easily be adapted to perform the calculations in the space of the arbitrary dimension n. Since we plan to use the hybrid version of the PSO algorithm with the DOE algorithm to solve the optimization problems in the search space D-2 (i.e., when $n = 2$), and, also, because of the good visibility of implementation of the DOE algorithm in this space, the further description of implementation of the DOE algorithm at the step 4 of the hybrid version of the PSO algorithm is given for the particular case in the search space D-2.

Step 1. To determine within the ranges boundaries $[d^j_{\min}, d^j_{\max}]$ ($j = 1,2$) 13 grid nodes (Figure 2, b the nodes of the first iteration of the DOE algorithm are noted with markers of circular shape in white, and the nodes of the second iteration are noted with markers of square shape in grey color, herewith, the nodes which participate in multiple iterations are noted with double markers of circular and square shape). The central node (the centroid of the grid) with coordinates χ^j ($j = 1,2$) (the example in Figure 2, b is the marker of circular shape with the selected contour) corresponds to the globally "best" particle of the swarm, and the width of the search ranges on the current iteration of the DOE algorithm is defined as $S^j = d^j_{\max} - d^j_{\min}$ ($j = 1,2$).

The coordinates of the grid nodes are defined as the following (when moving along the grid from the lower left node from bottom to top, from left to right):

$[\chi^1 - S^1/2, \chi^2 - S^2/2]$, $[\chi^1 - S^1/2, \chi^2 + S^2/2]$, $[\chi^1 + S^1/2, \chi^2 + S^2/2]$, $[\chi^1 + S^1/2, \chi^2 - S^2/2]$, $[\chi^1 - S^1/2, \chi^2]$, $[\chi^1, \chi^2 + S^2/2]$, $[\chi^1 + S^1/2, \chi^2]$, $[\chi^1, \chi^2 - S^2/2]$, $[\chi^1 - S^1/4, \chi^2 - S^2/4]$, $[\chi^1 - S^1/4, \chi^2 + S^2/4]$, $[\chi^1 + S^1/4, \chi^2 + S^2/4]$, $[\chi^1 + S^1/4, \chi^2 - S^2/4]$, $[\chi^1, \chi^2]$.

Step 2. To calculate the value of the objective function at each node of the grid and to find the coordinates φ^j ($j = 1,2$) of the node with the "best" value of the objective function.

Step 3. To override the width of the search ranges as $S^j/2$ ($j = 1,2$) and to use the calculated values as the new values of S^j ($j = 1,2$) for the next iteration of the DOE algorithm.

Herewith, the new boundaries of the search ranges are redefined for the next step as:

$$d^j_{\min} = \varphi^j - S^j/2, \tag{6}$$

$$d^j_{\max} = \varphi^j + S^j/2. \tag{7}$$

Step 4. Go to the step 1 if the number of iterations of the DOE algorithm is not exhausted, otherwise to complete the algorithm. Herewith, the values of coordinates of the "best" node φ^j ($j = 1,2$) found at the current iteration of the DOE algorithm are applied as the new coordinates of the center node of the grid χ^j ($j = 1,2$) (the example in Figure 2, b is marker of circular shape with the selected contour).

It should be noted that boundaries of the search ranges $[d^j_{\min}, d^j_{\max}]$ ($j = 1,2$) for the first iteration of the DOE algorithm are calculated based on formulas (4) and (5) at the step 3 of the hybrid version of the PSO algorithm, and for all other iterations of the DOE algorithm they are calculated on the base of formulas (6) and (7) at the step 3 of the DOE algorithm.

At implementation of the DOE algorithm the control for the acceptability of the new calculated boundaries of the search ranges is executed.

If at some of the current iteration of the DOE algorithm the coordinates of the "best" found node have been close to the current boundaries of the grid search ranges, then in case of building of the grid at the next iteration of the DOE algorithm the going beyond the originally defined (allowed) boundaries of the search ranges of the hybrid version of the PSO algorithm $[range^j_{\min}, range^j_{\max}]$ ($j=1,2$) is possible. If after the calculations according to the formulas (6) and (7) of the new boundaries of the grid search ranges $[d^j_{\min}, d^j_{\max}]$ ($j=1,2$) it is turned out that one of conditions $d^j_{\min} < range^j_{\min}$ for some $j=j^* \in \{1,2\}$ or $d^j_{\max} > range^j_{\max}$ for some $j=j^* \in \{1,2\}$ is produced, i.e., the going beyond the originally defined (allowed) boundaries of the search ranges of the hybrid version of the PSO algorithm takes place, the grid is narrowed to the new boundaries of the search ranges according to formulas:

if $d^j_{\min} < range^j_{\min}$ for some $j=j^* \in \{1,2\}$, then

$$d^{j^*}_{min} = \varphi^{j^*} - (\varphi^{j^*} - range^{j^*}_{min}), \tag{8}$$

$$d^{j^*}_{max} = \varphi^{j^*} + (\varphi^{j^*} - range^{j^*}_{min}). \tag{9}$$

If $d^j_{\max} > range^j_{\max}$ for some $j=j^* \in \{1,2\}$, then

$$d^{j^*}_{min} = \varphi^{j^*} - (range^{j^*}_{max} - \varphi^{j^*}), \tag{10}$$

$$d^{j^*}_{max} = \varphi^{j^*} + (range^{j^*}_{max} - \varphi^{j^*}). \tag{11}$$

As the result of implementation of this hybrid version of the PSO algorithm the search of the solution of one or another optimization problem can be carried out.

4 The results of experimental studies

The feasibility of application of the proposed hybrid algorithms is confirmed by the results of experimental studies. In particular, the problems of search of the optimum global solution of the several test functions and the problem of search of the optimum parameters values of the SVM classifier were considered.

The several versions of the PSO algorithm were used by performing experimental studies:

- the canonical PSO algorithm (hereinafter referred to as the basic PSO algorithm);
- the hybrid version of the basic PSO algorithm based on the classical GS algorithm (hereinafter referred to as the PSO-GS algorithm);
- the hybrid version of the basic PSO algorithm based on the DOE algorithm (hereinafter referred to as the PSO-DOE algorithm).

The software implementation of these algorithms was conducted by using a high level programming language Python (programming environment Python 3.5). Herewith, the SVM algorithm from the machine learning library Scikit-Learn was used.

The implementation of the optimization algorithms for the test functions. The comparative analysis of these three optimization algorithms was implemented within the framework of solving the problem of search of the global optimum of several test functions. In particular, the results of experimental studies for the Rastrigin, Rosenbrock and sphere objective functions are given in [7].

The obtained results [7] allow to say that the basic PSO algorithm is characterized by the worst values of the quality indicators, such as the average time of convergence, the average convergence rate, the average value of the objective function, the proportion of successful runs in comparison with the PSO-GS algorithm and the PSO-DOE algorithm.

Herewith, the PSO-DOE algorithm allows finding the global optimum of the test functions, on average, in less time than the PSO-GS algorithm does. A large proportion of successful launches is provided and a smaller error in calculating the values of the global optimum of the test functions is achieved by implementing of the PSO-DOE algorithm [7].

The implementation of the optimization algorithms for setting parameters values of the SVM classifier. The advisability of applying the PSO-GS algorithm and the PSO-DOE algorithm for problem solving of search of the optimum parameters values of the SVM classifier was confirmed experimentally.

The studies were conducted using data from Statlog project and from UCI machine learning library. The binary classification was performed for all datasets. The following datasets were used in the present work (Table 1):

- the dataset for medical diagnosis of heart disease – Heart (270 instances, 13 characteristics, the source is http://achive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/);
- the dataset for credit scoring of applications for consumer credits – Australian (690 instances, 14 characteristics; the source is http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/australian/);
- the testing dataset – МОТП12 (400 instances, 2 characteristics; the source is http://www.machinelearning.ru/wiki/index.php?title=Изображение:MOTP12_svm_example.rar).

The calculations using the hybrid versions of the PSO algorithm were performed with different total number of the grid nodes (i.e., in case of different total number of evaluations of the objective function in the grid nodes) calculating for the PSO-GS algorithm and the PSO-DOE algorithm correspondingly to formulas:

$$\gamma = (r+1)^2, \tag{12}$$

$$\gamma = 13 \cdot h, \tag{13}$$

where r is the number of splitting intervals on each j-th range of grid search $[d^j_{\min}, d^j_{\max}]$ ($j=1,2$); h is the number of iterations of the DOE algorithm.

The selection of the optimum parameters values of the SVM classifier was performed on the results of several experiments for different values of the parameters r and h (Figure 3, a special case for a dataset МОТП12). In the present work the following values were selected as optimum based on the criterion of minimum value of the time of the first finding of the optimum: $r = 5$ and $h = 5$.

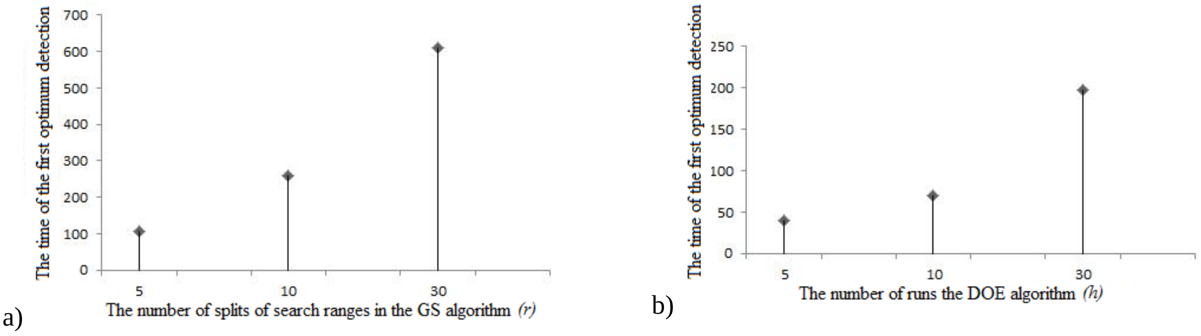

FIGURE 3 The determination of the optimal number of calculations on the grid with implementation of the hybrid versions of the PSO algorithm based on: a) the GS algorithm, b) the DOE algorithm

The radial basis kernel function (2) was used during the development of the SVM classifier. Therefore, the PSO algorithm and its hybrid versions were applied for searching the optimal values of two parameters of the SVM classifier: the regularization parameter C and the coefficient of the kernel function σ (i.e., the calculations were performed in the search space D-2). Herewith, it is supposed that the radial basic kernel function is a priori optimum in the context of solving classification problem.

The parameters values of the SVM classifier are relied as optimum if they provided the high classification accuracy and the minimum number of the support vectors in the training dataset.

The assessment of the classification quality can be performed using the different indicators of the classification quality: the overall accuracy (*Acc*), also called the total ratio of correct answers (the overall success rate, *OSR*); the sensitivity (*Se*), also called the indicator of completeness (the recall, *Re*); the specificity (*Sp*); the precision (*Pr*); the balanced F-measure (*F*1). These indicators are computed by the following formulas:

$$OSR = \frac{TP + TN}{TP + TN + FP + FN}, \quad (14)$$

$$Se = \frac{TP}{TP + FN}, \quad (15)$$

$$Sp = \frac{TN}{TN + FP}, \quad (16)$$

$$Pr = \frac{TP}{TP+FP}, \quad (17)$$

$$F1 = \frac{2 \cdot Pr \cdot Re}{Pr+Re}. \quad (18)$$

where TP is the number of true positive observations; TN is the number of true negative observations; FP is the number of false-positive observations (the error of type II); FN is the number of false negative observations (the error of type I); *Re*=*Se*.

The indicator of overall accuracy *OSR* presents the ratio of true predicted observations in relation to the total number of observations of the classifier.

The indicator of sensitivity *Se* presents which part of the total number of real positive observations is predicted as the positive, i.e. it shows how much the classifier is "pessimistic" in its assessments or how often it "throws off" the observations of the correct class (this occurs at low value of the indicator *Se*). This indicator is also called the indicator of completeness *Re*.

The indicator of specificity *Sp* presents which part of the total number of real negative observations is predicted as the negative.

The indicator of precision *Pr* presents how many of predicted positive observations are really positive, i.e. it shows how the classifier is optimistic in its assessments or how often it "prefers" (and this occurs at low value of the indicator *Pr*) to connect the observations of other classes to the specified.

The indicator of balanced F-measure (*F*1) calculates the harmonic mean between the indicator of precision *Pr* and the indicator of completeness *Re*. In formula (18) the same weight is assigned to both indicators.

To avoid the underfitting and the overfitting of the SVM classifier it was supposed that the high classification accuracy is achieved if the number of errors on the training and testing datasets is minimal, herewith, the number of errors of the SVM classifier on the training and testing datasets are virtually identical [7].

The same values of the parameters of the PSO algorithm and the same search ranges of the parameters values of the SVM classifier were defined for all runs of the optimization algorithms.

In order to ensure the objective comparison of the experiments results, the runs of the basic PSO algorithm and the proposed PSO-GS algorithm and the PSO-DOE algorithm for a particular dataset were initialized by the identical randomly generated initial population of particles.

Besides, the identical random partitions of the original dataset into the training and testing datasets are used. The size of the testing dataset was 20% of the original dataset during the process of the SVM classifier development.

The ROC analysis [6] was used for the quality assessment of the binary classification. The ROC curve, also known as the error curve, displays the ratio between the rate of correct positive classifications of the total number of positive classifications (true positive rate – *TPR*) and the rate of incorrect positive classifications of the total number of negative classifications (false positive rate – *FPR*). The AUC (the area under the ROC curve) gives quantitative interpretation of the ROC curve. It is believed that the higher AUC is, the better the classifier is.

The ROC curves for the SVM classifiers built by using data of the testing datasets for three original datasets described above and the AUC indicator for each SVM classifier are presented in Figure 4.

The parameters setting of the SVM classifiers was performed by using the basic PSO algorithm and its hybrid versions.

At first glance the results of ROC analysis including the results of the comparative analysis of the values of AUC indicator present that the differences of the SVM classifiers is quite small and it is difficult to determine the quality classification. However, presenting the classification results in the form of Table 1, which shows the number of correctly and wrongly classified objects, the advantage of the classification quality should be given to the hybrid versions of the basic PSO algorithm.

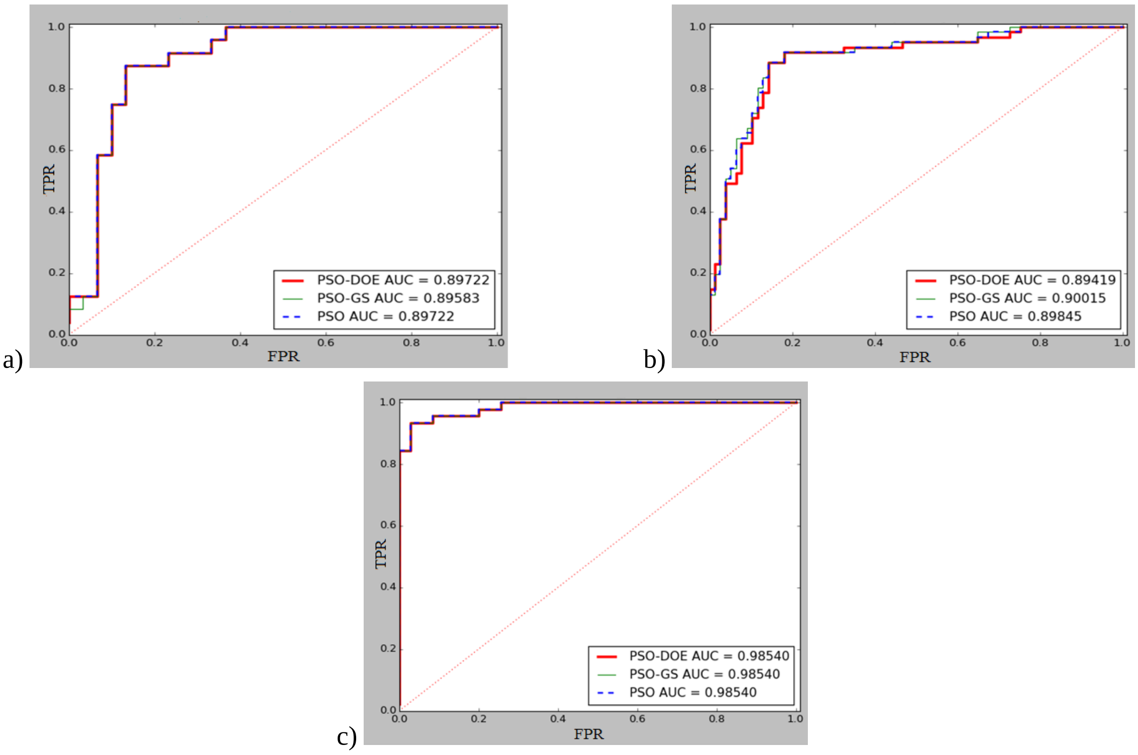

FIGURE 4 The ROC curves for the SVM classifiers built by using the basic PSO algorithm and its hybrid versions: a) for Heart dataset, b) for Australian dataset, c) for MOTП12 dataset

TABLE 1 The classification results

Dataset	Number of objects	Number of features	Version of PSO algorithm	Found parameters		Number of errors (class «1»/class «-1»)		Number of the support vectors	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure	Iteration of the first detection	Search time (sec.)	
				C	x_1	At the training	At the testing							Time of the first detection	total
Heart	270	13	basic PSO	8.87	0.05	6 (2/4)	7 (3/4)	108	95.19	96.67	93.33	0.9571	17	523	642
			PSO-GS	9.82	0.05	5 (1/4)	7 (3/4)	107	95.56	97.33	93.33	0.9605	9	362	714
			PSO-DOE	**9.98**	**0.05**	**5 (1/4)**	**7 (3/4)**	**107**	**95.56**	**97.33**	**93.33**	**0.9605**	**6**	**243**	**712**
Australian	690	14	basic PSO	9.48	0.13	11 (5/6)	18 (7/11)	276	95.80	96.09	95.56	0.9532	12	1546	2872
			PSO-GS	9.73	0.13	12 (5/7)	18 (7/11)	273	95.65	96.09	95.30	0.9516	5	1031	3481
			PSO-DOE	**9.99**	**0.13**	**10 (5/5)**	**18 (7/11)**	**273**	**95.95**	**96.09**	**95.82**	**0.9547**	**4**	**789**	**3292**
MOTP12	400	2	basic PSO	9.89	9.45	12 (5/7)	4 (3/1)	122	96.00	96.10	95.90	0.9610	8	171	441
			PSO-GS	9.89	9.49	12 (5/7)	4 (3/1)	121	96.00	96.10	95.90	0.9610	4	107	653
			PSO-DOE	**10**	**9.47**	**12 (5/7)**	**4 (3/1)**	**121**	**96.00**	**96.10**	**95.90**	**0.9610**	**1**	**40**	**509**

Based on Table 1 we can conclude that the PSO-GS algorithm and the PSO-DOE algorithm solve the problem of searching optimum parameters of the SVM classifier more efficient than the basic PSO algorithm does. The hybrid versions of the PSO algorithm allow reducing the search time of the optimal solution by 3-5 times and the best values of the quality indicators of the SVM classifier are achieved. In particular, we received the highest values of the overall accuracy *OSR*, the sensitivity *Se* and the specificity *Sp*, and the smaller values of the number of the support vectors.

Herewith, the using of the PSO-DOE algorithm provides the best rate of convergence to the optimal solution in most cases (i.e., less time of the first detection of the optimal solution).

5 Conclusion

The results of experimental studies confirm the feasibility of application of the proposed hybrid versions of the PSO algorithm in the framework of the solving the problem of the effective SVM classifier development. The advantage of hybridization of the basic PSO algorithm with the grid search algorithms is the reducing of time for searching the optimum parameters values of the SVM classifier, while maintaining, and in some cases improving, the quality of classification decisions.

The obtained results were achieved by the union of capabilities of the PSO algorithm with the positive features of the grid search algorithms. In particular, the additional search on the grid in the area of the potential globally best position of the particles in the swarm was implemented for updating the population of the particle swarm and removing the "worst" particles.

The further research may be associated with the development of the recommendations for applying of the hybrid optimization algorithms in the framework of the solving of the problem of the SVM classifiers development for unbalanced datasets.

References

[1] Joachims T 2005 A support vector method for multivariate performance measures *In Proceedings of the International Conference on Machine Learning (ICML)* 377–84

[2] Lean Yu, Shouyang Wang, Kin Keung Lai, Ligang Zhou 2008 *Bio-Inspired Credit Risk Analysis: Computational Intelligence with Support Vector Machines* Springer-Verlag: Berlin Heidelberg p 247

[3] Vapnik V 1998 *Statistical Learning Theory* John Wiley & Sons: New York p 740

[4] Ren Y, Bai G 2010 Determination of optimal SVM parameters by using GA/PSO *Journal of Computers* **5**(8) 1160–8

[5] Demidova L, Nikulchev E, Sokolova Y 2016 Big data classification using the SVM classifiers with the modified particle swarm optimization and the SVM ensembles *International Journal of Advanced Computer Science and Applications* **7**(5) 294-312

[6] Demidova L, Klyueva I, Sokolova Y, Stepanov N, Tyart N 2017 Intellectual Approaches to Improvement of the Classification Decisions Quality On the Base Of the SVM Classifier *Procedia Computer Science* **103** 222-230

[7] Demidova L, Klyueva I 2016 Razrabotka i issledovanie gibridnyh versij algoritma roja chastic na osnove algoritmov poiska po setke *Vestnik Rjazanskogo gosudarstvennogo radiotehnicheskogo universiteta* **3**(57) 107–17 *(in Russian)*

[8] Demidova L, Klyueva I, Pylkin A 2016 The Study of Characteristics of the Hybrid Particle Swarm Algorithm in Solution of the Global Optimization Problem *5th Mediterranean Conference on Embedded Computing (MECO)* 322–5

[9] Jun Sun, Choi-Hong Lai, Xiao-Jun Wu 2011 *Particle Swarm Optimisation: Classical and Quantum Perspectives* CRC Press p 419

AUTHORS

Liliya Demidova, 1968, Russia

Current position, grades: professor of Ryazan State Radio Engineering University and Moscow Technological Institute.
University studies: Ph.D (technical sciences, 1994), Dr.Sc (technical sciences, 2009).
Scientific interest: fuzzy set theory, evolutionary algorithms, machine learning.
Publications: more than 250 publications in the artificial intelligence.
Experience: more then 27 years.

Irina Klyueva, 1989, Russia

Current position, grades: post-graduate student of Ryazan State Radio Engineering University.
University studies: Ph.D candidate.
Scientific interest: evolutionary algorithms, machine learning.
Publications: more than 15 publications in the artificial intelligence.
Experience: more then 3 years.